Online Appendix

Prices, Plant Size, and Product Quality

Maurice Kugler
Eric Verhoogen

Review of Economic Studies, Manuscript 13423

A Data Appendix

This appendix provides additional details about the definitions of variables and the processing of the dataset.

A.1 Variable Definitions

Output unit value: Value of output of 8-digit product, divided by number of physical units 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 a yearly average) as the *output price*. In thousands of 1998 Colombian pesos. Average 1998 exchange rate: 1,546 pesos/US$1.

Input unit value: Value consumed of 8-digit product, divided by number of physical units 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 a yearly average) as the *input price*. In thousands of 1998 Colombian pesos. Average 1998 exchange rate: 1,546 pesos/US$1.

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 (i.e. non-casual, non-temporary), 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.

R&D and advertising intensity: 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.

Modified Gollop-Monahan measure of horizontal differentiation: The index is defined as follows:

$$GM_k = \sum_{j,k,t} w_{jt} \left(\sum_i \frac{|s_{ijkt} - \bar{s}_{ikt}|}{2} \right)^{\frac{1}{2}}$$

where i, j, k, and t index inputs, plants, 5-digit industries and years; s_{ijkt} is the expenditure share on input i of plant j in industry k in year t; $\bar{s}_{ikt}$ is the average expenditure share on input i by all plants in industry k in year t; w_{jt} is the share of revenues of plant j in year t in total revenues of all plants in all years in industry k. Following Bernard and Jensen (2007), here we are using just the "dissimilarity" component of the full Gollop-Monahan (1991) index. The term in parentheses gives a plant-specific measure of how different the input mix of plant j is from the average in its industry in the corresponding year. The measure then averages those plant-specific measures over plants and years, using revenues as weights.

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 industries using concordance from OECD, which generated some fractional values.

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 domestic 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.

A.2 Data Processing

Plant-level data from the *Encuesta Anual Manufacturera* (EAM) are available for 1977-2005, but product-level data are available only for 1982-2005.[1] The EAM is a census of all manufacturing plants in Colombia with 10 or more workers, with the following qualifications. Prior to 1992, the sole criterion for initial inclusion of a plant in the census was that the plant have a total of 10 or more employees (including unpaid and temporary employees).[2] 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 A.1) in excess of 65 million Colombian pesos (approx. US$95,000). 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.

[1] Note that we do not observe which plants purchase from which input suppliers.

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

Also, 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.[3]

The longitudinal links between plant-level observations we use are those that are reported directly by DANE or that can be constructed on the basis of name, address and telephone information. 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.[4]

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).[5] 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. Four-digit industry corresponds to the first four digits of the plant's 5-digit industry defined in this way. Our industry categories thus do not change over time within plant.

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.[6]

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

3. In the plant-level file, we "winsorized" the data within each year (Angrist and Krueger, 1999) for 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 file, 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).

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

[4]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.

[5]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].

[6]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.

5. In the product-level file, 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 in the plant-level file.

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.

We have recalculated results using a variety of different bounds for the winsorizing procedure as well as a number of different strategies for dealing with the remaining outliers, and have found the results we report to be robust.

As discussed in footnote 9, in order to carry out the estimation of plant-year effects in the two-step method in Section B.1, plants must be in a connected "network" of plants, where a plant is connected if it produces (consumes) a good that is also produced (consumed) 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 also use only the "connected" plants also when using the one-step procedure described by equation (1).

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 1982-1994 panel.[7]

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.

The U.S. FTC Line of Business Program, from which we draw the measure of R&D and advertising intensity, was in existence from 1974 to 1977 and was unique in that it required firms to break down advertising and R&D expenditures by industry, as opposed to reporting consolidated

[7]Information on exports and imported inputs is also available in 2000-2005, but the information is collected in a different way and the export measures are not directly comparable in the two periods.

figures at the firm level. We converted the information on advertising and R&D expenditures and sales from the FTC industry classification (which is similar to the 1972 U.S. Standard Industrial Classification) to the ISIC revision 2 4-digit level using verbal industry descriptions.

B Additional Results

This appendix presents a number of additional results and robustness checks related to the analysis in Section 3.

B.1 Two-Step Method

An alternative to the "one-step" econometric model in equation (1) is a two-step method, in which one estimates plant-year fixed effects in a first stage and then regresses the estimated fixed effects on plant size in a second stage. The econometric model is:

$$\ln p_{ijt} = \nu_t + \psi_{it} + \mu_{jt} + u_{ijt} \quad \text{(B1)}$$

$$\widehat{\mu}_{jt} = X_{jt}\gamma + \delta_{rt} + \xi_k + v_{jt} \quad \text{(B2)}$$

where μ_{jt} is a plant-year effect, u_{ijt} and v_{jt} are mean-zero disturbances, and other variables are defined as in (1) above.[8] The first-stage estimates $\widehat{\mu}_{jt}$ 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 log unit values of a given plant and log unit values of other plants producing (or consuming) the same products in the same year.[9] Columns 1-3 of Table A5 report the two-step estimates corresponding to (B1)-(B2). The estimates for the plant-average output price in Panel A are smaller than those in Panel A of Table 1, but the preferred estimates in Columns 2-3 are nonetheless positive and significant at the 95% level. The estimates for the plant-average input price in Panel B are nearly identical to those in Panel B of Table 1.[10] Column 4, Panel A presents an additional consistency check: not surprisingly, plants with high average output prices also pay high average input prices.[11] Overall, although not surprising, it is reassuring that the one-step and two-step methods are broadly consistent.

B.2 Results for Non-Exporters

A possible interpretation of the results in Section 3 is that larger plants tend to have higher output and input prices because they are more likely to be exporters and because plants sell higher-quality goods on the export market than on the domestic market. This might arise because of minimum quality standards in international markets (Hallak and Sivadasan, 2009), because consumers in

[8] Note that this two-step model differs from the one-step model above in the weights placed on each observation, with the one-step method effectively placing more weight on plant-years with a greater number of plant-product-year observations. See Baker and Fortin (2001, pp. 358-359) and Donald and Lang (2007) for a useful discussion of the relationship between such one-step and two-step estimators.

[9] 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 both plant and person effects (Abowd, Creecy, and Kramarz, 2002). 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 4% of plant-year observations in the sample.

[10] Note also that the output price-plant size and input price-plant size elasticities are of similar magnitudes, suggesting caution in interpreting the difference in magnitudes in the one-step method discussed in the main text.

[11] As a further check, omitted from the table to save space, we added the number of outputs as a co-variate to the regression in Column 2, Panel A. The coefficient on log employment was 0.013 and significant at just below the 95% level. It appears, in other words, that the fact that larger plants produce more outputs is not wholly responsible for the positive size-output price correlation.

export destination markets tend to be richer and hence more willing to pay for product quality (Verhoogen, 2008), or because per-unit transport costs lead plants to "ship the good apples out" (Hummels and Skiba, 2004). If this were the case, then our focus on plant size per se (as opposed to export status) might be misplaced. To investigate this issue, we re-estimate our baseline model using only data from non-exporting plants. Table A6 reports the results. Comparing to Table 1, 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 positive output and input price-plant size elasticities are entirely attributable to the fact that larger plants are more likely to be exporters.[12]

B.3 Wages

The one input for which prices are commonly observed in plant-level datasets is labor. To compare our results for material inputs to results for employee wages, Table A7 presents regressions that are similar to those in Panel B, Columns 1-3 of Table 2 but with average earnings of all employees, blue-collar employees, white-collar employees, and the white-collar/blue-collar earnings ratio, respectively, as the dependent variables (with all variables in logs). 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).

B.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 the main text, 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, measurement error in physical quantities will generate a spurious negative correlation between physical quantities and unit values. As discussed by Deaton (1988, pp. 422-423), this bias is separate from the standard attenuation bias due to classical measurement error. Column 1 of Table A8 reports regressions of the form of (1), 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 a clean instrument for physical quantities at the product level, log employment is available as an instrument at the plant level.[13] When we use log employment as an instrument for log physical

[12] These results leave open the possibility that larger plants ship outputs a greater distance domestically (and purchase domestic inputs from further away), and hence that *domestic* shipping of good apples is in part responsible for the observed price dispersion. Data on the distance of domestic shipment destinations are unfortunately not available in Colombia, and investigation of this issue will have to await future work in other countries. At the same time, we note that the domestic shipping-the-good-apples out story is not inconsistent with the broader conclusion that the plant-level price differences reflect differences in input and output quality.

[13] A potential instrument at the product level is the lag of physical quantity. In results available from the authors, we instrument log physical quantity with its lag and find a negative relationship between price and physical quantity, as in Column 1 of Table A8. Note, however, that the lag will not be a valid instrument if measurement error in physical quantity is serially correlated. We additionally regress log product-specific revenues (or expenditures) on

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 1) nor, indeed, from the reduced-form estimates in Column 2 of this table (which are the same as in Column 2 of Table 1). 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.

log physical quantity instrumented by its lag, and find coefficients on physical quantity that are significantly less than one. This suggests either (1) that plants that produce more physical units charge higher prices (and plants that purchase more physical units pay higher prices) or (2) that attenuation bias due to measurement error, even after instrumenting with the lag, remains important. A conservative interpretation of these results, from the point of view of our argument, is the latter, that measurement error in physical quantities is indeed serially correlated and that instrumenting with lagged physical output does not entirely solve the measurement error problem. It seems plausible that some plants misreport units of measurement and that they do so consistently over time, generating the serial correlation.

C Details for Model of Endogenous Quality Choice

This appendix presents details of the general equilibrium solution of our quality model, under the two variants of the quality production function. In both cases, the solutions closely follow Melitz (2003).

C.1 Variant 1: Complementarity between Plant Capability and Input Quality

In this case, to ensure that both the distribution of capability draws and the distribution of plant revenues in the final-good sector have finite means the assumption we need on the shape parameter of the Pareto distribution is $k > \max(\eta, 1)$.

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:

$$\pi_d(\lambda^*) = \frac{r_d^*(\lambda^*)}{\sigma} - f = 0 \tag{C1}$$

where $r_d^*(\cdot)$ represents revenues in the domestic market (given by (9d) when $Z = 0$). 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:

$$\pi_x(\lambda_x^*) = \frac{r_x^*(\lambda_x^*)}{\sigma} - f_x = 0 \tag{C2}$$

where $r_x^*(\cdot)$ represents revenues in the export market. 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:

$$[1 - G(\lambda^*)]\sum_{t=0}^{\infty}(1-\delta)^t\left\{\frac{E(r_d^*(\lambda))}{\sigma} - f\right\} + [1 - G(\lambda_x^*)]\sum_{t=0}^{\infty}(1-\delta)^t\left\{\frac{E(r_x^*(\lambda))}{\sigma} - f_x\right\} - f_e = 0 \tag{C3}$$

Using (C1), (C2), and the facts that $\frac{r_d^*(\lambda)}{r_d^*(\lambda^*)} = \left(\frac{\lambda}{\lambda^*}\right)^\eta$ and $\frac{r_x^*(\lambda)}{r_x^*(\lambda_x^*)} = \left(\frac{\lambda}{\lambda_x^*}\right)^\eta$, we have that, conditional on entering each market, $E(r_d^*(\lambda)) = \frac{k}{k-\eta}(\sigma f)$ and $E(r_x^*(\lambda)) = \frac{k}{k-\eta}(\sigma f_x)$. Then using (C3) we can solve for the entry cut-offs:

$$\lambda^* = \lambda_m \left\{\frac{f\eta}{f_e\delta(k-\eta)}\left[1 + \left(\frac{f}{f_x}\right)^{\frac{k-\eta}{\eta}}\right]\right\}^{\frac{1}{k}} \tag{C4}$$

$$\lambda_x^* = \lambda^* \left(\frac{f_x}{f}\right)^{\frac{1}{\eta}} \tag{C5}$$

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.

Let M_e be the mass of entrepreneurs who pay the investment cost, M be the mass of firms in business in the domestic market, and M_x be the mass of exporters. Total payments by final-good producers for material inputs are equal to total payments by intermediate-input producers for labor-hours. The per-period fixed costs, f and f_x, are also paid to workers. Given the wage normalization, payments to workers are equal to the number of labor-hours utilized. Thus the total effective utilization of labor-hours by existing final-good producers is the difference between total revenues and total profits of final-good producers, denoted Π. The labor market clearing condition is that total effective labor-hours utilization for final-good production plus labor-hours utilization for investment equals total labor supply:

$$L = [ME(r_d^*(\lambda)) + M_x E(r_x^*(\lambda)) - \Pi] + M_e f_e \tag{C6}$$

In steady state, the mass of new entrants in each country is equal to the mass of plants that die:

$$M_e\,(1 - G(\lambda^*)) = \delta M \tag{C7}$$

Combining this with the free-entry condition (C3), we have:

$$\Pi = M\left\{\left[\frac{E(r_d^*(\lambda))}{\sigma} - f\right] + \frac{1 - G(\lambda_x^*)}{1 - G(\lambda^*)}\left[\frac{E(r_x^*(\lambda))}{\sigma} - f_x\right]\right\} = M_e f_e \tag{C8}$$

Together (C6) and (C8) imply:

$$L = ME(r_d^*(\lambda)) + M_x E(r_x^*(\lambda)) \tag{C9}$$

Given the symmetry between countries, $M_x E(r_x^*(\lambda))$ is equal to domestic expenditures on foreign varieties as well as export revenue of domestic firms. Thus (C9) is also the clearing condition for the final-good market: total income (and hence total expenditures) of workers is equal to total revenues of final-good producers.

Using the fact that $\frac{M_x}{M} = \frac{1-G(\lambda_x^*)}{1-G(\lambda^*)} = \left(\frac{f}{f_x}\right)^{\frac{k}{\eta}}$, we can solve for the mass of final-good producers in steady state:

$$M = \frac{L(k-\eta)}{k\sigma f\left[1 + \left(\frac{f}{f_x}\right)^{\frac{k-\eta}{\eta}}\right]} \tag{C10}$$

The solution for M_x follows immediately.

C.2 Variant 2: Fixed Costs of Quality

In this variant, to ensure that the first moment of revenues is finite, we assume that $k > \frac{a(\sigma-1)}{\zeta}$, where k is the shape parameter of the Pareto distribution.

Analogous to (C1)-(C3), the entry cut-offs in this model are pinned down by three conditions. First, the marginal plant at the cut-off for entry into the domestic market (λ^*) earns zero profit:

$$\pi(\lambda^* | Z = 0) = 0 \tag{C11}$$

Second, the marginal plant at the cut-off for entry into the export market (λ_x^*) is indifferent between exporting and remaining solely in the domestic market:

$$\pi(\lambda_x^* | Z = 0) = \pi(\lambda_x^* | Z = 1) \tag{C12}$$

Third, the expected profit of a potential entrepreneur before she pays the entry cost and gets a capability draw is zero:

$$[G(\lambda_x^*) - G(\lambda^*)] \sum_{t=0}^{\infty} (1-\delta)^t E(\pi(\lambda)|Z=0) + [1 - G(\lambda_x^*)] \sum_{t=0}^{\infty} (1-\delta)^t E(\pi(\lambda)|Z=1) - f_e = 0 \quad \text{(C13)}$$

Using (11a)-(11e), profit can be written:

$$\pi(\lambda) = \frac{r(\lambda)}{\sigma} - f_q - f - Zf_x = \left(\frac{\zeta}{1-\zeta}\right) \{\alpha\Phi\}^{\frac{1}{\zeta}} \lambda^{\frac{a(\sigma-1)}{\zeta}} - f - Zf_x \quad \text{(C14)}$$

Conditions (C11) and (C12) then imply:

$$\frac{\lambda_x^*}{\lambda^*} = \left\{ \frac{f_x}{f\left(2^{\frac{1}{\zeta}} - 1\right)} \right\}^{\frac{\zeta}{a(\sigma-1)}} \quad \text{(C15)}$$

Under the assumption that $f_x > \left(2^{1/\zeta} - 1\right) f$, we have $\lambda_x^* > \lambda^*$.

Calculating expected profits separately for non-exporters and exporters using (C14), and combining with (C13) and (C15), we can solve for λ^* as a function of exogenous parameters:

$$\lambda^* = \lambda_m \left\{ \frac{f}{\delta f_e} \left(\frac{\frac{a(\sigma-1)}{\zeta}}{k - \frac{a(\sigma-1)}{\zeta}} \right) \left[\left(2^{\frac{1}{\zeta}} - 1\right)^{\frac{k\zeta}{a(\sigma-1)}} \left(\frac{f}{f_x}\right)^{\frac{k\zeta}{a(\sigma-1)}-1} + 1 \right] \right\}^{\frac{1}{k}} \quad \text{(C16)}$$

An expression for λ_x^* follows immediately. By the same argument as in appendix C.1, total profits equal total payments of the fixed entry cost ($\Pi = M_e f_e$) and the total wage bill is equal to total revenues:

$$L = M_d E(r(\lambda)|Z=0) + M_x E_x(r(\lambda)|Z=1) \quad \text{(C17)}$$

where M_d is the mass of non-exporting plants and M_x is the mass of exporters. Note that $\frac{M_d}{M_x} = \frac{G(\lambda_x^*) - G(\lambda^*)}{1 - G(\lambda_x^*)}$. Calculating expected revenues similarly to expected profits above, and using (C15), we have that:

$$M_d = \frac{\zeta L}{\sigma f} \left(\frac{k - \frac{a(\sigma-1)}{\zeta}}{k} \right) \left[\left(\frac{f_x}{f\left(2^{\frac{1}{\zeta}} - 1\right)} \right)^{\frac{k\zeta}{a(\sigma-1)}} - 1 \right] \left[\left(\frac{f_x}{f\left(2^{\frac{1}{\zeta}} - 1\right)} \right)^{\frac{k\zeta}{a(\sigma-1)}} - \frac{f_x}{f} \right]^{-1} \quad \text{(C18)}$$

That this expression is positive follows from $0 < \zeta < 1$ and $k > \frac{a(\sigma-1)}{\zeta}$. M_x follows immediately.

D "Quality" Interpretation of Melitz Model

This section spells out a "quality" interpretation of the Melitz (2003) model, which is alluded to in Melitz's original paper (p. 1699). As mentioned in the text, when $b = 0$ in the first variant or $\alpha = 0$ in the second, our model reduces to the Melitz (2003) model — to be precise, to the special case of the Melitz model with a Pareto distribution of productivity draws and zero transport costs. Letting $\varphi \equiv \lambda^a$, (2), equation (4) and either (9a)-(9d) or (11a)-(11e) become:

$$U \equiv X = \left[\int_{\varphi\in\Phi} x(\varphi)^{\frac{\sigma-1}{\sigma}} d\varphi\right]^{\frac{\sigma}{\sigma-1}} \tag{D1a}$$

$$P = \left[\int_{\varphi\in\Phi} p_O(\varphi)^{1-\sigma} d\varphi\right]^{\frac{1}{1-\sigma}} \tag{D1b}$$

$$p_I^*(\varphi) = c^*(\lambda) = q^*(\varphi) = 1 \tag{D1c}$$

$$p_O^*(\varphi) = \left(\frac{\sigma}{\sigma-1}\right)\frac{1}{\varphi} \tag{D1d}$$

$$r^*(\varphi) = (1+Z)\left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1} XP^\sigma \varphi^{\sigma-1} \tag{D1e}$$

which correspond to the equations in Melitz (2003). Now suppose that the above equations refer to measurements in *quality* units ("utils"), and that higher-φ plants produce goods with more utils per physical unit as given by:

$$\tilde{q}(\varphi) = \varphi^\epsilon \tag{D2}$$

Price in *physical* units is then:

$$\tilde{p}_O^*(\varphi) = p_O^*(\varphi)\,\tilde{q}(\varphi) = \left(\frac{\sigma}{\sigma-1}\right)\varphi^{\epsilon-1} \tag{D3}$$

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

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. If $\epsilon = 1$, then higher φ corresponds to higher quality but marginal cost and hence output price in physical units are constant, as discussed by Melitz (2003, p. 1699).

Second, this "quality" Melitz model is isomorphic to the quality model of Baldwin and Harrigan (forthcoming, 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 $\varphi^{\epsilon-1}$, their ψ corresponds to $\frac{1}{\epsilon-1}$, and their assumption that $\psi > 0$ corresponds to the assumption that $\epsilon > 1$.

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

Fourth, it is worth noting that in this quality Melitz model there is no endogenous quality choice by firms and the model is not well suited to analyzing endogenous variation in the extent of quality differentiation across sectors.

References

Abowd, J. M., R. Creecy, and F. Kramarz (2002): "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data," LEHD Technical Paper 2002-06.

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.

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.

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 (forthcoming): "Zeros, Quality and Space: Trade Theory and Trade Evidence," Forthcoming, *American Economic Journal: Microeconomics*.

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.

——— (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.

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

Deaton, A. (1988): "Quality, Quantity, and Spatial Variation of Price," *American Economic Review*, 78(3), 418 – 430.

Donald, S. G., and K. Lang (2007): "Inference with Difference-in-Differences and Other Panel Data," *Review of Economics and Statistics*, 89(2), 221–233.

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.

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.

Hallak, J. C., and J. Sivadasan (2009): "Productivity, Quality and Exporting Behavior under Minimum Quality Requirements," NBER working paper no. 14928.

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.

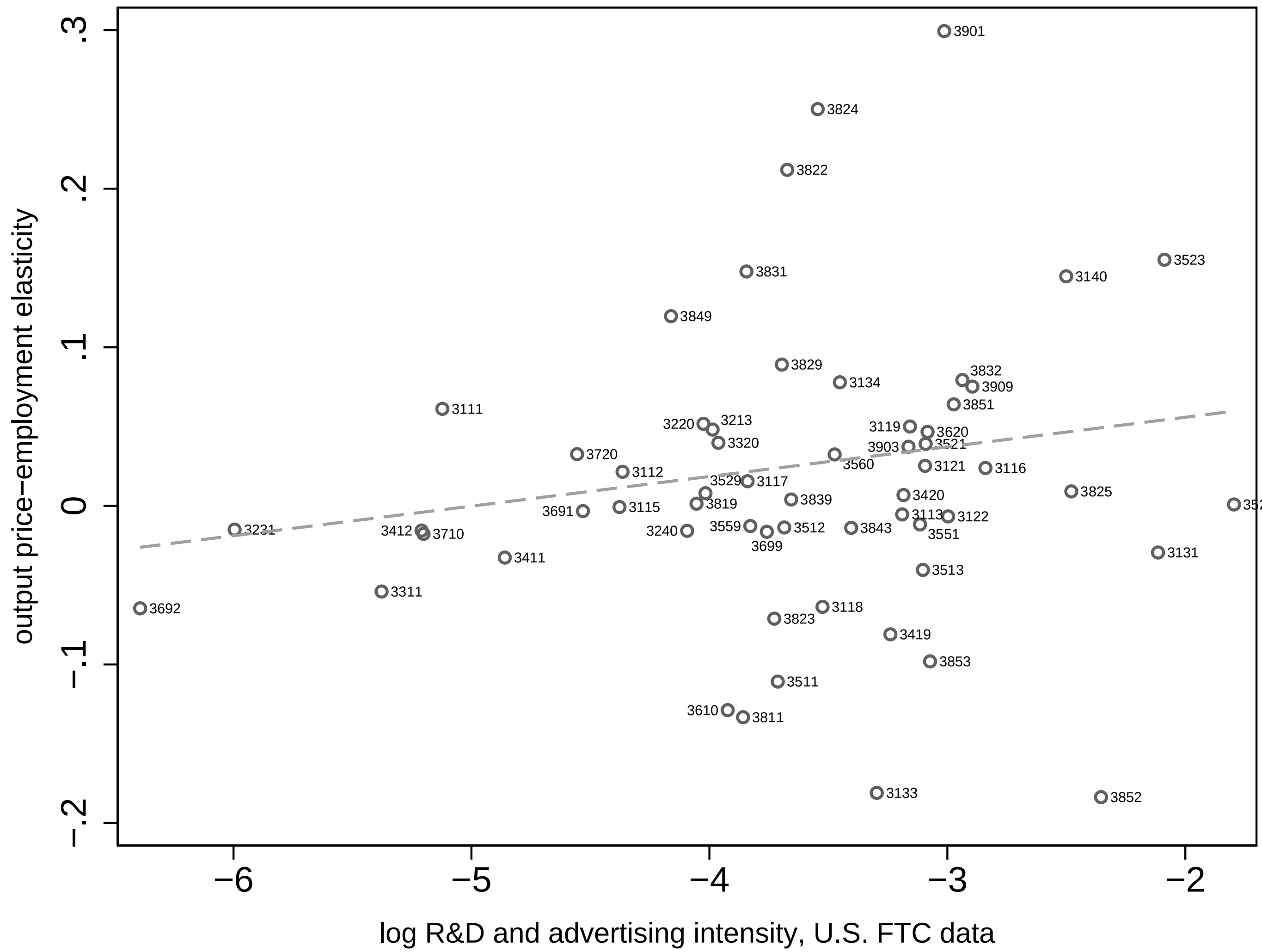
Melitz, M. J. (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71(6), 1695–1725.

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, Amsterdam.

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

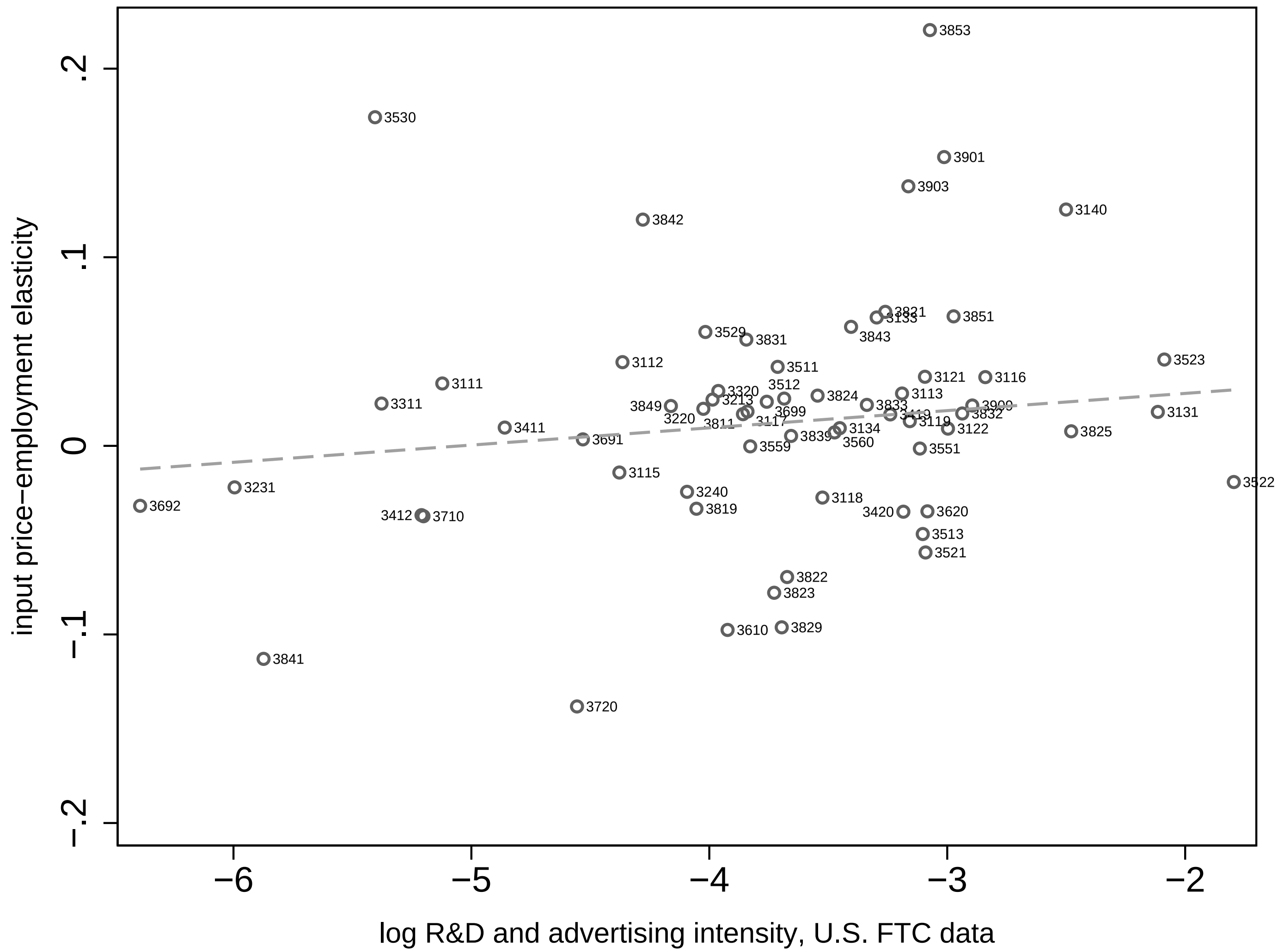
Verhoogen, E. (2008): "Trade, Quality Upgrading and Wage Inequality in the Mexican Manufacturing Sector," *Quarterly Journal of Economics*, 123(2), 489–530.

Figure A1. Output price-plant size elasticities vs. R&D and advertising intensity



Notes: Figure is identical to Figure 3 (for outputs) in main text, but with 4-digit industry used as plotting symbol. Descriptions of 4-digit sectors (and numbers of plant-product-year observations) in Table A2. For details, see notes to Figure 3.

Figure A2. Input price-plant size elasticities vs. R&D and advertising intensity



Notes: Figure is identical to Figures 3 (for inputs) in main text, but with 4-digit industry used as plotting symbol. Descriptions of 4-digit sectors (and numbers of plant-product-year observations) in Table A2. For details, see notes to Figure 3.

Table A1. Summary statistics, plant-level data

	1982-1994 panel			1982-2005 panel
	non-exporters	exporters	all plants	all plants
	(1)	(2)	(3)	(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. Earnings are annual, 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 A.1 for more detailed variable descriptions and Appendix A.2 for details of data processing.

Table A2. Summary statistics, product-level data

	product as output				product as input			
ISIC rev. 2 industry	# prod. (1)	avg. # plant-prod. obs. per yr (2)	w/in-prod. std. dev. log price (3)	w/in-prod-yr std. dev. log price (4)	# prod. (5)	avg. # plant-prod. obs. per yr (6)	w/in-prod. std. dev. log price (7)	w/in-prod-yr std. dev. log price (8)
3111 Meat products	50	395.2	0.53	0.48	52	431.4	0.67	0.60
3112 Dairy products	28	296.3	0.50	0.47	28	678.1	0.41	0.39
3113 Canned fruits/veg.	56	228.7	0.50	0.44	46	388.4	0.54	0.48
3114 Canned fish	19	27.7	0.57	0.44	13	10.1	0.61	0.44
3115 Veg./animal oils	58	175.7	0.46	0.35	70	861.6	0.48	0.43
3116 Grain mill products	50	624.9	0.38	0.32	51	869.4	0.44	0.39
3117 Bakery products	21	724.3	0.55	0.54	18	62.0	0.70	0.64
3118 Sugar refining	15	64.2	0.42	0.35	15	775.2	0.35	0.32
3119 Cocoa/chocolate	35	170.3	0.50	0.46	33	436.0	0.46	0.43
3121 Food products n.e.c.	70	358.8	0.67	0.58	74	2062.1	0.68	0.65
3122 Animal feed	18	193.9	0.42	0.37	16	152.0	0.60	0.53
3123 Dietary supplements	26	51.7	0.55	0.45	20	65.2	0.89	0.76
3131 Spirits	10	36.3	0.47	0.42	9	145.3	0.40	0.39
3132 Wine	6	42.4	0.40	0.38	7	55.2	0.52	0.47
3133 Malt liquors	10	33.1	0.62	0.55	8	35.2	0.64	0.58
3134 Soft drinks	6	126.6	0.50	0.44	5	9.7	1.68	1.20
3140 Tobacco products	5	8.7	0.35	0.29	6	7.8	0.77	0.60
3211 Spun textiles	45	101.9	0.57	0.49	61	2124.7	0.76	0.74
3212 Made-up textiles exc. apparel	31	143.2	0.93	0.83	20	603.4	1.10	1.08
3213 Knitting mills	46	338.7	0.64	0.59	25	585.6	0.99	0.97
3214 Carpets/rugs	7	24.3	0.79	0.64	5	10.9	0.61	0.54
3215 Rope/twine	13	31.7	0.85	0.73	14	142.4	1.04	1.01
3216 Woven cotton textiles	25	87.8	0.67	0.59	27	1084.8	0.58	0.57
3217 Woven wool textiles	14	22.9	0.56	0.47	8	171.4	0.58	0.56
3218 Woven synthetic textiles	21	75.5	0.77	0.71	20	585.8	0.65	0.63
3219 Textiles n.e.c.	25	29.8	0.85	0.70	31	326.2	0.90	0.86
3220 Apparel (w/o leather)	122	2066.4	0.57	0.54	15	2.9	0.29	0.02
3221 Apparel (w/ leather)	49	160.5	0.69	0.61	14	102.6	0.72	0.68
3231 Tanneries	14	63.3	0.84	0.49	16	454.3	0.75	0.52
3232 Fur dressing/dyeing	6	11.7	0.57	0.49	8	5.0	1.53	1.11
3233 Leather prod. exc. footwear	51	222.0	0.88	0.76	25	48.1	1.32	1.17
3240 Footwear	28	255.7	0.49	0.46	9	107.4	0.94	0.90
3311 Sawmills	27	223.8	0.89	0.82	22	436.1	0.71	0.68
3312 Wood/cane containers	7	8.9	1.08	0.83	7	76.8	1.28	1.23
3319 Wood/cork prod. n.e.c.	43	79.8	1.45	1.23	37	145.9	1.02	0.87
3320 Wood furniture	79	1395.3	0.89	0.85	21	19.0	0.88	0.61
3411 Manufacture of paper	49	85.5	0.61	0.49	53	1812.2	0.69	0.67
3412 Cardboard boxes	16	158.0	1.11	1.04	18	2260.0	1.06	1.04
3419 Paper products n.e.c.	73	171.8	1.00	0.77	74	810.4	0.89	0.84
3420 Printing/publishing	83	1323.6	1.22	1.15	57	1284.6	1.10	1.08
3511 Basic chemicals	196	282.0	0.74	0.60	349	7089.5	0.89	0.86
3512 Fertilizers/pesticides	21	88.2	1.07	0.98	20	51.1	1.38	1.20
3513 Resins/plastics	60	96.2	0.57	0.49	73	2018.4	0.65	0.62
3521 Paints	28	166.8	0.65	0.62	29	1261.4	0.64	0.63
3522 Drugs and medicines	32	25.3	0.97	0.55	112	679.7	1.05	0.94
3523 Cosmetics, cleaning prod.	52	468.5	0.91	0.88	40	210.5	0.68	0.62
3528 Various chemical prod.	20	119.6	0.86	0.83	17	166.4	1.02	0.99
3529 Chemical prod. n.e.c.	88	282.1	0.75	0.67	98	3194.3	0.89	0.86

Notes: Table continues on next page. See notes at end of table.

Table A2. Summary statistics, product-level data (cont.)

ISIC rev. 2 industry		product as output (1)	(2)	(3)	(4)	product as input (5)	(6)	(7)	(8)
3530	Petroleum refineries	29	8.6	0.89	0.28	44	537.1	0.87	0.83
3540	Misc. petroleum prod.	16	58.7	0.80	0.71	17	493.7	0.68	0.66
3551	Tires	14	38.3	0.59	0.50	14	48.4	0.82	0.68
3559	Rubber products n.e.c.	68	173.4	0.77	0.67	61	929.2	0.95	0.92
3560	Plastics n.e.c.	232	1019.8	1.00	0.87	215	5265.0	0.95	0.91
3610	Pottery	26	34.2	0.75	0.52	21	31.5	1.25	1.06
3620	Glass products (non-optical)	35	65.0	0.83	0.69	39	463.5	0.86	0.83
3621	Optical glass products	50	75.0	0.90	0.73	30	146.5	0.97	0.92
3691	Clay products	21	129.2	0.62	0.53	13	22.3	1.68	1.35
3692	Cement	8	52.8	0.62	0.57	8	222.8	0.59	0.52
3699	Non-met. min. prod. n.e.c.	81	263.0	0.77	0.67	68	480.9	1.00	0.94
3710	Iron/steel	61	161.9	0.93	0.81	70	2550.6	0.77	0.75
3720	Copper/aluminum prod.	37	61.6	0.80	0.63	41	716.7	0.73	0.70
3721	Lead/zinc products	22	15.1	0.70	0.42	26	133.5	0.66	0.54
3722	Tin/nickel products	31	33.7	0.78	0.63	46	476.5	0.77	0.72
3723	Precious metals	7	3.0	0.89	0.58	15	51.8	0.96	0.87
3811	Cutlery	77	185.2	1.14	0.99	53	563.5	1.15	1.11
3812	Metal furniture	73	700.8	0.87	0.83	25	29.4	1.13	1.01
3813	Structural metal prod.	61	385.0	1.11	1.03	45	62.5	1.14	0.81
3814	Plumbing/heating prod.	23	99.4	1.34	1.25	23	123.2	1.51	1.45
3819	Metal products n.e.c.	172	584.7	1.13	1.01	144	2800.9	0.94	0.89
3821	Engines/turbines	10	4.7	1.90	1.64	6	16.8	0.93	0.70
3822	Agr. machinery	31	54.4	1.20	1.02	7	3.4	0.80	0.49
3823	Metal/wood-working mach.	35	54.7	1.66	1.49	11	10.3	1.46	0.95
3824	Specialized machinery	58	55.8	1.16	0.89	6	1.3	0.88	0.00
3825	Office machinery	18	15.3	1.33	1.03	7	4.1	1.46	1.19
3826	Various non-elect. machinery	37	40.8	1.22	0.96	11	4.7	1.20	1.07
3827	Various non-elect. equipment	39	183.7	1.39	1.26	28	57.3	1.36	1.27
3829	Non-electric machines n.e.c.	57	171.2	1.24	1.13	40	195.6	1.40	1.35
3831	Elect. industrial machinery	45	149.5	1.74	1.59	44	713.1	1.15	1.09
3832	Radio/TV equip.	42	45.8	0.97	0.63	38	69.9	1.39	1.20
3833	Elect. appliances	24	55.5	1.21	1.08	12	8.9	1.31	1.02
3839	Elect. products n.e.c.	57	148.9	1.21	1.08	58	464.2	1.48	1.40
3841	Ship building	6	3.5	1.46	1.27	2	1.0	0.24	0.00
3842	Railroad equip.	3	3.2	1.58	1.09	1	1.0	.	.
3843	Motor vehicles	150	227.9	1.02	0.82	128	124.7	1.25	1.00
3844	Motorcycles/bicycles	16	24.2	0.58	0.43	15	16.9	0.70	0.52
3849	Trans. equip. n.e.c.	5	21.6	0.72	0.66	1	1.0	.	.
3851	Prof./scientific equip.	66	96.3	1.22	0.91	55	112.3	1.29	1.14
3852	Photographic products	8	4.2	0.47	0.24	9	4.5	1.71	0.38
3853	Watches/clocks	5	4.0	1.86	1.70	5	4.6	1.16	0.85
3901	Jewelry	12	23.0	1.16	0.90	7	9.7	1.66	1.10
3902	Musical inst.	6	3.0	0.56	0.43	2	1.0	0.80	0.00
3903	Sporting goods	26	22.7	1.02	0.73	7	1.5	0.59	0.26
3904	Various mfg. products	90	204.6	1.02	0.88	69	1143.4	0.94	0.89
3909	Mfg. products n.e.c.	38	109.3	1.38	1.24	25	33.1	1.28	1.01
	All sectors	3882	17548.6	0.87	0.79	3408	53131.6	0.87	0.83

Notes: Statistics based on 1982-2005 panel. Columns 1 and 5 contain number of distinct products with non-zero sales in any year. Columns 2 and 6 contain number of plant-product observations in industry per year, averaging across years. Columns 3, 4, 7, 8 report standard deviations of residuals from regressions of log real prices on sets of product effects (Cols. 3, 7) or product-year effects (Cols. 4, 8). Each plant-product-year observation assigned equal weight. See Appendix A.2 for details of data processing.

Table A3. Measures of differentiation, by industry

ISIC rev. 2 4-digit industry		R&D + advertising intensity (1)	Gollop-Monahan index (2)	Rauch (1999) index (3)
3111	Meat products	0.006	0.38	0.32
3112	Dairy products	0.013	0.29	0.06
3113	Canned fruits/veg.	0.041	0.48	0.32
3115	Veg./animal oils	0.013	0.40	0.05
3116	Grain mill products	0.058	0.22	0.00
3117	Bakery products	0.022	0.50	1.00
3118	Sugar refining	0.029	0.17	0.00
3119	Cocoa/chocolate	0.043	0.42	0.32
3121	Food products n.e.c.	0.045	0.55	0.58
3122	Animal feed	0.050	0.39	0.00
3131	Spirits	0.121	0.39	0.00
3133	Malt liquors	0.037	0.41	0.00
3134	Soft drinks	0.032	0.51	1.00
3140	Tobacco products	0.082	0.43	0.25
3213	Knitting mills	0.019	0.59	0.88
3220	Apparel (w/o leather)	0.018	0.60	1.00
3231	Tanneries	0.002	0.38	0.67
3240	Footwear	0.017	0.47	1.00
3311	Sawmills	0.005	0.56	0.58
3320	Wood furniture	0.019	0.60	1.00
3411	Manufacture of paper	0.008	0.54	0.00
3412	Cardboard boxes	0.005	0.57	0.44
3419	Paper products n.e.c.	0.039	0.50	0.56
3420	Printing/publishing	0.041	0.60	0.86
3511	Basic chemicals	0.024	0.54	0.07
3512	Fertilizers/pesticides	0.025	0.52	0.39
3513	Resins/plastics	0.045	0.51	0.30
3521	Paints	0.045	0.54	1.00
3522	Drugs and medicines	0.166	0.34	0.67
3523	Cosmetics, cleaning prod.	0.124	0.54	0.95
3529	Chemical prod. n.e.c.	0.018	0.59	0.88
3530	Petroleum refineries	0.004	0.00	0.09
3551	Tires	0.044	0.47	1.00
3559	Rubber products n.e.c.	0.022	0.49	1.00
3560	Plastics n.e.c.	0.031	0.57	0.79
3610	Pottery	0.020	0.25	1.00
3620	Glass products (non-optical)	0.046	0.35	1.00

Notes: Table continues on next page. See notes at end of table.

Table A3. Measures of differentiation (cont.)

ISIC rev. 2 4-digit industry		R&D + advertising intensity (1)	Gollop-Monahan index (2)	Rauch (1999) index (3)
3691	Clay products	0.011	0.42	0.50
3692	Cement	0.002	0.49	0.00
3699	Non-met. min. prod. n.e.c.	0.023	0.50	0.89
3710	Iron/steel	0.006	0.57	0.25
3720	Copper/aluminum prod.	0.011	0.27	0.02
3811	Cutlery	0.021	0.50	1.00
3819	Metal products n.e.c.	0.017	0.58	0.72
3821	Engines/turbines	0.038	0.00	1.00
3822	Agr. machinery	0.025	0.45	1.00
3823	Metal/wood-working mach.	0.024	0.48	1.00
3824	Specialized machinery	0.029	0.56	1.00
3825	Office machinery	0.084	0.30	1.00
3829	Non-electric machines n.e.c.	0.025	0.49	1.00
3831	Elect. industrial machinery	0.021	0.51	0.97
3832	Radio/TV equip.	0.053	0.47	0.99
3833	Elect. appliances	0.036	0.20	1.00
3839	Elect. products n.e.c.	0.026	0.49	0.66
3841	Ship building	0.003	0.00	1.00
3842	Railroad equip.	0.014	0.00	1.00
3843	Motor vehicles	0.033	0.60	1.00
3849	Trans. equip. n.e.c.	0.016	0.43	1.00
3851	Prof./scientific equip.	0.051	0.53	0.99
3852	Photographic products	0.095	0.38	1.00
3853	Watches/clocks	0.046	0.15	1.00
3901	Jewelry	0.049	0.47	0.43
3902	Musical inst.	0.036	0.00	1.00
3903	Sporting goods	0.042	0.15	1.00
3909	Mfg. products n.e.c.	0.055	0.57	0.98
	All sectors	0.030	0.52	0.74

Notes: Tables uses 1982-2005 panel. Table reports measures of differentiation for 4-digit sectors for which all three measures could be constructed. R&D and advertising intensity is defined as ratio of R&D and advertising expenditures to total industry sales from the U.S. Federal Trade Commission (FTC) 1975 Line of Business Survey. The Gollop-Monahan index is the dissimilarity component of the full index from Gollop and Monahan (1991), as defined in Appendix A.2, calculated for 5-digit industries and aggregated to 4-digit level. 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; values are then concorded to ISIC rev. 2 4-digit categories. Averages assign equal weight to each plant-product-year observation in the product-level data on outputs. See Appendix A.1 for more detailed variable descriptions and Appendix A.2 for details of data processing.

Table A4. Measures of concentration, by industry

ISIC rev. 2 4-digit industry		Herfindahl index (suppliers) (1)	Herfindahl index (purchasers) (2)
3111	Meat products	0.28	0.55
3112	Dairy products	0.32	0.26
3113	Canned fruits/veg.	0.36	0.62
3115	Veg./animal oils	0.34	0.30
3116	Grain mill products	0.17	0.29
3117	Bakery products	0.17	0.65
3118	Sugar refining	0.29	0.32
3119	Cocoa/chocolate	0.40	0.44
3121	Food products n.e.c.	0.34	0.41
3122	Animal feed	0.19	0.55
3131	Spirits	0.53	0.64
3133	Malt liquors	0.33	0.63
3134	Soft drinks	0.10	0.73
3140	Tobacco products	0.56	.
3213	Knitting mills	0.29	0.26
3220	Apparel (w/o leather)	0.17	0.93
3231	Tanneries	0.36	0.24
3240	Footwear	0.23	0.24
3311	Sawmills	0.29	0.50
3320	Wood furniture	0.14	0.83
3411	Manufacture of paper	0.60	0.24
3412	Cardboard boxes	0.22	0.09
3419	Paper products n.e.c.	0.52	0.40
3420	Printing/publishing	0.19	0.49
3511	Basic chemicals	0.58	0.26
3512	Fertilizers/pesticides	0.46	0.81
3513	Resins/plastics	0.67	0.35
3521	Paints	0.44	0.23
3522	Drugs and medicines	0.72	0.37
3523	Cosmetics, cleaning prod.	0.33	0.67
3529	Chemical prod. n.e.c.	0.46	0.34
3530	Petroleum refineries	0.88	0.28
3551	Tires	0.35	0.68
3559	Rubber products n.e.c.	0.45	0.36
3560	Plastics n.e.c.	0.34	0.29
3610	Pottery	0.56	0.92
3620	Glass products (non-optical)	0.51	0.37

Notes: Table continues on next page. See notes at end of table.

Table A4. Measures of concentration (cont.)

ISIC rev. 2 4-digit industry		Herfindahl index (suppliers) (1)	Herfindahl index (purchasers) (2)
3691	Clay products	0.30	0.82
3692	Cement	0.32	0.30
3699	Non-met. min. prod. n.e.c.	0.33	0.59
3710	Iron/steel	0.41	0.22
3720	Copper/aluminum prod.	0.62	0.34
3811	Cutlery	0.54	0.38
3819	Metal products n.e.c.	0.44	0.35
3821	Engines/turbines	0.73	0.95
3822	Agr. machinery	0.49	0.79
3823	Metal/wood-working mach.	0.47	0.83
3824	Specialized machinery	0.61	1.00
3825	Office machinery	0.67	0.87
3829	Non-electric machines n.e.c.	0.42	0.45
3831	Elect. industrial machinery	0.46	0.51
3832	Radio/TV equip.	0.52	0.75
3833	Elect. appliances	0.59	0.81
3839	Elect. products n.e.c.	0.48	0.51
3841	Ship building	0.84	1.00
3842	Railroad equip.	0.88	.
3843	Motor vehicles	0.51	0.76
3849	Trans. equip. n.e.c.	0.40	1.00
3851	Prof./scientific equip.	0.64	0.68
3852	Photographic products	0.62	.
3853	Watches/clocks	0.69	0.72
3901	Jewelry	0.57	0.98
3902	Musical inst.	1.00	.
3903	Sporting goods	0.73	1.00
3909	Mfg. products n.e.c.	0.36	0.88
	All sectors	0.30	0.42

Notes: Table uses 1982-2005 panel. Table reports measures of concentration for 4-digit sectors for which the R&D intensity, Gollop-Monahan and Rauch measures could all be constructed (see Table A3). Herfindahl index of suppliers is sum of squared market shares of producers of product as output, by 8-digit industry. Herfindahl index of purchasers is sum of squared expenditure shares of purchasers of product as input, by 8-digit industry. Averages assign equal weight to each plant-product-year observation in the product-level data on outputs. See Appendix A.1 for more detailed variable descriptions and Appendix A.2 for details of data processing.

Table A5. Plant-average prices vs. plant size

	OLS	Reduced form	2SLS	OLS
	(1)	(2)	(3)	(4)
A. Dependent variable: plant-average log output price				
log total output	0.010*		0.012**	
	(0.005)		(0.006)	
log employment		0.013**		
		(0.006)		
plant-avg. input price				0.439***
				(0.016)
industry effects	Y	Y	Y	Y
region-year effects	Y	Y	Y	Y
R^2	0.44	0.44		0.47
N (obs.)	114500	114500	114500	114500
N (plants)	13582	13582	13582	13582
B. Dependent variable: plant-average log 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: Table uses 1982-2005 panel. Plant-average output (input) price 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 (B1)-(B2) in Appendix B.1.) 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 A.1 for more detailed variable descriptions and Appendix A.2 for details of data processing.

Table A6. Product-level prices vs. plant size, non-exporters only

	OLS (1)	Reduced form (2)	2SLS (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: Table uses 1982-1994 panel, since export status is reported on a consistent basis only for those years. Specifications are the same as in Table 1, 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 A.1 for more detailed variable descriptions and Appendix A.2 for details of data processing.

Table A7. Wage variables vs. plant size and exporting variables

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: Table uses 1982-1994 panel, since export status and wages by occupation are reported on a consistent basis only for those years. 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 A.1 for more detailed variable descriptions and Appendix A.2 for details of data processing.

Table A8. Product-level prices vs. physical quantities

	OLS (1)	Reduced form (2)	2SLS (3)
A. Dependent variable: log real output unit value			
log physical quantity	-0.171***		0.033***
	(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	-0.138***		0.016***
	(0.001)		(0.004)
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: Table uses 1982-2005 panel. 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, 0.015 and 0.247 in Panel A and 0.744, 0.012 and 0.255 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 A.1 for more detailed variable descriptions and Appendix A.2 for details of data processing.