Materials Prices and Productivity

Enghin Atalay *

March 1, 2012

Abstract

There is substantial within-industry variation, even within industries that use and produce homogeneous inputs and outputs, in the prices that plants pay for their material inputs. I explore, using plant-level data from the U.S. Census Bureau, the consequences and sources of this variation in materials prices. For a sample of industries with relatively homogeneous products, the standard deviation of plant-level productivities would be 7% lower if all plants faced the same materials prices. Moreover, plant-level materials prices are both persistent across time and predictive of exit. After documenting these patterns, I discuss three potential sources of materials price variation: geography, differences in suppliers' marginal costs, and suppliers' price discriminatory behavior. Together, these variables account for 13% of the dispersion of materials prices. Finally, I demonstrate that plants' marginal costs are correlated with the marginal costs of their intermediate input suppliers.

1 Introduction

There is substantial within-industry variation, even within industries that use and produce homogeneous inputs and outputs, in the prices that establishments pay for their material inputs. This paper assesses the implications and sources of this variation in materials prices. When input prices differ across plants, some plants will have low marginal costs not

*PRELIMINARY AND INCOMPLETE. I thank Frank Limehouse, for his help with the data disclosure process, and Ali Hortaçsu and Chad Syverson, for providing me access to some of their datasets. In addition, I am indebted to Aditya Bhave, Ezra Oberfield, Marshall Steinbaum, and Stephane Wolton for their helpful comments on earlier drafts. Any comments, sent to atalay@uchicago.edu, would be greatly appreciated. Disclaimer: Any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.

only because they efficiently transform inputs into outputs, but also because they happen to purchase their intermediate inputs at particularly low prices. Using plant-level data from the U.S. Census Bureau, I document that intermediate input price dispersion is substantial, persistent across time, and an important factor in explaining dispersion in measured total factor productivity. I then provide suggestive evidence that plants' materials prices[1] are inversely related to their suppliers' productivities, and that intermediate input suppliers engage in price discriminatory behavior. Lastly, I relate plants' productivities to those of their suppliers. I show that these patterns can be matched with a heterogeneous-establishment industry model in which output and factor markets are imperfectly competitive.

Large, persistent, within-industry productivity differences motivate my analysis. Syverson (2004b), for example, estimates that, in the average 4-digit manufacturing industry, the 90th percentile plant has a total factor productivity that is approximately 90% higher than the 10th percentile plant.[2] Given the importance that a plant's productivity has on its growth and survival, as well as the strong relationship between countries' GDPs and the average productivities of their firms,[3] several papers have tried to explain why some plants are productive while others are not. This literature has argued that relatively productive plants are more likely to: employ high quality inputs (Sakellaris and Wilson (2004) and Fox and Smeets (2011)), patent (Balasubramanian and Sivadasan (2011)), expose themselves to export or import markets (Bernard and Jensen (1999), Eslava et al. (2004, 2009), and Halpern, Koren, and Szeidl (2011)), and use best-practice management techniques (Bloom and Van Reenen (2010)). In addition, productivity dispersion is larger in markets with less intense competition (Syverson (2004a)) and in countries with larger factor misallocations (Hsieh and Klenow (2009)).

In the cited studies, plants' productivities are calculated as the ratio of outputs to inputs. Usually data on input and output prices are not collected, meaning that–in most cases–real revenues are the measure of establishment outputs, while real input expenditures are the measure of establishment inputs.[4] With these measures, an establishment's measured

[1] Throughout this paper, I will use the terms "intermediate inputs" and "materials" interchangeably.

[2] Large within-industry productivity differences have been documented across a wide range of sectors and countries. See Bartlesman and Doms (2000) and Syverson (2011) for surveys.

[3] Prescott (1998), for example, argues that cross-country differences in (physical or human) capital stocks can only explain a small fraction of cross-country per capita GDP differences. The residual explanation for GDP per capita differences is that firms' TFP levels vary across countries.

[4] Five partial exceptions are Syverson (2004a), Eslava et al. (2004, 2009), Hsieh and Klenow (2009), and Ornhagi (2006).

In Syverson (2004a) and Hsieh and Klenow (2009), the authors utilize establishment-level output price data, but do not have establishment-level intermediate input price data. Ornhagi (2006), on the other hand, has data on materials prices. Her analysis, focuses on the estimation of input elasticities, instead of the distribution of plant-level productivities, which is the focus here.

Perhaps closest to the current paper, Eslava et al. (2004, 2009) use plant-level input and output-price productivity will depend on the output price it is able to charge and the factor prices it is charged. Potentially, an establishment's measured productivity may have no relationship with how efficient it is in transforming inputs into outputs. Quantifying the extent to which input price variation confounds the measurement of plants' technical efficiencies is one of the main contributions of the this paper.[5]

In empirically documenting a positive relationship between buyers' and suppliers' marginal costs, this paper contributes to two additional literatures. The first set of papers contain models that predict, but have not empirically confirmed, that firms/plants of similar productivities tend to trade with one another. Kremer (1993), for example, develops a model in which the productivity of different production units within a sequential production process is complementary. This model predicts that, as a result of the complementarity, units with similar productivity levels will choose to interact with one another. Oberfield (2011) constructs a model in which buyers meet supplier contacts stochastically, over time; each buyer chooses the supplier, among its contacts, that allows it to produce at the lowest marginal cost. In this model, as well, linked plants' measured productivities will be positively correlated.

Papers in the second literature explore the extent to which shocks to individual firms may have aggregate implications. An unknown, key parameter in these papers is the strength of the relationship between buyer and seller productivity. When buyer and seller productivities are strongly correlated, small differences between countries, in firm-level productivities, will lead to large differences in countries' GDPs (Jones (2011)). Similarly, when buyer and seller productivities are strongly correlated, shocks to a small set of firms will foster large aggregate fluctuations (Acemoğlu et al. (2011; Theorem 3)).

In the following section, Section 2, I sketch out a model in which plants sell output to a representative consumer and purchase intermediate inputs from suppliers in an upstream sector. The model is designed to match the observed correlations between plants' input prices, output prices, and different productivity measures, as well as the correlation between plants' suppliers' marginal costs and plants' own materials prices. In the model, estab-

data from Colombia to test the hypothesis that a trade liberalization stiffens the competitive environment, forces low-productivity plants to exit, and thus increases aggregate productivity.

Among these papers, only Syverson (2004a) restricts the sample to homogeneous-output industries. So, some of the variation in quantity total factor productivity in Eslava et al. (2004, 2009) and Hsieh and Klenow (2009) will be a result of differences in output or input quality.

[5]The potential confounding effects of input and output price variation in productivity estimation are already well known. Both Katayama, Lu, and Tybout (2009) and Gorodnichenko (2010) argue, in detail, how plant-level measured productivities may have little to do with plants' technical efficiencies. After arguing that conventional productivity estimates may be misleading, these papers propose structural estimators of establishments' cost and revenue functions, exploiting information derived from the solutions to their cost minimization and/or profit maximization problems.

lishments enjoy some degree of monopoly power in output markets and monopsony power in factor markets. As in Foster, Haltiwanger, and Syverson (2008), imperfect competition in the goods market yields a positive relationship between revenue productivity and output prices and a negative relationship between quantity productivity and output prices. What is new, compared to most other models of heterogeneous-establishment industries, is a discussion of how intermediate input prices are determined.[6] I assume that each plant has access to only a small set of potential suppliers. Plants that are fortunate to be connected to low-marginal cost suppliers pay lower-than-average materials prices. Because buyers and suppliers of the intermediate input share in the gains from the buyer-supplier relationship, the marginal costs of linked counterparties will correlated with one another.

I begin Section 3 by describing the two plant-level datasets—the Census of Manufacturers and the Commodity Flow Survey—employed in this article, as well as the set of industries that comprise my sample. Building heavily off of Foster, Haltiwanger, and Syverson (2008), I focus on the few industries–such as gasoline, ready-mix concrete, and cardboard boxes–for which plants' output prices and materials prices can be computed and meaningfully compared across establishments. In Section 3.3, I specify five assumptions that allow me to compute plant-level productivity measures. As much as possible, I choose assumptions that are common in the literature on estimation of plants' productivities. The one key difference with the literature is a relaxation of the assumption that all plants within each industry purchase their intermediate inputs at the same price.[7] In Section 3.4, I define three different measures of plant productivity. These three measures differ according to their inclusion or exclusion of plant-level prices. Revenue total factor productivity (TFPR) is computed using industry-level price indices for both plants' outputs and intermediate inputs. Quantity total factor productivity (TPFQ) uses plant-level output prices, but industry-level price indices for intermediate inputs. Finally, technical efficiency (which I denote Φ) uses both plant-level materials and output prices.

I establish, in Section 4.1, the hypothesized negative correlation between TFPQ and materials prices. In my benchmark sample, the correlation between the logarithm of TFPQ and materials prices is -30%. In Section 4.2, I compute the fraction of TFPQ dispersion that is due to differences in materials prices: the standard deviation would be 7% lower, and the 90/10 ratio would be 15% lower, in a counterfactual world in which all plants face the

[6] Several papers model the allocations and profits that emerge out different networks of buyer-supplier relationships. Three examples are de Fontenay and Gans (2007), Abreu and Manea (2011), and Manea (2011). Unlike these papers, the analysis in Section 2.2 invokes convenient assumptions on suppliers' productivities and the method by which prices are determined. In combination, these assumptions yield relatively simple expressions for the distribution of transaction prices.

[7] However, I maintain the conventional assumption that there is no within-industry dispersion in the prices at which plants purchase their labor, capital, or electricity inputs; see Section 3.3.

same materials prices. To give context, 7% to 15% is larger than the fraction of productivity dispersion explained by the competitive environment (Syverson (2004a)), but smaller than fraction explained by differences in labor quality (Fox and Smeets (2011)).

As I demonstrate in Sections 4.3 and 4.4, plant-level intermediate inputs prices are both persistent across time and important predictors of plant survival. The 5-year autocorrelation of the logarithm plants' material prices is 31%. Thus, the serial correlation of the logarithm of input prices is comparable to that of the logarithm of TFPR, TFPQ, or output prices, which are 31%, 35%, and 43%, respectively. In addition, plants that pay more for their material inputs are also significantly more likely to exit the industry: a one standard deviation increase in plants' materials prices corresponds to a 1.1 percentage point increase in the probability that the plant will exit in the next five years. At the same time, there seems to be no difference between entrants and incumbents in the prices that they pay for intermediate inputs.

In Section 5, I address the question of why materials prices differ across the plants of a given industry. An analysis of the sources of materials price dispersion is critical to understanding plants' contributions to social welfare. Plants that, for example, have low materials prices from exploiting monopsony power are contributing nothing to social welfare. On the other hand, plants that search for low-marginal-cost suppliers are contributing to social welfare, even though having low-marginal-cost suppliers do not increase plants' own technical efficiencies.

I offer three potential explanations for within-industry differences in materials prices. First, plants in particular geographic regions enjoy particularly low input prices due, for example, to the abundance of primary materials with which the intermediate input is produced. Second, plants will pay relatively little for their intermediate inputs when their suppliers are exceptionally productive: productive upstream plants are passing on some of their low marginal costs through to their buyers. Thirdly, suppliers tend to charge different prices for their outputs across different destinations. I take this last finding as evidence of price discriminatory behavior, a third source of price variation in intermediate goods markets. Together, these three sources account for 13% of the price variation, for a pooled sample of ready-mix concrete and cardboard box manufacturers. In terms of explaining variation in materials prices, both the across-supplier component (low-marginal-cost suppliers charge, on average, low prices) and the within-supplier component (a given supplier will charge different prices to different downstream plants) are consequential.

Section 6 concludes. Proofs (Appendix A), other useful formulas (Appendix B), a more detailed description of the construction of the benchmark sample (Appendix C), and robustness checks (Appendix D) are all relegated to appendices.

2 Theoretical Motivation

In this section, I motivate the empirical analysis of Sections 4 and 5, using an amended version of the model given in Section 2 of Foster, Haltiwanger, and Syverson (2008). The model consists of an industry of plants, which sell output to a representative consumer, and purchase intermediate inputs from an upstream industry.

The analysis is comprised of two, largely independent, parts. In Section 2.1, the prices that plants pay for intermediate inputs are exogenously specified. Besides facing different materials prices, plants are heterogeneous in their technical efficiencies and appeal to consumers. The main result of this subsection is Proposition 2, which contains a series of relationships among plants' productivity measures, input prices, and output prices.

Section 2.2 describes how materials prices are determined. A key feature of the model is that plants have access to different sets of heterogeneous-cost suppliers. Plants share, with their suppliers, the surplus generated by the buyer-supplier relationship. As a result, plants with low-marginal-cost suppliers will purchase their intermediate inputs more cheaply, generating a positive relationship between the marginal costs of buyers and suppliers.

2.1 Equilibrium conditions, taking materials prices as given

The analysis centers around an industry, with $\tilde{I}$ plants, each producing a differentiated good.

The expenditure function of the representative consumer is given by:[8]

$$\log \Upsilon = \tilde{\Delta}_0 + \sum_{i=1}^{\tilde{I}} \left(\tilde{\Delta}_i \log P_i^{out} - \frac{\gamma\left(\tilde{I}-1\right)}{\tilde{I}} \left(\log P_i^{out}\right)^2 \right) + \frac{1}{2}\frac{\gamma}{\tilde{I}} \sum_{i=1}^{\tilde{I}} \sum_{k=1,k\neq i}^{\tilde{I}} \log P_i^{out} \cdot \log P_k^{out} \tag{1}$$

In this equation, Υ is the total revenues of plants in the industry, P_i^{out} is the price that plant i charges for its output, $\tilde{\Delta}_0$ is a demand shifter for the industry's output, and $\tilde{\Delta}_i$ are idiosyncratic demand shocks for the output of plant i.[9] The parameter γ (restricted to be

[8] Consumer preferences are specified differently in Foster, Haltiwanger, and Syverson (2008). In that paper, demand for each variety is linear in prices. Linear demand yields predicted correlations among the *levels* of prices and productivities. Since, in Section 4, I will correlate the *logarithm* of plants' input prices, output prices, and productivity measures, I prefer the translog specification: this set of preferences generates predictions on the relationships among the logarithm of prices and productivity measures.

I eschew the simpler CES specification of consumer preferences, as this specification–in combination with monopolistic competition–would generate the counterfactual prediction that revenue total factor productivities are unrelated to marginal costs.

[9] The $\tilde{\Delta}_i$ are restricted to sum to 1: $\sum_{i=1}^{\tilde{I}} \tilde{\Delta}_i = 1$.

greater than 0) reflects the degree to which the representative consumer is able to substitute across different plants' outputs.

In general, some subset of the $\tilde{I}$ plants may decide to not produce, in equilibrium. (Below, I will provide an inequality describing when plants are able to profitably produce.) Without loss of generality, label plants so that only the first $I < \tilde{I}$ plants produce. Then, the representative consumer's expenditure function becomes (see Theorem 1 of Bergin and Feenstra (2009)):

$$\log \Upsilon = \Delta_0 + \sum_{i=1}^{I} \left(\Delta_i \log P_i^{out} - \frac{\gamma (I-1)}{I} \left(\log P_i^{out} \right)^2 \right) + \frac{1}{2} \frac{\gamma}{I} \sum_{i=1}^{I} \sum_{k=1, k \neq i}^{I} \log P_i^{out} \cdot \log P_k^{out} \quad (2)$$

$$\text{where } \Delta_0 = \tilde{\Delta}_0 \cdot \left(\sum_{i=I+1}^{\tilde{I}} \left(\tilde{\Delta}_i \right)^2 - \frac{1}{I} \left(\sum_{i=I+1}^{\tilde{I}} \tilde{\Delta}_i \right)^2 \right) \text{ and } \Delta_i = \tilde{\Delta}_i + \frac{1}{I} \left(1 - \sum_{i=1}^{I} \tilde{\Delta}_i \right)$$

Using Shephard's Lemma, and differentiating Equation 2 with respect to $\log P_i^{out}$, yields a simple expression for R_i, the share of industry revenues that go to plant i:

$$R_i = \frac{d \log \Upsilon}{d \log P_i^{out}} = \Delta_i - \gamma \frac{I-1}{I} \ln P_i^{out} + \frac{\gamma}{I} \sum_{k=1, k \neq i}^{I} \ln P_k^{out} \quad (3)$$

As γ increases, plants' revenues are increasingly responsive to their prices.

I assume that each plant, i, produces with a constant returns to scale Cobb-Douglas production function, with intermediate inputs (N_i) and labor (L_i) as the two inputs:

$$Q_i = \Phi_i \left(\frac{L_i}{1-\sigma} \right)^{1-\sigma} \left(\frac{N_i}{\sigma} \right)^{\sigma} \quad (4)$$

While factor elasticities are the same throughout the industry, each plant has a unique technical efficiency, Φ_i. I assume that the wage, W, is the same for all plants in the industry, but the price of the intermediate input, P_i^{in}, may differ. The process by which P_i^{in} is determined will be specified in the next subsection. For now, it suffices to assume that P_i^{in} is constant in N_i (e.g., there are no quantity discounts). Each plant's marginal cost of production equals:

$$MC_i = \frac{1}{\Phi_i} W^{1-\sigma} \left(P_i^{in} \right)^{\sigma} \quad (5)$$

Given the preferences defined by Equations 1-3 and the production function defined by Equation 4, plants' pricing behavior takes a simple form.

Proposition 1 *When the revenue share, R_i, of each plant is sufficiently close to 0, the output price, P_i^{out}, is approximated by the following equation:*[10]

$$\begin{aligned}\log P_i^{out} &\approx \frac{1}{2}\left(\log MC_i + \overline{\log MC}\right) + \frac{1}{2\gamma}\left(\Delta_i + \overline{\Delta}\right) \\ &\approx \frac{1}{2}\left(\sigma \log P_i^{in} - \log \Phi_i + \sigma \overline{\log P^{in}} - \overline{\log \Phi}\right) + \frac{1}{2\gamma}\left(\Delta_i + \overline{\Delta}\right)\end{aligned} \tag{6}$$

where, for any variable X, $\overline{X} \equiv \frac{1}{I}\sum_{i=1}^{I} X_i$.

Proof. See Appendix A on page 39. ■

Output price variation has three sources: Plants with low output prices are technically efficient (Φ_i is large), have low material prices (P_i^{in} is small), or have low demand shocks (Δ_i is small). Even though low marginal costs are correlated with low output prices, plants with low marginal costs also have larger than average markups.

Using Equation 6, I can write plants' revenue and quantity total factor productivities in terms of Φ, P^{in}, and Δ.

$$\log TFPQ_i \equiv \log\left(\frac{Q_i}{Q_i \cdot MC_i}\right) = \log \Phi_i - \sigma \log\left(P_i^{in}\right) - (1-\sigma)\log W \tag{7}$$

$$\begin{aligned}\log TFPR_i &= \log TFPQ_i + \log P_i^{out} \\ &\approx \frac{1}{2}\left(\log \Phi_i - \overline{\log \Phi} - \sigma \log\left(P_i^{in}\right) + \sigma \overline{\log\left(P^{in}\right)} + \frac{1}{\gamma}\left(\Delta_i + \overline{\Delta}\right)\right)\end{aligned} \tag{8}$$

Φ, which embodies plants' ability to transform inputs into outputs, is independent of conditions in factor and output markets. $TFPQ$ reflects both the ability to transform inputs into outputs, as well as conditions in factor markets. Finally, a plant's $TFPR$ is determined by its technical efficiency, conditions in factor markets, and conditions in output markets.

Plugging Equation 6 into Equation 3 yields the following expression for the revenue share of plant i:

$$R_i \approx \frac{1}{2}\left(\Delta_i + \overline{\Delta}\right) - \frac{\gamma}{2}\left(\log MC_i - \overline{\log MC}\right) \tag{9}$$

In combination, Equations 8 and 9 imply that plants will have a positive revenue share ($R_i > 0$) if and only if their $\log TFPR$ is greater than 0.

[10] Equation 6 is derived by invoking the approximation $\log(1+R_i) \approx R_i$, which becomes more accurate as $R_i \to 0$.

Finally, for plants $i \in \{1, ..., I\}$, profits equal:

$$\begin{aligned} \Pi_i &= \frac{\Upsilon \cdot R_i}{P_i^{out}} \left(P_i^{out} - MC_i\right) \\ &\approx \Upsilon \left[\frac{1}{2}\left(\Delta_i + \overline{\Delta}\right) - \frac{\gamma}{2}\left(\log MC_i - \overline{\log MC}\right)\right] \left[e^{\frac{1}{2}\left(\overline{\log MC} + \frac{\bar{\Delta}}{\gamma}\right)} - (MC_i)^{1/2} e^{-\frac{1}{2\gamma}\Delta_i}\right] \end{aligned} \tag{10}$$

Plants $i \in \left\{I+1, ..., \tilde{I}\right\}$ earn zero profits. Total industry sales, Υ, can be computed by plugging in the optimal pricing decisions (Equation 6) into Equation 2.[11]

To close the model, and solve for $\tilde{I}$, I assume that there is a large pool of potential entrants, who may enter only after paying a sunk cost. Potential entrants will weigh expected profits to the cost of entry. To compute the expected profitability, one would integrate Equation 10 over the joint distribution, $\varphi(\Delta, P^{in}, \Phi)$, of Δ_i, P_i^{in}, and Φ_i.

Using Equations 6-8, one may express relationships between plants' productivities and input/output prices. These relationships, which are collected in Proposition 2, depend on the relationships among the three exogenous variables (Δ, Φ, and P^{in}). In turn, the correlations between Δ, Φ, and P^{in} will depend on φ as well any correlation that is induced by selection on $TFPR$. Plants will profitably produce if Δ or Φ is sufficiently large, or if P^{in} is sufficiently low. Even if φ is multiplicatively separable in Δ, Φ, and P^{in}, the correlation between Δ and Φ will be negative in the observed sample. Likewise, selection on $TFPR$ will tend to induce a positive correlation between Δ and P^{in}, and a positive correlation between P^{in} and Φ.

In the following proposition, I assume that these correlations are not too far from 0 in the observed sample. Without making this assumption, it would be difficult to say anything meaningful about the correlations between plants' productivity measures and input/output prices.

Proposition 2 *Assume that a) Cov(*$\log P^{in}$, $\log\Phi$*),Cov(*$\log P^{in}$, Δ*), and Cov(*Δ, $\log\Phi$*) are sufficiently close to 0 in the observed sample, and b)* $R_i \to 0$ *for all plants i. Then, the following relationships hold:*

1. $Cov(\log TFPR, \log TFPQ) = \frac{1}{2}Var(\log TFPQ) + \frac{1}{2\gamma}\left[Cov(\Delta, \log\Phi) - \sigma Cov(\Delta, \log P^{in})\right] > 0$

2. $Cov(\log TFPQ, \log\Phi) = Var(\log\Phi) - \sigma Cov(\log P^{in}, \log\Phi) > 0$

[11] One can show that (see Appendix B.1 on page 41):

$$\log\Upsilon = \Delta_0 + \frac{I\gamma}{8}\left[2 \cdot Cov\left(\frac{\Delta}{\gamma}, \log MC\right) + 3 \cdot Var\left(\frac{\Delta}{\gamma}\right) - Var(\log MC)\right] + \frac{1}{2}\left[E\left[\log MC\right] + E\left[\frac{\Delta}{\gamma}\right]\right]$$

.

3. $Cov(\log TFPQ, \log P^{in}) = -\sigma Var(\log P^{in}) + Cov(\log P^{in}, \log \Phi) < 0$

4. $Cov(\log TFPR, \log P^{out}) = \frac{1}{4} Var(\log TFPQ) + \frac{1}{4\gamma^2} Var(\Delta) > 0$

5. $Cov(\log TFPQ, \log P^{out}) = \frac{1}{2\gamma}\left[Cov(\Delta, \log \Phi) - \sigma Cov(\Delta, \log P^{in})\right] - \frac{1}{2} Var(\log TFPQ) < 0$

6. $Cov(\log P^{in}, \log P^{out}) = \frac{1}{2}\left[\sigma Var(\log P^{in}) - Cov(\log P^{in}, \log \Phi) + \frac{Cov(\log P^{in}, \Delta)}{\gamma}\right] > 0$

7. $Var(\log TFPQ) - Var(\log \Phi) = \sigma^2 Var\left(\log P^{in}\right) - 2\sigma Cov\left(\log \Phi, \log P^{in}\right) > 0$

In Section 4, I check that these seven predicted relationships hold for a sample of homogeneous-goods industries.[12] Moreover, consistent with Equation 8, I document that plants with low technical efficiencies and high input prices tend to exit with greater probability. What is unclear, so far, is why intermediate input prices differ across plants. I develop a theory of materials price dispersion in the following subsection.

2.2 Materials prices

There are two aims of this subsection. The first is to show that plants' marginal costs are positively related to those of their suppliers: plants that are lucky enough to be connected to an exceptionally low-cost supplier will purchase their intermediate inputs at a low price, and thus have low marginal costs. The second aim of this subsection is to decompose materials price variation into a between-supplier and a within-supplier component. Between-supplier variation in materials prices reflects differences in marginal costs, across suppliers, whereas within-supplier variation occurs because of differences in the bargaining position–determined by the web of buyer-supplier relationships–between buyers and suppliers.

I assume that manufacturers of the intermediate input, the upstream plants $j \in \{1, ..., J\}$, produce at constant marginal cost, MC_j. The intermediate input produced by the upstream industry plants are not vertically differentiated: the downstream plant will choose the supplier that provides the input at the lowest price. Each downstream plant, i, has access to $\xi \geq 2$ potential suppliers. While the number of potential suppliers is the

[12] As I will show in Section 4, $TFPQ$ is significantly more disperse than $TFPR$. The model that I have presented in this subsection yields an ambiguous prediction on the relative dispersions of $TFPQ$ and $TFPR$:

$$\begin{aligned} Var(\log TFPQ) - Var(\log TFPR) &= -Var(\log P^{out}) - 2Cov\left(\log P^{out}, \log TFPQ\right) \\ &= \frac{3}{4} Var\left(\log TFPQ\right) + \frac{3}{4\gamma} Cov\left(\log TFPQ, \Delta\right) - \frac{1}{4\gamma^2} Var\left(\Delta\right) \\ &\lesseqgtr 0 \end{aligned}$$

same, the identities of these potential suppliers differ across the downstream plants.[13] Let $MC_{i(k)}$ denote the marginal cost of the k^{th} most efficient supplier of plant i. Finally, let $G(\cdot)$ denote the cumulative distribution (with associated probability distribution function, $g(\cdot)$) corresponding to the marginal cost of a randomly drawn supplier.

Suppliers engage in Bertrand competition for the patronage of each downstream plant, i. Given this assumption, each buyer purchases from the lowest-marginal-cost supplier at price $P_i^{in} = MC_{i(2)}$. The p.d.f. of the the materials price, P^{in}, conditional on the marginal cost of the supplier, MC^{up}, equals:[14]

$$h(P^{in}|MC^{up}) = (\xi - 1)\frac{g\left(P^{in}\right)}{1 - G(P^{in})}\left(\frac{1 - G(P^{in})}{1 - G(MC^{up})}\right)^{\xi - 1} \text{ for } P^{in} > MC^{up} \qquad (11)$$

By integrating Equation 11, moments of any function, $f(\cdot)$, of the materials price that i pays, conditional on the marginal cost of the best potential supplier, are given by the following expression:

$$E\left[f\left(P^{in}\right)|MC^{up}\right] = \frac{\xi - 1}{\left(1 - G(MC^{up})\right)^{\xi - 1}}\int_{MC^{up}}^{\infty} f(z)g(z)\left(1 - G(z)\right)^{\xi - 2} dz \qquad (12)$$

According to Equations 11 and 12, the conditional distribution of the materials price shifts to the right as the supplier's marginal cost increases. Combining this finding with the positive relationship between i's marginal cost and its own materials price (Proposition 2, part 3) yields a positive relationship between the marginal costs of establishments that trade with one another. These two findings are stated in the following proposition:

Proposition 3 *Under the same assumptions as in Proposition 2:*

1. *The materials price is increasing, in the sense of first order stochastic dominance, in the marginal cost of the supplier.*

2. *Plants' marginal costs are positively correlated to their suppliers' marginal costs.*

Proof. See Appendix A on page 40. ■

[13]The idea that plants are exogenously restricted in whom they can potentially trade with is taken from de Fontenay and Gans (2007), Abreu and Manea (2011), and Manea (2011). The restriction may have any number of origins. For example, the set of counterparties that each plant can trade with may be the result of a costly search process, as outlined in Chaney (2011).

[14]In the appendix I show that $H(P^{in}|MC^{up}) = 1 - \left(\frac{1-G(P^{in})}{1-G(MC^{up})}\right)^{\xi-1}$. Differentiating this expression, with respect to P^{in}, gives the desired result.

Proposition 3 generates qualitative predictions on the relationships between plants' marginal costs, their materials prices, and their suppliers' marginal costs. Next, I will argue that the predicted price dispersion is substantial, and that both the within-supplier and across-supplier components (which I will define below) are important drivers of the observed variation in materials prices.

For future reference, one can show that the probability distribution functions of the first and second order statistics are:

$$h_1(x) = \xi \cdot g(x) \cdot (1 - G(x))^{\xi - 1} \tag{13}$$

$$h_2(y) = (\xi - 1) \cdot G(y) \cdot g(y) \cdot (1 - G(y))^{\xi - 2} \tag{14}$$

I assume that suppliers' quantity productivities are drawn from the Pareto distribution, with shape parameter, ζ.[15] Then, for $MC \in [0, 1]$, $G(MC) = MC^{\zeta}$ and $g(MC) = \zeta MC^{\zeta - 1}$.[16]

In the left panel of Figure 1, I plot $SD\left(\log P^{in}\right)$, which is computed by integrating over the probability distribution function of prices, as given in Equation 14. For all values of ξ, the variance of observed prices is of the same order of magnitude as the variance of the potential suppliers' marginal costs.

The right panel of Figure 1 displays the fraction of materials price variation that originates from the differences in suppliers' marginal costs.[17] Variation in materials prices, $Var\left[\log\left(P^{in}\right)\right]$, equals the sum of $Var\left[E\left[\log\left(P^{in}\right)\middle| MC^{up}\right]\right]$ and $E\left[Var\left[\log\left(P^{in}\right)\middle| MC^{up}\right]\right]$. $Var\left[E\left[\log\left(P^{in}\right)\middle| MC^{up}\right]\right]$ reflects the variation in materials prices due to the differences in MC^{up} that occur across upstream plants, while $E\left[Var\left[\log\left(P^{in}\right)\middle| MC^{up}\right]\right]$ reflects the variation in prices due to differences in the relative bargaining position of the buyers and sellers of the intermediate good. The takeaway from this figure is that, for all values of ξ, both the within and between-supplier components contribute substantially to the observed variation in materials prices.

In summation, when buyers and suppliers have limited trading opportunities, plants with access to low-marginal-cost suppliers will pay below average prices for their intermediate inputs. This, in turn, generates a positive correlation between the measured productivities of plants that transact with one another. In a similar vein, materials prices will be spatially correlated to the extent that upstream plants' marginal costs can be explained by geogra-

[15] The proceeding results are similar under the assumption of lognormal quantity productivities.

[16] Taking 1 to be the upper limit of the support of potential suppliers' marginal costs is only a normalization. One could re-define the units of the intermediate good so that the highest-possible marginal cost of the supplier is any positive constant.

[17] I plot only one line, because the fraction of materials price variation from the across-supplier component is unchanged in ζ.

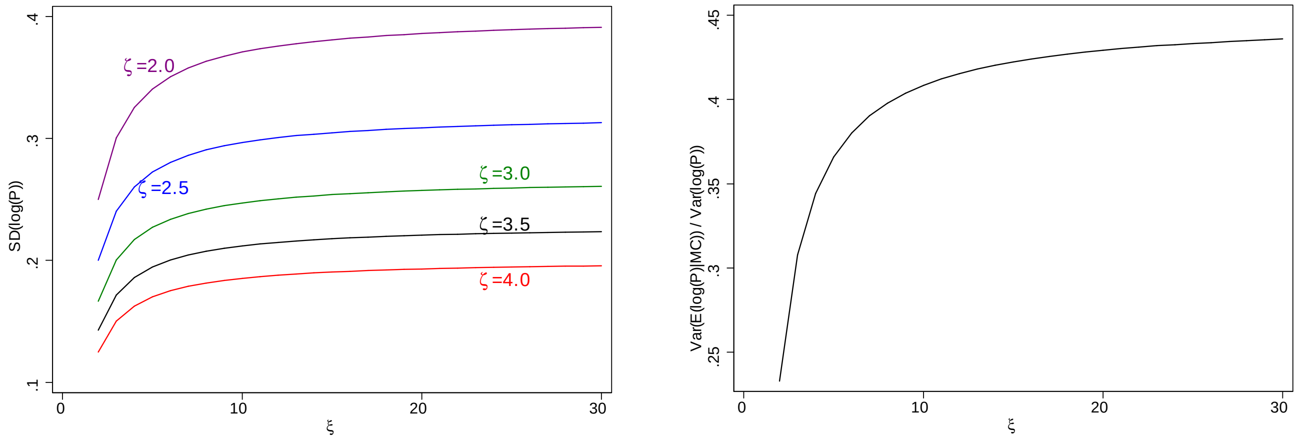

Figure 1: Theoretical results. The left panel gives the standard deviation of $log(P^{in})$, for different combiantions of ζ and ξ. The right panel displays the fraction of materials price variation that originates from the across-supplier component.

phy. While a large fraction of materials price dispersion can be explained by the marginal cost of the supplier, suppliers also price discriminate across their customers, based on the downstream plants' outside options. These findings are verified in Section 5: I show that plants' materials prices are, indeed, related to the marginal costs of their suppliers, and that individual sellers do sell the same output at different prices to different downstream plants. I also find some evidence that suppliers and buyers' marginal costs are correlated.

3 Data and Definitions

3.1 Data sources

For this paper, the main data sources are the Commodity Flow Survey and the Census of Manufacturers. Both datasets are collected and maintained by the U.S. Census Bureau.

The Census of Manufacturers is conducted every five years, in years ending in '2' or '7'. This dataset contains information on manufacturing establishments' productive characteristics, including its employment of production and nonproduction workers, measured in hours; the book value of its building and machine capital; and expenditures on electricity. For certain industries, establishments with five or more employees are asked additional questions about the products that they produce and the materials that they consume. Of particular importance for the current paper, establishments list both the quantity and the value of each of the products they produce (at the 7-digit level) and the quantity and value of each of the materials they consume (at the 6-digit level).[18] For this paper, I use the 1972, 1977, 1982, 1987, 1992, and 1997 versions of the Census of Manufacturers.

The Commodity Flow Survey is necessary to impute buyer-supplier relationships, as I do in Section 5.1. The Commodity Flow Survey was first conducted in 1993, and thereafter in years ending in '2' or '7'. Surveyed establishments are asked to list 25-40 shipments that they make each quarter.[19] Each observation includes information on: the weight and value of the shipment; a five-digit code, specifying the commodity that was shipped; the method of transport (air, truck, rail, courier service, etc...); and the destination zip code. Finally, the Census uses the same plant-level identifier in the Commodity Flow Survey and the Census of Manufacturers. This allows me to combine the information on production and input utilization from the Census of Manufacturers with information on shipment patterns from the Commodity Flow Survey. In Section 5, I employ the 1993 and 1997 Commodity Flow Surveys.

3.2 Sample

In this subsection, I describe the construction of the sample. Further details are provided in Appendix C on page 43.

Similar to Roberts and Supina (1996) and Foster, Haltiwanger, and Syverson (2008), the analysis centers around industries for which outputs and inputs are relatively homogeneous. In industries with heterogeneous inputs or outputs, differences in quality may be a main source of the variation in the prices that different firms charge. I would like, as much as possible, to rule out quality as a source of input or output price variation. An additional restriction is that both the inputs and outputs should be measured in units that are comparable across establishments.[20]

The ten industries[21] that comprise the main sample are packaged milk, white wheat flour, ready-mix concrete, grey cotton yarn, cardboard boxes (with the years 1972-1987 and

[18] To give the reader some idea of the scope of a 7-digit product, ready-mix concrete (3273000) is one of the larger product groups. Some of the smaller product-groups include corrugated boxes, used as containers of chemicals and drugs, including paints, varnishes, cosmetics, and soaps (2653021); or self-rising family white flour (2044126). For 1992, a description of the product codes can be found at http://www.census.gov/prod/2/manmin/mc92-r-1.pdf.

[19] Not all manufacturers are surveyed in the Commodity Flow Survey. In 1993, approximately 60 thousand (out of the 350 thousand existing manufacturing plants) were surveyed, while, in 1997, approximately 30 thousand plants were surveyed.

[20] This second restriction rules out industries like oak, hardwood rough lumber (7-digit product code= 2421163). For this industry, output is measured in units of board feet, but different plants manufacture lumber with different thickness. For this reason, it is difficult to compare different plants' output prices, productivities, or other plant-level characteristics.

[21] When referring to the different subsamples of the benchmark sample, I will use the terms "industry" and "product" interchangeably.

Sample	Units of Output	Material Inputs	N
Milk-Bulk	1000 lbds.	Unprocessed whole milk (88%)	127
White Wheat Flour	50 lbd. sacks	Wheat (90%)	503
Ready-Mix Concrete	1000 cubic yards	Cement (53%), Sand/Gravel (28%)	3708
Carded Cotton Yarn	1000 lbds.	Cotton Fibers (80%), Polyester tow (10%)	431
Boxes, Year≤1987	Short tons	Paper/paperboard (90%)	1820
Gasoline	1000 barrels	Crude petroleum (84%)	692
Ground Coffee	1000 lbds.	Green coffee beans (80%)	300
Raw Cane Sugar	Short tons	Sugar cane (93%)	177
Boxes, Year≥1992	Square feet	Paper/paperboard (89%)	646
Milk-Packaged	1000 quarts	Unprocessed whole milk (72%)	2099
Pooled	-		10503

Table 1: Description of the ten industries in the benchmark sample. The percentages that appear in the Material Inputs column are the fraction of intermediate input expenditures that go to each particular material input. The Material Inputs column shows the inputs that represent greater than 6% of the average plant's total material purchases.

1992-1997 analyzed separately), gasoline, ground coffee, raw cane sugar, and bulk milk; see Table 1.[22,23] In addition to producing a product from one of these ten industries, to be in the benchmark sample, the manufacturers must fill out the materials and production supplements. These supplemental forms, which the Census sends out to larger establishments, are necessary to compute the unit values of manufacturers' outputs and materials purchases.

Thus, there are two sources of sample selection. First, I have chosen industries based on the characteristics of the outputs they produce and the inputs they purchase. These industries tend to use intermediate materials more intensively than the broader manufacturing sector. Since the scope for material price differences to cause measured productivity dispersion increases with the intensity of intermediate input usage (see Equation 22), it is likely that the decline in total factor productivity dispersion is larger for the ten industries in my

[22] A problem similar to the one explained in footnote 20 exists for the post-1992 cardboard box industry. Beginning in 1992, the units of output switch from thousands of pounds to thousands of square feet. I detail my response to this potential problem in Appendix C.

[23] Cardboard boxes, raw cane sugar, gasoline, ground coffee, and ready-mix concrete are included in both the current paper and Foster, Haltiwanger, and Syverson (2008). I could not include carbon black, block ice, or processed ice, as there were insufficiently many plants that filled out both the production and materials supplements. I do not include hardwood flooring or plywood, the last two industries that Foster, Haltiwanger, and Syverson (2008) include; large output price dispersions seem to indicate that the outputs of these industries are not sufficiently homogeneous.

sample than for the broader manufacturing sector.

Second, the plants that are the focus of this study tend to be larger and more productive, compared to the other plants from their respective industries. The average plant in my sample employs roughly 4 times more employees, and has revenues that are 9 times larger, compared to the average plant of their respective industry. (For more details, see Table 14, in Appendix C). Furthermore, since the probability of exit tends to decrease with size, the plants in my benchmark sample are relatively more likely to survive: plants in the benchmark sample have a 5-year survival rate of 87%, compared to the average survival rate for plants in their industries, 70%.

These sample selection issues limit the generalizability of the results given in Sections 4 and 5. However, by sacrificing generality, I am able to isolate the effect of differences in materials prices on intra-industry productivity dispersion.

3.3 Assumptions

I make five assumptions regarding plants' production technologies and the way in which labor, capital, and electricity are supplied. The dual objectives of these assumptions are to a) take the predictions of Section 2's model to the data, and to b) highlight the importance of material price dispersion in the measurement of plant-level productivities. With these goals in mind, I will, as much as possible, adhere to conventional assumptions made in the literature on plant-level production function estimation.

Assumption 1: Plants within an industry have Cobb-Douglas production functions, with labor, capital, electricity, and materials as the inputs. Furthermore, the factor shares are common across all plants within an industry-year combination. Also, plants' production functions exhibit constant returns to scale.

There are three components to the first assumption: a unitary elasticity of substitution, common factor shares within an industry, and constant returns to scale. The unitary elasticity of substitution is common in studies of plants' production functions, mainly for convenience. However, several authors have estimated an elasticity of substitution between labor and capital that is less than 1 (see Raval (2011)). For the objects of interest, the Cobb-Douglas assumption seems to have little effect on the dispersion of measured productivity. When measuring the fraction of productivity dispersion that can be explained by differences in labor quality, Fox and Smeets (2011; Section 5) find that productivity dispersions are remarkably similar, whether the production function is assumed to be Cobb-Douglas or translog.

The other parts of Assumption 1 are also rather innocuous. In Syverson (2004b), the relationships between within-industry productivity dispersion and other industry characteristics is robust to using plant-specific factor shares when estimating plants' TFPs.[24] Related to the constant-returns-to-scale component of Assumption 1, Syverson (2004a) estimates that the returns to scale are indistinguishable from 1 for plants in the ready-mix concrete industry, the industry that contains roughly one-third of the plants in my sample.[25]

Assumption 2: The unit input costs of capital and electricity are the same for all plants within an industry-year combination. In addition, the marginal cost of each of these inputs is constant in the amount purchased.

Data limitations necessitate the first part of Assumption 2. For capital and electricity, it is impossible to observe plant-level input prices. Since I can directly observe the quantity of labor input, I do not need to assume that wages are the same across plants.[26]

Assumptions 1 and 2 are consistent with the set-up of Section 2's model. Assumptions 3-5 deal with the fact that plants may produce multiple outputs and purchase multiple intermediate inputs. These assumptions were not addressed in Section 2, since plants were assumed to produce and purchase only one type of good.

Assumption 3: The fraction of each input employed in producing a particular product equals the plant's share of revenue coming from that product.

The need for Assumption 3, an assumption also made by Foster, Haltiwanger, and Syverson (2008), stems from a limitation of the dataset. In particular, for plants that produce multiple goods, it is impossible to know exactly how much of each input that is used in the production of each output. I make the simplest possible assumption, and assume that each input is allocated in proportion to plant's sales of each product. For example, for a hypothetical plant that employs L units of labor and sells Y_g dollars of good g, for $g \in \{1, ...G\}$, the amount of labor used in the production of good g is

$$L \cdot \frac{Y_g}{\sum_{\hat{g}=1}^{G} Y_{\hat{g}}}. \tag{15}$$

[24] Syverson does find, however, that the measured within-industry differences in TFP are smaller when computed using plant-specific factor shares.

[25] Baily, Hulten, and Campbell (1992) estimate returns to scale for a broader set of industries and find the same result.

[26] Differences in labor quality, across plants, may muddle the interpretation of plants' productivities. Using hours worked as the measure of labor means that plants with exceptionally skilled workers would appear to be highly productive. If workers' wages reflect differences in worker skill (as opposed to, for example, workers' bargaining power), it would be preferable to measure labor inputs by the wages paid by each plant. In an unreported robustness check, I reproduce Tables 2 and 3 using the wage bill, instead of hours worked, as the measure of labor inputs. The results are essentially identical when using this different measure of labor inputs.

Similar to Foster, Haltiwanger, and Syverson (2008), I argue that the dispersion of productivity is robust to the way in which inputs are allocated to outputs, mainly because the plants in my sample tend to be heavily specialized in the goods they manufacture.

In addition to Assumptions 1-3, which are common in papers that estimate plant-level productivities, I make two assumptions on the substitutability among different material inputs. Together, Assumptions 4 and 5 will allow me to compute plant-specific material prices from the data at hand. While restrictive, they are much less so than the common presumption that all plants face the same intermediate input prices.

Assumption 4: If multiple intermediate inputs are observed, the elasticity of substitution between the materials is 0.

This assumption is pertinent only for the two industries, plants producing ready-mix concrete or yarn, for which I observe multiple intermediate inputs being employed. I show, in Appendix D.3 on page 51, that the level of productivity dispersion is extremely robust to moderate levels of substitutability among the different intermediate inputs.

Assumption 5: The elasticity of substitution, between plants' "priced" and "non-priced" materials is 1. In addition, the elasticity of substitution between "non-priced" materials and capital, labor, and materials is also 1.

Here, "priced materials" are the materials that most plants in the industry purchase. For instance, in the case of yarn manufacturers, cotton fibers and polyester tow are the "priced materials." The non-priced materials are purchased by only a few plants in the industry. Again, turning to the yarn industry, approximately 10% of the expenditures on intermediate inputs go to purchases of materials other than cotton fibers (see the 'Material Inputs' column of Table 1). Some of these yarn-producing plants purchase silk fibers; others purchase nylon tow. Since only a few plants purchase these materials, it is difficult to ascertain if plants are purchasing these inputs relatively cheaply or expensively. I treat the "non-priced" materials as if they were any other input for which I do not observe unit prices, such as capital or electricity, and assume that there is a unitary elasticity of substitution between "non-priced" materials and "priced" materials, labor, capital, and electricity.

3.4 Definitions

In this subsection, I define plants' materials and output prices, as well as the three plant-level productivity measures: $TFPQ$, $TFPR$, and Φ. The first two productivity measures are exactly as in Foster, Haltiwanger, and Syverson (2008). The productivity measure that is new to this paper, Φ, aims to isolate plants' abilities to transform inputs into outputs. In particular, this measure, Φ, should not reflect plants' abilities to sell their output at a relatively high price, or to purchase their intermediate inputs at a relatively low price.

I begin by defining plants' input and output prices. The price, P_{ijt}^{out}, that plant i charges for product j in year t is simply the ratio of sales, Y_{ijt}, to physical quantity shipped, Q_{ijt}:

$$P_{ijt}^{out} \equiv \frac{Y_{ijt}}{Q_{ijt}}. \tag{16}$$

Before defining plant-level input prices, I introduce some notation. Let M_{ijt} be the expenditures on materials of plant i in the production of product j in year t. Plant i's purchases consist of "non-priced" materials, which I denote using M_{ijt}^0, and "priced" materials, which I denote using M_{ijt}^1 (and M_{ijt}^2 if j is produced using two material inputs). Let, $s_{\mu jt}$ denote the average fraction–across plants in my sample in industry j and year t–of materials expenditures that are spent on material μ.[27] Finally, let S_{jt} denote the average fraction of materials expenditures, in industry j in year t, that go to "priced" materials.[28]

For plants in industries that use only one type of "priced" material (i.e., all industries except for ready-mix concrete and yarn manufacturing), the input price is given by the ratio of materials expenditures of the first good, M_{ijt}^1, to the physical quantity, N_{ijt}^1, consumed of the first good.

$$P_{ijt}^{in} \equiv \frac{M_{ijt}^1}{N_{ijt}^1} \tag{17}$$

To construct plant-level materials prices for ready-mix concrete and yarn manufacturers, I begin by defining a unit of the intermediate input bundle as follows:

$$\begin{aligned} N_{ijt} &= \min\left\{\frac{s_{1jt}}{S_{jt}}\frac{N_{ijt}^1}{\bar{N}_{ijt}^1}, \frac{s_{2jt}}{S_{jt}}\frac{N_{ijt}^2}{\bar{N}_{ijt}^2}\right\} \\ &= \lim_{\varrho\to 0}\left(\frac{s_{1jt}}{S_{jt}}\left(\frac{N_{ijt}^1}{\bar{N}_{ijt}^1}\right)^{\frac{\varrho-1}{\varrho}} + \frac{s_{2jt}}{S_{jt}}\left(\frac{N_{ijt}^2}{\bar{N}_{ijt}^2}\right)^{\frac{\varrho-1}{\varrho}}\right)^{\frac{\varrho}{\varrho-1}} \end{aligned} \tag{18}$$

In Equation 18, N_{ijt} is the number of units of the intermediate input bundle purchased by plant i in industry j in year t. Because the units of N_{ijt} have no natural interpretation, it is necessary to normalize by the average input utilization of each of the intermediate goods, $\bar{N}_{ij}^1$

[27] For example, for j=concrete and μ =cement, $s_{\mu jt}$ would be approximately 0.53, slightly more in some years, slightly less in others.

[28] Continuing with the example from the previous footnote, S_{jt} would be approximately $0.81 (= 0.28 + 0.53)$ for ready-mix-concrete manufacturers.

and $\bar{N}^2_{ij}$.[29] Assumption 4 pins down how the two different materials are combined to form the composite intermediate input; relaxing Assumption 4 would involve allowing $\varrho > 0$.

Let $\bar{P}^1_{ij}$ and $\bar{P}^2_{ij}$ be the average unit price of materials 1 and 2 in industry j. Then, the materials bundle's ideal price index equals the value-weighted average of the individual inputs' prices:

$$P^{in}_{ij} \equiv \frac{s_{1jt}}{S_{jt}} \frac{P^{in}_{1ijt}}{\bar{P}^{in}_{1jt}} + \frac{s_{2jt}}{S_{jt}} \frac{P^{in}_{2ijt}}{\bar{P}^{in}_{2jt}} \tag{19}$$

Having defined plant-level materials and output prices, I am now able to state how plant-level productivities are computed. For each plant, i, producing in industry j, define its quantity total factor productivity ($TFPQ$) in year t as the ratio between the physical quantity produced and the inputs it utilized in the production of this product.[30]

$$TFPQ_{ijt} \equiv Q_{ijt} \left(L_{ijt}\right)^{-\lambda_{jt}} \left(K_{ijt}\right)^{-\kappa_{jt}} \left(E_{ijt}\right)^{-\epsilon_{jt}} \left(M_{ijt}\right)^{-\sigma_{jt}} \tag{20}$$

In Equation 20, L_{ijt}, K_{ijt}, and E_{ijt} denote the amount of labor, capital, and energy used in the production of product j. From Assumption 3, each input listed in Equation 20 equals the total amount purchased of that input multiplied by the share of plant i's revenues derived from producing its main output. As in Foster, Haltiwanger, and Syverson (2008), labor is stated in terms of hours, and capital is computed by summing plants' reported book values of equipment and structures. The industry-year specific cost shares are computed as in Foster, Haltiwanger, and Syverson (2008). Note that, because of Assumption 1, the factor elasticities, λ_{jt}, κ_{jt}, ϵ_{jt} and σ_{jt}, are the same for all plants within an industry-year pair. In addition, $\lambda_{jt} + \kappa_{jt} + \epsilon_{jt} + \sigma_{jt} = 1$ for all j, t pairs. To emphasize, since $M_{ijt} = P^{in}_{ijt} \cdot N_{ijt}$, low materials prices are a factor associated with high $TFPQ_{ijt}$.

Revenue total factor productivity (TFPR) captures a plant's ability to transform a given bundle of inputs into revenue. As Equation 21 makes clear, plants will have a high TFPR for one of two reasons: Either they have high $TFPQ$, or they sell their output at a

[29] Klump, McAdam, and William (2011) comprises a discussion on the necessity of normalizing CES production functions when $\varrho \neq 1$. (When $\varrho = 1$, the units can be factored out into a multiplicative constant.)

[30] There are many ways to estimate plant-level productivity. I would have liked to compare the estimates generated by Equations 20-22 to those computed using other estimation methodologies. Like Foster, Haltiwanger, and Syverson (2008), I am unable to compute plant-level productivities using the methods outlined in Olley and Pakes (1995), Blundell and Bond (2000), Levinsohn and Petrin (2003), and Ackerberg, Caves, and Frazer (2006). These methods generally require annual observations. Unfortunately information on quantities of output produced or intermediate inputs purchased exist only for years in which the Census of Manufacturers is conducted. Most likely, my results would not change if other productivity measures were used. Van Biesebroeck (2008) reports that, unlike estimates of input elasticities, which are sensitive to the estimation methodology, plant-level productivity estimates are highly correlated across different estimation methodologies.

particularly high price.

$$\begin{aligned} TFPR_{ijt} &\equiv Y_{ijt}\left(L_{ijt}\right)^{-\lambda_{jt}}\left(K_{ijt}\right)^{-\kappa_{jt}}\left(E_{ijt}\right)^{-\epsilon_{jt}}\left(M_{ijt}\right)^{-\sigma_{jt}} \\ &= TFPQ_{ijt}\cdot P_{ijt}^{out} \end{aligned} \tag{21}$$

Finally, plants' technical efficiencies (Φ_{ijt}) purge out the effect of variation in the prices that plants pay for their intermediate inputs:

$$\begin{aligned} \Phi_{ijt} &\equiv Q_{ijt}\cdot\left(L_{ijt}\right)^{-\lambda_{jt}}\cdot\left(K_{ijt}\right)^{-\kappa_{jt}}\cdot\left(E_{ijt}\right)^{-\epsilon_{jt}}\cdot\left(M_{ijt}^{0}\right)^{-\sigma_{jt}\cdot(1-S_{jt})}\cdot\left(N_{ijt}\right)^{\sigma_{jt}\cdot S_{jt}} \\ &= Q_{ijt}\cdot\left(L_{ijt}\right)^{-\lambda_{jt}}\cdot\left(K_{ijt}\right)^{-\kappa_{jt}}\cdot\left(E_{ijt}\right)^{-\epsilon_{jt}}\cdot\left(M_{ijt}\right)^{-\sigma_{jt}}\cdot\left(P_{ijt}^{in}\right)^{\sigma_{jt}\cdot S_{jt}} \\ &= TFPQ_{ijt}\cdot\left(P_{ijt}^{in}\right)^{\sigma_{jt}\cdot S_{jt}} \end{aligned} \tag{22}$$

The equality of the first and second lines of Equation 22 follows from Assumption 5: because of the unitary elasticity of substitution between "priced" and "non-priced" materials, $M_{ijt}=\left(M_{ijt}^{0}\right)^{(1-S_{jt})}\cdot\left(M_{ijt}^{1}+M_{ijt}^{2}\right)^{S_{jt}}$. The equality of the second and third lines follows from the definition of $TFPQ$. Equation 22 states that plants will have high $TFPQ_{ijt}$ for one of two reasons: either the plant is technically efficient (Φ_{ijt} is large), or intermediate input prices are low (P_{ijt}^{in} is low).[31]

So that I can compare observations across industries and years, all quantities will be stated relative to the mean for that industry×year. I use lower-case letters to denote the logged, de-meaned values. For any plant-level statistic, X_{ijt}, define:

$$x_{ijt}\equiv\log\left(X_{ijt}\right)-\frac{\sum_{k:k\in\text{i's industry in year t}}\log X_{kjt}}{\left\|\{k:k\in\text{i's industry in year t}\}\right\|} \tag{23}$$

4 Implications of Materials Price Dispersion

In this section, I explore some of the implications of price dispersion in intermediate inputs markets. Correlations among plant-level statistics are given in Section 4.1. In Section 4.2, I estimate that 7 to 16 per cent of the variation in $tfpq$ is attributable to differences in the materials prices that plants face. In Section 4.3, I argue that the price that plants face when purchasing their materials is persistent across time. Then, in Section 4.4, I document that materials prices are slightly higher for entrants (compared to incumbents)

[31] Of course, there may be within-industry differences in the factor market conditions for labor, capital, and electricity. Because of Assumption 2, these differences would be incorrectly labeled as differences in "technical efficiencies."

and substantially higher for exiting plants (compared to survivors).

4.1 Correlations

I begin by computing the correlations among different plant-level statistics. All plant-level variables are de-meaned by industry-year according to Equation 23. Correlations are weighted by plants' real sales.[32] The correlations, which are collected in Table 2, match the predictions made in parts 1-6 of Proposition 2.

The correlation coefficients between $tfpq$, $tfpr$, and p^{out} are similar to those computed in Foster, Haltiwanger, and Syverson (2008). Plants with higher $tfpq$ have a lower marginal cost and pass on some of this lower marginal cost to their consumers (generating a low p^{out}). In addition, $tfpq$ and $tfpr$ are highly correlated, as are $tfpr$ and p^{out}. Furthermore, the standard deviation of $tfpr$ is less than that of $tfpq$: the distribution of revenue-based productivity is more compressed than that of quantity-based productivity. The differences in the dispersions of the two productivity measures are related to the negative correlation between $tfpq$ and p^{out}.[33]

The variables that are new to this study are ϕ and p^{in}, the logged technical efficiencies and materials prices. First, plant-level materials prices, p^{in}, are negatively correlated with $tfpq$. Plants that purchase inputs cheaply appear to be more productive. At the same time, $tfpq$ and ϕ are highly correlated with one another, while the correlation between ϕ and $tfpr$ is similar to the correlation between $tfpr$ and $tfpq$. These last correlations should bring some solace to researchers who do not have access to plant-level materials and output prices, and who wish to use $tfpr$ as a proxy for ϕ.

Finally, plant-level input and output prices are correlated with one another.[34] There are several possible explanations for this positive relationship. First, the positive correlation may reflect that some differences in input and output quality still remain (despite my best efforts to choose a sample of industries with outputs and material inputs that are comparable across plants). Independent of quality differences, the positive correlation between input and output prices may be due to imperfections in output or factor markets, as Proposition 2 would predict. A third possible explanation is that a selection mechanism, one on plant survival, may be causing us to observe a positive relationship between p^{out} and p^{in} (see

[32] Throughout this section, real revenues are computed using 4-digit-industry output price indices from the NBER Productivity Database. Unweighted calculations are presented in Appendix D.5.

[33] Since $tfpr = tfpq + p^{out}$, a negative correlation between $tfpq$ and p^{out} is a necessary, but not sufficient, condition for the standard deviation of $tfpr$ to be less than that of $tfpq$. A necessary and sufficient condition is that $SD(tfpr) < SD(tfpq) \iff 2\rho(tfpq, p^{out}) \cdot SD(tfpq) + SD(p^{out}) < 0$.

[34] Eslava et al. (2004) document this empirical regularity, as well, using a broad sample of Colombian manufacturing establishments.

	p^{in}	p^{out}	$tfpq$	ϕ	$tfpr$
p^{out}	0.278				
$tfpq$	-0.303	-0.653			
ϕ	0.141	-0.549	0.899		
$tfpr$	-0.098	0.212	0.601	0.581	
Std. Dev.	0.167	0.186	0.227	0.219	0.176
Skewness	0.497	0.169	0.279	0.446	0.848

Table 2: Correlations and sample statistics of plant-level characteristics, pooled across all products. All variables are de-meaned by year and product. N=10,503. Note: In the current draft, the correlations, standard deviations, and skewness coefficients are computed using the unweighted distribution. In a future draft, I will weight observations by plants' revenues, to be consistent with the computations that are given in Table 3.

Equations 8 and 9).

Correlations for each of the ten industries are presented in Appendix D.4. It seems that, within the benchmark sample, the sugar industry is something of an outlier: for the 177 plant-year observations in the industry, the correlation between p^{in} and $tfpq$ is positive. Except for sugar, the range of correlations between p^{in} and $tfpq$ lies between -28% (packaged milk) and -43% (post-1992 cardboard boxes).

4.2 Implications for productivity dispersion

In this subsection, I compare the dispersions of the distributions of $tfpq$ and ϕ. In so doing, I provide a measure of the fraction of $tfpq$ dispersion that can be explained by differences in intermediate input prices. The main finding, that dispersion in $tfpq$ is greater than the dispersion of ϕ, is the prediction of part 7 of Proposition 2.

Pooling across the 10 industries in the sample, the standard deviation of ϕ is 15.8% ($= e^{0.147}$), while the standard deviation of $tfpq$ is 17.0%, 7% larger than the standard deviation of ϕ. So, by purging out the effect of differences in materials prices, the observed distribution of productivities would be approximately 7% lower.

I also give, in Table 3, two other measures of dispersion, the 90/10 ratio and the 75/25 ratio. The difference between the dispersion of $tfpq$ and the dispersion of ϕ is even greater with these two alternate measures: 15% for the 90/10 ratio, and 16% for the 75/25 ratio.

Again, even though I have chosen industries based on the homogeneity of the inputs and outputs, it is likely that at least some of the variation in materials and output prices is due to differences in quality. Variation in input/output quality attenuates the negative correlation between $tfpq$ and p^{in} (see Appendices D.1 and D.2.) As a result, within-industry

Sample	Dispersion of $tfpq$ 90/10	75/25	SD	Dispersion of ϕ 90/10	75/25	SD	Percent Decline 90/10	75/25	SD	N
Milk-Bulk	0.608	0.302	0.256	0.522	0.266	0.227	18.0%	14.5%	13.5%	127
Flour	0.355	0.187	0.141	0.341	0.156	0.147	4.0%	22.2%	-3.7%	503
Concrete	0.517	0.250	0.222	0.478	0.233	0.213	8.5%	7.4%	4.5%	3708
Yarn	0.582	0.272	0.259	0.647	0.314	0.257	-9.6%	-12.5%	1.0%	431
Boxes, Year$\leq$1987	0.377	0.174	0.167	0.361	0.174	0.165	4.7%	0.1%	1.2%	1820
Gasoline	0.289	0.139	0.128	0.265	0.124	0.119	9.5%	13.4%	8.2%	692
Coffee	0.760	0.361	0.279	0.594	0.301	0.238	32.1%	22.2%	19.0%	300
Sugar	0.604	0.304	0.287	0.793	0.349	0.325	-21.1%	-12.1%	-11.0%	177
Boxes, Year$\geq$1992	0.569	0.294	0.225	0.528	0.277	0.212	8.2%	6.4%	6.6%	646
Milk-Packaged	0.529	0.259	0.230	0.501	0.244	0.221	5.9%	6.2%	4.1%	2099
Pooled	0.341	0.157	0.157	0.299	0.136	0.147	15.1%	16.3%	7.1%	10503

Table 3: Dispersion of $tfpq$ and ϕ. Observations are weighted by plant revenue. See Table 23 for the unweighted computations. Due to Census' rules regarding data confidentiality, I am prohibited from reporting the actual quantiles of any empirical distribution (i.e., output prices). The quantiles (but not the standard deviations) are computed in a two-step process. First, using a kernel density estimator, I produce a smoothed version of the empirical cumulative distribution function of the variable of interest. I then report the quantile of this smoothed distribution. The decrease in productivity dispersion—between $tfpq$ and ϕ—is not substantially affected by this smoothing procedure.

variation in quality will lead to a downward bias in the measured difference between the dispersion of $tfpq$ and the dispersion of ϕ. In other words, the 7% to 16% decline in dispersion most likely underrepresents the actual fraction of $tfpq$ dispersion that is due to differences in materials prices.

To provide some context, I relate the 7% to 16% decline in dispersion to dispersion declines reported in two other papers. First, Syverson (2004a) hypothesizes that, in markets for which competitive forces are exceptionally strong, low-productivity plants are relatively more apt to exit the industry, thus leading to a more compressed productivity distribution. Within the ready-mix concrete industry, Syverson characterizes areas with high densities of construction activity as highly competitive markets. Compared to low-density markets, the interquartile range of total factor productivity is 2 to 3% smaller in high-density markets. In a second example, Fox and Smeets (2011) compute the fraction of measured productivity dispersion that can be explained by differences in worker quality. For a sample of Danish firms in eight industries, the 90/10 ratio of the distribution of TFP declines by approximately 20% after including controls for worker quality and the wage bill.

	Revenue-weighted?	$tfpq$	ϕ	$tfpr$	y	p^{in}	p^{out}
β	Yes	0.351	0.299	0.306	0.895	0.309	0.432
$s.e.$	Yes	(0.022)	(0.023)	(0.024)	(0.010)	(0.025)	(0.024)
$\beta^{1/5}$	Yes	0.811	0.786	0.789	0.978	0.791	0.846
β	No	0.210	0.216	0.239	0.878	0.336	0.308
$s.e.$	No	(0.076)	(0.068)	(0.081)	(0.030)	(0.048)	(0.045)
$\beta^{1/5}$	No	0.732	0.736	0.751	0.974	0.804	0.790

Table 4: Coefficient estimates and robust standard errors of the regression defined by Equation 24.

4.3 Persistence

A long stream of research, beginning with Baily, Hulten, and Campbell (1992), has documented the persistence of plant level characteristics. Using regressions of the form,

$$x_{i,j,t+5} = \alpha + \beta \cdot x_{ijt} \ + \varepsilon_{ijt}, \tag{24}$$

Foster, Haltiwanger, and Syverson (2008) compute the 1 and 5-year autocorrelation coefficients for different plant-level statistics, x. They compute that plant-level productivities and output prices have a 1-year autocorrelation coefficient of approximately 80%, while plant revenues are much more persistent. I replicate these findings in Table 4. The novel components of Table 4 appear in the second, fourth, and penultimate columns, where I compute the autocorrelation of ϕ, y, and p^{in}. I find that the persistence of ϕ is similar to that of the two other plant-level productivity measures, and that the persistence of p^{in} not statistically different from the persistence of p^{out}. By far, the most persistent plant-level characteristic is log revenues, y: the 1-year autocorrelation coefficient of approximately 98%. For the most part, the estimated autocorrelation coefficients are larger when observations are weighted by revenues, indicating that plant-level characteristics are more persistent for larger plants.

4.4 Characteristics of entering and exiting plants

Along key dimensions, plants that are about to exit are different from surviving plants, while plants that are entering an industry are different from incumbents. Several studies have documented a negative relationship between productivity (or size) and survival. Moreover, while entrants are smaller than the average plant in their industry, they are not less productive.

I replicate these results in Table 5, using regressions defined by Equations 25 and

Coefficient on:	Fixed Effects	$tfpq$	ϕ	$tfpr$	y	p^{in}	p^{out}
Entry	Year×Product	0.015	0.009	0.004	-0.732*	-0.011	-0.011
Entry	Year×Product	(0.024)	(0.021)	(0.019)	(0.103)	(0.018)	(0.018)
Entry	Year×Product×Division	0.008	0.002	-0.012	-0.751*	-0.009	-0.020
Entry	Year×Product×Division	(0.022)	(0.018)	(0.018)	(0.108)	(0.017)	(0.016)
Exit	Year×Product	-0.058*	-0.043*	-0.041	-0.599*	0.020	0.017
Exit	Year×Product	(0.019)	(0.018)	(0.024)	(0.193)	(0.013)	(0.014)
Exit	Year×Product×Division	-0.058*	-0.045*	-0.042	-0.600*	0.017	0.016
Exit	Year×Product×Division	(0.020)	(0.020)	(0.023)	(0.189)	(0.013)	(0.014)

Table 5: Comparison of plant-level statistics and entry/exit status. In the first four rows, each cell gives the coefficient estimate, or standard error, of β_1 in Equation 25. In the final four rows, each cell gives the coefficient estiamte, or standard error, of β_2 in Equation 26. In each regression, observations are revenue-weighted. A division is a Census-defined region, of which there are 9 in the United States. Stars denote significance at the 5% level. N=10,503.

26. Like Foster, Haltiwanger, and Syverson (2008), I find that entrants have above-average $tfpq$ and $tfpr$, but below-average revenues. These authors also estimate that, compared to other plants in the industry-year, exiting plants have lower $tfpq$ and $tfpr$. In addition to these already-known empirical regularities, I compute, in the second-to-last column of Table 5, that exiting plants pay, approximately 1.5% more per unit of the intermediate input than the average plant in their industry-year. Finally, the technical efficiency, ϕ, is 2% higher for entrants (compared to incumbents), and 1.5% lower for exiting plants (compared to survivors).

$$x_{ijt} = \alpha_{jt} + \beta_1 \mathbb{I}\left\{i \in \text{Entrant at year } t\right\} + \varepsilon_{ijt} \tag{25}$$

$$x_{ijt} = \alpha_{jt} + \beta_2 \mathbb{I}\left\{i \in \text{Exit before } t+5\right\} + \varepsilon_{ijt} \tag{26}$$

Interactions among different selection mechanisms are explored in Table 6. This table presents the results of a logit regression, in which the dependent variable equals 1 if the plant exits before the subsequent Census of Manufacturers. Besides industry-year fixed effects, the regressions include characteristics that affect plants' probabilities of survival. The effects of increasing measured productivity or materials prices are considered, in isolation, in the first four columns of Table 6. A one standard deviation increase in $tfpq$, $tfpr$, and ϕ is associated with a 1.2%, 1.2%, or 0.8% decline in the probability of exit, while a one standard deviation increase in p^{in} is associated with a 1.1% increase in the probability of exit. Including plant-level input prices, as I do in the fifth and sixth columns, attenuates the estimates of the $tfpr$ and $tfpq$ terms, since some component of the selection-on-productivity mechanism is now being captured by the p^{in} variable. Note that no such attenuation occurs

$tfpr$	-0.594*				-0.552*		
	(0.179)				(0.180)		
$tfpq$		-0.458*				-0.380*	
		(0.133)				(0.139)	
ϕ			-0.315*				-0.369*
			(0.138)				(0.139)
p^{in}				0.530*	0.472*	0.375*	0.594*
				(0.180)	(0.182)	(0.189)	(0.181)

Table 6: Coefficient estimates and standard errors (clustered at the plant level) from a logit regression: the depedent variable equals 1 provided the plant exits before the following Census of Manufacturers. Stars denote significance at the 5% level. N=10,503.

between the third/fourth and seventh columns.

5 Sources of Materials Price Dispersion

I discuss three explanations for the observed within-industry dispersion of intermediate input prices. The sources of materials price dispersion have implications for the social benefits provided by each plant. Plants that pay low materials prices by taking advantage of monopsonistic power are not providing any societal benefit: low materials prices are a transfer of profits from supplier to buyer. On the other hand, if plants pay low materials prices because their suppliers are exceptionally productive, low materials prices represent a positive impact on social welfare.

To calculate the relative importance of these different sources of materials price dispersion, I need to impute, for each manufacturer, the identities of the its suppliers. I outline, in Section 5.1, the algorithm that I use to impute buyer-supplier relationships. In Section 5.2, I compute the fraction of dispersion in $tfpq$ and p^{in} that can be explained by plants' geographic location and their suppliers' marginal costs. A positive correlation between plants' marginal costs and their suppliers' marginal costs stimulates the following question: If plants with low-marginal cost suppliers are more productive, and having low marginal costs is so advantageous, what prevents plants from purchasing their materials inputs from the low marginal cost suppliers? In Section 5.3, I argue that buyer-supplier relationships are persistent, suggesting that there is some inertial force that inhibits all plants from switching to low marginal cost suppliers. I close, in Section 5.4, by showing that suppliers charge different prices to different customers, *prima facie* evidence of price discriminatory behavior. I calculate the fraction of materials price dispersion that can be explained by this price discriminatory behavior.

5.1 Imputation of buyer-supplier relationships

To impute buyer-supplier relationships, I use the algorithm introduced in Atalay, Hortaçsu and Syverson (2012). The algorithm generates a list of establishments that could potentially receive any shipment that is observed in the Commodity Flow Survey. Consider a hypothetical shipment of commodity, c, made by establishment, h, to zip code, z.[35] The establishments, i, that could potentially receive this shipment are those who are located in zip code, z, and are members of an industry that use commodity, c. For example, the potential recipients of a shipment of Portland cement to z would be all plants in that zip code who were engaged in ready-mix concrete manufacturing, concrete block manufacturing, and wholesalers of brick, stone and related materials. If there are multiple potential recipients of the shipment, and one of these establishments is owned by the same firm as the sending establishment, then I presume that the shipment was received by the same-firm establishment.[36] Otherwise, I assign each potential recipient, i, to be downstream of plant h.[37]

So that I can compare suppliers' and buyers' productivities, I require both the upstream and downstream industries to be from the manufacturing sector. Of the ten industries in the benchmark sample, only two–ready-mix concrete and cardboard boxes–have a main input that is produced by a manufacturer. For Portland cement producers, I look for shipments in the Commodity Flow Survey for which the commodity code is that of cement.[38] The industries with establishments that could potentially receive Portland cement (STCC=32411) are concrete brick and block manufacturers (SIC=3271), road construction firms (SIC=1610-1619), ready-mix concrete manufacturers (SIC=3273), and wholesalers of brick, stone and related materials (SIC=5032). For paper and paperboard manufacturers, I look for shipments in the Commodity Flow Survey for which the commodity code is that of

[35] The commodity code used in the 1993 Commodity Flow Survey is the Standard Transportation Commodity Code (STCC). A list of STCC codes can be found in pages 117 to 167 of "Reference Guide for the 2008 Surface Transportation Board Carload Waybill Sample," published by Railinc. Since 1997, the Commodity Flow Survey has used the Standard Classification of Transported Goods (SCTG) classification of commodity codes. Documentation related to SCTG codes can be found on the Census website.

[36] Atalay, Hortaçsu and Syverson (2012) make the same assumption. This assumption is motivated by the finding that establishment h is much more likely to ship to zip codes that contain an establishment from the same firm.

[37] By assigning all potential recipients, i, to be downstream of plant h, I am undoubtedly overcounting the number of buyer-supplier relationships. In Appendix D.6, I reproduce the analysis of Section 5.2 on a restricted sample. In the restricted sample, I include only plants, i, for which there is only 1 other potential recipient in zip code, z, that could receive h's shipment.

Alternatively, one could weight observations by the inverse of the number of potential recipients in the destination zip code. It turns out that doing so leaves the results of Section 5 unchanged.

[38] Productivity data for ready-mix-concrete manufacturers are unavailable in 1997, so, for cement and concrete manufacturers, I only look at buyer-supplier relationships in the 1993 Commodity Flow Survey.

paperboard (STCC=26311 in 1993, SCTG=27319-27320 in 1997). The industries with establishments that could potentially receive these shipments of paperboard are manufacturers of corrugated and solid fiber boxes (SIC=2653) and folding paperboard boxes (SIC=2657).

For within-firm shipments, surveyed establishments do not report the actual market value of the transaction. Instead, the establishments are asked to estimate what the value of the shipment would have been if it had been sold to some other firm. Since it is unclear what these values actually represent, I remove downstream establishments who receive a substantial fraction, 15% or more, of the relevant input from other plants from the same firm.[39]

Mismeasurement of buyer-supplier relationships will primarily affect the estimated strength of the relationship between buyers' and suppliers' productivities. In other exercises, I a) compare downstream plants' materials prices and upstream plants' productivities, and b) examine the persistence of buyer-supplier relationships. These exercises rely only on information about the sending establishments, and thus are not as sensitive to mismeasurement of buyer-supplier relationships.

5.2 Materials prices and supplier marginal costs.

I document the following two relationships. First, ready-mix concrete and cardboard box manufacturers that have exceptionally productive suppliers purchase their cement or paperboard at lower than average prices. This finding, along with the negative relationship between a plant's input prices and its $tfpq$ (which is documented in Section 4.2), combines to generate the prediction of a second relationship: the $tfpq$ of a plant and the $tfpq$ of its suppliers should be positively correlated. These relationships match the predictions contained in Proposition 3.

I begin with some notation. Let χ_{hit} denote the total mass (in thousands of pounds) of shipments sent by plant h to plant i in year t, and let ω_{hit} denote the total value (in thousands of real dollars) of shipments sent by plant h to plant i in year t. Then, the

[39] While varying the 15% cutoff down to 0% or up to 25% does not affect this section's results, the relationship between input prices and supplier productivity begins to disappear once the cutoff exceeds 25 or 30%.

Bernard, Jensen, and Schott (2006) show that reported prices on cross-border shipments, for which the sender and receiver are part of the same firm, are manipulated to take advantage of the different tax policies of the destination and source countries. Even though such an incentive to mis-report does not exist in the Commodity Flow Survey data, I argue that one should not put too much weight on input prices of the plants that buy a substantial fraction of their inputs from within the firm.

(f.o.b.)[40] price that plant h charges plant i is simply the ratio of the value to the price:

$$P_{hit}^{CFS} = \frac{\omega_{hit}}{\chi_{hit}} \tag{27}$$

The "CFS" superscript will denote prices computed using the Commodity Flow Survey data (as opposed to the prices that are computed in Section 4, using data from the Census of Manufacturers.)[41]

A supplier's average output price equals a value-weighted average of the prices that it charges to the plants that we observe it supplying to:

$$P_{ht}^{out,CFS} = \frac{\sum_{i:h \to i} \omega_{hit} P_{hit}^{CFS}}{\sum_{i:h \to i} \omega_{hit}} \tag{28}$$

Plants' average input prices are defined similarly. For each downstream plant, i, I take the value-weighted average over all plants, h, that I observe i purchasing from:

$$P_{it}^{in,CFS} = \frac{\sum_{h:h \to i} \omega_{hit} P_{hit}^{CFS}}{\sum_{h:h \to i} \omega_{hit}} \tag{29}$$

Note that, because it does not include freight charges, $P_{it}^{in,CFS}$ will be less than what plants pay for their intermediate inputs. With this in mind, I define a second plant-level input price, that includes freight charges:

$$\tilde{P}_{it}^{in,CFS} = \frac{\sum_{h:h \to i} \omega_{hit} \left(P_{hit}^{CFS} + \tau_{hit}\right)}{\sum_{h:h \to i} \omega_{hit}} \tag{30}$$

I estimate transportation costs, τ_{hi}, from the mileage of the shipment and the mode of transport.[42,43]

[40] F.o.b. stands for free on board (or freight on board). Unlike the c.i.f. (cost, insurance, and freight) price, the f.o.b. price does not include freight or insurance charges.

[41] The Commodity Flow Survey has, up to now, been an unexploited source of data on plants' output prices. With this in mind, I compare plants output prices, derived from the Commodity Flow Survey to the prices derived from the better-known Census of Manufacturers. For the 53 cement manufacturers in this section's sample, the correlation between $p_h^{out,CFS}$ and p_h^{out} is 21%. For the 60 paperboard manufacturers in the sample, the correlation between the two plant-level output prices is 41%.

[42] The Bureau of Transportation Statistics collects information on ton-mile freight charges for shipments sent along different transport modes. Since the Commodity Flow Survey contains information on the weight of each shipment, as well as the distance that the shipment traveled, it is straightforward to estimate the shipment freight charge.

See http://www.bts.gov/publications/national_transportation_statistics/html/table_03_21.html for the data on freight rates.

[43] For the corrugated-box manufacturing industry, I relate $\tilde{p}_{it}^{in,CFS}$ and p_{it}^{in}. (Remember that p_{it}^{in} cannot be computed in 1992 or 1997 for ready-mix-concrete manufacturers.) The strength of this relationship, between the materials prices computed from the two data sources, indicates the success (or lack thereof)

Sample	Boxes	Concrete	Pooled	Boxes	Concrete	Pooled
$\overline{tfpq}_{it}$	-0.257*	-0.196*	-0.245*	-0.233*	-0.141	-0.214*
	(0.059)	(0.092)	(0.051)	(0.059)	(0.086)	(0.059)
N	190	131	321	190	131	321
Adjusted R^2	0.111	0.039	0.093	0.193	0.428	0.224
Division F.E.?	No	No	No	Yes	Yes	Yes

Table 7: Coefficient estimates and robust standard errors, from the regressions defined by Equation 32. Stars indicate significance at the 5% level.

Finally, for any concrete or cardboard box manufacturer, i, that is identified by the algorithm outlined in Section 5.1, I compute average supplier productivity, $\overline{TFPQ}_{it}$ as follows:

$$\overline{TFPQ}_{it} \equiv \frac{\sum_{h:h\to i} \omega_{hit} TFPQ_{ht}}{\sum_{h:h\to i} \omega_{hit}} \tag{31}$$

Similar to the analysis of Section 4, all plant-level statistics are logged and stated relative to the average value for the industry-year. Again, the logged, de-meaned variables are written using lower-case letters.

With these definitions in hand, I am now able to compare the price that a plant pays for its material inputs to the average productivity of the plant's suppliers.[44]

$$\tilde{p}_{it}^{in,CFS} = \beta_{\text{division}} + \beta_1 \cdot \overline{tfpq}_{it} + \varepsilon_{it} \tag{32}$$

The results are presented in Table 7. A 10 percent increase in the marginal cost of plants' suppliers corresponds to a 2 to 2.5 percent increase in plants' materials prices. The estimated effect of supplier productivity on materials prices is stronger for boxes than it is for ready-mix concrete. (This difference is not statistically significant.) Including fixed effects for the geographic region of the downstream plant slightly decreases the estimate of β_1: one reason why intermediate input prices are spatially correlated is that the plants within a particular region are likely to have the same suppliers. (An alternate explanation is that marginal costs of plants in the supplying industry are spatially correlated.)

Each cell in Table 8 presents the unexplained variation–measured as the (revenue-weighted) standard deviation of the residuals, when $\tilde{p}_{it}^{in,CFS}$ is regressed on different com-

of the imputation procedure outlined in Section 5.1. The correlation between $\tilde{p}_{it}^{in,CFS}$ and p_{it}^{in} is 25%, meaning that I am mismeasuring many buyer-supplier relationships, but that the imputation algorithm yields a workable dataset.

[44] Using $p_{it}^{in,CFS}$, instead of $\tilde{p}_{it}^{in,CFS}$, as the dependent variable of Regression 32 generates a similar estimate of β_1.

Sample	Include Division Fixed Effects ?	No	Yes	No	Yes
	Include $\overline{tfpq_{it}}$?	No	No	Yes	Yes
	Sample Size				
Boxes	190	0.202	0.197	0.190*	0.187*
Concrete	131	0.374	0.351	0.365	0.345
Pooled	321	0.224	0.220	0.213*	0.210*

Table 8: Each cell gives the real-revenue-weighted standard deviation of the residuals in a particular regression; the full specification is given in Equation 32. Across the columns of the table, different combinations of independent variables are included in the regressions. Across the rows of the table, different samples are used. Stars indicate that the decline in dispersion is significantly more than the decline that would occur from simply including "fake" random variables on the right hand side of Equation 32. See Appendix E for details.

binations of the right-hand side variables of Equation 32. Comparing the first and second columns of Table 8, I calculate that approximately 2% $\left(= \frac{0.224-0.220}{0.224}\right)$ of the dispersion in input prices can be explained by division fixed effects. Suppliers' productivities explain approximately 5% $\left(= \frac{0.224-0.213}{0.224}\right)$ of the dispersion in input prices, while the two sets of variables jointly explain 6% of the dispersion in materials prices. Stars in Table 8 indicate that the decline in dispersion is significantly more than what would be expected by including randomly-generated regressors.[45]

A prediction that emerges from two previously documented relationships–the negative relationship between p^{out} and $tfpq$, and the negative relationship between $\tilde{p}_{it}^{in,CFS}$ and $\overline{tfpq_{it}}$–is that plant level $tfpq$ should be positively correlated to the $tfpq$ of the plants' suppliers. I test this prediction by running the regression specified by Equation 33.

$$tfpq_{it} = \beta_{\text{division}} + \beta_1 \cdot \overline{tfpq_{it}} + \varepsilon_{it} \tag{33}$$

The results of this regression, which are displayed in Table 9, confirm that there is, indeed, a positive relationship between plants' marginal costs and the marginal costs of plants' suppliers. The relationship between $tfpq_{ti}$ and $\overline{tfpq}_{it}$ is weaker, however, than the relationship between $tfpq_i$ and $\tilde{p}_{it}^{in,CFS}$ that was estimated in Table 7.

Table 10 shows the unexplained variation of $tfpq_{it}$ that remains after regressing on different combinations of explanatory variables. Only 2% of the variation in $tfpq_{it}$ can

[45] Any set of variables–for example, a random variable drawn from a standard normal distribution, or a set of 12 dummy variables that sum up to 1–will necessarily explain some positive fraction of the variation in $\tilde{p}_i^{in,CFS}$. In Appendix E on page 56, I test whether the decline in dispersion is significantly greater than what would be expected from including different combinations of "fake" random variables on the right hand side of Equation 32.

Sample	Boxes	Concrete	Pooled	Boxes	Concrete	Pooled
$\overline{tfpq_{it}}$	0.078	0.130*	0.089	0.088	0.136*	0.097*
	(0.060)	(0.045)	(0.049)	(0.060)	(0.046)	(0.049)
N	190	131	321	190	131	321
Adjusted R^2	0.004	0.057	0.010	0.009	0.064	0.018
Division F.E.?	No	No	No	Yes	Yes	Yes

Table 9: Coefficient estimates and robust standard errors of the regressions defined by Equation 33. Stars indicate significance at the 5% level.

	Include Division Fixed Effects ?	No	Yes	No	Yes
	Include $\overline{tfpq_{it}}$?	No	No	Yes	Yes
Sample	Sample Size				
Boxes	190	0.219	0.211	0.218	0.210
Concrete	131	0.210	0.203	0.203*	0.197
Pooled	321	0.218	0.211	0.217	0.211

Table 10: Each cell gives the real-revenue-weighted standard deviation of the residuals in a particular regression; the full specification is given in Equation 33. Across the columns of the table, different combinations of independent variables are included in the regressions. Across the rows of the table, different samples are used. Stars indicate that the decline in dispersion is significantly more than the decline that would occur from simply including "fake" random variables on the right hand side of Equation 33. See Appendix E for details.

be accounted for, using division fixed effects and suppliers marginal costs. Moreover, the decline in dispersion does not seem, in most cases, to be significantly more than would occur from including randomly generated covariates.

5.3 Persistence of relationships

Buyer-supplier relationships are persistent across time. The estimated persistence of buyer-supplier relationships suggests that there is some force that inhibits intermediate inputs purchasers from switching suppliers. Whether this inhibiting force reflects some extra profitability that is conferred by repeated interaction, or some idiosyncratic match-specific productivity, it will prevent all buyers from switching to the lowest-cost intermediate goods suppliers.

To provide some empirical evidence for the persistence of buyer-supplier relationships, I explore the shipments sent by cement and paperboard manufacturers in the 1993 and 1997 Commodity Flow Surveys. Again, from the Commodity Flow Survey, I cannot observe the identity of the downstream buyer. Instead, I proxy for the identity of the down-

Sample	Cement	Cement	Cement	Paperboard	Paperboard	Paperboard
Did the plant sell to		2.075	2.009		2.988	2.661
the zip code in 1993?		(0.105)	(0.105)		(0.067)	(0.069)
Log Mileage	-2.835	-2.594	-2.593	-1.025	-0.867	-0.881
	(0.050)	(0.050)	(0.051)	(0.024)	0.026	(0.026)
N	106795	106795	106795	75360	75360	75360
Number of zip codes	2015	2015	2015	1256	1256	1256
Number of plants	53	53	53	60	60	60
Adjusted R^2	0.687	0.713	0.718	0.148	0.257	0.290
Unconditional probability of entry	0.021	0.021	0.021	0.030	0.030	0.030
Include control for firm presence in z?	No	No	Yes	No	No	Yes

Table 11: Coefficient estimates and standard errors, from the regression defined by Equation 35. The sample is comprised of cement and paperboard plants that were included in the sample of Regressions 32, 33, and 38. For a zip code to be in the sample, at least one plant must have shipped to the zip code.[47]

stream buyer using the destination zip code. I run a conditional logit regression, described by Equation 35; the dependent variable equals 1 if the cement/paperboard plant, i, ships to zip code, z, in 1997. The explanatory variable of interest is an indicator, which equals 1 if the plant shipped to the zip code in 1993. Destination zip-code level fixed effects, supplier fixed effects, and the log distance between plant i and zip code z are additional explanatory variables.[46]

$$\begin{aligned} \mathbb{I}\{i \to z \text{ in } 1997\} = \beta_z + \beta_i + \beta_1 \log(\text{distance } i \to z) + \beta_2 \mathbb{I}\{i \to z \text{ in } 1993\} \\ + \beta_3 \mathbb{I}\{\text{plant of i's firm is located in } z \text{ in } 1997\} + \varepsilon_{iz} \end{aligned} \tag{35}$$

The results are presented in Table 11. Both cement and paperboard suppliers' decisions on which destinations to ship to are persistent across time. If plant i sells to zip code z in 1993, the probability that i will zip to z in 1997 is much higher, approximately 8 times larger for cement manufacturers and 10 times larger for paperboard manufacturers.

[46] Equation 35 can be derived from a discrete choice model in which establishments choose which zip code to sell to. Suppose that establishments can sell to at most one zip code and suppose that the profits that establishment i would earn from selling to zip code z are given by:

$$\begin{aligned} u_{izt} &= \alpha_i + \alpha_z - \alpha_1 \log(\text{distance } i \to z) \\ &\quad -\alpha_2 \mathbb{I}\{i \to z \text{ in year } t-1\} - \alpha_3 \mathbb{I}\{\text{plant of i's firm} \in z\} + \eta_{izt}. \end{aligned} \tag{34}$$

When η_{izt} is distributed according to the Type 1 Extreme Value distribution, establishments choices will be consistent with Equation 35.

[47] In addition, I restrict the sample to establishments that were sampled in both the 1993 and 1997

There are two distinct interpretations of the positive estimate on, β_2, the coefficient on the persistence of buyer-supplier relationships (see Heckman (1981), or Dubé, Hitsch, and Rossi (2010)).[48] In the first interpretation, establishments' profitability of working with a counterparty increases from having transacted with that counterparty in the past. There are many reasons why the profitability of a buyer-supplier relationship might increase as buyers and suppliers interact with one another. Kellogg (2011), for instance, documents that oil production companies and drillers become more productive as they gain experience working with one another. Other explanations for persistence of buyer-supplier relationships are that: a) long-term contractual relationships overcome underinvestment in relationship-specific capital (Crawford (1988)); and b) the promise of long-term relationships allow buyers and suppliers to overcome problems associated with post-contractual opportunistic behavior (Klein, Crawford, and Alchian (1978) and Tesler (1978)).

Alternatively, according to the second interpretation, some establishments happen to find it more profitable to work with certain counterparties for idiosyncractic reasons, other than geographic proximity. The estimate of the persistence term, β_2, will be positive provided these idiosyncratic factors display some persistence.

Unfortunately, the data that I have at hand do not permit me to distinguish between these two interpretations. Either interpretation, however, is consistent with downstream establishments that decide to remain matched with high-marginal-cost suppliers.

5.4 Price discriminatory behavior of suppliers

A third explanation for price variation lies in differences in the relative bargaining power of the suppliers and buyers of any given material input. As the model in Section 2.2 shows, plants may have different relative bargaining powers because they occupy different positions within the buyer-supplier network. In this subsection, I document that suppliers charge different prices–for the same goods–across destinations. I then show that these within-supplier price differences explain a large fraction of the dispersion in the prices that downstream plants pay for their intermediate inputs.

For each buyer-supplier relationship, I define the *within-supplier price deviation*, ψ_{hit}

Commodity Flow Surveys. Secondly, in order to comply with Census disclosure rules, I restrict the sample to plants that are members of firms, f, such that the following three criteria hold: a) there exists at least one i, z pair for which plant i (owned by firm, f) shipped to zip code z in 1993, but not in 1997; b) there exists at least one i, z pair for which plant i shipped to zip code z in 1997, but not in 1993; and c) there exists at least one i, z pair for which plant i shipped to zip code z in both 1993 and 1997. The coefficient estimates are similar when the sample is constructed without this second restriction.

[48] Spurious state dependence occurs when the η_{izt} in 34 are serially correlated, or if there are omitted variables in Equation 35.

as:

$$\psi_{hit} \equiv \log\left(\frac{P_{hit}^{CFS}}{P_{ht}^{out,CFS}}\right), \tag{36}$$

ψ_{hit} is the price that i pays for h's output, compared to the other plants that buy intermediate inputs from h; ψ_{hit} is positive provided plant i purchases its material inputs from h at a higher price than $P_{ht}^{out,CFS}$, the average output price of supplier, h.

Motivated by the right panel of Figure 1, Figure 2 decomposes the price distribution into two separate components. Any buyer-supplier-specific price, p_{hit}^{CFS}, is the sum of suppliers' average output prices, $p_{ht}^{out,CFS}$, and the within-supplier price deviation, ψ_{hit}. The price, p_{hit}^{CFS}, that a supplier charges a buyer for intermediate inputs can be, mechanically, high for one of two reasons: either the supplier has a high average price, $p_{ht}^{out,CFS}$, or the supplier charges i a higher price than its other customers (i.e., ψ_{hit} is large). For my sample of cement and paperboard manufacturers, the distributions of p_{hit}^{CFS}, ψ_{hit}, and $p_{ht}^{out,CFS}$ are depicted in Figure 2.[49] The standard deviation of p_{hit}^{CFS} is 0.35, larger than the standard deviation of suppliers' average output prices ($SD(p_{ht}^{out,CFS}) = 0.28$), and almost twice as large as the standard deviation of the within-supplier deviations ($SD(\psi_{hit}) = 0.19$).

The *price-discrimination coefficient,* ψ_{it}, measures the extent to which plant i pays its supplier a higher materials price than the other customers of its suppliers. It is a weighted average, over i's suppliers, of the ψ_{hit}. If, on average, plant i pays a materials price, p_{hit}, that is larger than the average price, $p_{ht}^{out,CFS}$, that supplier h receives for its output, then ψ_{it} will be positive.

$$\psi_{it} \equiv \frac{\sum_{h:h\to i} \omega_{hit}\psi_{hit}}{\sum_{h:h\to i} \omega_{hit}} = \frac{\sum_{h:h\to i} \omega_{hit}\left(p_{hit} - p_{ht}^{out,CFS}\right)}{\sum_{h:h\to i} \omega_{hit}} \tag{37}$$

I regress buyers' average materials prices on their suppliers' productivities and their own ψ_{it}.

$$\tilde{p}_{it}^{in,CFS} = \beta_{\text{division}} + \beta_1 \cdot \overline{tfpq}_{it} + \beta_2 \cdot \psi_{it} + \varepsilon_{it} \tag{38}$$

The regression results are given in Table 12. As in Table 7, the estimate of β_1 is positive and statistical significant: plants with low-marginal cost suppliers pay lower than average prices for their intermediate inputs. In addition, there is a strong, positive relationship between $\tilde{p}_{it}^{in,CFS}$ and ψ_{it}.

A positive, significant coefficient estimate of β_2 is not surprising. A mechanical relationship between ψ_{it} and $\tilde{p}_{it}^{in,CFS}$ exists, as higher-than-average-priced shipments will

[49] Figure 2 looks similar, whether one uses the sample of cement manufacturers, the sample of paperboard manufacturers, or the pooled sample of paperboard and cement manufacturers.

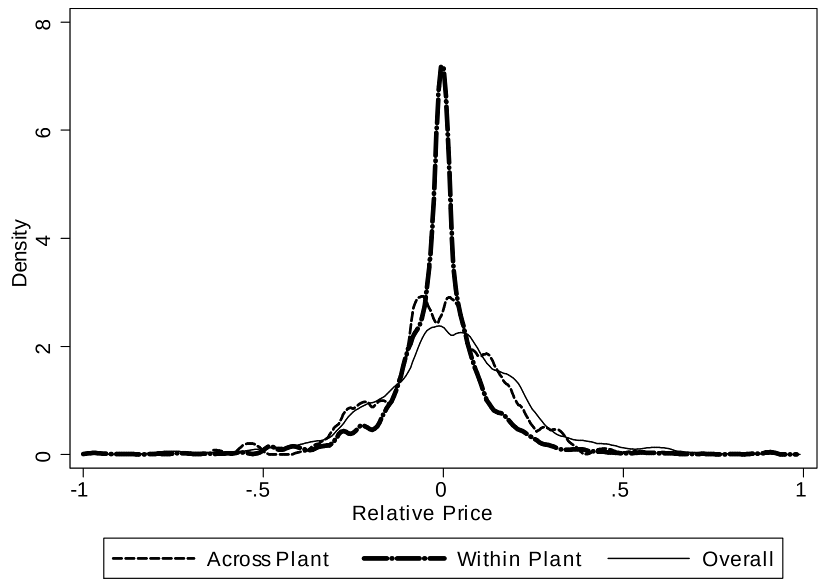

Figure 2: Value-weighted price distributions. The solid, thin line gives the distribution of $log(P_{hi}^{CFS})$. The thick, dash-dot line gives the disitribution of within-supplier price deviations, ψ_{hi}, while the thin, slashed line gives the distribution of suppliers' average output prices. The sample includes all shipments sent by cement and paperboard manufacturers that comprised the sample of the regressions defined by Equations 32 and 33.

Sample	Boxes	Concrete	Pooled	Boxes	Concrete	Pooled
$\overline{tfpq}_{it}$	-0.250*	-0.190	-0.236*	-0.233*	-0.146	-0.212*
	(0.057)	(0.111)	(0.048)	(0.056)	(0.102)	(0.047)
ψ_{it}	0.325*	0.866*	0.423*	0.330*	0.903*	0.422*
	(0.097)	(0.284)	(0.123)	(0.104)	(0.230)	(0.126)
N	190	131	321	190	131	321
Adjusted R^2	0.106	0.085	0.097	0.192	0.507	0.227
Division F.E.?	No	No	No	Yes	Yes	Yes

Table 12: Coefficient estimates and robust standard errors of the regressions defined by Equation 37. Stars indicate significance at the 5% level.

Sample	Include Division Fixed Effects ?	No	No	Yes	No	Yes
	Include $\overline{tfq}_{it}$?	No	No	No	Yes	Yes
	Include ψ_{it} ?	No	Yes	Yes	Yes	Yes
	Sample Size					
Boxes	190	0.202	0.191*	0.186*	0.181*	0.177*
Concrete	131	0.374	0.286*	0.257*	0.280*	0.252*
Pooled	321	0.224	0.206*	0.202*	0.197*	0.194*

Table 13: Each cell gives the real-revenue-weighted standard deviation of the residuals in a particular regression; the full specification is given in Equation 38. Across the columns of the table, different combinations of independent variables are included in the regressions. Across the rows of the table, different samples are used. Stars indicate that the decline in dispersion is significantly more than the decline that would occur from simply including "fake" random variables on the right hand side of Equation 38. See Appendix E for details.

generate a large value for $\tilde{p}_{it}^{in,CFS}$ (see Equation 30) and a large value of ψ_{it} (see Equation 37). The more interesting calculations, which are given in Table 13, compute the fraction of the dispersion in $\tilde{p}_{it}^{in,CFS}$ that can be explained by ψ_{it}. A regression with only ψ_{it} on the right hand side explains 7% ($=\frac{0.224-0.206}{0.224}$) of the variation in $\tilde{p}_{it}^{in,CFS}$. A combination of ψ_{it} and $\overline{tfpq}_{it}$ explains 12% ($=\frac{0.224-0.197}{0.224}$)of the variation of plants' materials prices, while the full combination of right-hand-side variables explains 13% ($=\frac{0.224-0.194}{0.224}$) of the variation in $\tilde{p}_{it}^{in,CFS}$.

6 Conclusion

In this paper, I have studied the consequences and sources of materials price dispersion. Variation in materials prices explains a substantial fraction of the variation in plants' marginal costs, revenue total factor productivities, and probabilities of survival. Moreover, one reason why some plants have low materials prices is because they have access to suppliers with low marginal costs.

These results indicate that establishments' survival and growth prospects are directly related to those of their customers and/or suppliers. In future work, I hope to investigate the causal relationship between establishments' growth and the growth rates of their counterparties. Such an investigation will be an important building block in understanding the propensity with which shocks to a small set of firms have the potential to cascade throughout the economy and produce aggregate fluctuations.

Appendix

A Proofs

Proposition 1 (Reminded) When the revenue share, R_i, of each plant is sufficiently close to 0, the output price, P_i^{out}, is approximated by the following equation

$$\begin{aligned}\log P_i^{out} &\approx \frac{1}{2}\left(\log MC_i + \overline{\log MC}\right) + \frac{1}{2\gamma}\left(\Delta_i + \overline{\Delta}\right) \\ &\approx \frac{1}{2}\left(\sigma \log P_i^{in} - \log \Phi_i + \sigma \overline{\log P^{in}} - \overline{\log \Phi}\right) + \frac{1}{2\gamma}\left(\Delta_i + \overline{\Delta}\right)\end{aligned} \tag{39}$$

Proof.

Write the profits of firm i as the product of a) quantity b) and the difference between price and marginal costs:

$$\Pi_i = \frac{\Upsilon R_i}{P_i^{out}}\left(P_i^{out} - MC_i\right) = \Upsilon R_i \left(1 - \frac{MC_i}{P_i^{out}}\right)$$

Take first order conditions and re-arrange. Doing so results in a markup, $\frac{P_i^{out}}{MC_i}$, that equals $1 - \left[\frac{\partial \log R_i}{\partial \log P_i^{out}}\right]^{-1}$. Using the expression for $\frac{\partial \log R_i}{\partial \log P_i^{out}} = -\frac{I-1}{I}\frac{\gamma}{R_i}$ (which can computed from Equation 3), plant i's output price is given by the following equation:

$$\frac{P_i^{out}}{MC_i} = 1 + \frac{R_i I}{\gamma(I-1)} \tag{40}$$

Assuming that R_i is sufficiently close to 0, $\log\left(1 + \frac{R_i I}{\gamma(I-1)}\right)$ is approximately equal to $\frac{R_i}{\gamma}\frac{I}{I-1}$. Thus, Equation 40 leads to

$$\log P_i^{out} \approx \log MC_i + \frac{R_i}{\gamma}\frac{I}{I-1}. \tag{41}$$

Furthermore, the revenue share of plant i equals (see Equation 3)

$$R_i = \Delta_i - \frac{\gamma(I-1)}{I}\ln P_i^{out} + \frac{\gamma}{I}\sum_{k=1, i\neq k}^{I} \ln P_k^{out}$$

yielding:

$$
\begin{aligned}
\frac{R_i I}{\gamma(I-1)} &= \Delta_i \frac{I}{\gamma(I-1)} - \log P_i^{out} + \frac{1}{I-1}\sum_{k\neq i}\log P_k^{out} \\
&= \Delta_i \frac{I}{\gamma(I-1)} - \frac{I}{I-1}\log P_i^{out} + \overline{\log P^{out}}. \qquad (42)
\end{aligned}
$$

Plugging Equation 41 into Equation 42 leads to

$$
\begin{aligned}
\log P_i^{out} &\approx \log MC_i + \Delta_i \frac{I}{\gamma(I-1)} - \frac{I}{I-1}\log P_i^{out} + \overline{\log P^{out}}. \\
\log P_i^{out} &\approx \frac{1}{2}\left(\log MC_i\right) + \frac{\Delta_i}{\gamma} + \frac{\overline{\log P^{out}}}{2} \qquad (43)
\end{aligned}
$$

The equation

$$
\frac{1}{I}\sum_{i=1}^{I}\log P_i = \frac{1}{2I}\sum_{i=1}^{I}\left(\log MC_i + \overline{\log P}\right) + \frac{1}{2\gamma}\frac{1}{I}\sum_{i=1}^{I}\Delta_i,
$$

is equivalent to

$$
\overline{\log P^{out}} \approx \overline{\log MC} + \frac{1}{\gamma}\overline{\Delta}. \qquad (44)
$$

Thus, plugging 44 into Equation 43 yields

$$
\log P_i \approx \frac{1}{2}\left(\log MC_i + \overline{\log MC}\right) + \frac{1}{2\gamma}\left(\Delta_i + \bar{\Delta}\right),
$$

which is the desired result. ■

Proposition 3 (Reminded) Assume that a) Cov($\log P^{in}$, $\log \Phi$),Cov($\log P^{in}$, Δ), and Cov(Δ, $\log \Phi$) are sufficiently close to 0 in the observed sample, and b) $R_i \to 0$ for all plants i. Then, the following relationships hold:

1. The materials price is increasing, in the sense of first order stochastic dominance, in the marginal cost of the supplier.

2. Plants' marginal costs are positively correlated to their suppliers' marginal costs.

Proof.

1. Let $G(\cdot)$ denote the cumulative distribution function of suppliers' marginal costs, and let $H(P^{in}|MC^{up})$ denote the cumulative distribution function of the price that i pays

for the intermediate input, conditional on the marginal cost of the lowest-marginal-cost supplier.

Then:

$$\begin{aligned} 1 - H(P^{in}|MC^{up}) &= \Pr\left\{\xi - 1 \text{ draws each } \geq P^{in}|\xi - 1 \text{ draws each} \geq MC^{up}\right\} \\ &= \left(\frac{1 - G(P^{in})}{1 - G(MC^{up})}\right)^{\xi - 1} \end{aligned}$$

Re-arranging:

$$H(P^{in}|MC^{up}) = 1 - \left(\frac{1 - G(P^{in})}{1 - G(MC^{up})}\right)^{\xi - 1}$$

The desired result follows from the fact that the right hand side of the previous equation is decreasing in, MC^{up}, the marginal cost of the best supplier.

2. Since, from part 1, $\frac{\partial E\left[\log P^{in}|MC^{up}\right]}{\partial MC^{up}} > 0$, $Cor(\log P^{in}, \log MC^{up}) > 0$. Since I assumed that $Cov(\log MC^{up}, \log \Phi) \approx 0$

$$\begin{aligned} Cov(\log MC, \log MC^{up}) &= Cov(\sigma \log P^{in}, \log MC^{up}) - Cov(\log \Phi_i, \log MC^{up}) \\ &> 0 \end{aligned}$$

■

B Other Calculations

B.1 An expression for industry sales in terms of marginal costs and demand shocks

As a reminder, expenditures on the industry output equals (see Equation 2):

$$\log \Upsilon = \Delta_0 + \sum_{i=1}^{I} \Delta_i \log P_i^{out} - \frac{\gamma}{2} (\log P_i)^2 + \frac{\gamma}{2I} \sum_{i=1}^{I} \sum_{k=1}^{I} \log P_i^{out} \log P_k^{out}$$

This is equivalent to:

$$\begin{aligned} \log \Upsilon &= \Delta_0 + I \cdot E\left[\Delta \log P^{out}\right] - \frac{I\gamma}{2} E\left[\left(\log P^{out}\right)^2\right] + \frac{\gamma I}{2} \left(\overline{\log P^{out}}\right)^2 \qquad (45) \\ &= \Delta_0 + I \cdot E\left[\Delta \log P^{out}\right] - \frac{I\gamma}{2} Var\left(\log P^{out}\right) \end{aligned}$$

Next, I derive the following expression for $Var(\log P^{out})$, in terms of moments of the joint distribution of MC and Δ. Note that these moments are taken over the sample of plants that we observe selling positive amounts of the good.

$$\begin{aligned} Var(\log P^{out}) &= Var\left(\frac{1}{2}(\log MC) + \frac{1}{2\gamma}(\Delta)\right) \\ &= \frac{1}{4}Var\left(\log MC + \frac{\Delta}{\gamma}\right) \\ &= \frac{1}{4}Cov\left(\log MC + \frac{\Delta}{\gamma}, \log MC + \frac{\Delta}{\gamma}\right) \\ &= \frac{1}{4}Var(\log MC) + \frac{1}{4\gamma^2}Var(\Delta) + \frac{1}{2\gamma}Cov(\log MC, \Delta) \end{aligned} \tag{46}$$

Similarly, I derive the following expression for $E\left[\Delta \log P^{out}\right]$, in terms of the moments of the joint distribution of MC and Δ.

$$\begin{aligned} I \cdot E\left[\Delta \log P^{out}\right] &= \frac{I}{2}E\left[\Delta \log MC + \frac{\Delta^2}{\gamma}\right] + \frac{I\bar{\Delta}\overline{\log MC}}{2} + \frac{I\left(\bar{\Delta}\right)^2}{2\gamma} \\ &= \frac{I}{2}E\left[\Delta \log MC\right] + \frac{I}{2\gamma}E\left[\Delta^2\right] + \frac{\overline{\log MC}}{2} + \frac{\bar{\Delta}}{2\gamma} \\ &= \frac{I}{2}\left[E\left[\Delta \log MC\right] + \frac{1}{\gamma}E\left[\Delta^2\right]\right] + \overline{\log MC} + \frac{\bar{\Delta}}{\gamma} \end{aligned} \tag{47}$$

Finally, I plug Equations 46 and 47 into Equation 45, which leads to the following expression for Υ, in terms of the moments of the joint distribution of MC and Δ:

$$\begin{aligned} \log \Upsilon &= \Delta_0 + \frac{I}{2} \cdot Cov(\Delta, \log MC) + \frac{I}{2\gamma}Var\left(\Delta\right) + \frac{\overline{\log MC}}{2} \\ &\quad + \frac{\bar{\Delta}}{2\gamma} - \frac{I\gamma}{8}Var(\log MC) - \frac{I}{8\gamma}Var(\Delta) - \frac{I}{4}Cov(\log MC, \Delta) \\ &= \frac{I\gamma}{8}\left[2 \cdot Cov\left(\frac{\Delta}{\gamma}, \log MC\right) + 3 \cdot Var\left(\frac{\Delta}{\gamma}\right) - Var(\log MC)\right] \\ &\quad \Delta_0 + \frac{1}{2}\left[E\left[\log MC\right] + E\left[\frac{\Delta}{\gamma}\right]\right] \\ \Upsilon &= \exp\left\{\frac{I\gamma}{8}\left[2 \cdot Cov\left(\frac{\Delta}{\gamma}, \log MC\right) + 3 \cdot Var\left(\frac{\Delta}{\gamma}\right) - Var(\log MC)\right]\right\} \\ &\quad \times \exp\left\{\Delta_0 + \frac{1}{2}\left[E\left[\log MC\right] + E\left[\frac{\Delta}{\gamma}\right]\right]\right\} \end{aligned}$$

The final expression for Υ is the value that appears in footnote 11.

C Construction of the Sample

The benchmark sample consists of 10 industries (collections of 7-digit products) for which both inputs and outputs display minimal levels of quality differentiation. The construction of the sample consists of plants for which the following five conditions hold. First, I discard any plants that have missing data on labor inputs, capital stocks, electricity bills, or materials bills. Second, I discard any plants that do not fill out either the Census of Manufacturers Materials Supplement (containing information on purchases of intermediate inputs) or the Census of Manufacturers Productivity Supplement (containing information on products produced). Third, I throw out plants that have imputed values for quantities of materials purchased or products produced. Fourth, I require that the plants in the benchmark sample earn at least half of their revenues from one of the 10 main industries. Fifth, I discard any plant that has an output price (defined by p^{out}, as in Equation 16), an input price (defined by p^{in}, as in Equation 17 or 19), or a quantity total factor productivity (defined by $tfpq$, as in Equation 20) that is more than 3 units away than the average for that industry$\times$year.

Industries are defined as the collection of 7-digit products in the following manner.

Bulk milk is the combination of fluid whole milk, bulk sales (2026112) and fluid skim milk, bulk sales (2026115). The units of bulk milk are thousands of pounds.

White wheat flour is the combination of the ten 7-digit products: white flour, shipped for export (2041105 and 2041107); bakers' and institutional white bread-type flours (2041111 and 2041113); bakers' and institutional soft wheat flour (2041115 and 2041117); family white flour, other than self-rising (2041121 and 2041123); self-rising family white flour (2044126); and flour shipped to blenders or other processors (2041128 and 2041129). The units of white wheat flour are 50-pound sacks.

Ready-mix concrete consists of the single 7-digit product (3273000). In 1972 and 1977 some ready-concrete plants were producing a product with a code of 3273011. The units of output are cubic yards. Production data do not exist for 1997; materials data do not exist for 1992 or 1997. Because of this, for the analysis in Section 4, the sample period for ready mix concrete is 1972-1987. The sample period for the analysis of Section 5, in which I use the Commodity Flow Survey but not the Census of Manufacturers' materials data, is 1992. In addition to the five criteria listed in the first paragraph of this section, I require ready-mix concrete plants to have positive purchases of both cement and sand/gravel. (I need positive purchases of both of the main intermediate inputs so that I can estimate the elasticity of substitution between cement and sand/gravel, using Equation 48.)

Yarn is comprised of the two 7-digit products, spun gray (22811100) and yarn, spun and finished in the same establishment (2281187). The units of output are thousands of pounds.

Gasoline is comprised of the following three 7-digit products: motor gasoline (2911131), distillate fuel oil (2911412), and No. 4 type light fuel oil (2911414). The units of output are thousands of barrels.

Coffee consists of two 7-digit products, (2095111) whole bean coffee and ground coffee (2095115). The units of output are thousands of pounds.

Sugar consists of the single 7-digit product, raw cane sugar (2061011). The units of output are short tons.

Packaged milk consists of the following three 7-digit products: fluid whole milk (2026212), low fat milk (202623), and skim milk (2026225). The units of output are thousands of quarts.

Corrugated boxes is a combination of nine 7-digit products, with products being classified by their end use. These end-uses are containers of food and beverages (2653012); carry-out boxes for retail food (2653014); containers of paper and allied products (2653013); containers of glass, clay, and stone products (2653015); containers of metal products, machinery, equipment, and supplies (2653016); containers of electrical machinery, equipment, supplies, and appliances (2653018); containers of chemicals and drugs, including paints, varnishes, cosmetics, and soaps (2653021); containers of lumber and wood products, including furniture (2653022); all other end uses not specified (2653030). From 1972 to 1987, the units of output for corrugated boxes were thousands of pounds. From 1992 on, the units of output for corrugated boxes have been thousands of square feet.

Measuring corrugated boxes in terms of area, instead of mass, is somewhat problematic. The density of boxes depend on their final use. In particular, the density of boxes is lower for those boxes that are used as containers of food, beverages, paper and allied products, glass, clay, stone, or metal, while the density of boxes is higher for boxes that are used as containers of machinery, electronics, chemicals, lumber, and other products. Since the total cost of producing corrugated boxes seems to be more closely related to the mass–instead of surface area–of the amount produced, measured quantity total factor productivity for low-density box manufacturers began to exceed, in 1992, the measured quantity total factor productivity of high-density boxes.

To mitigate the impact of this measurement problem, I de-meaned, according to Equation 23, plant-level statistics separately for the high-density (those plants that produced output with a product code between 2653016 and 2653030) and low-density (those plants that

	Employment		Total Value of Shipments		N		Main
Sample	Benchmark	Main Ind.	Benchmark	Main Ind.	Benchmark	Main Ind.	Ind.
Bulk Milk	3.197	3.193	9.082	8.023	127	7661	2026
Flour	3.900	2.786	9.854	7.942	503	2073	2041
Concrete	2.805	2.214	7.682	6.827	3708	20956	3273
Yarn	5.171	4.597	9.508	8.645	431	2233	2281
Boxes, Yr.≤'87	4.554	3.717	9.426	8.404	1820	7742	2653
Gasoline	5.741	4.532	12.977	11.157	692	1706	2911
Coffee	3.975	3.154	9.920	8.736	300	874	2095
Sugar	4.813	4.356	10.000	9.301	177	301	2061
Boxes, Yr.≥'92	4.624	3.717	9.799	8.404	646	7742	2653
Packaged Milk	4.299	3.193	9.465	8.023	2099	7661	2026
Pooled	3.934	2.572	9.119	6.929	10503	87180	

Table 14: Descriptive statistics for the main sample. Variables are stated in logs. The final column refers to the 4-digit SIC industry of which the product is a memeber.

produced output with a product code between 2653012 and 2653015) box manufacturers.[50]

In Table 14, I provide some descriptive statistics of the benchmark sample. The average log employment for plants is 3.93 (i.e., roughly $51 \approx e^{3.93}$ employees work in the average plant.) Plants that produce ready-mix concrete are one-third the size of the average benchmark-sample plant, while plants engaged in gasoline production employ approximately 6 ($\approx e^{5.741-3.934}$) times as many workers as the average plant.

Compared to the universe of plants that are in the same 4-digit SIC industry, the plants in the benchmark sample employ $4(\approx e^{3.934-2.572})$ times as many employees and have revenues that are $9(\approx e^{9.12-6.93})$ times larger. The difference is due to the Census Bureau's survey methodology: the largest plants tend to receive the survey questionnaires on the products they produce or the materials they consume.

For a particular intermediate input to be included in the analysis, expenditures of the material must make up at least 6% of total materials expenditures for that product group. As the cut-off expenditure share decreases, additional intermediate inputs are included in the analysis. Setting the cut-off too low results in the inclusion of intermediate inputs that are purchased only by a few plants, hindering cross-plant comparisons of materials prices. Setting the cut-off too high means that important components of plants' materials prices are ignored. The 6% cut-off seems like a good compromise between these two considerations.

In some instances for which I combine groups of similar 6-digit products to form a

[50] Dropping the "Boxes, Year≥1992" subsample does not change any of the results from Section 4. I find it worth the trouble to keep the "Boxes, Year≥1992" subsample, since corrugated box manufacturers purchase one of their main inputs–namely, paperboard–from the manufacturing sector, and thus can be included in the analysis of Section 5.

given "material input".[51] For example, I combine material 131111 (domestic crude petroleum) and 131112 (foreign crude petroleum). The presumption when deciding to combine two materials is that the manufacturer is indifferent between the two 6-digit materials. The way in which I combined these 6-digit products is given below.

For milk (either bulk or packaged), the sole material is unpasteurized whole milk (024111).

For white wheat flour, the sole material is wheat (011111).

In the production of ready-mix concrete, the two materials are cement (which was coded as 324101 in 1982 and 1992 and 324102 in other years) and sand/gravel aggregate (144201).

In the production of yarn, the two materials are raw cotton fibers (013101) and a combination of polyester staple and tow (282425) and acrylic staple and tow (282426).

In the production of gasoline, I have combined foreign and domestic crude petroleum into one material.

In the production of raw cane sugar, the sole material is sugar cane (013321).

Green coffee beans (017921) are the sole material used in the production of ground/whole bean coffee.

Finally, in 1992-1997, the sole material used in the production of corrugated boxes is coded 260003 ("Paper and Paperboard"). In 1987, the material "Paper/Paperboard" is the combination of 262104 ("Paper, Cellulosic Wadding") and 262108 (Paper). Earlier than this, "Paper/Paperboard" is the combination of materials 262102, 262103, and 262105.

D Robustness Checks

D.1 Industries with heterogeneous quality outputs

In this subsection, I reproduce the empirical analysis of Sections 4.1 and 4.2 for a set of industries that display substantial output quality variation. The four industries that I chose for this exercise are wine, softwood cut stock, cucumber pickles, and sausages.

Industries are defined as the collection of 7-digit products in the following manner:

Wine is a combination of the following three products: white grape wine (2084012), red grape wine (2084014), and rose grape wine (2084016). The units of output are thousands of gallons.

[51] For 1992 and 1997, a description of the 6-digit material codes can be found by downloading MC92F7.dbf from the following Census web-page:
ftp://ftp2.census.gov/econ1992/MC92/.

Sample	Units of Output	Material Inputs	N
Wine	1000 gallons	Fresh grapes (41%) Purchased wines (23%) Glass containers (19%)	330
Softwood Cut Stock	1000 board feet	Softwood dressed lumber (75%) Softwood logs (8%) Hardwood dressed lumber (8%)	160
Pickles	1000 gallons	Cucumbers (43%) Glass containers (28%)	145
Sausages	1000 lbds.	Fresh and frozen pork (34%) Fresh and frozen beef (30%) Meat, unknown species (13%)	621
Pooled	-		1256

Table 15: Description of the four industries that have heterogeneous-quality outputs. The Material Inputs column gives the inputs that represent greater than 6% of the average plants' total material purchases. The percentages that appear in the Material Inputs column are the fraction of intermediate input expenditures that go to each particular material input.

Softwood cut stock is a combination of two product groups: furniture cut stock (2421711) and industrial cut stock (2421751). The units of output are thousands of board feet.

Pickles are a combination of four products: dill pickles (2035211), sour pickles (2035213), sweet pickles (2035215), and refrigerated pickles (2035219). The units of output are thousands of gallons.

Sausages are a combination of six products: fresh sausage (2011711 and 2013711); dry or semi-dry sausages (2011717 and 2013717); and frankfurters (2011721 and 2013721). The units of output are thousands of pounds.

As with the benchmark sample, materials that make up more than 6% of intermediate input expenditures are included as "priced" materials. A summary of the characteristics of the Output Quality Variation sample are given in Table 15.

Correlations among plant-level characteristics are presented in Table 16. Compared to the benchmark sample, the standard deviations of most plant-level characteristics are larger, while the skewness coefficients are smaller. (A positive skewness coefficient of, for example, $tfpr$ could reflect the existence of a productivity threshold $\widetilde{tfpr}$, below which no establishments may profitably operate.) In other words, when outputs vary in quality, the distribution of plant-level characteristics both becomes less skewed and more disperse. Moreover, compared to the benchmark sample, the correlations among the different productivity measures are weaker.

	p^{in}	p^{out}	$tfpq$	ϕ	$tfpr$
p^{out}	0.232				
$tfpq$	-0.236	-0.799			
ϕ	0.305	-0.655	0.850		
$tfpr$	-0.030	0.225	0.406	0.382	
Std. Dev.	0.355	0.375	0.399	0.410	0.246
Skewness	0.175	-0.016	0.080	0.256	0.303

Table 16: Correlations among plant-level characteristics, pooled across the four industries of the "Quality Variation Sample." N=1256.

	Dispersion of $tfpq$			Dispersion of ϕ			Percent Decline			
Sample	90/10	75/25	SD	90/10	75/25	SD	90/10	75/25	SD	N
Wine	1.343	0.734	0.491	1.321	0.769	0.499	1.7%	-4.4%	-1.5%	330
Softwood Cut Stock	1.624	0.886	0.568	1.340	0.729	0.489	23.6%	24.1%	17.6%	160
Pickles	0.944	0.473	0.355	0.950	0.436	0.353	-0.6%	8.8%	0.4%	145
Sausages	0.733	0.400	0.298	0.755	0.369	0.305	-2.9%	8.6%	-2.2%	621
Pooled	1.001	0.512	0.393	1.022	0.507	0.393	-1.9%	1.0%	0.0%	1256

Table 17: Dispersion of $tfpq$ and ϕ, for the "Quality Variation Sample." Observations are weighted by revenue. Due to Census' rules regarding data confidentiality, I am prohibited from reporting the actual quantiles of any empirical distribution (e.g., ϕ). See the caption of Table 3 for a description of the imputation of the quantiles of $tfpq$ and ϕ.

The dispersions of $tfpq$ and ϕ are given in Table 17. For the pooled sample, the dispersions of the two distributions are essentially the same. Looking across the four industries, there is a significant decline in productivity dispersion for one of the industries, softwood cut stock, and no difference for the other three industries.

In summation, output quality variation has the potential to severely attenuate the difference between the dispersions of ϕ and $tfpq$. To the extent that any quality variation exists in the benchmark sample, the difference between the dispersions of ϕ and $tfpq$, as reported in Table 3, may be downward biased.

D.2 Variation in input quality

One of the main presumptions of the empirical analysis has been that variation in input quality is not an important source of variation for input prices or productivities. I have chosen industries to try to minimize the role of input quality differentiation. There is one specific industry, ready-mix concrete, for which there is reason to suspect that input quality differences could be contaminating some of the results. In this subsection, I explain the reasons why input quality varies across plants, and then determine how big of an effect input quality variation has on the observed relationships between input prices and different productivity measures.

Portland cement, the main intermediate input used in the production of ready-mix concrete, comes in four types, labeled type I, II, III, or IV.[52] Sales of type I and II cement constitute over 90% of the expenditures on cement, with the majority of sales coming from type I cement (U.S. Department of Interior (1989)). In areas where the soil has high sulfate concentrations, type II cement may be preferable to type I cement, since ready-mix concrete produced using type I cement is susceptible to *sulfate attack* (cracking or loss of strength in the presence of sulfate). Since high-sulfate concentrations exist only in the soil of parts of the western third of the United States, one should observe type I and type II cement plants in the western United States, and only type I cement plants in the remainder of the United States.[53]

So, because of differences in soil composition, input quality variation should be greater in the western United States than in the eastern United States. With this in mind, I split the sample of ready-mix concrete plants into two subsamples: plants residing in Census divisions 1-7, and plants located in Census divisions 8-9.[54] In the left panel of Figure 3, I plot the distribution of p^{in} separately for the two halves of the United States. As expected, ready-mix concrete plants in divisions 8 and 9 pay a higher price for their intermediate inputs, as some of these plants pay a higher price to purchase type II Portland cement. Moreover, the dispersion of p^{in} is larger in divisions 8-9, as some ready-mix concrete plants purchase the low-price type I cement, while others must purchase the high-price type-II cement. In contrast, in the eastern United States, virtually all ready-mix concrete plants purchase type I cement, leading to a more compressed p^{in} distribution.

In the right panel of Figure 3, I plot the relationship between p^{in} and $tfpq$, separately for the two subsamples. The negative relationship between $tfpq$ and p^{in} is somewhat stronger in the eastern United States, while correlation relationship between p^{in} and p^{out} is greater in the western United States. These geographic differences are consistent with greater cement quality variation in the western United States.

Finally, in Table 19, I compute the dispersion of $tfpq$ and ϕ for the ready-mix concrete subsamples. The decline in dispersion is larger for each of the two subsamples than

[52] The standards for the different types of Portland cement are set by the American Society for Testing and Materials (ASTM). See the ASTM webpage for more information on the distinguishing factors of different types of Portland cement: http://www.astm.org/Standards/C150.htm

[53] I confirm this prediction using the Census of Manufacturers production file.

[54] Census division 8 is made up of Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming, while Census division 9 is made up of Alaska, California, Hawaii, Oregon, and Washington. See http://www.census.gov/geo/www/us_regdiv.pdf for a correspondence between states and Census divisions.

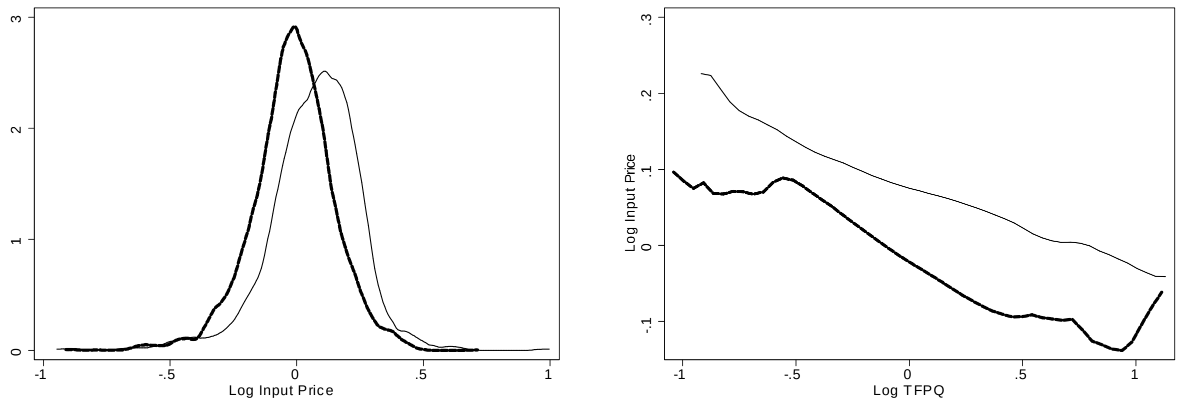

Figure 3: In the left panel, I plot the distribution of p^{in} separately for ready-mix concrete platns located in divisions 1-7 (thick, dashed line) and plants located in divisions 8-9 (thin, solid line). In the right panel, I plot the relationship between $tfpq$ and p^{in}.

Sample	$\rho(p^{in},\, tfpq)$	$\rho(p^{in},\, \phi)$	$\rho(p^{in},\, p^{out})$	$\rho(\,\phi,\, tfpq)$	$\rho(\,tfpq,\, tfpr)$	$\rho(tfpq,\, p^{out})$	N
Entire U.S.	-0.271	0.107	0.234	0.928	0.776	-0.458	3708
Divisions 1-7	-0.317	0.064	0.234	0.926	0.783	-0.420	3049
Divisions 8-9	-0.272	0.077	0.284	0.938	0.739	-0.570	659

Table 18: Correlations between different plant-level statistics of ready-mix concrete producers. Each row gives the correlations for different subsamples of ready-mix concrete plants.

	Dispersion of $tfpq$			Dispersion of ϕ			Percent Decline			
	90/10	75/25	SD	90/10	75/25	SD	90/10	75/25	SD	N
Divisions 1-7	0.489	0.238	0.211	0.441	0.220	0.198	11.6%	8.5%	6.4%	3049
Divisions 8-9	0.593	0.291	0.250	0.546	0.240	0.234	8.9%	23.9%	7.0%	659
Entire U.S.	0.517	0.250	0.222	0.478	0.233	0.213	8.5%	7.4%	4.5%	3708

Table 19: Dispersion of $tfpq$ and ϕ, for different subsamples of plants in the ready-mix concrete industry. Due to Census' rules regarding data confidentiality, I am prohibited from reporting the actual quantiles of any empirical distribution (e.g. ϕ). See the caption of Table 3 for a description of the imputation of the quantiles of $tfpq$ and ϕ.

it is for the pooled sample of 3708 ready-mix concrete plants.

In summation, there is almost no variation in the quality of cement purchased by ready-mix concrete plants in the eastern two-thirds of the United States. For this subsample, the difference between the dispersion of $tfpq$ and the dispersion of ϕ is 1 to 3 percentage points larger than the differences that were reported in Table 3. So, moderate amounts of intermediate input quality variation is probably causing me to somewhat underreport the fraction of productivity dispersion that is due to differences in materials prices.

D.3 Substitution across material inputs

Throughout the paper, I have assumed that the elasticity of substitution between different material inputs–for industries that use different material inputs–is 0 (see Assumption 4 on page 18). For plants that produce ready-mix concrete, I assess the importance of the assumption that plants may not substitute across different material inputs.

There is some evidence that ready-mix concrete manufacturers substitute across the two intermediate inputs. I regress plants' utilization of cement, relative to sand/gravel, against the ratio of plant-level cement and sand/gravel prices. The results, which are given in Table 20, indicate that the elasticity of substitution is significantly greater than 0.[55]

$$\log\left(\frac{N_{ijt}^{Cement}}{N_{ijt}^{Sand/Gravel}}\right) = \beta_{\text{division}\times\text{year}} + \beta \cdot \log\left(\frac{P_{ijt}^{Cement}}{P_{ijt}^{Sand/Gravel}}\right) + \varepsilon_{ijt} \tag{48}$$

[55]A common problem when running regressions similar to Equation 48 is the presence of simultaneous-equation bias: If a particular concrete manufacturer has a production technology that uses cement exceptionally intensively, the cement supplier may respond by increasing the price that it charges. The simultaneous-equation bias would case my estimated of β to under-represent the actual elasticity of substitution.

In unreported regressions, I instrument for $\frac{P_{ijt}^{Cement}}{P_{ijt}^{Sand/Gravel}}$ using local employment in the limestone industry (limestone is an intermediate input in cement manufacturing, but not in sand/gravel). The first-stage regressions are consistent with a negative relationship between local limestone employment and $\frac{P_{ijt}^{Cement}}{P_{ijt}^{Sand/Gravel}}$. The estimate of β, while substantially larger, is much less precisely estimated.

Fixed Effects	Sample	β	$s.e.$	Adjusted R^2	N
Year	Entire U.S.	-0.178	0.019	0.026	3708
Year$\times$Division	Entire U.S.	-0.158	0.021	0.135	3708
Year$\times$County	Entire U.S.	-0.209	0.035	0.467	3708
Year	Divisions 1-7	-0.092	0.022	0.010	3049
Year$\times$Division	Divisions 1-7	-0.131	0.024	0.068	3049
Year$\times$County	Divisions 1-7	-0.184	0.043	0.436	3049

Table 20: Each row presents the results of a regression, defined by Equation 48. The first three rows give estimates of the elasticity of substitution, using data from the entire U.S., while the last three rows give estimates computed using data only from the geographic divisions in which only type I cement is used.

When the elasticity of substitution between gravel/sand and cement is constant (but not necessarily 0), the price of a bundle of material inputs is:

$$P_{ijt}^{in} \equiv \left[\frac{s_{Gravel,jt}}{s_{Gravel,jt} + s_{Cement,jt}} \left(\frac{P_{Gravel,ijt}^{in}}{\bar{P}_{Gravel,jt}^{in}} \right)^{1-\varrho} + \frac{s_{Cement,jt}}{s_{Gravel,jt} + s_{Cement,jt}} \left(\frac{P_{Cement,ijt}^{in}}{\bar{P}_{Cement,jt}^{in}} \right)^{1-\varrho} \right]^{\frac{1}{1-\varrho}} \tag{49}$$

In Equation 49, $s_{Gravel,jt}$ refers to the share of materials expenditures that go to gravel, $P_{Gravel,ijt}^{in}$ is the price that plant i pays per 1000 pounds of gravel in year t, $\bar{P}_{Gravel,ijt}^{in}$ is the geometric average of the price paid by all ready-mix-concrete producing plants in year t, and σ is the elasticity of substitution between cement and sand/gravel. In the baseline analysis, I had set $\varrho = 0$.

Using Equation 49, I compute ready-mix concrete plants' materials prices. I then re-compute Φ_{ijt}, using Equation 22, and compute ϕ_{ijt}, using 23. The dispersion of ϕ is given in Table 21. As ϱ increases, the price of a bundle of intermediate inputs decreases for plants that have exceptionally cheap input prices for one of the two intermediate inputs. Also, as ϱ increases, the relative price increases for plants that pay roughly the same relative price for the two intermediate inputs. It turns out that these two effects has almost no impact on the overall dispersion of ϕ.

D.4 More correlations

Table 22, below, presents correlations between plant-level characteristics for each of the 10 industries in the benchmark sample.

For several of the correlations, the subsample of raw cane sugar manufacturing plants is anomalous. For this industry, plants' marginal costs are *negatively* related to their ma-

Sample	Revenue-weigthed?	90/10	SD	90/10	SD	90/10	SD	N
ϱ		0.1	0.1	0.3	0.3	0.5	0.5	
Concrete	No	0.5223	0.2302	0.5223	0.2302	0.5222	0.2303	3708
	Yes	0.4784	0.2127	0.4782	0.2127	0.4777	0.2127	3708
Pooled	No	0.4930	0.2190	0.4929	0.2190	0.4928	0.2191	10503
	Yes	0.2988	0.1470	0.2988	0.1470	0.2988	0.1470	10503

Table 21: Dispersion of ϕ, which is computed using Equations 22 and 23. Due to Census' rules regarding data confidentiality, I am prohibited from reporting the actual quantiles of any empirical distribution (e.g., ϕ). See the caption of Table 3 for a description of the imputation of the quantiles of $tfpq$ and ϕ.

Sample	$\rho(p^{in},tfpq)$	$\rho(p^{in},\phi)$	$\rho(p^{in},p^{out})$	$\rho(\phi,tfpq)$	$\rho(tfpq,tfpr)$	$\rho(tfpq,p^{out})$	N
Milk-Bulk	-0.280	0.033	0.382	0.950	0.522	-0.789	127
Flour	-0.344	0.454	0.256	0.681	0.254	-0.715	503
Concrete	-0.271	0.107	0.234	0.928	0.776	-0.458	3708
Yarn	-0.354	0.117	0.297	0.887	0.455	-0.788	431
Boxes, Yr.$\leq$'87	-0.406	0.153	0.404	0.841	0.425	-0.824	1820
Gasoline	-0.353	0.203	0.125	0.844	0.840	-0.396	692
Coffee	-0.293	0.227	0.300	0.862	0.517	-0.645	300
Sugar	0.055	0.459	0.116	0.912	0.889	-0.405	177
Boxes, Yr.$\geq$'92	-0.428	0.142	0.366	0.834	0.127	-0.873	646
Milk-Packaged	-0.282	0.054	0.237	0.943	0.456	-0.753	2099
Pooled	-0.303	0.141	0.278	0.899	0.601	-0.653	10503

Table 22: Correlations between plant-level characteristics. All variables are de-meaned by year and product.

terials prices. Moreover, the correlation between input prices and technical efficiencies is much stronger (46%) than most other subsamples. These patterns are somewhat puzzling. Most likely, the goods produced by raw cane sugar manufacturers is as not homogeneous as Foster, Haltiwanger, and Syverson (2008) and I have hypothesized.

Except for the raw cane sugar, correlations among plant-level characteristics are qualitatively similar across the different industries in the benchmark sample. The correlation between materials prices and marginal costs is moderately negative for the 9 other industries, while the correlation between marginal costs and output prices is strongly negative (ranging between -40% and -87%). Finally, the three productivity measures are always highly correlated with one another, with the correlation between ϕ and $tfpq$ being larger than the correlation between $tfpq$ and $tfpr$.

D.5 Unweighted results

In this section, I present the unweighted versions of Tables 3, 5, and 17. In the benchmark calculations, observations are revenue-weighted.

Compared to the revenue-weighted versions, the unweighted dispersions of $tfpq$ and ϕ are larger (see Table 23 for the benchmark sample and Table 24 for the "Output Quality Variation" sample). The larger dispersions have two sources. First, weighting by revenue gives more importance to high revenue-per-plant industries. Since gasoline, which by far has the largest average sales per plant, has more compressed $tfpq$ and ϕ distributions, revenue-weighting causes the pooled dispersion to be larger in the unweighted calculations. Second, the unweighted calculations give relatively weight, within industries, to the low-productivity, low employment plants, again causing unweighted dispersions to be larger than the weighted dispersions.

The decline in dispersion, of the pooled distribution of productivities, is smaller when observations are not weighted by revenue. For example, compared to the 15.1% decline that is given in Table 3, the 90/10 ratio of $tfpq$ is only 7.2% larger than the 90/10 ratio of ϕ. The difference, between the unweighted and weighted calculations, is due to differences in the weight that particular industries get. When observations are not weighted by revenue, the ready-mix-concrete (which had a particularly small decline in productivity dispersion) is relatively more important in the calculations. On the other hand, when observations are weighted by revenue, the gasoline industry (which has a particularly large decline in productivity dispersion) is relatively more important in the calculations. Note that, weighting by revenue, does not cause the within-industry declines in dispersion to be systematically larger or smaller. For the sample of industries with substantial variation in output quality, there are no systematic differences between the weighted and unweighted calculations.

In Table 25, I compare plant-level characteristics between entrants and incumbents and between exiting plants and surviving plants. Unlike Table 5, each observation receives the same weight. The standard errors are approximately 2-3 times larger when observations are unweighted. In the unweighted regressions, entrants are no longer significantly more productive than incumbents. In addition, the materials prices of exiting plants and surviving plants are no longer significantly different from one another.

D.6 Sample restriction for regression 33

One concern about the regressions in Section 5.2 is that mismeasurement of buyer-supplier relationships may be contaminating the coefficient estimates.

Sample	Dispersion of $tfpq$ 90/10	75/25	SD	Dispersion of ϕ 90/10	75/25	SD	Percent Decline 90/10	75/25	SD	N
Milk-Bulk	0.809	0.306	0.316	0.681	0.322	0.303	20.6%	-4.8%	4.4%	127
Flour	0.404	0.205	0.163	0.396	0.176	0.172	2.0%	17.9%	-5.3%	503
Concrete	0.558	0.275	0.238	0.522	0.260	0.230	7.1%	5.8%	3.4%	3708
Yarn	0.620	0.310	0.262	0.629	0.312	0.248	-1.4%	-0.6%	5.6%	431
Boxes, Year≤1987	0.475	0.204	0.199	0.409	0.196	0.185	17.3%	4.1%	8.1%	1820
Gasoline	0.309	0.151	0.147	0.296	0.137	0.141	4.6%	10.7%	4.8%	692
Coffee	0.635	0.321	0.257	0.569	0.272	0.249	12.3%	19.6%	3.3%	300
Sugar	0.692	0.330	0.313	0.807	0.352	0.352	-13.2%	-6.3%	-10.6%	177
Boxes, Year≥1992	0.617	0.318	0.242	0.548	0.278	0.221	13.4%	15.7%	10.1%	646
Milk-Packaged	0.564	0.284	0.235	0.531	0.262	0.226	6.5%	8.9%	4.2%	2099
Pooled- Main Sample	0.527	0.253	0.227	0.493	0.238	0.219	7.2%	6.5%	3.9%	10503

Table 23: Dispersion of $tfpq$ and ϕ. Due to Census' rules regarding data confidentiality, I am prohibited from reporting the actual quantiles of any empirical distribution (e.g. ϕ). See the caption of Table 3 for a description of the imputation of the quantiles of $tfpq$ and ϕ.

Sample	Dispersion of $tfpq$ 90/10	75/25	SD	Dispersion of ϕ 90/10	75/25	SD	Percent Decline 90/10	75/25	SD	N
Wine	1.407	0.735	0.502	1.444	0.779	0.536	-2.6%	-5.4%	-6.1%	330
Softwood Cut Stock	1.379	0.675	0.490	1.310	0.651	0.489	5.4%	3.6%	0.2%	160
Pickles	0.890	0.441	0.346	0.962	0.471	0.368	-7.3%	-6.2%	-5.8%	145
Sausages	0.800	0.409	0.316	0.730	0.358	0.308	10.0%	15.3%	2.5%	621
Pooled	1.028	0.500	0.400	1.021	0.477	0.410	0.7%	4.8%	-2.5%	1256

Table 24: Dispersion of $tfpq$ and ϕ. Due to Census' rules regarding data confidentiality, I am prohibited from reporting the actual quantiles of any empirical distribution (i.e., ϕ). See the caption of Table 3 for a description of the imputation of the quantiles of $tfpq$ and ϕ.

Coefficient on:	Fixed Effects	$tfpq$	ϕ	$tfpr$	y	p^{in}	p^{out}
Entry	Year×Product	0.017*	0.020*	0.015*	-0.456*	0.005	-0.002
Entry	Year×Product	(0.008)	(0.007)	(0.006)	(0.032)	(0.006)	(0.006)
Entry	Year×Product×Division	0.014	0.019*	0.014*	-0.446*	0.009	0.000
Entry	Year×Product×Division	(0.008)	(0.007)	(0.006)	(0.032)	(0.005)	(0.006)
Exit	Year×Product	-0.025*	-0.016*	-0.020*	-0.534*	0.014*	0.005
Exit	Year×Product	(0.007)	(0.007)	(0.006)	(0.031)	(0.005)	(0.006)
Exit	Year×Product×Division	-0.024*	-0.015*	-0.019*	-0.528*	0.015*	0.005
Exit	Year×Product×Division	(0.007)	(0.007)	(0.006)	(0.031)	(0.005)	(0.006)

Table 25: Comparison of plant-level statistics and entry/exit status. In the first four rows, each cell gives the coefficient estiamte of β_1 in Equation 25. In the final four rows, each cell gives the coefficient estiamte of β_2 in Equation 26. In each regression, observations are weighted equally. Stars denote significance at the 5% level. N=10,503.

Sample	Boxes	Concrete	Pooled	Boxes	Concrete	Pooled
$\overline{tfpq_i}$	0.112	0.034	0.105	0.120	0.039	0.115
	(0.065)	(0.074)	(0.060)	(0.065)	(0.076)	(0.060)
N	148	43	191	148	43	191
Adjusted R^2	0.013	-0.018	0.013	0.029	-0.040	0.029
Division F.E.?	No	No	No	Yes	Yes	Yes

Table 26: Coefficient estimates and robust standard errors of the regressions defined by Equation 33. Stars indicate significance at the 5% level.

The imputation algorithm fails, most obviously, if there are multiple candidate recipients of a shipment. If there are multiple plants in the destination zip code that can purchase the good being shipped, my algorithm will assign each candidate plant to be a buyer of the sending establishment. Thus, the algorithm is over-counting the number of buyer-supplier relationships. To assess the magnitude of this problem, I re-estimate regressions 33 using only a subsample of buyer-supplier relationships. I restrict the sample so that the downstream plant is in a zip code with at most 1 other potential receiving plant.

Table 26 presents the results of regression 33, run on the restricted sample. Restricting the sample according to the number of potential recipients in the buyer's zip code has a larger effect for the "Ready-Mix Concrete" subsample. The sample restriction removes roughly two thirds of the ready-mix concrete plants in the sample, but only one quarter of the cardboard box manufacturers. Compared to the results presented in Table 9, the coefficient estimates are larger for the "Boxes" subsample, and smaller for the "Concrete" subsample. Overall, the standard errors are larger, especially for the subsample of ready-mix concrete manufacturers.

In summation, restricting the sample–to limit the number of buyer-supplier relationships that I am spuriously imputing–increases the estimated standard errors, but does not substantially change the coefficient estimates that are presented in Sections 5.2. Since there are more plants per zip code that can potentially purchase cement, the sample restriction has a larger effect on the subsample of ready-mix-concrete manufacturers.

E Statistical Significance of the Decline in Dispersion

In any regression, including additional explanatory variables mechanically lowers the unexplained variation of the dependent variable. In this section, I check whether the declines in dispersion that are reported in Tables 8, 10, and 13 are significantly more than would be expected by simply adding independent variables. To do so, I will create "fake" versions of the right hand side variables, and compute the unexplained variation that remains after including these "fake" variables as the only independent variables.

Consider, as an example, a particular version of the regression defined by Equation 32:

$$\tilde{p}_{it}^{in,CFS} = \beta_{\text{division}\times\text{year}} + \varepsilon_{it} \tag{50}$$

The standard deviation of the residuals in this regression is 0.220; see Table 8. This is 0.004 less than the standard deviation of $\tilde{p}_{it}^{in,CFS}$. To see whether this decline is greater than what would be expected, by chance alone, I implement the following algorithm 200 times:

1. Define a set of right hand side variables, $\mathcal{P}\left(\beta_{\text{division}\times\text{year}}\right)$, which are a random permutation of the set of the division-year fixed effects.

2. Regress $\tilde{p}_{it}^{in,CFS}$ against $\mathcal{P}\left(\beta_{\text{division}\times\text{year}}\right)$ and compute the standard deviation, s_ι, of the residuals from this regression.

Finally, I compute the mean, $\bar{s} = \frac{1}{200}\sum_{\iota=1}^{200} s_\iota$, and the standard deviation, $\bar{\sigma} = \sqrt{\frac{1}{199}\sum_{\iota=1}^{200}\left(s_\iota - \bar{s}\right)^2}$, of the sample of s_ι. The confidence interval, which appears in the first three rows of the second column of Table 27, is $\bar{s} \pm 1.96\bar{\sigma}$.

I follow a similar procedure for the following regression:

$$\tilde{p}_{it}^{in,CFS} = \overline{tfpq}_{it} + \varepsilon_{it} \tag{51}$$

Letting t_i be a set of normally distributed random variables (with the same standard deviation as $\overline{tfpq}_{it}$), I regress $\tilde{p}_{it}^{in,CFS}$ against t_{it}, 200 times. Again, I compute the standard deviation of the residuals, for each regression, and then compute a confidence interval using these standard deviations. The confidence intervals are given in the first three rows of the third column of Table 27.

The remaining cells of Table 27 are computed in a similar manner. In rows 4 to 6 of Table 27, the cells display the confidence intervals of the standard deviations of the residuals of different versions of Regression 33. The final three rows give the confidence intervals of the standard deviations of the residuals of different versions of Regression 38. (In the regression defined by Equation 38, ψ_{it} is an explanatory variable. In the final three rows of Table 27, I always include u_{it}, which is a normally distributed random variable with the same standard deviation of ψ_{it}.)

Include Division Fixed Effects?		No	Yes	No	Yes
Include $\overline{tfpq_{it}}$?		No	No	Yes	Yes
Sample Size	Sample				
131	Boxes		[0.1932;0.2022]	[0.2003;0.2031]	[0.1927;0.2017]
190	Concrete		[0.3194;0.3802]	[0.3635;0.3781]	[0.3168;0.3772]
321	Pooled		[0.2164;0.2238]	[0.2218;0.2250]	[0.2154;0.2236]
131	Boxes		[0.2066;0.2200]	[0.2166;0.2206]	[0.2060;0.2194]
190	Concrete		[0.1922;0.2106]	[0.2063;0.2117]	[0.1913;0.2093]
321	Pooled		[0.2075;0.2189]	[0.2160;0.2192]	[0.2067;0.2185]
131	Boxes	[0.1998;0.2034]	[0.1912;0.2022]	[0.1984;0.2032]	[0.1902;0.2016]
190	Concrete	[0.3647;0.3827]	[0.3151;0.3845]	[0.3588;0.3828]	[0.3123;0.3817]
321	Pooled	[0.2219;0.2251]	[0.2152;0.2234]	[0.2210;0.2250]	[0.2145;0.2231]

Table 27: 95% Confidence intervals. The first 3 rows give the confidence intervals corresponding to the dispersions given in Table 8. Rows 4-6 give the confidence intervals corresponding to the dispersions given in Table 10 . The final 3 rows give the confidence intervals corresponding to the dispersions given in Table 13.

References

Abreu, Dilip and Mihai Manea, 2011. "Markov Equilibria in a Model of Bargaining in Networks," *Games and Economic Behavior,* forthcoming.

Acemoğlu, Daron, Vasco M. Carvalho, Asuman Ozdağlar, and Alireza Tahbaz-Salehi, 2011. "The Network Origins of Aggregate Fluctuations," mimeo.

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer, 2006. "Structural Identification of Production Functions," mimeo.

Atalay, Enghin, Ali Hortaçsu, and Chad Syverson, 2012. "Why Do Firms Own Production Chains?" mimeo.

Baily, N. Martin, Charles Hulten, and David Campbell, 1992. "Dynamics in Manufacturing Plants," *Brookings Papers on Economic Activity: Microeconomics,* 1992, 187-249.

Balasubramanian, Natarajan and Jagadeesh Sivadasan, 2011. "What Happens When Firms Patent? New Evidence from U.S. Economic Census Data," *Review of Economics and Statistics*, 93(1): 126–146.

Bartelsman, Eric J., and Mark Doms, 2000. "Understanding Productivity: Lessons from Longitudinal Microdata," *Journal of Economic Literature*, 38(3): 569–594.

Bergin, Paul R., and Robert C. Feenstra. "Pass-Through of Exchange Rates and Competition between Floaters and Fixers," *Journal of Money, Credit, and Banking,* 41(1): 35-70.

Bernard, Andrew B. and J. Bradford Jensen, 1999. "Exceptional Exporter Performance: Cause, Effect, or Both?" *Journal of International Economics*, 47(1): 1–25.

Bernard, Andrew B. J. Bradford Jensen, and Peter Schott, 2006. "Transfer Pricing by U.S.-based Multinational Firms," mimeo.

Bloom, Nicholas, and John Van Reenen. 2010. "Why Do Management Practices Differ Across Firms and Countries?" *Journal of Economic Perspectives*, 24(1): 203–224.

Blundell, Richard, and Stephen Bond, 2000. "GMM Estimation with Persistent Panel Data: An Application to Production Functions," *Econometric Reviews*, 19(3): 321-340.

Chaney, Thomas, 2011. "The Gravity Equation in International Trade: An Explanation," mimeo.

Crawford, Vincent P., 1988. "Long-Term Relationships Governed by Short-Term Contracts," *American Economic Review,* 78(3): 485-499.

de Fontenay, Catherine C., and Joshua S. Gans, 2007. "Bilateral Bargaining with Externalities," mimeo.

Dubé, Jean-Pierre, Günter J. Hitsch, and Peter E. Rossi, 2010. "State Dependence and Alternative Explanations for Consumer Inertia," *RAND Journal of Economics,* 41(3): 417-445.

Eslava, Marcela, John Haltiwanger, Adriana Kugler, and Maurice Kugler, 2004. "The Effects of Structural Reforms on Productivity and Profitability Enhancing Reallocation: Evidence from Colombia," *Journal of Development Economics*, 75(2): 333-371.

Eslava, Marcela, John Haltiwanger, Adriana Kugler, and Maurice Kugler, 2009. "Trade Reforms and Market Selection: Evidence from Manufacturing Plants in Colombia," NBER Working Papers, 14935.

Foster, Lucia, John Haltiwanger, and C.J. Krizan, 2001. "Aggregate Productivity Growth: Lessons from Microeconomic Evidence," in New Developments in Productivity Analysis, University of Chicago Press.

Foster, Lucia, John Haltiwanger, and Chad Syverson, 2008. "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review*: 98(1): 394–425.

Fox, Jeremy and Valerie Smeets, 2011. "Does Input Quality Drive Measured Differences in Firm Productivity?" *International Economic Review,* 52(4): 961-989.

Gorodnichenko, Yuriy, 2010. "Using Firm Optimization to Evaluate and Estimate Productivity and Returns to Scale," mimeo.

Halpern, László , Miklós Koren, and Adam Szeidl, 2011. "Imported Inputs and Productivity," mimeo.

Heckman, James. J., 1981. "Statistical Models for Discrete Panel Data," in C.F. Manski and D. McFadden, eds., Structural Analysis of Discrete Data with Econometric Applications.

Hsieh, Chang-Tai, and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 124(4): 1403–1448.

Jones, Chad, 2011. "Intermediate Goods and Weak Links in the Theory of Economic Development," *American Economic Journal: Macroeconomics*, 3 (2): 1-28.

Katayama, Hajime, Shihua Lu, and James R. Tybout, 2009. "Firm-Level Productivity Studies: Illusions and a Solution," *International Journal of Industrial Organization*, 27(3): 403-413.

Kellogg, Ryan, 2011. "Learning by Drilling: Inter-Firm Learning and Relationship Persistence in the Texas Oilpatch," *Quarterly Journal of Economics,* 126(4):1961-2004.

Klein, Benjamin, Robert G. Crawford, and Armen A. Alchian, 1978. "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process," *Journal of Law and Economics*, 21(2): 297-326.

Klump, Rainer, Peter McAdam, and Alpo William, 2011. "The Normalized CES Production Function: Theory and Empirics," ECB Working Paper #1294.

Kremer, Michael, 1993. "The O-Ring Theory of Economic Development," *Quarterly Journal of Economics,* 108(3): 551-575.

Levinsohn, James and Amil Petrin, 2003. "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 70(2): 317-341.

Malik, Henrick J., 1967. "Exact Moments of Order Statistics from a Power Function Distribution," *Scandinavian Actuarial Journal,* 50(1-2): 64-69.

Manea, Mihai, 2011. "Bargaining in Stationary Networks," *American Economic Review,* 101(5): 2042-2080.

Oberfield, Ezra, 2011. "Business Networks, Production Chains, and Productivity," mimeo.

Olley, G. Steven and Ariel Pakes, 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64(6): 1263–1297.

Ornaghi, Carmine, 2006. "Assessing the Effects of Measurement Errors on the Estimation of Production Functions," *Journal of Applied Econometrics*, 21(6): 879–891.

Prescott, Edward C., 1998. "Needed: A Theory of Total Factor Productivity,"*International Economic Review*, 39 (3): 525–551.

Raval, Devesh, 2011. "Estimation of a CES Production Function with Factor Augmenting Technology," mimeo.

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

Sakellaris, Plutarchos and Daniel J. Wilson, 2004. "Quantifying Embodied Technical Change," *Review of Economic Dynamics,* 7(1): 1-26.

Syverson, Chad, 2004a. "Market Structure and Productivity: A Concrete Example," *Journal of Political Economy*, 112(6): 1181–1222.

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

Syverson, Chad, 2011. "What Determines Productivity?" *Journal of Economic Literature*, 49(2): 326–365.

Tesler, L.G., 1987. "A Theory of Self-Enforcing Agreements," *Journal of Business,* 53(1): 27-44.

U.S. Department of the Interior, 1989. *Cement Mineral Yearbook*.

U.S. Department of Transportation, Bureau of Transportation Statistics, 2009. Table 3-21: Average Freight Revenue Per Ton-mile.

Van Biesebroeck, Johannes, 2008. "The Sensitivity of Productivity Estimates," *Journal of Business and Economic Statistics*, 26(3): 311-328.