Quantifying Non-monotonic Relations Between Input Size and
Productivity: Theory with Applications

*****Working Paper*****
Sasan Bakhtiari
University of New South Wales
Sydney, Australia
March 2009

Abstract

It is stylized in the economic literature that productivity and input size should relate positively and monotonically in the long run. In this paper, I point out and emphasize that under optimality assumption, depending on the demand and production structure, this relation is most likely a bell-shape and possibly a multi-modal form. I use a very general production model to quantify a simple condition on demand and production elasticities that governs the relation between productivity and input size. This condition is especially useful in inferring facts about an industry’s demand structure based on observables such as size and productivity. Possibility of building wider monotonicity is suggested by letting plants trade with several markets. Supportive evidence is also 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 Scale


*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, its program sponsors or data providers. Some or all of the data used in this paper are confidential data from the LEHD Program. 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 (Jeremy.S.Wu@census.gov, http://lehd.did.census.gov).

*I am indebted to Prof. John Haltiwanger for his continuous 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. Through a very general production problem, I show that with most general cost and revenue functions the relationship between productivity and input size can easily turn non-monotonic, 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.

Under optimality and mild assumptions on demand and production, I show that the relevant condition governing how the input level changes with productivity under optimality condition is

\[
\varepsilon_{MR} + \varepsilon_{\nu} = -1, \quad (1.1)
\]

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. This formula summarizes the influence of production and demand on the shape of the relation and 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 also be used to build 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 unobservable demand curve by utilizing observables such as employment, sales, and productivity.

The first implication of my results is naturally a need to re-examine how size distributions are formed as a result of productivity differences. Rossi-Hansberg & Wright (2007) cite the intensity of capital utilization as a reason for difference in the thickness of upper tail in the size distributions of different industries. This paper points to differences in market structures as another major reason. 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 more than previously thought, as the opening of new markets cause vast shifts in employment and output shares to the high-productivity producers, who were constrained otherwise.

Literature on size distribution probably dates back to Lucas (1978), in which differences in management skills generate distributions of productivity and size. 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 the market, reallocating...
### Table 1: Correlations between employment and log productivity (revenue total factor productivities is used).

<table>
<thead>
<tr>
<th>SIC</th>
<th>corr(E,rtfp)</th>
<th>SIC</th>
<th>corr(E,rtfp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.106</td>
<td>30</td>
<td>0.079</td>
</tr>
<tr>
<td>22</td>
<td>-0.088</td>
<td>31</td>
<td>-0.019</td>
</tr>
<tr>
<td>23</td>
<td>-0.049</td>
<td>32</td>
<td>0.211</td>
</tr>
<tr>
<td>24</td>
<td>-0.017</td>
<td>33</td>
<td>-0.060</td>
</tr>
<tr>
<td>25</td>
<td>-0.006</td>
<td>34</td>
<td>-0.026</td>
</tr>
<tr>
<td>26</td>
<td>0.160</td>
<td>35</td>
<td>-0.112</td>
</tr>
<tr>
<td>27</td>
<td>0.100</td>
<td>36</td>
<td>0.064</td>
</tr>
<tr>
<td>28</td>
<td>0.128</td>
<td>37</td>
<td>0.015</td>
</tr>
<tr>
<td>29</td>
<td>-0.345</td>
<td>38</td>
<td>0.075</td>
</tr>
<tr>
<td>All</td>
<td>0.007</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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\(^1\). 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\(^2\). 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 then derive a condition on the elasticities of marginal revenue and RTS that specifies the slope of productivity-input relation at any point (Equation (1.1)). Geometrically, this condition shows why it is hard to achieve monotonicity: to have a fully monotonic relation between productivity and

---

\(^1\) See, for example, Bartelsman & Dhrymes (1998).

\(^2\) Section 3 describes how productivities are computed.
input, the revenue function, especially, has to grow at least as fast as a logarithmic function, on average, while almost all revenue functions are bounded above because of finite consumer population. Higher RTS close the range of possibilities and diminishing RTS demand a faster growing revenue function, for a monotonic relation to be possible. The theory can also be extended to encompass industries with more global markets, through a simple trade model where more monotonic relations between input and productivity are possible with producers self-selecting into trading with several other markets.

Condition 1.1 also finds application in the design and identification of demand. With fixed RTS, the design problem is to build productivity-input relations of arbitrary form. I approach this problem by setting $1 + \varepsilon_{MR}$ to a polynomial and by strategically placing its zeros to make the desired form happen. Alternatively, when the productivity–input relation can be observed (or estimated) in the data, it is possible to roughly identify demand that is implied by the observed data. I put this approach to practice with the concrete industry, where I can almost replicate the observed relations and bind them to candid demand curves.

Other 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 and definition. 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 structure is not presented as a final verdict in the analysis of non-monotonicity, though the empirical results of this paper draw a fairly favorable picture for market 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 goes over the theoretical model
and describes conditions under which input size falls with productivity and presents some examples and applications. Following Melitz (2003), a trade model is also introduced that helps to create more monotonicity in relations. Section 3 describes my data. In Section 4 I investigate the empirical relationship between productivity and employment for the concrete industry. Robustness to different market types is checked in Section 5. The paper is then concluded.

2 A General Production Problem

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 \( q \geq 0 \) and \( \phi \geq 0 \):

(i) \( C(q, \phi) > 0 \),

(ii) \( C_q > 0 \) and \( C_\phi < 0 \),

(iii) \( C(.,.) \) is multiplicative separable, i.e., \( C(q, \phi) = c_1(q)c_2(\phi) \) for some functions \( c_1 \) and \( c_2 \) 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. 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\(^3\). Section 2.4 looks at cases where trading with other markets is possible and points to

\(^3\)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.
differences and similarities with the main result. 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 aggregates 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 \( \nu(q) = \frac{dq}{dx} \frac{w}{q} \). Eliminating \( \nu \) between this relation and (2.1) and solving the resulting partial differential equation yields

\[
4x = \frac{C(q, \phi)}{w}.
\]

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 \( \varepsilon_{\nu} = \frac{q\nu}{\nu} \).

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 straightforward as depicted by the following proposition.

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

\footnote{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

\[
(R_q - C_q) \frac{dq}{d\phi} = C_{q\phi}.
\]

By second-order optimality condition we have \( R_q - C_{qq} < 0 \). 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 one-to-one mapping of productivity. Hence, in most discussions, productivity and output play an equivalent role. The next lemmas establish the behavior of \( x(\phi) \) in a neighborhood of \( \phi \).

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

Proof: Note that \( C_{qq} \) can be written as follows

\[
C_{qq} = \frac{\partial}{\partial q} \left( \frac{\partial C}{\partial q} \right) = \frac{\partial}{\partial q} \left( \frac{C}{q^\nu(q)} \right) = \frac{C_q}{q^\nu(q)} - \frac{C}{q^\nu(q)} \frac{\nu_q}{\nu(q)} - \frac{C}{q^\nu(q)} \frac{1}{q}.
\]

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

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

The second-order condition requires that \( R_{qq} < C_{qq} \). Replacing with (2.3) into this condition yields

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

which can be rewritten as

\[
\frac{q R_{qq}}{R_q} + \varepsilon_\nu < \frac{1}{\nu(q)} - 1 .
\]

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.
Lemma 2.2. Input and output are related in the following way

\[ \frac{\partial x}{\partial q} \leq 0 \iff \epsilon_{MR} + \epsilon_{\nu} \leq -1. \]

**Proof:** With \( x(\phi) = C(q(\phi), \phi) \) and (2.1), it is possible to write \( x = \nu(q)qC_q \). But the first-order condition demands that \( C_q = R_q \). Replacing \( C_q \) gives \( x = \nu(q)qR_q \). Now, take partial derivatives with respect to \( q \) to get

\[ \frac{\partial x}{\partial q} = \nu(q)qR_{qq} + \nu(q)R_q + \nu_qqR_q \]
\[ = \nu(q)R_q \left( 1 + q\frac{\nu_q}{\nu(q)} + q\frac{R_{qq}}{R_q} \right). \]

Replacing with Definitions 2.1 and 2.2 yields

\[ \frac{\partial x}{\partial q} = \nu(q)R_q(1 + \epsilon_{\nu} + \epsilon_{MR}). \]

With \( \nu(q) > 0 \) and \( \frac{\partial R}{\partial q} > 0 \), the sign only depends on the last term above. Results immediately follow. ■

The following proposition is the main result specifying conditions that induce 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

\[ -1 < \epsilon_{MR} + \epsilon_{\nu} < \frac{1}{\nu} - 1. \]

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

\[ \epsilon_{MR} + \epsilon_{\nu} < -1. \]

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

**Proof:** Since \( \frac{dx}{d\phi} = \frac{\partial x}{\partial q} \frac{dq}{d\phi} \) and \( dq/d\phi > 0 \), it is clear that when \( \epsilon_{MR} + \epsilon_{\nu} > -1 \) then \( \frac{dx}{d\phi} > 0 \) and vice versa. Alternatively, when \( \epsilon_{MR} + \epsilon_{\nu} < -1 \) then \( \frac{dx}{d\phi} < 0 \). Accounting for the upper limit on \( \epsilon_{MR} + \epsilon_{\nu} \) set by Lemma 2.1, the results follow. ■

Figure 1 illustrates the relevant ranges of elasticities in which input size increases or decreases with
productivity\textsuperscript{5}. Proposition 2.2 is a description of how the demand-side and supply-side influence the 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

\textsuperscript{5}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 condition from Proposition 2.2 changes to

\[
\frac{dx}{d\phi} \geq 0 \iff \varepsilon_{MR} + \varepsilon_{\nu} \geq \frac{1}{\nu} (\Xi - 1) - 1,
\]

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.

then be written as

$$-\frac{1}{q} < \frac{R(q) - R(q_*)}{R_q(q_*)} < \frac{1}{q} - 1$$

(2.4)

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

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

(2.5)

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 increase at least as fast as a logarithmic function on average 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 (2.4). 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 than a logarithmic function. With the same reasoning, if the RTS go up with $q$, both limits in (2.5) shift downward, with the upper limit moving faster. In this case, the range in (2.5) closes quickly.
2.3 Applications of Proposition 2.2

For simplicity, let’s assume that \( \nu \) is fixed and constant and plants produce according to

\[
q = \phi x^\nu.
\]  

(2.6)

It is easy to check that the cost function associated with this production is \( C(q, \phi) = (q/\phi)^{\frac{1}{\nu} - 1} \) and satisfies Assumption 2.1. With fixed \( \nu \), the threshold condition in Proposition 2.2 becomes \( \varepsilon_{MR} = -1 \).

The immediate application of Proposition 2.2 is to determine the shape of \( x(\phi) \) when the demand function is known. With a linear inverse demand function of the form \( p(q) = p_0 - p_1q \), the production problem can be solved analytically to yield

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

(2.7)

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

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

(2.8)

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

Yet, a more creative way of using Proposition 2.2 is to design an arbitrary form for \( x(\phi) \) using a counter-factual \( \varepsilon_{MR} \) as the starting point. The design of counter-factual relations for \( x(\phi) \) is possible by setting \( 1 + \varepsilon_{MR} \) equal to a polynomial such as

\[
1 + \varepsilon_{MR} = a \prod_{i=1}^{n} (Z_i - q),
\]  

(2.9)

and strategically placing zeros, \( Z_i \), to shape the form of \( x(\phi) \) in the desired ways. Some constraints also apply. First of all, general forms of relations observed in the data suggest that \( n \) should be odd and \( \prod_{i=1}^{n} Z_i > 0 \), so that the input level rises in the start and falls eventually with productivity. 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 in between them can be generated.

---

*These numbers come from calibrating this model to the concrete data in Section 4.4.*
To see an example of a multi-modal relation designed in this way, let $1 + \varepsilon_{MR} = \frac{1}{1000}(10 - q)(20 - q)(40 - q)$, where the generated 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.

This approach to building input relations by initiating a demand function also suggests a reverse identification possibility: using observable data on a plant’s productivity and input level, one can guess the demand structure facing an industry by looking at the input relation. $n$ in (2.9) can be determined from the number of peaks in the relation. To have $m$ peaks and a falling upper tail, $n = 2m - 1$ because there are $m$ maxima and $m - 1$ minima. On the other hand, if the upper tail is rising, then $n = 2m$.

A simple but preliminary way to place zeros would be to use the first-order condition. (2.9) can be integrated to obtain a candidate $R_q$ and then be set equal to $1/\phi$ (in the case of CR TS production). That yields

$$Bq^{-1+AP}e^{-AP}\prod_{n_{i=1}}^{1}Z_{i} = \frac{1}{\phi},$$

where $P(q)$ is a polynomial resulted from integrating $\frac{1}{q}((\prod_{1}^{n_{i=1}}(Z_{i} - q) - \prod_{1}^{n_{i=1}}Z_{i})$. It is desirable to bypass using $q$ in practice as output is normally measured in dollars and not in physical terms. Inserting $q = \phi x$ into the above, parameters can then be found by applying MM or GMM to the observed or estimated data points. This identification process is preliminary so far and is merely used to show application prospects. Further work is required to build a robust estimation process that accurately extracts aggregate demand using production-side data. For demonstration purposes, I later use an MM to fit the peak of bell-shapes and get similar curves to those estimated from the data.

### 2.4 More Monotonicity with Trade

So far, markets were assumed localized in the sense that revenue function was fixed for everyone and more productive units could not improve their demand by trading with other markets. Predictions of such models are not fully applicable to industries with a more global outreach. When costly trade is possible, Melitz (2003) shows that more productive units engage in trade and increase their market share as a result. At the same time, less productive units supply only domestically. In a similar work, Melitz & Ottaviano (2008) subject plants to a linear demand to the same effect. However, they do not include any input size in their model, hence avoiding any discussion of productivity-input relations. For expositional purposes, I consider a parsimonious trade model similar to that of Melitz (2003) that is able to combine bell-curves in different markets to generate a more monotonic relation to the advantage of high-productivity plants. Let trade be possible with $N$ other markets ($N + 1$ total markets) at a marginal
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.
Figure 7: The shape of productivity–output and productivity–input relations when it is possible to trade with five other markets.

cost of $\tau_i$ for outside market $i$. For simplicity, assume production is constant returns to scale (CRS) 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:

$$\pi = (p_0 - p_1 q^D)q^D + \sum_{i=1}^{N} (p_0 - p_1 q^O_i)q^O_i - \frac{q^D}{\phi_j} - \sum_{i=1}^{N} \tau_i q^O_i, \quad (2.11)$$

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

$$q_j = \frac{p_0 \phi_j - 1}{2p_1 \phi_j} I[\phi_j > \phi^D_j] + \sum_{i=1}^{N} \frac{(p_0 - \tau_i) \phi_i - 1}{2p_1 \phi_j} I[\phi_j > \phi^O_i], \quad (2.12)$$

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

$$\phi^D_j = \frac{1}{p_0}, \quad \phi^O_i = \frac{1}{p_0 - \tau_i}, i = 1, \ldots, N. \quad (2.13)$$

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 let trade be possible with five other markets whose trading costs are $\tau_i = \{3, 3.5, 4, 5, 6\}$. Note that a huge increase in market shares of high-productivity producers still does not rule out a bell-shaped relation, though a wider bell is generated as a result of trading.

---

*I use the term “outside” to signify that other markets need not be international.*
3 Data

3.1 The Concrete Industry as Test Bench

Results of the previous section 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 that make it a suitable case for study. First, due to high costs of transportation, concrete is not shipped very far compared to many other products. 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 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 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, entries and exits. 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.

*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.
of the average plant\textsuperscript{9}.

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.

### 3.2 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\textsuperscript{10}. 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\textsuperscript{11} 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 estimations. This leaves me with 2,027 sample concrete plants.

The revenue Total Factor Productivity (rTFP) is used as a measure of productivity, which is 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

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

where lower case letters label variables in logs. Here \( q \) is the nominal output deflated by the industry-specific 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

\footnotemark[9]\textsuperscript{9}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.

\footnotemark[10]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).

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 (2.6) as if it depended on employment only.

3.3 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\footnote{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.}. 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 CBSA level and matched by CBSA-year\footnote{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.}. There are 667 markets that match to my subsample. More detailed statistics for this market definition can be found in Table 2.

Population density of construction workers is also used afterwards for robustness check. To compute worker density, county areas are obtained from the Census Bureau’s County and City Databook and aggregated up to CBSA level. The main difference between worker population and its density is that, with population density, the physical extents of an urban area are also taken into account. With transportation costs weighing on concrete’s market, I test whether the productivity–employment relations stays mostly
the same with both market definition. A preliminary prediction based on correlation between the two
types suggests a very strong correspondence (Table 2).

Census Bureau has made the City and County Databook 2000 available that also provides me with
the year 2000 estimates of resident population and its density at CBSA level, which can be regarded as
two other measures of market size. Though the panel falls outside my time-frame, strong correlations
reported in Table 2 confirm that rankings of market sizes is not very different. These market definitions
will also be used for robustness check.

|                   | Worker Population | Worker Density | Resident Population | Resident Density |
|-------------------|-------------------|                |                     |                  |
| Mean              | 43,173.4          | 9.14           | 2,541,081.9         | 534.4            |
| Std.Dev.          | 58,022.6          | 9.79           | 3,793,412.7         | 632.2            |
| Min.              | 48                | 0.04           | 12,457              | 3.6              |
| Median            | 16,600            | 5.99           | 924,786             | 301.5            |
| Max.              | 327,397           | 48.76          | 18,747,320          | 2,792.2          |

<table>
<thead>
<tr>
<th>Correlations</th>
<th>Pop.Worker</th>
<th>Density Worker</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>0.852</td>
<td>0.975</td>
</tr>
<tr>
<td></td>
<td>0.822</td>
<td>0.955</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.871</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Summary statistics for different definitions of market size.

4 Empirical Results

In this section, I undertake a series of exercises with three major goals. First, I want to showing that
a bell-shaped productivity–employment relation best fits the concrete data, using a range of parametric
to non-parametric methods. Second, I will try to explain the low correlations in the data by non-
monotonicity in the productivity–input relation. Finally, I will apply a similar approach to (2.10) to the
concrete industry to identify a candid demand curve. The effect of the market size on the productivity–
employment relation is also highlighted in the exercises.

4.1 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

\[
TE_j = H_1(\log(\phi_j)) + \epsilon_j, \tag{4.1}
\]

\[
Q_j = H_2(\log(\phi_j)) + \zeta_j. \tag{4.2}
\]
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 $H_i(\cdot)$, $i = 1, 2$ 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 estimated 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. There seems to be an interval of plant concentration along the productivity axis, while the density of observations becomes very sparse at the upper and lower ends, where the estimation error is expected to be larger. Overall, both pictures show ranges where both input and output fall with productivity with a good degree of confidence. However, with output, the range where the relation is positive seems to be much larger than with input. This is expected, as theory predicts that output should positively relate to input all the way. Also, the relation seems to show robustness to using different market size classes (I am using worker population of 3,000 to break the data).

### 4.2 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. In this section, I try to overcome these issues by estimating a semi-parametric model with a polynomial of predetermined degree in the log 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
Figure 8: Kernel regression estimates of productivity–employment and productivity–output relations and
the estimated distribution of productivity in the concrete industry.

Form of this model is

$$\log(T E_{jt}) = \sum_{p=0}^{P} \beta_1 p \log(\phi_{jt})^p + H_1(L_{jt}, t) + \epsilon_{jt},$$  \hspace{1cm} (4.3)

$$\log(Q_{jt}) = \sum_{p=0}^{P} \beta_2 p \log(\phi_{jt})^p + H_2(L_{jt}, t) + \zeta_{jt},$$  \hspace{1cm} (4.4)

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 $H_i(.,.,)$

$$S_{\lambda} = \frac{1}{n} \sum_{j=1}^{n} \sum_{t} c_{jt}^2 + \lambda J_2(H_i(L, t)).$$  \hspace{1cm} (4.5)

$J_2(H_i(.,.,))$ 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
Table 3: Standard deviation of error in the estimated semi-parametric models.

<table>
<thead>
<tr>
<th>Dep. Var.</th>
<th>$\log(TE_{jt})$</th>
<th>$\log(Q_{jt})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{\text{error}}$</td>
<td>1.127</td>
<td>1.119</td>
</tr>
</tbody>
</table>

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

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 (4.3), 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 (4.3).

To demonstrate the estimation results, predicted values were generated for three representative market sizes: 1,000, 10,000, and 100,000. The estimated curves for productivity–employment and productivity–output relations are shown in Figure 9. Results are very similar to those from the kernel regression. Due to the nature of estimation, market size enters as an additive perturbation, which does not change the polynomial shape but scales it up or down non-uniformly.

On another issue, smooth estimations of the relationship are done at the price of losing cutoff information. To account for that to some extent, the plots were shown in the range of productivities observed
in the data. However, readers should notice that the main purpose of this exercise is to demonstrate the non-monotonic nature and the qualitative shape of the relation at higher productivities, rather than point out the already known facts about low productivity plants.

As a final comment, note that, especially with employment data, the predicted values in the far tails of the estimated bell-curves shoot up, which seems to be a result of truncation on the degree of polynomial used in estimation. As Figure 8(a) shows, the density of plants at rTFP more than 1,000 is basically non-existent and estimates in that area are subject to large errors.

4.3 A Candid Demand Curve for Concrete

The one-peaked form of the relation suggests \(1 + \varepsilon_{M R} = A(Z - q)\) as a suitable form. I estimate parameters for each of the market sizes 1,000, 10,000, and 100,000 by making (2.10) hold at the peak. There are three free parameters to estimate. I set \(B = 1\), then \(A\) and \(Z\) for each market size are estimated using (2.10) and reported in Table 4.

Figure 10 shows the estimated demand curves and productivity–employment relations for each of the market sizes. In computing the demand, there is one more integration involved, and the integration constant can be used to create different demand scales corresponding to different market sizes. Alternatively, I set the constant so that, in all markets, demand curves start from the same point, as a way to compare curvatures rather than scales. In this case, larger markets seem to show slightly higher elasticity of demand at lower prices.

4.4 Correlations with Productivity

It is also worth calibrating a model of linear demand with production function (2.6) 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 CRTS assumption for production is realistic enough and also lets me solve for an analytical solution, which is of the form

\[
q_j = \frac{p_0 - \frac{1}{\phi_j}}{2p_1}, \quad x_j = \frac{q_j}{\phi_j}
\]  

(4.6)
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 formed and employment and its correlation with productivity are computed using (4.6) and the estimated $p_0$ and $p_1$. The correlations with data and bootstrapping are reported in Table 5.

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

\[ p_j = p(q_j, L) + \delta_j, \]  

(4.7)

where \( \delta_j \) is a 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 5 reports the simulated correlations with different ranges of shocks and using the same linear demand 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 upwards.

5 Robustness to Different Market Definitions

In the data section, four different types of market definitions were discussed. Here, I test the robustness of the bell-shape and implications for employment size by looking across these different market types and estimating the relationship separately with each one. In all cases, I run the model (4.3) for the bottom and top decile market sizes using each type. The estimation results for both market sizes and different productivity measures are shown in Figure 11.

First, in both small and large markets, using the population of construction workers or residents does not seem to really matter as the estimated plots almost overlap. This was expected as a result of almost perfect correlation between the two (Table 2). The same thing can be said about population densities. More importantly, the estimated curves with population density, worker or resident, mostly predict smaller employments in larger markets than the ones estimated using populations. I can associate that with the existence of large cities (in both population and area), where the high transportation cost of concrete takes its toll on the concrete plants due to a more dispersed population. Other than that, the qualitative implications of populations or their densities are the same for the relationship.
6 Conclusion

It is common practice in the current economic literature to let producers have access to infinite demand, or use a constant elasticity of demand. In both cases, the relation between input, say employment, and productivity turns out positive and one-to-one. Starting with Dixit & Stiglitz (1977), using constant elasticity of demand has been particularly popular due to the possibility of aggregating the consumer behavior. 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. Consequently, future models of heterogeneous producers have to consider the possibility of non-monotonic relations, especially when discussing productivity dispersion and size distributions.

On the other hand, the sign of the input–productivity slope is summarized in a simple 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, as I have shown with an example and real data exercise. This paper also introduces a basic approach to identifying demand (unobservable) from available data on employment, sales, capital stock, and productivity (observable). In summary, 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


