

AUSTRALIAN SCHOOL OF BUSINESS™

www.business.unsw.edu.au

THE UNIVERSITY OF NEW SOUTH WALES

# School of Economics Australian School of Business UNSW Sydney NSW 2052 Australia

http://www.economics.unsw.edu.au

# Quantifying Non-monotonicity in Productivity-Input Relations

Sasan Bakhtiari

School of Economics Discussion Paper: 2009/11

The views expressed in this paper are those of the authors and do not necessarily reflect those of the School of Economic at UNSW.

ISSN 1837-1035 ISBN 978-0-7334-2836-4

# Quantifying Non-monotonicity in Productivity-Input Relations

Sasan Bakhtiari Australian School of Business University of New South Wales Sydney NSW 2052, Australia Phone: (+61 2) 9385 3962 Fax. : (+61 2) 9313 6337 Email: s.bakhtiari@unsw.edu.au

March 2009

#### Abstract

It is stylized that productivity and input size should relate positively and monotonically in the long run. In this paper, I present a theory that unifies the role of demand and production to investigate conditions that make this relation a bell-shape. Under the optimality assumption and when establishments operate in the same market, I quantify a simple algebraic condition on demand and production elasticities that governs the relation between productivity and input size. The case where establishments face different demands is also considered using a simple trade model and the implications are shown to be qualitatively the same. Supportive evidence is obtained from plant-level data on ready-mix concrete. Findings of this paper have important implications on how productivity dispersion and size distribution are formed within industries.

Keywords: Productivity, Size Distribution, Demand Elasticity, Returns to ScaleJEL Classification: D21, D24, L11, L22, L25, L61.

<sup>\*</sup>Disclaimer: This document reports the results of research and analysis undertaken by the U.S. Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This document is released to inform interested parties of research and to encourage discussion. This research is a part of the U.S. Census Bureau's Longitudinal Employer-Household Dynamics Program (LEHD), which is partially supported by the National Science Foundation Grants SES-9978093 and SES-0427889 to Cornell University (Cornell Institute for Social and Economic Research), the National Institute on Aging Grant R01 AG018854-02, and the Alfred P. Sloan Foundation. The views expressed on statistical issues are those of the author and not necessarily those of the U.S. Census Bureau supports external researchers' use of these data through the Research Data Centers (see www.ces.census.gov). For other questions regarding the data, please contact Jeremy S. Wu, Program Manager, U.S. Census Bureau, LEHD Program, Room 6H141, 4600 Silver Hill Road, Suitland, MD 20746, USA (Jeremey.S.Wu@census.gov).

<sup>\*</sup> I am indebted to Prof. John Haltiwanger for his invaluable support. I am also deeply grateful to Hyowook Chiang for providing me with his data. I gratefully acknowledge John Shea, Daniel Vincent, Arghya Ghosh, Kevin Fox, Shiko Maruyama, and Yoonsoo Lee for their insightful comments and advice.

## 1 Introduction

This study mainly addresses the formation of size distribution among heterogeneous producers in a new light by deviating from the traditional view that mapping from productivity distribution to that of input size is one-to-one. I present a theory that unifies the role of both demand and production sides in turning this relation into a non-monotonic one, most likely bell-shaped with one or even several peaks. Besides, I rely on plant-level data to show this fact and its implications empirically. It turns out that the relevant condition governing how input level changes with productivity is

$$\varepsilon_{MR} + \varepsilon_{\nu} = -1, \tag{1.1}$$

where  $\varepsilon_{MR}$  is the elasticity of marginal revenue and  $\varepsilon_{\nu}$  is the elasticity of returns to scale (RTS), both with respect to output. Prima facie, this formula can be useful in two ways: 1) it helps to envision the shape of input size relations with a known demand function; 2) it can suggest counter-factual demand functions that exhibit certain features in their input size relation. The latter application can also find use in identifying an industry's unobserved aggregate demand curve by utilizing observables such as employment, sales, and productivity.

The first implication of my result is naturally a need to re-examine how size distributions are formed as a result of productivity differences. In addition to the intensity of capital utilization discussed by Rossi-Hansberg & Wright (2007), differences in market structures can also be a reason for differences in the thickness of upper-tail in size distributions across industries when mid-productivity plants are the contributors. The conclusions of this study also offer new perspectives into causes of productivity dispersion. With a bell-shaped relationship, in particular, ranges of productivity differences can be created and shaped at any given input level, even in the absence of shocks and uncertainties. Trade gains can also be far greater than previously thought. The opening of new markets would now have two expansionary effects: as Melitz (2003) argues, it causes a shift in employment and output shares to the incumbent producers with higher productivities, and it also lets very productive plants to act "normally" by significantly expanding their workforce in response to the increase in demand. Welfare losses as a result of employment frictions, such as Hopenhayn & Rogerson (1993) adjustment costs, can also be smaller than thought as small plants might not be suffering from those frictions but from demand constraints.

From the literature, however, the standard picture is always a monotonic one. In an early work, Lucas (1978) uses difference in management skills to generate distributions of productivity and size, and at the end more productive firms hire larger labor. In Jovanovic (1982), Hopenhayn (1992), Ericson & Pakes (1995), and Luttmer (2007) growth happens as producers accumulate idiosyncratic shocks to their productivity. The evolution of industry happens as less efficient producers, or those hit by a string of bad shocks, realize that they can never be profitable and exit market, reallocating their resources to entrants and/or more productive units. More productive units, nevertheless, grow fast and become large in the long run. Bontemps, Robin & Van Den Berg (2000), and Bertola & Garibaldi (2001) achieve similar results with the search and matching models of employment. Constant-elasticity demand is a favorite among theorists and is one demand function that can create monotonic relation between input and productivity. A large body of theoretical works, such as Melitz (2003) and Luttmer (2007), walk along this line. Melitz & Ottaviano (2008) use a linear demand in their analysis which, by the analysis in this paper, would cause non-monotonic productivity-input relation, thought they avoid any discussion of input sizes. This attitude towards a positive relation between input and productivity seems to have become so stylized that thinking otherwise attracts harsh criticism.

Interestingly enough, the available empirical works provide very few details on the exact shape of possible productivity-size relations. It is widely documented that average productivity is (slightly) higher in larger employment classes<sup>1</sup>. But, if we believe that this trend is enough evidence that employment should positively relate to productivity, then I have to confront that with Table 1. In all two-digit manufacturing industries, establishments demonstrate very weak, and sometimes negative, correlations between their employments and productivities<sup>2</sup>. Bakhtiari (2008) shows that focusing on plants older than six years does not change the implications significantly, underlining the need for a long-run explanation. In this paper, the contradiction is addressed by pointing out the combined role of both the supply and demand side in forming a non-monotonic relation between productivity and input size, which in turn explains why correlations observed in the data are so low.

My approach is mostly a rerun of the firm production problem, but I rely on very general forms of revenue and cost function with no peculiar assumption attached to a specific market or production process. I first assume that all establishments operate in the same market, possibly in a monopolistic competition, and derive an algebraic condition on the elasticities of marginal revenue and RTS that specifies the slope of productivity-input relation at any point (Equation (1.1))<sup>3</sup>. Using a simple trade model, I then extend the theory to encompass the cases where establishments face different demands and show that despite the fact that more monotonicity can be achieved in the relation, the qualitative

<sup>&</sup>lt;sup>1</sup>See, for example, Bartelsman & Dhrymes (1998).

<sup>&</sup>lt;sup>2</sup>Section 4 describes how productivities are computed.

<sup>&</sup>lt;sup>3</sup>The reason that the condition is stated using the elasticity of marginal revenue is that using the more familiar elasticity of demand or revenue results in a differential equation which complicates further discussion.

SIC	$\operatorname{corr}(\mathbf{E}, \operatorname{rtfp})$	SIC	$\operatorname{corr}(E, rtfp)$
20	0.106	30	0.079
22	-0.088	31	-0.019
23	-0.049	32	0.211
24	-0.017	33	-0.060
25	-0.006	34	-0.026
26	0.160	35	-0.112
27	0.100	36	0.064
28	0.128	37	0.015
29	-0.345	38	0.075
All	0.007		

Table 1: Correlations between employment and log productivity (revenue total factor productivities is used).

features of a bell-shape still hold.

Geometrically, the elasticity condition derived here explains why it is hard to achieve monotonicity: to have a fully monotonic relation between productivity and input, the revenue function, especially, has to stay above a logarithmic function, while almost all revenue functions are bounded above because of finite consumer demand. Higher RTS close the range of possibilities and diminishing RTS demand a faster growing revenue function, for a monotonic relation to be possible.

Empirical exercises are focused on demonstrating existence of a bell-shaped relation among the concrete plants. I use the concrete industry as my pilot study for several reasons. The localized structure of the market for concrete proves very useful in characterizing market type and demand size. The average concrete plant is very mature, and homogeneity of concrete provides me with a more accurate measure of productivity. However, product homogeneity does not rule out productivity dispersion because concrete is a spatially differentiated product (Syverson 2004). Using non-parametric and semi-parametric models, the productivity–employment relation in concrete is shown to resemble a bell-shape and the form of relation is shown to be robust to changes in market size. A positive correlation between average size and market demand is also revealed in the results. As a testable implication, I show that bootstrapped simulations assuming a bell-shape relation between productivity and employment push the corresponding correlation in the correct direction and come close to the numbers observed in the data.

Lastly, market and production structures are not presented as a final verdict in the analysis of nonmonotonicity, though the empirical results of this paper draw a fairly favorable picture of their influence. In practice, many producers might be affected by suboptimal decision making, also causing the input– productivity relationship to deviate from its monotonic form. It is not clear if the inefficiency of decisions is spread amongst plants consistently to induce bell-shaped relations; detailed study of those effects currently falls out of the scope for this paper. However, assuming that inefficiency and its extent are randomly and uniformly distributed among plants, input and output trends should not change, making the effect of suboptimal decisions trivial in my study.

The rest of the paper is organized as follows. The following section derives the condition under which input size falls with productivity when plants all operate in the same market. It also presents some examples. Section 3 discusses the relation when plants can trade across markets and face different demands. Section 4 describes my data, and in Section 5 I investigate the empirical relationship between productivity and employment for the concrete industry. The paper is then concluded.

# 2 Size Relations: A Single Market

Plants produce output level q by incurring a cost of  $C(q, \phi)$ .  $\phi$  is a productivity parameter so that higher values of  $\phi$  correspond to lower costs of production. Function C accounts for variable costs of production as well as any possible fixed overhead costs. Assume C(.,.) is smooth enough, and let subscripts denote partial derivatives with respect to that argument. General properties of the cost function are set out below as obvious physical facts.

**Assumption 2.1.** Cost function  $C(q, \phi)$  is such that at every  $q \ge 0$  and  $\phi \ge 0$ :

- (i)  $C(q, \phi) > 0$ ,
- (ii)  $C_q > 0$  and  $C_{\phi} < 0$ ,
- (iii) C(.,.) is multiplicative separable, i.e., C(q, φ) = c<sub>1</sub>(q)c<sub>2</sub>(φ) for some functions c<sub>1</sub> and c<sub>2</sub> satisfying
  (i) and (ii).

Properties (i) and (ii) are natural for any valid cost function. Assumption (i) reiterates the "no free-lunch" condition and also implies that there might be some fixed costs of operation (especially if  $C(0, \phi) > 0$ ). By Assumption (ii), producing every extra unit of output involves a positive cost, with productivity acting as cost reducing. Item (iii) ensures that relative changes in cost can be decoupled into additive output and productivity effects, as is the case with most popularly used cost functions. An important result of such decoupling is that the production returns to scale (RTS) will be independent of productivity and will be a function of the output level only. To see that, notice that RTS relate to cost function as (Zellner & Ryu 1998)

$$\nu(q) = \frac{C(q,\phi)}{qC_q(q,\phi)}.$$
(2.1)

Plants also make revenue of R(q) by producing output level q, which is the same for all plants in the

same market<sup>4</sup>. The revenue function is also assumed smooth enough, with properties described in the next assumption.

Assumption 2.2. Revenue function R(q) is such that:

- (i) R(0) = 0 and  $R_q(0) > 0$ ,
- (ii)  $R_{qq} < 0.$

The assumption establishes that revenue function is concave and increasing in output level in a range  $q \in [0, \bar{q})$ , where  $\bar{q}$  can be finite or infinite and defines the maximum feasible output under the optimality condition. A plant's profit function is simply  $\pi = R(q) - C(q, \phi)$ . In a textbook manner, a bounded solution for q exists where marginal cost equals marginal revenue. Plants hire x units of a composite input to meet their production. This input is thought of as aggregating the role of different production factors (such as labor, capital, material, intangibles, etc.) Zellner & Ryu (1998) show that the RTS can also be expressed as  $\nu(q) = \frac{dq}{dx} \frac{x}{q}$ . Eliminating  $\nu$  between this relation and (2.1) and solving the resulting partial differential equation yields<sup>5</sup>

$$x = \frac{C(q,\phi)}{w}.$$
(2.2)

where w is the unit price of input and is normalized to 1 henceforth.

#### 2.1 Local Non-Monotonicity

The following definitions prove useful in the coming analysis.

**Definition 2.1.** Let the output-elasticity of marginal revenue be  $\varepsilon_{MR} = \frac{qR_{qq}}{R_q}$ .

**Definition 2.2.** Let the output-elasticity of returns to scale be  $\varepsilon_{\nu} = \frac{q\nu_q}{\nu}$ .

Note that by Assumption 2.2,  $\varepsilon_{MR}$  is always negative. However, the sign of  $\varepsilon_{\nu}$  is kept ambiguous to make results applicable to a variety of RTS functions.

Local properties of  $q(\phi)$  and  $x(\phi)$  are discussed by determining the signs of derivatives with respect to  $\phi$  under the optimality conditions. The general form of the relation  $q(\phi)$  is rather straight-forward as depicted by the following proposition.

**Proposition 2.1.** Let R(.) and C(.,.) satisfy Assumptions 2.1 and 2.2. Then, more productive plants produce more output, i.e.  $\frac{dq}{d\phi} > 0$ .

 $<sup>{}^{4}</sup>$ A common revenue function for all producers in a market can be thought of as the outcome of a monopolistic competition, where each producer is atomistic, hence everybody responds to market aggregates rather than strategic interactions.

 $<sup>^{5}</sup>$ Assuming that cost function is linear homogeneous in input prices, Shephard's lemma leads to the same result.

**Proof:** Starting from the first order condition  $R_q(q) - C_q(q, \phi) = 0$  and taking derivatives with respect to  $\phi$  gives

$$(R_{qq} - C_{qq})\frac{dq}{d\phi} = C_{q\phi}.$$
(2.3)

By second-order optimality condition we have  $R_{qq} - C_{qq} < 0$ . Therefore, to show that output is an increasing function of  $\phi$ , it only requires that  $C_{q\phi} < 0$ . But, from (2.1),  $C_q = \frac{C}{q\nu}$  and because of Assumption 2.1(iii) the only explicit dependence on  $\phi$  appears in C(.,.). Therefore

$$C_{q\phi} = \frac{\partial}{\partial \phi} C_q = \frac{\partial}{\partial \phi} \left( \frac{C}{q\nu(q)} \right) = \frac{C_{\phi}}{q\nu(q)} < 0,$$

which completes the proof.

One important implication of Proposition 2.1 is that, in equilibrium, output level is a unique and oneto-one mapping of productivity. Hence, in most discussions, productivity and output play an equivalent role.

**Lemma 2.1.** Let Assumptions 2.1 and 2.2 hold. Then  $\varepsilon_{MR} + \varepsilon_{\nu} < \frac{1}{\nu(q)} - 1$ .

**Proof:** Note that  $C_{qq}$  can be written as follows

$$C_{qq} = \frac{\partial}{\partial q}(C_q) = \frac{\partial}{\partial q} \left(\frac{C}{q\nu(q)}\right)$$
$$= \frac{C_q}{q\nu(q)} - \frac{C}{q\nu(q)}\frac{\nu_q}{\nu(q)} - \frac{C}{q\nu(q)}\frac{1}{q}$$

But, from (2.1),  $\frac{C}{q\nu(q)} = C_q$ . The first-order condition also requires that  $C_q = R_q$ . Replacing both in the above equation results in

$$C_{qq} = \frac{R_q}{q} \left( \frac{1}{\nu(q)} - \varepsilon_{\nu} - 1 \right).$$
(2.4)

The second-order condition requires that  $R_{qq} < C_{qq}$ . Replacing with (2.4) into this condition yields

$$R_{qq} < \frac{R_q}{q} \left(\frac{1}{\nu(q)} - \varepsilon_{\nu} - 1\right),$$

which can be rewritten as

$$\frac{qR_{qq}}{R_q} + \varepsilon_\nu < \frac{1}{\nu(q)} - 1$$

The proof is complete by applying the definition of  $\varepsilon_{MR}$ .

The above lemma defines the feasible region for the elasticities where solutions to the production

\_

problem exist and are finite. The following proposition is the main result specifying conditions that result in non-monotonicity in  $x(\phi)$  relation:

**Proposition 2.2 (Local Non-Monotonicity).** More productive plants hire larger input in a neighborhood of  $\phi$ , i.e.  $\frac{dx}{d\phi} > 0$ , if and only if

$$-1 < \varepsilon_{MR} + \varepsilon_{\nu} < \frac{1}{\nu} - 1.$$

Conversely, more productive plants hire smaller input in a neighborhood of  $\phi$ , i.e.  $\frac{dx}{d\phi} < 0$ , if and only if

$$\varepsilon_{MR} + \varepsilon_{\nu} < -1.$$

The relation  $x(\phi)$  peaks where  $\varepsilon_{MR} + \varepsilon_{\nu} = -1$ .

**Proof:** Since  $x(\phi) = C(q(\phi), \phi) = c_1(q(\phi))c_2(\phi)$ , then for  $x(\phi)$  to be decreasing, it must be that

$$\frac{dx(\phi)}{d\phi} = c_1'(q)c_2(\phi)\frac{dq(\phi)}{d\phi} + c_1(q)c_2'(\phi) < 0,$$

where primes denote derivatives. Take the inequality  $c'_1(q)c_2(\phi)\frac{dq(\phi)}{d\phi} + c_1(q)c'_2(\phi) < 0$ . Note that, from (2.3),  $dq/d\phi = c'_1(q)c'_2(\phi)/(R_{qq}-C_{qq})$ . Replace this into the inequality and multiply by  $q(R_{qq}-C_{qq})\frac{c_2(\phi)}{c'_2(\phi)}$ . Note that the direction of inequality does not change because  $c'_2(\phi) < 0$  and  $R_{qq} - C_{qq} < 0$  (second-order condition), hence the multiplier is positive. Use  $c_1(q)c_2(\phi) = C(q,\phi)$  and  $R_q = C_q$  (first-order condition) and simplify the inequality to get

$$\frac{qR_{qq}}{R_q} - \frac{qC_{qq}}{C_q} + \frac{qC_q}{C} < 0.$$

By definition, the following results hold

$$\epsilon_{MR} = \frac{qR_{qq}}{R_q}, \quad \epsilon_{\nu} = \frac{qC_q}{C} - \frac{qC_{qq}}{C_q} - 1.$$

Now, replace these results into the inequality to get  $\epsilon_{MR} + \epsilon_{\nu} < -1$ . The condition for  $x(\phi)$  increasing can also be derived by changing the direction of inequalities. The upper bound on  $\epsilon_{MR} + \epsilon_{\nu}$  comes from Lemma 2.1.

Figure 1 illustrates the relevant ranges of elasticities in which input size increases or decreases with  $productivity^{6}$ . Proposition 2.2 is a description of how the demand-side and supply-side influence the

<sup>&</sup>lt;sup>6</sup>In (2.2), input price does not have to be fixed with productivity. A  $w(\phi)$  with  $w_{\phi} > 0$  means that more productive firms, for instance, are willing to offer higher wages or higher rental prices. In that case, it is possible to show that the



Figure 1: Ranges of values for elasticities and the direction of change for input and output.



Figure 2: If  $\nu_2 > \nu_1$ , then x increases when area B is larger than area A.

direction of change for  $x(\phi)$ . An incrementally more productive plant produces more output at a lower price. Keeping marginal revenue constant, plants are bound to hire less input with higher RTS. At the same time, cost savings increase the output level. Proposition 2.2 states that input level will increase with productivity if the latter effect is larger than the former (Figure 2). However, marginal revenue is not constant. An incrementally more productive plant will optimally increase its input level only if it receives a large enough boost in revenue as a result. With diminishing marginal revenue, this condition becomes less probable as the productivity level gets higher. Very high RTS at any point also close the range of possibilities for a positive  $x(\phi)$  relation.

#### 2.2 Geometric Interpretation

To show the possibility range of revenue functions that can achieve a positive one-to-one relation between productivity and input, let's assume  $\nu$  is fixed (I generalize to varying  $\nu$  afterward). Proposition 2.2 can then be written as

$$-\frac{1}{q} < \frac{R_{qq}}{R_q} < \frac{\frac{1}{\nu} - 1}{q}.$$
 (2.5)

condition from Proposition 2.2 changes to

$$\frac{dx}{d\phi} \ge 0 \quad \Leftrightarrow \quad \varepsilon_{MR} + \varepsilon_{\nu} \ge \frac{1}{\nu} (\Xi - 1) - 1,$$

where  $\Xi \in (0, 1]$  and depends on elasticities of input price and cost to productivity. In particular, with fixed  $w, \Xi = 1$ .



Figure 3: (a) The permissible range of normalized revenue functions for  $x(\phi)$  to be positive monotonic. (b) Change in the permissible range with diminishing RTS.

Let  $x(q(\phi))$  be monotonically increasing in the interval  $[q_*, \infty)$ . Then, the revenue function has to satisfy

$$q_* \log\left(\frac{q}{q_*}\right) < \frac{R(q) - R(q_*)}{R_q(q_*)} < \nu q_*^{1-\frac{1}{\nu}} q^{\frac{1}{\nu}} - \nu q_*.$$
(2.6)

The middle term is the normalized revenue function so that it starts from zero with slope 1 at  $q_*$ .  $q_*$  can be regarded as the cutoff output level. With  $C(0, \phi) > 0$ , plants producing output levels below some  $q_*$  make negative profit and leave the market in equilibrium. With a one-to-one mapping between productivity and output, there exists a corresponding cutoff productivity  $\phi_*$  for which plants do not produce if  $\phi < \phi_*$ . Plants with  $\phi = \phi_*$  are indifferent between producing quantity  $q_*$  and exit.

The upper and lower ranges of a revenue function that can generate a monotonic  $x(\phi)$  are illustrated in Figure 3(a). The upper limit is the feasibility restriction, so that bounded solutions exist. At the same time, revenue function has to stay above a logarithmic function to stay within those bounds. If the RTS diminishes with q, so that  $\varepsilon_{\nu} < 0$ , a positive term is added to both sides of (2.5). In addition, the upper limit is also affected by  $\nu$  falling. Together, they help shift both limits upwards, with the upper limit shifting by a larger value. This effect is depicted in Figure 3(b). In this case, revenue has to increase with output, on average, even faster. With the same reasoning, if the RTS go up with q, both limits in (2.6) shift downward, with the upper limit moving faster. In this case, the range in (2.6) closes quickly.

#### 2.3 Some Examples

For simplicity, let's assume that  $\nu$  is fixed and constant and plants produce according to

$$q = \phi x^{\nu}.\tag{2.7}$$

It is easy to check that the cost function associated with this production is  $C(q, \phi) = (q/\phi)^{\frac{1}{\nu}}$  and satisfies Assumption 2.1. With fixed  $\nu$ , the relevant condition becomes  $\varepsilon_{MR} = -1$ . Testing this condition with a linear inverse demand function of the form  $p(q) = p_0 - p_1 q$  gives

$$\frac{dx}{d\phi} \leq 0 \quad \text{if} \quad q(\phi) \geq \frac{p_0}{4p_1}. \tag{2.8}$$

Using the one-to-one mapping of output to productivity in this example, one can equivalently express the above condition in terms of productivity as follows

$$\frac{dx}{d\phi} \leq 0 \quad \text{if} \quad \phi \geq \left(\frac{2}{\nu p_0}\right)^{\nu} \left(\frac{p_0}{4p_1}\right)^{1-\nu}.$$
(2.9)

Figure 4 shows  $q(\phi)$  and  $x(\phi)$  generated using  $p_0 = 7.8$  and  $p_1 = 0.0335$ <sup>7</sup> and for three different values of  $\nu = 0.8$ , 1, 1.2. The figure also features the cutoff productivity for each case.

It can be shown that  $x(\phi)$  relation need not necessarily be unimodal. A careful choice of demand function can create an arbitrary form of  $x(\phi)$ . To show that, I reverse engineer a specific demand function by assuming constant  $\nu$  and setting  $1 + \varepsilon_{MR}$  equal to the following polynomial

$$1 + \varepsilon_{MR} = a \prod_{i=1}^{n} (Z_i - q), \qquad (2.10)$$

and then strategically placing the zeros,  $Z_i$ , to shape the form of  $x(\phi)$  in desired ways. Some constraints also apply. First of all, to have a falling upper tail, n has to be odd and  $\prod_{i=1}^{n} Z_i > 0$ . The position of each zero and the distance between zeros,  $|Z_{i+1} - Z_i|$ , determine curvature. The location of peaks in  $x(\phi)$  can be controlled with each  $Z_i$ . Also, the closer two zeros are, the smaller the trough in between the corresponding peaks will be. If zeros are farther apart, then a larger trough can be generated. For example, let  $1 + \varepsilon_{MR} = \frac{1}{11000}(10 - q)(20 - q)(40 - q)$ . The corresponding revenue, inverse demand, productivity-output relation, and productivity-input relation for three different values of  $\nu = 0.8$ , 1, 1.2 are shown in Figures 5 and 6. In this case, using three zeros, a bimodal  $x(\phi)$  relation is created. Notice that the demand function in Figure 6 looks almost indistinguishable from a CES demand, but with totally different implications.

<sup>&</sup>lt;sup>7</sup>These numbers come from calibrating this model to the concrete data in Section 5.3.



Figure 4: The output–productivity and input–productivity relations with linear demand.



Figure 5: Revenue and demand functions generating a multi-modal input relation.



Figure 6: Resulting output and input relations with the demand in Figure 5.

### 3 Size Relations: Several Markets

So far, all plants were supposed to be operating in the same market and facing the same demand. When costly trade is possible, Melitz (2003) shows that more productive units actually engage in trade and increase their demand as a result. At the same time, less productive units supply only domestically. Consequently, the very presence of trade possibilities subjects producers to different demands depending on their production efficiency. Using a parsimonious trade model similar to that of Melitz (2003), I show that differences in demand combines bell-curves for different markets to generate a more monotonic relation to the advantage of high-productivity plants. However, in the big picture, the relation still has a falling upper tail.

Let trade be possible with N other markets (N + 1 total markets) at a marginal cost of  $\tau_i$  for outside market  $i^{8}$ . For simplicity, assume constant  $\nu$  and that all markets face the same linear demand curve. Also assume that w = 1 in all markets, both before and after trade opens. Then, profit function for plant j will be:

$$\pi = (p_0 - p_1 q^D) q^D + \sum_{i=1}^N (p_0 - p_1 q^O_i) q^O_i - \frac{q^D + \sum_{i=1}^N q^O_i}{\phi_j} - \sum_{i=1}^N \tau_i q^O_i,$$
(3.1)

where D and O superscripts refer to the domestic and outside markets, respectively. Solving the first order conditions yields

$$q_j = \frac{p_0 \phi_j - 1}{2p_1 \phi_j} I[\phi_j > \phi_*^D] + \sum_{i=1}^N \frac{(p_0 - \tau_i)\phi_j - 1}{2p_1 \phi_j} I[\phi_j > \phi_{*i}^O],$$
(3.2)

where I[] is the indicator function, and domestic and foreign cutoff productivities are

$$\phi_*^D = \frac{1}{p_0}, \quad \phi_{*i}^O = \frac{1}{p_0 - \tau_i}, i = 1, \dots, N.$$
 (3.3)

Figure 7 is based on using a linear demand and demonstrates how trade with multiple markets adds more peaks to  $x(\phi)$ , resulting in monotonicity over a wider range of productivities compared to the non-trading case. In this example, I am still using  $p_0 = 7.8$  and  $p_1 = 0.0335$ , and I let trade be possible with five other markets with trading costs  $\tau_i = \{3, 3.5, 4, 5, 6\}$ . Note that a huge increase in market shares of highproductivity producers still does not rule out a bell-shaped relation, though a wider bell is generated as a result.

 $<sup>^{8}\</sup>mathrm{I}$  use the term "outside" to signify that other markets need not be international.



Figure 7: The shape of productivity–output and productivity–input relations when it is possible to trade with five other markets.

#### 4 Data

#### 4.1 The Concrete Industry as Test Bench

Results of Section 2 rely on the characterization of a market, so that all plants belonging to the same market face the same, or very similar, demands. One class of industries where a market is easier to identify is the localized-market industries. In industries whose products are mostly traded locally, the center of trade is mainly an urban or industrial area with boundaries conveniently defined by already available geopolitical borders such as county and state lines. In practice, even within a localized-market industry, trade can still cross these borders for a number of traders, but the concentration of activity makes defining markets for these industries a more feasible task compared to other industries.

Among industries with a localized market, the ready-mix concrete (SIC 3273) has many attractive features to make it suitable for study. First, due to the high costs of transportation, concrete is not shipped very far compared to many other products<sup>9</sup>. Therefore, it qualifies as a localized-market industry.

Second, concrete is a very homogeneous product. As a result, the magnitude of revenue variation due to quality or taste differences is largely minimized, leaving mostly physical productivity to drive differences in revenue productivity across plants. Foster, Haltiwanger & Syverson (2008) demonstrate this fact empirically by showing that revenue and physical productivities behave mostly the same in several industries, including concrete, where output is mostly homogeneous. This characteristic of concrete is especially useful since most data, including mine, lack information on input and output prices and can provide estimates of revenue productivity only.

Third, the homogeneity of concrete does not rule out the presence of productivity dispersion, even at

<sup>&</sup>lt;sup>9</sup>The US Bureau of Transportation Statistics' Commodity Flow Survey reports that concrete plants shipped their products to an average radius of 64 miles in 1993 and 82 miles in 1997.

the equilibrium. Syverson (2004) shows that, because transportation costs for concrete are high, customers make purchase decisions based not only on efficiency of production but also on physical distance. This finding illustrates that concrete is a diverse product, not by variety, but by spacial differentiation. As a result of this diversity, a wide range of productivities are present in the data, enabling me to study the productivity–size relation.

Finally, more than 86% of all concrete plants and about 76% of those plants with less than 10 employees are at least four years old (the average age is about 15 years). This fact is very likely caused by the spatial differentiation of products, which limits competition, entry and exit. As for my results, I benefit from the fact that the effect of entries and dynamics of young plants is largely minimized due to the maturity of the average plant<sup>10</sup>.

I will define market size as the population of construction workers in an urban area. Syverson (2004) discusses the suitability of such a definition by arguing that the construction industry is the main consumer of ready-mix concrete, while the cost of concrete is a small share of construction costs. This makes the demand measure reasonably with productivity shocks to concrete.

#### 4.2 Data on the Concrete Industry

The source for my data is the US Center for Economic Studies' Census of Manufactures (CM) panels 1982, 1987, 1992, and 1997. The CM spans the universe of manufacturing plants in the US, with plant defined as an individual physical place of production and identified with a Plant Permanent Number (PPN). Some of the reported variables in the CM are total shipment value, employment for production and non-production workers, total hours worked, book values of and investment in machinery and structures and costs of energy and materials. For each plant, the four-digit Standard Industry Classification (SIC), product class, and location (state-county) are also reported in the CM<sup>11</sup>. The location information, especially, enables me to link each plant geographically to its corresponding market defined below. I use the real values for input and output constructed by Chiang (2005). Specifically, Chiang uses the 4-digit deflators available from the NBER/CES Productivity Database<sup>12</sup> and estimates real equipment and structure capital from a perpetual inventory model and the NBER estimated depreciation rates.

Many of the CM records are flagged as administrative records, for which all data except employment is imputed. The quality of the imputed data is in serious doubt. For that reason, I use the weighted CM subsamples for my analysis and estimates. This leaves me with 2,027 sample concrete plants.

 $<sup>^{10}</sup>$ Davis, Haltiwanger & Schuh (1996) discuss how job creation and destruction rates change sharply from two to four year-old plants, yet change very slowly as plants get older than four years.

<sup>&</sup>lt;sup>11</sup>Some of the state-county data were missing or erroneous. These were fixed by matching the CM to the Census Bureau's Standard Statistical Establishment List (SSEL).

 $<sup>^{12}\</sup>mathrm{Refer}$  to Bartelsman & Gray (1996) for more details.

The revenue Total Factor Productivity (rTFP) is used to measure productivity based on a Cobb-Douglas production function and is computed using input cost shares and the deflated revenue as real output. Formally, for plant j at time t, rTFP is defined as

$$rtfp_{jt} = q_{jt} - \alpha^h h_{jt} - \alpha^{eq} k_{jt}^{eq} - \alpha^{st} k_{jt}^{st} - \alpha^e e_{jt} - \alpha^m m_{jt},$$

$$\tag{4.1}$$

where lower case letters label variables in logs. Here q is the nominal output deflated by the industryspecific price index. h is labor input (total hours worked), and  $k^{eq}$  and  $k^{st}$  are the equipment and structures capital stocks, respectively. e is energy and m is material input. The  $\alpha$  coefficients for concrete are computed using the cost share indexes described by Chiang (2005). To make productivities comparable over the selected range of years, I use residuals from regressing productivity values on year dummies. I then re-adjust the mean value of the residual productivities to be equal to the original total mean.

Finally, in the coming empirical results, instead of measuring a composite input, I will measure the input size of plants by their total employment (TE) as defined by Davis et al. (1996, Appendix A.3.1). Employment is easily observed for each plant and has reasonably low measurement error compared to estimates of a composite input. In defense of this shift, I find that the correlation between total employment and total hours for concrete plants is 0.95. Besides, if the relative intensity of productive factors is assumed constant, the optimal choice of each input factor will be a constant proportion of total hours, or alternatively a constant proportion of total employment. This enables me, at least for the concrete industry, to treat the production function (2.7) as if it depended on employment only.

#### 4.3 Demand Market

Due to availability of detailed data and required crosswalks, I use Core-Based Statistical Areas (CBSA) as markets for concrete plants. A CBSA is a functional region around an urban center. The CBSA system includes a mix of micro- and metropolitan areas in the United States, providing me with a sufficiently large range of market sizes<sup>13</sup>. Economic activity is mostly concentrated within a CBSA, making it a suitable candidate for market analysis, though the degree of market isolation can still depend on the physical proximity of CBSA's.

Market size is measured as the population of construction workers (SIC 15– to 17–) aggregated to the CBSA level. Construction employment is obtained from the County Business Patterns aggregated to the

 $<sup>^{13}</sup>$ The US Office of Management and Budget's definition of a metropolitan area is an urban area with the population of at least 50,000. Micropolitan areas are those with the population between 10,000 and 50,000.

Mean	Std.Dev.	Min.	Median	Max.
43,173.4	58,022.6	48	$16,\!600$	327,397

Table 2: Summary statistics for population of construction workers in different markets.

CBSA level and matched by CBSA-year<sup>14</sup>. There are 667 markets that match to my subsample. More detailed statistics for this market definition can be found in Table 2.

# 5 Empirical Results

In this section, I undertake a series of exercises with two major goals. First, I show that a bellshaped productivity-employment relation best fits the concrete data, using a range of parametric to non-parametric methods. Second, I explain the low correlations in the data by non-monotonicity in the productivity-input relation. The effect of the market size on the productivity-employment relation is also highlighted in the exercises<sup>15</sup>.

#### 5.1 Non-Parametric Estimation of the Relationship

To see how input and output relate to productivity in concrete, while imposing the least constraints, I estimate the following non-parametric relations

$$TE_j = H_1(\log(\phi_j)) + \epsilon_j, \qquad (5.1)$$

$$Q_j = H_2(\log(\phi_j)) + \zeta_j. \tag{5.2}$$

TE is the total employment at plant j and Q is the deflated value of output. I am leaving out time effects for the moment to increase the number of observations used in estimation. Later, the time effect will be included when estimating the relation semi-parametrically.

I estimate  $H_i(.)$ , i = 1, 2 using the Nadaraya–Watson kernel regression with a Gaussian kernel (Simonoff 1996). In my preferred setting, I choose a fixed bandwidth of 0.4 (for log productivity). This choice enables me to demonstrate the qualitative nature of both relations, while filtering excess variations due to noise and disturbances. Estimation is done for 1,000 points spaced logarithmically along the productivity axis and using all the available observations on concrete plants. Figure 8 shows the es-

<sup>&</sup>lt;sup>14</sup>The employment data for some of the counties is suppressed to protect confidentiality of the data. I follow Syverson's method to impute those data. Basically, since the number of employers in several different size groups is being reported, I will multiply the number by mid point of the size range and sum up to generate the impute. Also, the County Business Pattern reports data as early as 1986. For that reason, I link my 1982 panel to the 1987 data on the worker population.

 $<sup>^{15}</sup>$ All the empirical exercises in this section are also repeated (but not reported here) using revenue Labor Productivity (rLP) for robustness check. rLP is especially less noisy but less detailed in describing production. On supporting side, the implications turn out very much the same with both rLP and rTFP.



Figure 8: Kernel regression estimates of productivity–employment and productivity–output relations and the estimated density of plants in the concrete industry.

timated relations. The estimation error for these plots is inversely related to the probability distribution of productivity (Bierens 1994). For this purpose, the KDE (with Gaussian kernel and bandwidth 0.4 for log productivity) of plant concentration is shown at the bottom of Figure 8. The productivity range 1 to 20 seems to host most of the plants, while the density of observations becomes very sparse at the upper and lower ends, where the estimation error is expected to be large. Focusing on the rTFP interval [1,20], input is mostly falling with productivity, whereas output is mostly increasing. These observations seem robust to the size of market, where market sizes are classified into two groups and KDE for each case is shown in Figure 8 (I am using worker population of 3,000 to break the data).

#### 5.2 Semi-Parametric Estimation of the Relationship

In the non-parametric estimates, the varying density of plants along the productivity axis and the presence of outliers can undermine confidence that the overall picture of productivity–employment relationship is that of a bell-shape. Also, in non-parametric estimation, data is not sliced by time so that a reasonable number of observations are available for the method. More importantly, the shape of the relation in markets of a certain size needs to be shown to be bell-shaped. In this section, I try to overcome these issues by estimating a semi-parametric model with a polynomial of predetermined degree in the log

Dep.Var.	$\log(TE_{jt})$	$\log(Q_{jt})$
$\sigma_{error}$	1.127	1.119

Table 3: Standard deviation of error in the estimated semi-parametric models.

of productivity to approximate the relationship. I also estimate the productivity-output relationship in the same way and make comparisons. The effects of time and market size are secondary and will be approximated in both relations non-parametrically by fitting thin-plate splines (Moussa & Cheema 1992). The general form of this model is

$$\log(TE_{jt}) = \sum_{p=0}^{P} \beta_{1p} \log(\phi_{jt})^p + H_1(L_{jt}, t) + \epsilon_{jt}, \qquad (5.3)$$

$$\log(Q_{jt}) = \sum_{p=0}^{P} \beta_{2p} \log(\phi_{jt})^p + H_2(L_{jt}, t) + \zeta_{jt}, \qquad (5.4)$$

where TE is total employment, and Q is deflated shipment value.  $L_{jt}$  is the market size for plant j at time t.  $\phi_{jt}$  is measured as rTFP. To minimize the computational burden and to reduce running time down to a reasonable length, market size is classified into discrete values by rounding its log to the nearest 0.5. P is the degree of the polynomial term used in the model.

The estimates are computed using a penalized least-squares method that minimizes the following function with respect to  $\beta_p$ 's and a proper choice of function  $H_i(.,.)$ 

$$S_{\lambda} = \frac{1}{n} \sum_{j=1}^{n} \sum_{t} \epsilon_{jt}^2 + \lambda J_2 \big( H_i(L,t) \big).$$

$$(5.5)$$

 $J_2(H_i(.,.))$  is a measure for the roughness of the fit, and here it is defined as the integral of the square of the second derivative of  $H_i$  with respect to its arguments.  $\lambda$  is the penalty parameter, whose choice is a trade-off between accuracy of the fit and its smoothness. s is the number of observations. My actual choice of value for  $\lambda$  proves not to be very crucial as the estimation result remains practically unchanged for values of  $\lambda$  within a wide range from 0.1 to 10. I report results when I set  $\lambda$  equal to 1.

The choice of polynomial degree in model (5.3), however, seems critical. A small value of P will not capture enough curvature, and high values of P will add in noise and cause instability of estimates. In an experimental stage, I added polynomial powers one by one, until the estimates started to become unstable. The most stable predictions are achieved when P = 4. Table 3 reports the standard errors from estimating models in (5.3).

To demonstrate the estimation results, predicted values were generated for three representative market sizes: 1,000, 10,000, and 100,000. The results of the previous section suggest using the productivity



Figure 9: Estimated productivity–employment and productivity–output relationship in the concrete industry.

range 1 to 20 to avoid having to interpret the segments caused by outliers. The estimated curves for productivity–employment and productivity–output relations are shown for this range in Figure 9. Results are very similar to those from the kernel regression.

#### 5.3 Correlations with Productivity

It is also worth calibrating a model of linear demand with production function (2.7) to investigate the extent to which I have been able to reduce productivity-employment correlations as a result of a bell-shaped relation. With Syverson (2004) estimating the returns to scale in the concrete industry around 0.996, the constant returns to scale assumption is realistic enough and also lets me solve for an analytical solution, which is of the form

$$q_j = \frac{p_0 - \frac{1}{\phi_j}}{2p_1}, \quad x_j = \frac{q_j}{\phi_j}.$$
 (5.6)

Applying nonlinear least-squares to the data on plant-level employment and rTFP, the model parameters are estimated as  $p_0 = 7.797(0.280)$  and  $p_1 = 0.034(0.002)$  (numbers in parentheses are standard deviations). 1,000 bootstrapped distributions of productivity are generated and employment and its correlation with productivity are computed using (5.6) and the estimated  $p_0$  and  $p_1$ . The resulting correlations are reported in Table 4.

Data shows weak negative correlations between productivity and employment, an indication that the

Source	corr(TE,rTFP)	$\operatorname{corr}(Q, rTFP)$
Data	-0.035	-0.008
Simulation	-0.298(0.109)	0.193(0.106)
Simulation with		
$\delta_j = U[-0.25p_0, 0.25p_0]$	-0.259(0.095)	0.020(0.018)
Simulation with		
$\delta_j = U[-0.5p_0, 0.5p_0]$	-0.197(0.075)	0.010(0.017)
Simulation with		
$\delta_j = U[-0.9p_0, 0.9p_0]$	-0.132(0.052)	0.005(0.016)

Table 4: Actual and bootstrapped correlations between productivity and employment and between productivity and output. (Standard deviations are shown in parentheses.)

most productive plants are not necessarily the largest. Meanwhile, output has a very weak correlation with productivity, but more positive than that with employment, as also indicated by the results of previous sections.

The bootstrapped results, on the supporting side, show that a bell-shaped relationship can bring down the correlations with employment into the negative territory, while still keeping correlations with output positive. However, looking at correlation levels, the parametric model seems to have overdone its purpose. With market size deemed important in this relation, I suspect that the parametric model is estimated for the aggregate industry and, hence, misses demand variations due to differences in market size. Also, my theoretical model ignores the presence of demand shocks that might be caused by shifting construction activity or economic conditions. To partly account for these effects, I rewrite my inverse demand function as

$$p_j = p(q_j, L) + \delta_j, \tag{5.7}$$

where  $\delta_j$  is an idiosyncratic demand shifter and summarizes the effect of change in market size as well as demand shocks. Random shocks are drawn from a uniform probability distribution independently for each single plant and in each run of the bootstrap process. Table 4 reports the simulated correlations with different ranges of shocks and using the same estimated parameters as before. Correlations actually move closer to those of data as the range of possible shifts widens to cover the whole range of demand sizes from zero and upwards.

## 6 Conclusion

Using models that generate non-monotonic relations between input size and productivity are subject to harsh criticism. Constant-elasticity demand has been very popular with researchers since Dixit & Stiglitz (1977) showed the nice aggregation properties that this demand function exhibits. However, in practice, this assumption does not seem to hold very well. The input-productivity relation is non-monotonic within concrete and also within most other industries, judging by their low productivity-employment correlations and also shown by Bakhtiari (2008) for the general class of localized-market industries. Consequently, future models of heterogeneous producers have to consider the possibility of non-monotonic relations, especially when discussing productivity dispersion and size distributions.

This paper unifies the role of demand and production in predicting the sign of input-productivity slope in a simple algebraic condition. The simplicity of the condition opens new avenues for a creative mind to wander into new territories. Arbitrary forms of input-productivity relations can be generated and be associated with a corresponding demand structure. Also, the possibility of identifying demand (unobserved) from available data on employment, sales, capital stock, and productivity (observable) exists. The availability of micro-level data on size and productivity might actually be a practical bridge to the demand side, and the results of this paper act as a prelude to harnessing the wealth of information already available to us but hidden in the data.

# References

- Bakhtiari, S. (2008), Productivity Dispersion, Plant Size, and Market Structure, Ph.D. dissertation, University of Maryland.
- Bartelsman, E. J. & Dhrymes, P. J. (1998), 'Productivity dynamics: U.S. manufacturing plants 1972– 1986', Journal of Productivity Analysis 9(1), 5–34.
- Bartelsman, E. J. & Gray, W. (1996), The NBER manufacturing productivity database, Technical Report 205, NBER Working Paper series.
- Bertola, G. & Garibaldi, P. (2001), 'Wages and the size of firms in dynamic matching models', *Review of Economic Dynamics* 4(2), 335–368.
- Bierens, H. J. (1994), Topics in Advanced Econometrics: Estimation, Testing, and Specification of Cross-Section and Time-Series Models, Cambridge University Press, New York, NY.
- Bontemps, C., Robin, J.-M. & Van Den Berg, G. J. (2000), 'Equilibrium search with continuous productivity dispersion: Theory and nonparametric estimation', *International Economic Review* 41(2), 305– 358.
- Chiang, H. (2005), Essays on Employer–Employment Relationships and Firm Performance, Ph.D. dissertation, University of Maryland.

- Davis, S. J., Haltiwanger, J. & Schuh, S. (1996), *Job Creation and Destruction*, Massachusetts Institute of Technology.
- Dixit, A. K. & Stiglitz, J. E. (1977), 'Monopolistic competition and optimum product diversity', American Economic Review 67(3), 297–308.
- Ericson, R. & Pakes, A. (1995), 'Markov-perfect industry dynamics: A framework for epirical work', The Review of Economic Studies 62(1), 53–82.
- Foster, L., Haltiwanger, J. & Syverson, C. (2008), 'Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?', American Economic Review 98(1), 394–425.
- Hopenhayn, H. (1992), 'Entry, exit, and firm dynamics in long run equilibrium', *Econometrica* **60**(5), 1127–1149.
- Hopenhayn, H. & Rogerson, R. (1993), 'Job turnover and policy evaluation: A general equilibrium model', Journal of Political Economy 101(5), 915–938.
- Jovanovic, B. (1982), 'Selection and the evolution of industry', *Econometrica* **50**(3), 649–670.
- Lucas, R. E. (1978), 'On the size distribution of firms', Bell Journal of Economics 9(2), 508-523.
- Luttmer, E. G. (2007), 'Selection, growth, and the size distribution of firms', *Quarterly Journal of Economics* **122**(3), 1103–1144.
- Melitz, M. J. (2003), 'The impact of trade on intra-industry reallocations and aggregate industry productivity', *Econometrica* 71(6), 1695–1725.
- Melitz, M. J. & Ottaviano, G. (2008), 'Market size, trade, and productivity', *Review of Economic Studies* 75, 295–316.
- Moussa, M. A. & Cheema, M. Y. (1992), 'Non-parametric regression in curve fitting', *The Statistician* **41**(2), 209–225.
- Rossi-Hansberg, E. & Wright, M. (2007), 'Establishment size in the aggregate economy', American Economic Review 97(5), 1639–1666.
- Simonoff, J. S. (1996), Smoothing Methods in Statistics, Springer-Verlag New York, Inc.
- Syverson, C. (2004), 'Market structure and productivity: A concrete example', Journal of Political Economy 112(6), 1181–1222.

Zellner, A. & Ryu, H. (1998), 'Alternative functional forms for production, cost and returns to scale functions', Journal of Applied Econometrics 13, 101–127.