On Productivity Dispersion and Plant Size: The Role of Market Structure

Sasan Bakhtiari*
University of Maryland, College Park, USA

January 15, 2008

Abstract

This paper focuses on the study of productivity dispersion and, especially, analyzes the relationship between productivity dispersion and the producer size. The US Census of Manufactures reveals significant productivity dispersion at each employment level. Moreover, this productivity dispersion falls with employment size within most manufacturing industries. It will be shown that plant-level births and deaths and simple statistical aggregation do not explain this pattern convincingly. This paper introduces a market structure model that uses demand size and market localization as constraining forces to generate a bell-shaped relationship between input size and productivity. The non-monotonicity of the relation suggests that the selection process relies on both productivity and profitability, where low productivity plants are forced out of the market by competition and very productive plants do not hire inputs due to the constrained market. This is a major departure from models where only low productivity plants leave the market and productivity is monotonically increasing in the size of operating plants in equilibrium. In addition, the proposed model predicts significant productivity dispersion at any level of input size and implies that this dispersion should decrease with the size, exactly as is observed in the data. Also in the data I find a bell-shaped relationship between employment and productivity in industries with localized markets but not the industries with more globalized markets. Throughout the paper, a distinction is made between physical and revenue productivities and the theoretical implications of both measures are shown to be qualitatively the same.

*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 (Jeremey.S.Wu@census.gov, http://lehd.did.census.gov). You can contact the author by emailing sasanb@umd.edu

* My special thanks to Prof. John Haltiwanger for his invaluable support all along. I am also deeply grateful to Hyowook Chiang for providing me with the data. I would also want to thank John Shea, Michael Pries, Daniel Vincent, Ingmar Prucha, Carlos Vegh, and Arghya Ghosh for their insightful comments and guidance.

1 Introduction

Micro-level data suggest that productivity at the establishment level is dominated by idiosyncratic factors. Productivity heterogeneity does not disappear by narrowing down the scope of industry. Haltiwanger, Lane & Spletzer (2000) find remarkable heterogeneity in a very narrowly defined universe of restaurants and plumbing establishments in the state of Maryland. Reducing measurement error or product heterogeneity does not take the productivity differences away, as shown in the case of Cement industry (Syverson 2004). These facts, per se, indicate that there are frictions affecting the reallocation of resources, both within and across industries and even among plants belonging to the same firm (Chew, Clark & Bresnahan 1990).

Much research has been attempted to generate the observed size and productivity distributions. Decreasing returns to scale, per se, prevent infinite growth of the more productive producers, giving a productivity advantage to smaller less productive units. Bartelsman & Doms (2000) offer a review of other important factors such as management, quality of workforce, technology differences, and international exposure that are capable of creating significant productivity dispersion. Investment decisions and R&D activities can also play a role in shaping the productivity distribution (Dunne, Foster, Haltiwanger & Troske 2000, Nelson 1981). Regulatory environments can pose an obstacle to or facilitate resource allocation and thus increase or decrease productivity heterogeneity (Olley & Pakes 1996).

A direct consequence of productivity heterogeneity is an endogenous size distribution among producers. The Jovanovic (1982) model of industry evolution predicts that in the process of selection, less productive plants realize that they are never going to be profitable and exit the market. Their resources, labor or capital, are allocated to entrants and/or more productive plants. As a result, more productive plants grow faster and become larger as they age. This monotonic size ranking of producers by their productivity is typical of most existing models: It is most commonly reasoned that if producer A is more productive than B, then A must be larger than B in the long run. Alongside with Jovanovic (1982), Bontemps, Robin & Van Den Berg (2000) and Bertola & Garibaldi (2001), predict a monotonically increasing relation between producer's size and productivity in equilibrium using job search and job matching approaches, respectively. A one-to-one correspondence between size and productivity immediately implies zero productivity dispersion for plants of a given size. The Ericson & Pakes (1995) model of uncertain investment outcomes offers an equilibrium in which producers can reverse ranks, depending on the shocks they receive, but more productive producers can immediately adjust themselves and are larger both in size and output.

On the contrary, I will show that in the US Census of Manufactures (CM) productivity differences are present at any level of employment size, and this is true even when plants belong to the same industry. Furthermore, as plants get larger, they become less dispersed in their productivities. This pattern still persists after controlling for aggregate and industry specific factors. Most interesting, the decreasing relationship between productivity dispersion and employment is mostly due to old and stable plants and is barely affected by entrants at all. These old plants have a much lower probability of failure than

the young ones, so one can reason that exit is not playing a major role either. Therefore, the observed behavior of productivity dispersion seems to be a consequence of long-run equilibrium rather than industry evolution. This size-dependent behavior is not stranger to the current economic literature that extensively documents how size affects different measures of performance: Rates of growth and failure fall with size (Hall 1987, Evans 1987). Larger producers offer higher wages (Brown & Medoff 1989) and generally set up their own research divisions and invest more in R&D activities (Acs & Audretsch 1991). Also larger producers need more complicated and sophisticated management and organizational structures to control their operations (Churchill & Lewis 1983, Greiner 1998).

A simple statistical aggregation is the natural direction to explore first. In general, Fisher (1993) emphasizes that the additivity of plant behaviors does not hold unless under very restrictive conditions[1]. I will demonstrate the inadequacy of statistical aggregation in explaining the observed pattern of productivity dispersion by running a counter-factual test. As expected, this approach fails mostly because productivity dispersion falls to zero very fast; consequently, it cannot generate heterogeneity in the size-dispersion relationship across industries such as is observed in the data.

In theory, in presence of supply side shocks, any friction that makes the adjustment of labor and capital infrequent compared to the arriving rate of shocks will be able to produce a sustaining distribution of productivity for each level of employment. Especially, the Hamermesh (1995) or the Abel & Eberly (1996) type costs can produce ranges of productivity dispersion that are functions of the adjustment costs. However, the range of productivity dispersion that is observed for smaller plants is about 12:1, which requires adjustments to be immensely costly. Also, the more important question that still remains is how the falling productivity dispersion by employment can be simulated using a constant cost of adjustment. The same argument more or less applies to other cases of supply-side frictions. Instead, I attempt a market structure approach to explaining the observed behavior of productivity dispersion.

Recent works suggest that market structure is a considerable force in shaping the distribution of productivity. Syverson (2003) shows that the level of product substitutability affects both toughness of competition and demand elasticities, thus affecting the way selection works. When faced with high substitutability of products, less productive plants withdraw from market because they are not able to produce the minimum required quality or quantity. As products get more differentiated, low productivity plants get a chance to stay in the market and produce, thereby, increasing the productivity dispersion within the industry. However, Syverson's model predicts the same productivity dispersion among any rich enough subset of operating plants within the same industry, irregardless of their size. Melitz & Ottaviano (2005) analyze the impact of market size in a model of trade liberalization, showing that in larger markets productivity dispersion goes down in response to tougher competition. These works suggest that market structure may shape productivity dispersion in a way that can explain the observed patterns from the

[1] Fisher (1993) shows that the additivity property will basically hold if it can be assumed capital and labor are completely homogeneous and have perfect mobility across plants, and the production functions have very specific functional properties. Most of the conditions are shown to be very hard to satisfy in practice.

CM.

The producer profits depend on both their efficiency of production and their demand structure. Foster, Haltiwanger & Syverson (2006) emphasize these factors as they show that both profitability and productivity affect the selection process. The plants that survive are not necessarily those that are more productive, but also those that face more favorable demand conditions. I will further reason that when the industry faces a limited localized market, the market size becomes an important parameter in determining the plant behavior. In response to limited demand, high productivity plants must choose low input levels to stay profitable. Only when demand becomes large or inelastic enough can the top productivity plants grow large and make enough profit to justify their growth.

The main objective of this paper is to build a model that relates market demand size to the observed relationship between size and productivity dispersion. A product differentiation framework with localized market structure endogenously creates a bell-shaped relation between a plant's productivity and its size defined in terms of a composite input factor. This relationship, in turn, produces more productivity dispersion for plants with lower input sizes. Allowing for a distribution of markets with different sizes (say a segmentation of a national market) helps to make the productivity distribution denser at any given input size and to bring the results closer to reality. In my empirical work, I select a subset of industries that feature both market localization and product differentiation and compare data moments to those predicted by the model. I find that such industries have a bell-shaped relationship between input size and productivity. Conversely, industries exposed to more global markets behave in a more conventional way, i.e., input size increases monotonically with productivity. Plants in global markets in effect serve an infinitely large market, making demand size irrelevant to the analysis.

The remainder of the paper is composed as follows: Section 2 offers some motivating facts from the data about patterns of productivity dispersion. Section 3 presents the theoretical framework for the analysis. The relevant data is discussed in Sections 4, and Section 5 compares model predictions to the data. A version of the model with constant returns to scale (CRTS) production is calibrated and simulated in Section 6. Finally, Section 7 concludes the paper.

2 Motivating Facts

Most previous analysis has focused on variations of mean productivity by input or output size, leaving out possible variation of higher moments. It is commonly assumed that entrants draw their productivity from a single distribution, and then industry selection reshapes both size and productivity distribution as producers survive and age. In related work, Davis & Haltiwanger (1991) look at wage variation within manufacturing industries and find that wage dispersion is significant for all levels of employment size and, moreover, decreases with employment for both production and non-production workers. Strong correlation of wage and productivity dispersions shown in parctice suggests a similar relationship between

productivity dispersion and employment size.

In this study, revenue Total Factor Productivity (rTFP) is the basis for analysis and is computed based on a Cobb-Douglas production function. The revenue productivity is computed using input cost shares and the deflated revenue as real output. Some recent literature has emphasized the distinction between revenue productivity and physical output productivity[2]. In particular, revenue measures of productivity are driven by variations in factor prices, demand shocks and price heterogeneity, as well as the efficiency of production.

For a plant j belonging to industry i at time t, rTFP is defined as

$$rtfp_{ijt} = q_{ijt} - \alpha_i^h h_{ijt} - \alpha_i^{eq} k_{ijt}^{eq} - \alpha_i^{st} k_{ijt}^{st} - \alpha_i^e e_{ijt} - \alpha_i^m m_{ijt}, \tag{2.1}$$

where lower case letters here label variables in logs. Here q is the nominal output deflated by industry-specific price indices. 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 α coefficients are computed using the cost share indices described by Chiang (2004). rTFP provides a detailed measure of productivity taking into account various productive factors. However, it produces relatively noisy estimates due to inaccuracies in, or in some cases unreported, data on capital or other input factors.

For robustness, I also compute revenue Labor Productivity (rLP) which follows the standard definition

$$rlp_{ijt} = q_{ijt} - h_{ijt}. \tag{2.2}$$

Again, lower case letters denote variables in logs. In comparison to rTFP, rLP is less detailed in assessing the contribution of different inputs, but it is also less prone to noise and measurement error. Over time, most of industries show a strong correlation between their rTFP and rLP estimates. The qualitative similarity of the two measures enables me to emphasize the empirical results using rLP rather than rTFP, because the rLP results are more stable.

Panels 1982, 1987, 1992, and 1997 from the Census of Manufactures are used to compute both measures of productivity. More details on the composition of the data is available in Section 4. Both productivity measures have to be comparable over a range of industries and years. Hence, I construct residual productivities by regressing logs of rTFP and rLP on year and year by industry dummies.

These measures allow me to assess the relation between productivity dispersion and employment size. Grouping plants into employment classes, productivity dispersion does not vanish for any size class but decreases as employment size gets larger (Figure 1(a,b))[3]. Both measures of productivity suggest the same relationship, though the behavior of rTFP is more unstable perhaps for the reason explained earlier. Results are different when using output as an indicator of size (Figure 1(c,d)), especially, productivity

[2]See Katayama, Lu & Tybout (2003) or Foster et al. (2006).

[3]Employment size classes are 1-19, 20-49, 50-99, 100-249, 250-499, 500-999, 1000-2499, 2500-4999, and 5000+.

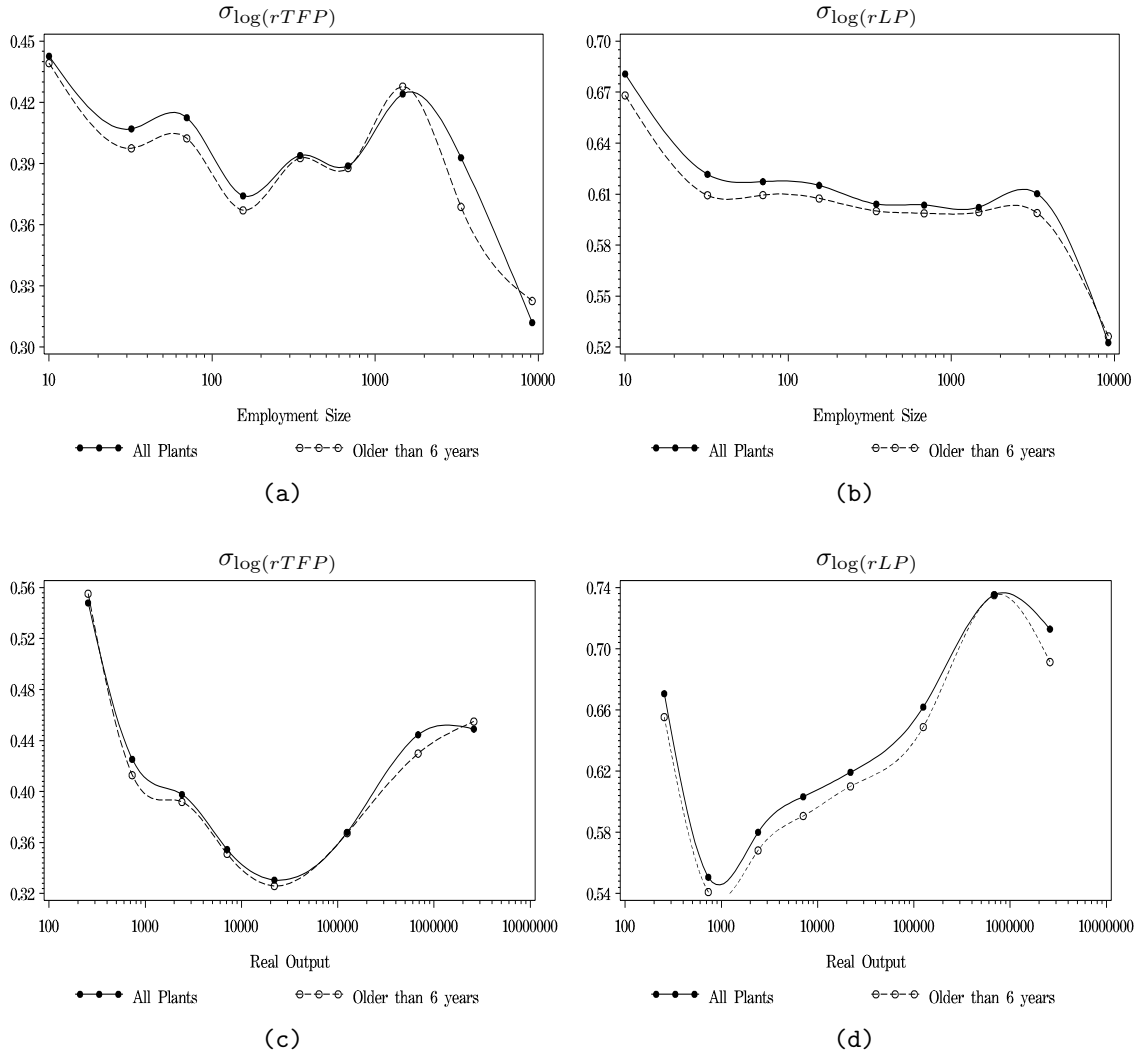


Figure 1: Standard deviation of log productivities by employment and output size classes, using pooled samples (solid line) and only plants older than 6 years (dashed line).

dispersion seems to increase with employment, though the implications with rTFP and rLP are not quite the same. The observed patterns are not a result of entrants distorting the productivity distribution at smaller sizes. Figure 1 also shows productivity dispersion curves computed using only plants that are at least 6 years old. The discrepancy is clearly minimal. Evans (1987) shows that these older plants were about 27% less likely to fail than their younger counterparts, so exit is also unlikely to have a major effect on the shape of the curves too.

On the other hand, such observation immediately raises the question whether some behavioral aggregation is driving the results. For that to happen, the plant behaviors must be additive, i.e., a large plant can be thought of as a collection of several smaller plants bound together. This additivity of behavior was largely rejected by Fisher (1993), when he shows very restrictive conditions need to be satisfied for that to hold. To further confirm the inadequacy of aggregation, I will test a model of statistical aggregation that suggests that larger plants are less dispersed in their productivities because they aggregate on a larger number of (labor) shocks, hence, can operate close to their mean productivity. To check this explanation, I

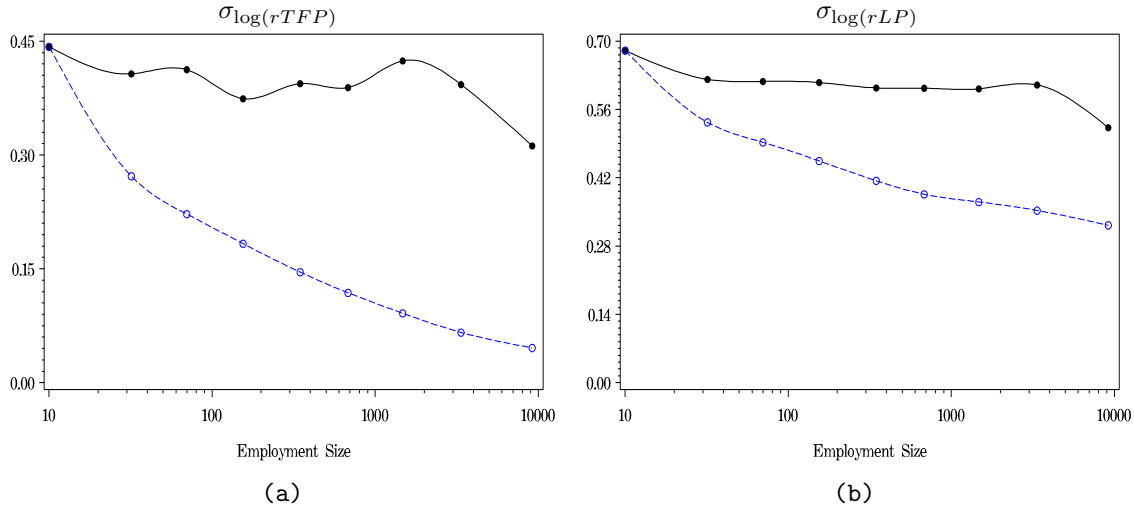


Figure 2: Counter-factual test of statistical aggregation. Solid line is the actual productivity dispersion curve and the dashed line is the counter-factual one.

bootstrapped distributions of larger counter-factual plants by aggregating actual plants with 1-20 employees. 100 bootstrapped distributions were created for each size class and the average standard deviation of productivities for bootstrapped plants was compared against the observed counterpart for actual plants. Results for both measures of productivity are shown in Figure 2. As the plots show, statistical aggregation results in a very steep decline in the productivity dispersion as employment size gets larger[4]. Statistical aggregation basically drives productivity dispersion to zero too fast to play an important role explaining the observed pattern[5].

Using statistical aggregation to explain productivity dispersion in this case poses another problem. The pattern of productivity dispersion by employment is not uniform across industries and can vary significantly. To demonstrate how the pattern changes with industry, I define the following measure for industry i:

$$r_i = \sigma_{Large}/\sigma_{Small}, \tag{2.3}$$

where σ is the standard deviation of the productivity measure, either rTFP or rLP. *Large* refers to plants in the upper employment quartile of their industry. Similarly, *Small* refers to plants in the lower employment quartile of their industry. Very small values of r_i correspond to industries for which larger plants exhibit much less productivity dispersion than small plants. As r_i grows towards 1, productivity dispersion is expected to level out across small and large plants.

I estimate r_i for 365 four-digit industries. I use both rTFP and rLP measures and compare results. For more comparability, the productivity measures are purged of year effects. A KDE estimate of the distribution of r_i demonstrates that heterogeneity among industries is wide and significant (Figure 3). However,

[4]This is not surprising as pure statistical aggregation predicts that the standard deviation must fall at a rate $1/\sqrt{l_1}$, with l_1 being the number of labor units employed at a plant

[5]The counter-factual curve in Figure 2(d) does not converge to near zero as the number of aggregated units goes up. This is perhaps due to presence of other productive factors in residual rLP, hence leaving some covariation between productivities. In that case the limit dispersion will be 2ζ, where ζ is the covariance between idiosyncratic rLP's.

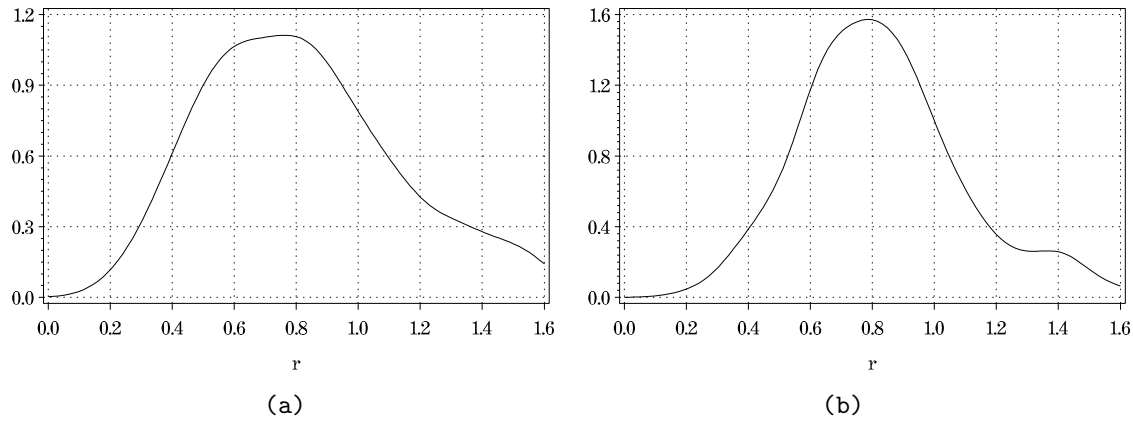


Figure 3: KDE plot of r_i using (a) rTFP, (b) rLP.

about 75 percent of manufacturing industries have r_i less than one under both measures of productivity. Therefore, an industry specific explanation for the relation between employment and productivity dispersion seems to be more appropriate than a simple statistical aggregation mechanism.

In the next section, I analyze a model of market structure and show that it creates relationships between input size and productivity dispersion that depend on factors such as market size and demand elasticities that plausibly vary across industries.

3 The Theoretical Model

The theoretical framework used here is the same as in Syverson (2003) and Melitz & Ottaviano (2005). The theory is based on the differentiated product model developed by Dixit & Stiglitz (1977), and plants are assumed to operate in localized markets and to have monopoly power over their demand. The model has several advantages for my analysis. First, market size and the elasticity of demand can be incorporated into the model easily through choice of utility function. Also, the model is static and therefore tractable. Since I am interested in productivity dispersion as a long-run equilibrium phenomenon, not as a transient process, a long-run model is both simpler and more appropriate than a model with dynamics.

I add to the model a general single input constant or decreasing returns production function. This production function creates a connection between productivity, output and size. The analysis will be more complicated than the regular framework, where measured productivity is summarized in production cost. The payoff is that I obtain strong results. Most importantly, the relationship between plant productivity and its input size is bell-shaped. This result, in turn, offers an explanation for a declining pattern of productivity dispersion by input size.

3.1 Consumers

A market is composed of L identical consumers. There is a continuum of producers, each producing a distinct variety of product indexed by j. The set of available products in each market is J, which is

a subset of total possible varieties J^*. Let N be the measure of set J [6]. As in Syverson (2003), the representative consumer's utility function is

$$\begin{aligned} U &= y + \alpha \int_J q_j^c dj - \frac{1}{2}\eta \left(\int_J q_j^c dj \right)^2 - \frac{1}{2}\gamma \int_J (q_j^c)^2 dj \\ &= y + \alpha \int_J q_j^c dj - \frac{1}{2}\left(\eta + \frac{\gamma}{N}\right) \left(\int_J q_j^c dj \right)^2 - \frac{1}{2}\gamma \int_J (q_j^c - \bar{q})^2 dj, \end{aligned} \tag{3.1}$$

where y is the numeraire consumption, q_j^c is the consumption of each variety by the representative consumer, and $\bar{q}^c = \frac{1}{N}\int_J q_j^c dj$. The utility function parameters α, η, and γ are all non-negative. The utility function has a general quadratic form with parameters α and η determining the consumption of each variety relative to numeraire and γ determining the degree of distinction between varieties. For $\gamma = 0$, there is no variety distinction and the consumer cares only about the aggregate consumption. For higher values of γ, the consumer gains utility by smoothing consumption across different varieties. Compared to other utility functions used within this framework, the utility function of (3.1) has the advantage that it generates a linear demand curve whose elasticity varies by market size, and hence is more suitable for my analysis.

Utility maximization yields the following inverse demand curve per consumer for each variety

$$p_j = \alpha - \eta N \bar{q} - \gamma q_j^c. \tag{3.2}$$

Total demand for each variety is $q_j = L q_j^c$, so each plant faces the following demand curve

$$p_j = \frac{\alpha\gamma + \eta N \bar{p}}{\gamma + \eta N} - \frac{\gamma}{L} q_j. \tag{3.3}$$

Note that (3.2) requires that all prices be bounded above by α; consequently, we will have $\bar{p} \leq \alpha$. The ratio L/γ in (3.3) will have a direct controlling effect on the elasticity of demand and many of the theoretical results discussed in the next sections.

3.2 Producers

Plants operate in a single sector and produce distinct products facing the demand curve (3.3). Upon entry, each plant makes a random draw of its productivity ϕ_j from a known cumulative distribution $G(\phi)$ with support $\phi \in [0, \phi_M]$. There is also a continuous distribution of markets with different sizes. Plants incorporate randomly in a particular market. The markets are assumed to be local; thus plants do not have access to other markets to broaden their demand.

Plants use a single composite input factor x_j for production and choose their input size optimally to maximize profit. The rental price of input w is assumed exogenous and constant within each market and

[6] Equivalently, N will be a measure of plants operating in the market, and it is determined endogenously by the equilibrium conditions discussed later.

also across markets. Section 6.3 examines the effect of price variations across markets and shows that most of the model implications are robust to reasonable amount of price variation. Plants produce according to a production function of the form

$$q_j = \phi_j x_j^{\nu}, \tag{3.4}$$

where ν is assumed to be a fixed value in the range (0,1]. A revenue-based measure of productivity will be useful later to compare the theoretical results to their empirical counterparts and is defined as

$$\theta_j = \frac{p_j q_j}{x_j}, \tag{3.5}$$

Each plant faces a profit function of the form

$$\pi_j = p_j(q_j)q_j - wx_j - f, \tag{3.6}$$

where f is the fixed cost of operation, which is the same for all plants and all markets. Plants are profit maximizing, and the quantity of output that maximizes profit for each plant in that market is the solution to the first order condition

$$\frac{2\gamma}{L}q_j + \frac{w}{\nu\phi_j^{1/\nu}}q_j^{1/\nu-1} = \frac{\alpha\gamma + \eta N\bar{p}}{\gamma + \eta N}. \tag{3.7}$$

The solution to (3.7) is not trivial in general due to non-linearity. However, the set of possibilities can be narrowed down to simplify further analysis.

Proposition 1 *There exists a unique positive solution to* (3.7).

All proofs are in Appendix A. In a broad view, the uniqueness of the solution rules out multiple equilibria or solution jumps and solution bifurcations as ν changes continuously. In the coming sections, I assume that q_j is the unique optimal output produced by each plant j.

3.3 Free Entry Equilibrium

In equilibrium, plants must be indifferent between entering the market or staying out. If the fixed cost of entering the market is f_E, then the equilibrium requires that the expected profit be equal to this fixed cost to prevent an influx of new entry. In presence of entry costs, only plants operating above a certain cutoff productivity ϕ_* will be profitable and will stay in the market. Put formally, it must be that

$$\int_{\phi_*}^{\phi_M} \pi_j(\phi)dG(\phi) = f_E. \tag{3.8}$$

Plants operating at the cutoff productivity ϕ_* are expected to earn zero profit, i.e.,

$$\pi_j(\phi_*) = 0. \tag{3.9}$$

The equilibrium conditions (3.8) and (3.9) together with (3.6) and (3.7) determine an implicit relation between the cutoff productivity ϕ_* and model parameters. As one observation, note that ϕ_* is always less than ϕ_M, since $\phi_* = \phi_M$ is a clear contradiction to (3.8) when $f_E > 0$.

Finally, N can be determined endogenously when ϕ_* is known. Finding a closed form solution for N has proved to be difficult, although this does not limit my ability to assess the model's implications for productivity dispersion. Later, in Section 6, I look at a CRTS production function under which I have closed form solutions for all endogenous variables in terms of parameters. For that reason, I defer further discussion of N to that section.

3.4 Analytical Results

In this section I seek to describe the plant behavior within a market under the assumption that markets are localized. Comparative statics are also presented that define the distribution of input size and productivity within each market and across markets. For now, without loss of generality, focus on a single market.

Taking the partial derivatives of (3.7) with respect to ϕ_j provides the first general result

$$\frac{\partial q_j}{\partial \phi_j} = \frac{L w q_j^{1/\nu - 1}}{2\gamma\nu^2 \phi_j^{1/\nu-1} + (1-\nu) L w \phi_j q_j^{1/\nu-2}} > 0, \tag{3.10}$$

Proposition 2 *More productive plants produce more. However, there is an upper limit on output size that varies by market size.*

Lemma 1 *It is that $d\theta_j/d\phi_j > 0$.*

In the absence of shocks, revenue productivity is a monotonic and one-to-one transformation of physical productivity. This transformation consists of a scaling (non-uniform unless $\nu = 1$) and a shift. This result proves useful, because any model implications with respect to physical productivity can be immediately generalized to revenue productivity too. For this reason, in the coming propositions, I will refer to both measures of productivity simply as "productivity".

Continuing with the analysis, I combine (3.4) and (3.7) to get the following relation between input and output size:

$$x_j = \left(\frac{q_j}{\phi_j}\right)^{1/\nu} = \frac{\nu}{w}\left(\frac{\alpha\gamma + \eta N \bar{p}}{\gamma + \eta N} - \frac{2\gamma}{L} q_j\right) q_j. \tag{3.11}$$

Taking partial derivatives in (3.11) with respect to ϕ_j and using (3.10), it can be shown that

$$\begin{cases} \partial x_j/\partial \phi_j \geq 0 & \text{if } q_j \leq \frac{L(\alpha\gamma+\eta N\bar{p})}{4\gamma(\gamma+\eta N)}, \\ \partial x_j/\partial \phi_j < 0 & \text{Otherwise.} \end{cases} \tag{3.12}$$

Proposition 3 *Under the localized market assumption and when $\gamma > 0$, the relationship between input size and productivity is bell-shaped. The market elasticity of demand affects the peak and the width of the*

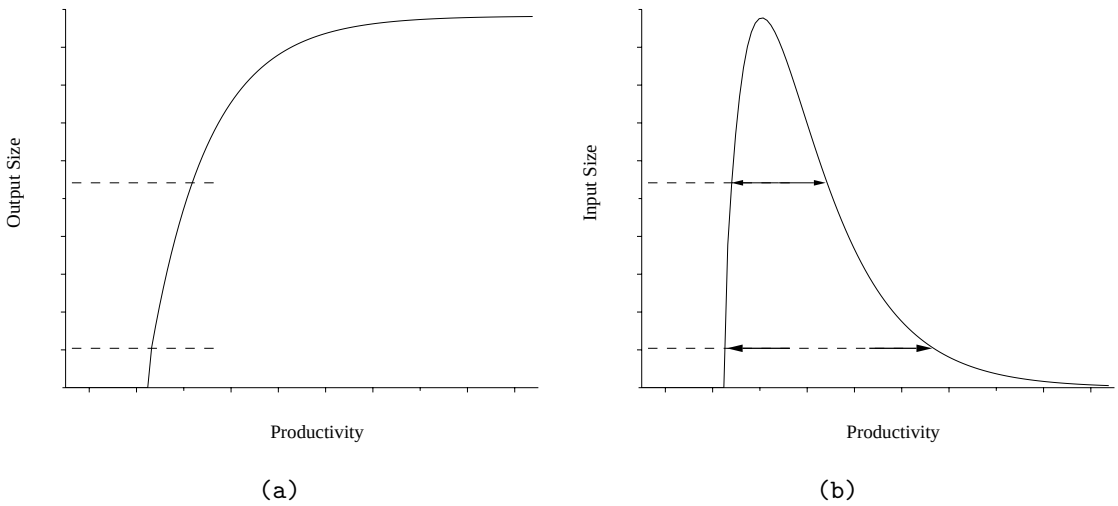


Figure 4: The bell-shaped relation between (a) output and productivity and (b) input and productivity. The arrows demonstrate the range of productivity dispersion in small and large plants.

bell curve.

Proposition 3 follows from continuity of the solutions plus (3.12). This result is a major departure from standard models, for it asserts that the input size need not grow monotonically with productivity. In my model, input size inside a market goes up only to the extent permitted by demand limitations. The maximum value of input size provides some measure of "size opportunities" in that particular market. Using (3.11) and (3.12) this maximum value is

$$x_{\max} = \frac{L\nu}{8w\gamma} \frac{\alpha\gamma + \eta N\bar{p}}{\gamma + \eta N}. \tag{3.13}$$

It seems that the peak of the bell-shaped curve should get higher and move to the right as markets get larger, though the endogeneity of N and $\bar{p}$ require some caution when making such statements. The simulation results in Section 6.2 and Appendix C, nevertheless, are clearly consistent with this assertion.

Given Proposition 3, the bell-shaped relationship between input size and productivity is no surprise. With output bounded from above, more productive plants are able to produce the limit output by hiring smaller inputs. The more productive they get, the less amount of input they need to produce that output (Figure 4). It is also useful to repeat that the localization of markets is essential to Proposition 3. This assumption makes it impossible for the more productive plants to improve their demand by trading with other markets, thus market size became a parameter in determining the plant performance. The following proposition characterizes the behavior of plants when their markets become globalized.

Proposition 4 *As $L \to \infty$, the relationship between input size and productivity converges to a monotonic one.*

Figure 5 shows the limit behavior of plants when L goes to infinity for both constant and decreasing returns to scale production functions. With decreasing returns production function ($\nu < 1$), the relation-

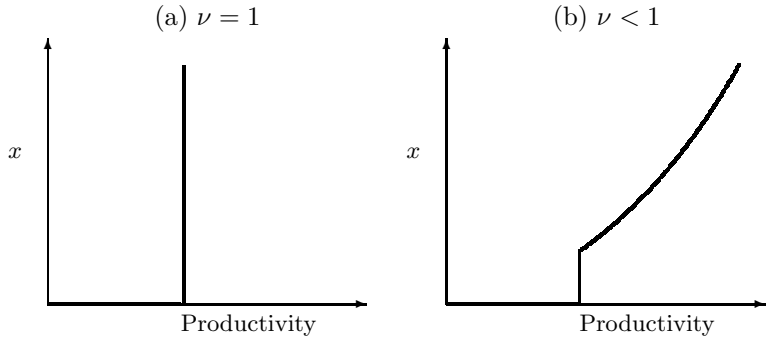


Figure 5: The limit behavior of plants when $L \to \infty$.

ship between input size and productivity converges to a strictly increasing exponential relation. With constant returns production function ($\nu = 1$), since the marginal productivity is not affected by size, the input sizes all go to infinity in the limit, rendering a not so strictly monotonic relation. In both cases, market size does not play a role in the plant performance anymore. Note that as $\gamma \to 0$, the implication is not the same as Proposition 4. With $\gamma = 0$, the products are perfectly substitutable, and the most productive plant in market can offer the lowest price and take over the whole market. Because of that, we shall have $\bar{p} = 0$ and $N = 0$ and a degenerate distribution of size and productivity in the market [7].

The bell-shaped relationship between productivity and input size in localized markets is what allows this model to generate higher productivity dispersion at lower levels of input size. For any given market, the gap between productivity differences closes as the level of input goes higher, and at the maximum input level of (3.13) the dispersion goes to zero (Figure 4). Later, in Section 6, a continuum of markets of different sizes is used to generate a dense productivity distribution at any given input size and whose dispersion still falls with the level of input. On the other hand, at any given output size within a market the productivity dispersion is zero. This a direct consequence of the monotonicity of the relationship between output size and productivity. However, when markets of different sizes are present, productivity dispersion can still be generated.

The cutoff productivity ϕ_* is another variable in this model that affects the distribution of productivity within a market, especially for plants with lower input sizes. Therefore, it is useful to know how the cutoff productivity varies by market size. Let q_* and x_* be the output and size for the plant operating at cutoff productivity ϕ_*. With some algebra, it can be shown that

$$\frac{\partial \phi_*}{\partial L} = \frac{\frac{\gamma}{L^2}\left(\frac{\gamma}{L}\frac{2\nu-1}{\nu}q_* + \frac{f}{\nu q_*}\right)\int_{\phi_*}^{\phi_M} q(q-q_*)dG(\phi)}{\frac{wx_*}{\nu^2\phi_*}\left(\frac{f}{q_*^2} + (2\nu-1)\frac{\gamma}{L}\right)\int_{\phi_*}^{\phi_M} qdG(\phi)}. \tag{3.14}$$

Proposition 5 *When $\nu \geq 0.5$, then it is that $\partial\phi_*/\partial L > 0$.*

Proposition 5 is in line with the findings of Melitz & Ottaviano (2005), who find that in larger markets

[7] Notice that with a continuous measure, measure of a singleton is zero.

tougher competition drives out the less productive plants and raises average productivity in the market. It can also be shown that $\partial\phi_*/\partial\gamma < 0$, consistent with the finding of Syverson (2003) that more product substitutability, or equivalently smaller γ, leads to higher cutoff productivities.

Summarizing, (3.12) and (3.14) together provide two instruments by which distributions of productivity across markets can be analyzed and compared. The generated patterns of productivity dispersion in the later sections of this paper will be a direct application of both of these findings. As the last comment, it should also be emphasize that the restriction $1 \geq \nu \geq 0.5$ merely helped to resolve signs on the comparative statics. The model outcome is by no means limited to decreasing or constant returns to scale. In fact, because of the continuity of all relations, the propositions and model implications still hold at some neighborhood of $\nu = 1$, which also includes increasing returns to scale. As the analysis of next sections will show, there is no reason to believe that the returns to scale in industries is very far from constant.

4 Data Considerations

The main source of data for this study is the US Center for Economic Studies' Census of Manufactures (CM) covering census years 1982, 1987, 1992 and 1997. For many small businesses the CM data are gathered from administrative sources or imputed, causing serious distortions. For that reason, only the weighted sample from CM is used. McGuckin & George A. Pascoe (1988) detail how the CM panels are composed and sampled. The unit of observation in the data is plant, defined as an individual physical place of production and identified with a Plant's Permanent Number (PPN). The CM panel provides information on plant observables such as industry, product class, employment, labor hours, location, capital stock, energy and material consumptions, and shipment value. Plants larger than 50000 employees and plants whose industry code was other than manufacturing were dropped. A few super-large plants with sizes in excess of 1 million employees, assumed to be erroneous records, fell into this group. Industry deflators and capital composition information from BLS are used to adjust for changing prices of input and output and to find estimates of equipment and structures capital (Chiang 2004). Estimates of plant age are supplemented using data from Haltiwanger, Jarmin & Schank (2003). The Standard Statistical Establishment List (SSEL) is also used to correct geographical information for some plants.

To limit myself to well-defined industries, I drop plants belonging to any four-digit SIC code ending in 9. These codes collect plants that could not be classified under any other detailed classification in the same group. Also I exclude two-digit SIC code 21 (Tobacco Industry) from my analysis. SIC 21 plants are disproportionately large when compared to other manufacturing industries and do not constitute an interesting industry. Only plants belonging to the US 50 states are kept in the analysis. This reduces the data set used for the analysis to 365 industries and 202593 establishments. My primary measure of size is total employment as defined by Davis, Haltiwanger & Schuh (1996, Appendix A.3.1).

	Census Years			
Shares	1982	1987	1992	1997
From total number of Plants	0.108	0.092	0.098	0.093
From total sales	0.174	0.126	0.127	0.116
# localized industry codes	32			

Table 1: Share of localized market industries from total manufacturing.

5 Plant Behavior in Localized Markets

The theoretical model focuses on market localization as one way of explaining the observed behavior of productivity dispersion. In this section, first I set to identify the localized market industries within my data. Then, to test the model's implications, I will select a subset of those industries that show a high degree of product differentiation. Under such circumstances, the theoretical model predicts that the bell-shaped relation between input size and productivity is more clearly defined. I supplement my data with the US Transportation Department's Commodity Flow Survey that provides me with information on value, tonnage, ton-miles shipped and average shipment distance for disaggregated commodities. The commodity descriptions are matched to four-digit Standard Industry Classification (SIC) codes as closely as possible using US Dept. of Labor descriptions[8].

I use the average shipment distance (DISTANCE) from Commodity Flow Survey as my measure of market localization. Notice that this measure is not with its shortcomings as it assumes that plants within an industry ship their products to the same radius, ignoring the heterogeneity of plants in targeting their markets and their variety of products. However, a low value of DISTANCE strongly signals that the dominant trend among plants is targeting localized markets. From now on I will identify the localized market industries as those four-digit industries that shipped their products on average no more than 150 miles away. Restricting attention to these industries leaves me with 32 industries (out of a total of 365), which is less than 10% of all the four-digit industries included in my data. However, as Table 1 shows, during the four census years in my data, these industries account for about 10% of the total number of plants in the data, and even produce more than 10% of the total output in nominal sales.

Under market localization assumption, the theoretical model predicts that the most productive plants will stay small in response to their demand limits. This creates a bell-shaped relationship between productivity and input size which, in turn, generates a decreasing relationship between productivity dispersion and input size. The rest of this section will estimate whether such a bell-shaped relationship is observed in the data for plants operating in localized markets. In this section and the remainder of the paper, I will measure the size of plants using employment. Employment is easily observed for each plant and has reasonably low measurement error, especially compared to estimates of the composite input. In defense of this shift, note that if the intensity of productive factors is constant within an industry, the optimal choice of each input factor will be a constant proportion of employment size, so that the composite input

[8] Available at `http://www.osha.gov/pls/imis/sic_manual.html` .

will be a linear function of employment. This enables me to treat the production function (3.4) as if it depended on labor only. Therefore, all the model's implications should still hold with employment as the measure of size.

I use both non-parametric and semi-parametric approaches to examine how employment responds to productivity, and how this response changes with market size. I use a similar semi-parametric method to examine the relationship between the number of plants and the size of the market.

Before proceeding with the results, there are a few points to note about the observed data.

1. Demand elasticity is an important factor that shapes the relationship between input size and productivity. The theoretical results of Section 3 provide guidance on which type of industries should demonstrate a more defined relationship between employment and productivity. Specially, choosing industries that demonstrate high values of γ causes plants to be more monopolistic and respondent to their market structure. As a measure of market differentiation, I am using the diversity index (DIVINDX) defined by Gollop & Monahan (1991). This index accounts for diversity not only by looking at the number of different products in an industry, but by how different the products are and how unequal the distribution of products is across production lines. I use the PPC product codes from the CM to distinguish different products in each industry. This index can be affected by over- or under-classification of output types in each four-digit industry. However, the values computed for DIVINDX showed a high degree of correlation with the expected degree of differentiation within most industries.

2. The employment size correspondence to productivity in the data is not one-to-one, as in the theory. In the data, demand shocks and other perturbations create a distribution of employment levels at each productivity level. To get a smooth relation between employment and productivity, the effect of perturbations must be taken out by application of some filtering technique.

Comment 1 outlines the type of industries that are best for the analysis, and comment 2 targets the methodology. In light of comment 1, I select industries based on the following conditions:

1. Satisfactory level of market localization: As before, I restrict attention to industries with DISTANCE no more than 150 miles.

2. Satisfactory level of product differentiation: The computed value of DIVINDX among industries ranges from 0 to 1.046. I restrict attention to industries with DIVINDX greater than 0.5, which is almost the middle point.

Conditioning on the above rules leaves me with 7 industries, listed in Table 2. Summary statistics show that on average these industries shipped their product 123.6 miles and had a product diversity index of 0.68. Most of the industries listed are food industries, providing some degree of homogeneity

Description	SIC	#Obs	DISTANCE	DIVINDX	Mean Size
Natural, Processed, and Imitation Cheese	2022	521	90	0.544	80
Dry, Condensed, and Evaporated Dairy Products	2023	262	90	0.829	103
Bread and Other Bakery Products	2051	1605	145	0.650	140
Cookies and Crackers	2052	352	145	0.528	191
Frozen Bakery Products	2053	184	145	0.543	103
Soft Drinks and Carbonated Waters	2086	1135	45	0.805	115
Concrete Products	3272	1494	91	0.789	29
Total		5553			

Table 2: List of industries and summary statistics on the industries used in the analysis.

among the products. The only exception is Concrete Products (SIC 3272), without which the method of Section 5.2 couldn't converge because of lack of observations. Therefore I am keeping SIC 3272 and use industry dummies to purge the industry effect. I should add that due to reliability concerns the industry code 3915 belonging to the "Miscellaneous" industry was dropped from the list, though it qualified as an industry with localized market and highly differentiated products. As before, I use the CM panels of years 1982,1987,1992, and 1997 to get a rich enough dataset.

The analysis is conducted for markets of different sizes to see the effect of market size on the relationship between employment and productivity. Measures of market size are supplemented from the US Census Bureau's County and City Databook for Core Based Statistical Areas (CBSA). This data, along with the cross-walks, provide the area and the list of counties belonging to 1784 urban centers in the USA. Combining it with county population estimates, two measures of market size, namely CBSA population and population density (per Sq.Miles), are constructed and used together to check the robustness of results. While population measures the number of consumers and, hence, the size of the demand, it ignores the physical extent of the market. Population density takes this into account and provides a measure of market size adjusted for physical distances. Availability of CBSA coding is very advantageous to my analysis because it provides statistics on both metro- and micropolitan areas, thus providing me with a larger range of market sizes to analyze[9]. Markets are grouped into four classes "Small", "Medium-Small", "Medium-Large", and "Large" based on their size measure. The size quartiles listed in Table 3 are chosen as break points for each class.

Residual rLP and rTFP are computed for each plant by running a regression of productivity on year and the interaction of industry and year. The total mean productivity across those plants in Table 2 is then added back to reset the scale.

[9]US Office of Management and Budget's definition of a metropolitan area is an urban area with population of at least 50,000. Micropolitan areas are those with population between 10,000 and 50,000.

Market Size	Mean	Std.Dev.	Min.	1st Qrtl.	Median	3rd Qrtl.	Max
Pop.	1058493.8	2549636.8	11638	208356	905266	3909954	18747320
Pop.Density	265.9	407.6	1.8	142.9	313.6	712.9	2792.2

Table 3: Summary statistics on the market sizes with both CBSA population and population density.

5.1 Size and Productivity: The Non-Parametric Approach

To estimate a smooth relationship between employment and productivity, I must first apply a certain degree of filtering to the data. However, there is worry that the findings will be driven by too much filtering rather than the behavior of plants. To address this concern, I first attempt a non-parametric fuzzy logic approach that uses a minimal level of filtering and helps me visualize the outline of the relations I am looking for. I use these results to guide a more structured semi-parametric estimation in the next section to estimate smooth moments of input size and productivity. The discussion in this section and next rely on revenue productivity θ instead of physical output productivity ϕ. Both measures rTFP and rLP are used to check the robustness of results.

Calculating mean of employment at each level of productivity is the most straightforward way of estimating their relation. However, in practice, the density of plants along the productivity axis is dispersed and non-uniform. Using percentiles to define the productivity classes misses the important aspects of the curvature, since plants are not uniformly concentrated along different parts of the curve. Experiments with different definitions of intervals proved to be nontrivial and time consuming. Choosing the number of classes posed a challenge because several points were needed to produce the steep parts of the curve, while using more classes automatically made the plots look noisy and unreliable.

Instead, I use a fuzzy-logic approach to the mean estimation. Fuzzy logic provides "soft" transitions between neighboring productivity classes by defining membership functions that generate a continuum of grades of membership for nearby observations according to some heuristic belief about how much they qualify to be a member of that class (Zadeh 1965). The fuzzy logic was meant to mimic a subjective or human-like way of classifying objects, rather than the Boolean logic commonly used by computers. Here, the fuzzy technique provides plots that capture curvature at the expense of losing some of the variation in the estimated employment sizes. The general form of the membership function I use is

$$\mu_x = \frac{1}{1 + \frac{(x-\theta)^2}{a}}, \quad a > 0, \tag{5.1}$$

where x spans the range of (revenue) productivities, θ is the centrality parameter that sets the productivity at which employment is to be estimated, and μ_x is the weight that measures the degree to which x qualifies as θ. The parameter a sets the bandwidth within which productivities strongly qualify as θ [10]. Smaller values of a narrow the bandwidth and include only those plants with productivities very close to θ, while larger values of a include plants with more distant productivities. Figure 6 shows how changing the

[10] The function's bandwidth is $2\sqrt{a}$.

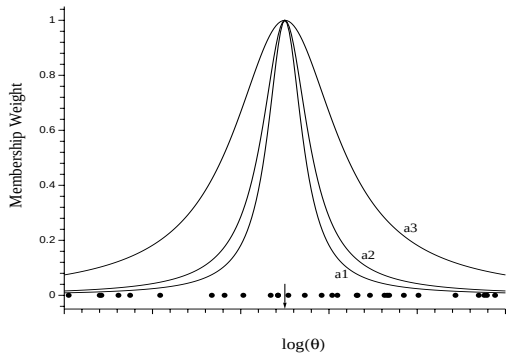


Figure 6: Shape of the membership function with three different bandwidth parameters $a3 > a2 > a1$. The solid dots represent some data points, and the center is marked by an arrow.

bandwidth parameter helps to include or exclude the observed data points. Obviously larger values of a provide smoother estimates, while causing remarkable loss of curvature. Therefore, it is reasonable to use smaller values of a for productivity levels where the plant density is high and to use larger values of a where plants are very dispersed. For my purpose, after some experimenting, I use the rule of thumb relation $a = 1/(5\sqrt{n_x})$, where n_x is the sum of weights used at the previous point estimate, assuming that the density of points does not change abruptly.

Mean employment is estimated at each log productivity level rather than the actual θ and is the weighted mean of all points, with weights assigned according to the membership function (5.1), centrality parameter set to $\log(\theta)$, and bandwidth chosen by the given rule above[11]. The size of employment is estimated for 150 uniformly distributed log productivity levels and for each market size class. Figure 7 summarizes the findings. A bell-shaped relationship between size and productivity is clearly seen for both measures of productivity and both measures of market size. The curves show a good degree of monotonicity with respect to market size in the sense that the curves move upwards as the size of market gets larger, although the peaks occur at around the same level of productivity, a fact that is revisited with revenue productivity in Section 6. It is also important to remember that these curves are somewhat flattened by the application of the membership function (5.1) so that both tails have an added offset and the peaks have been lowered. The adaptive bandwidth is somewhat responsible for this bias, assigning larger bandwidths on both tails of the curve as shown in Figure 8. At the same time, looking at the figure, the bandwidth selection mechanism seems to do a good job of detecting points of curvature.

The fuzzy non-parametric method was able to provide an outline of the relation between employment and productivity. The next section will apply a more structured semi-parametric method to produce smooth moments that will later be used to pinpoint the model parameters.

[11] This is the Center of Gravity defuzzification method (Zimmermann 1996).

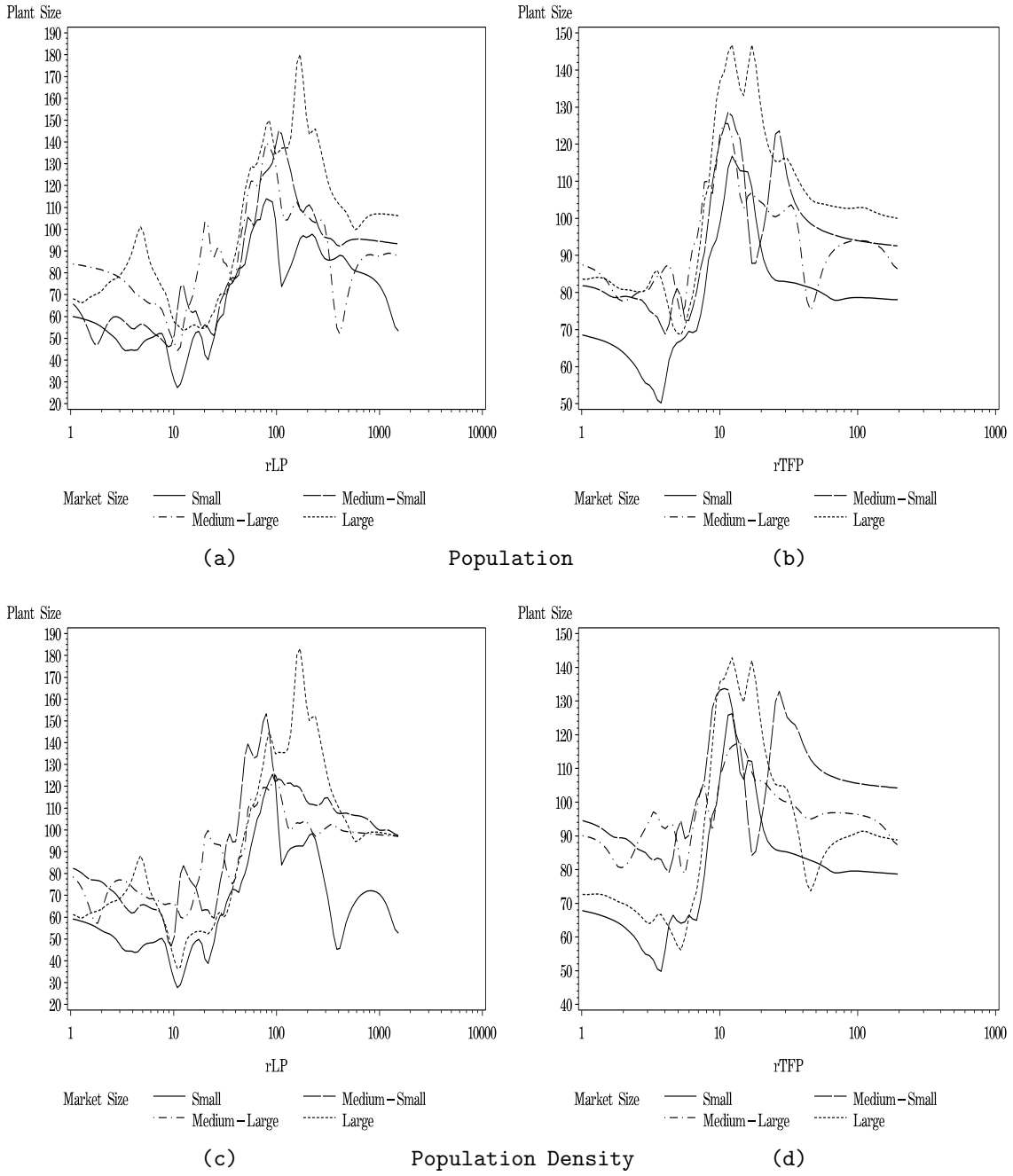


Figure 7: Non-parametric estimation of employment-productivity correspondences using (a) rLP and market population, (b) rTFP and market population, (c) rLP and market population density, and finally (d) rTFP and market population density.

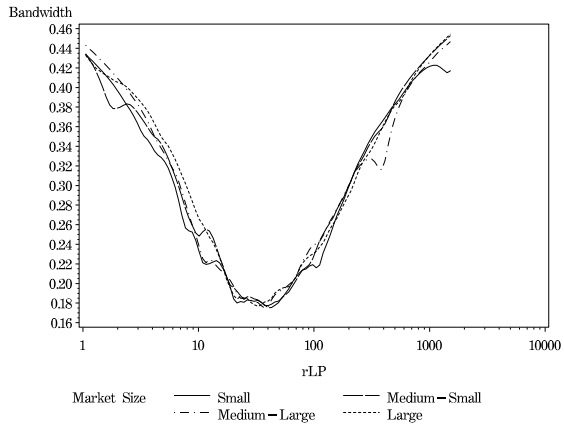


Figure 8: Plot of computed bandwidths to estimated employment-productivity relationship using rLP and market population measures.

5.2 Size and Productivity: The Semi-Parametric Model

In calculus, every smooth function can be written as an infinite polynomial sum. Truncating the sum at any point gives a corresponding approximation of the function. In particular, when approximating noisy data, it is desirable to filter the noise by eliminating the high frequency components; here high power terms. In this section a semi-parametric model is fitted in order to estimate a smooth functional relationship between employment and productivity. The general form of the model is

$$\log(l_{ijtm}) = \sum_{p=0}^{P} \alpha_{mp} \log(\theta_{ijtm})^p + h(i,t) + \epsilon_{ijtm}, \quad m = 1,2,3,4, \tag{5.2}$$

where i is the industry index, j is the plant index, t indexes time and m is the market class defined earlier by size quartiles. l is the employment size. P is the degree of the polynomial term used in the model. The model is estimated for each market class m separately to produce four vectors of estimates of the size-productivity relation.

Due to the presence of significant noise, a penalized least squares method is used that computes thin-plate splines to approximate a smooth curvature. The penalized least squares function is defined as

$$S_\lambda = \frac{1}{s_{ijtm}} \sum_{i,j,t} \epsilon^2_{ijtm} + \lambda J_2(h). \tag{5.3}$$

$J_2(h)$ is the integral of the square of the second derivative of h and is the penalty on the roughness of the fit. λ is the penalty parameter, whose choice is a trade-off between accuracy of the fit and its smoothness. s is the number of observations used. After some experimenting, λ was set to 10000 in my preferred specification.

In an experimental stage, polynomial powers were added one by one, until the estimates started to become unstable. The most stable predictions were achieved when $P = 3$. The estimated model was used to predict employment size at each (revenue) productivity. The results are shown in Figure 9. The plots

again confirm that the bell-shaped relationship exists between employment and productivity. The curves shift up as the size of market gets larger, except for the "Large" market class, whose peak does not stand on top and seems out of sync with other curves. As expected from the beginning, predictions with the rLP measure are more stable than those with rTFP.

As an extra test, I look at industries with globalized markets and compare their behavior to the localized market industries. In light of Proposition 4, industries with globalized markets should have their employment size increase monotonically with productivity, especially with decreasing returns production function. For comparison, the relationship between employment and productivity is estimated using (5.2) and (5.3) for pooled plants from two groups of industries (1) localized market and differentiated product industries (DIVINDX$\geq$ 0.5 and DISTANCE$\leq$ 150), (2) global market industries (DISTANCE$\geq$ 500). Plants of group 1 are the same ones listed in Table 2. Several industries in the data qualified for the second group and their list and descriptions appear in Appendix B. Figure 10 shows results using both rLP and rTFP. Industries facing global markets demonstrate a monotonic relation between their employment and productivity, and the implication with both measures of productivity are very similar.

The theoretical model of Section 3, suggests zero productivity dispersion at any given employment level for globalized market industries. The same inference is obtained from the monotonic relations for globalized market industries in Figure 10. In the data, however, plants with globalized markets show a mixed behavior in comparison with their localized market counterparts (Figure 11). Their productivity dispersion is still significant at any employment level, and the dispersion seems to fall with employment size, but not monotonically all the way and not with the same average rate. First thing to note is that many more industries were pooled to generate productivity dispersion for globalized market industries, and even after purging the industry effect, still some residual dispersion should be observed. That in part can explain the observed behavior. In a related work, Bakhtiari (2007) shows that for localized market industries the productivity dispersion falls faster with employment size. This finding is not affected by including other industry-specific factors that are capable of generating different productivity dispersions in small and large plants. A combination of those factors may also be the explanation behind the observed behavior of productivity dispersion for globalized market industries, whose study is beyond the goal of this paper.

5.3 Number of Plants per Market

The empirical relationship between number of plants and market size is another piece of instrument that will be needed in Section 6 to estimate model parameters, therefore this section is dedicated to the empirical estimation of such relationship. The theoretical model of Section 3 did not provide a clear picture of how N should vary with market size. In practice, larger markets offer larger demand and should have the capacity to accommodate more production plants. This fact seems especially likely under decreasing

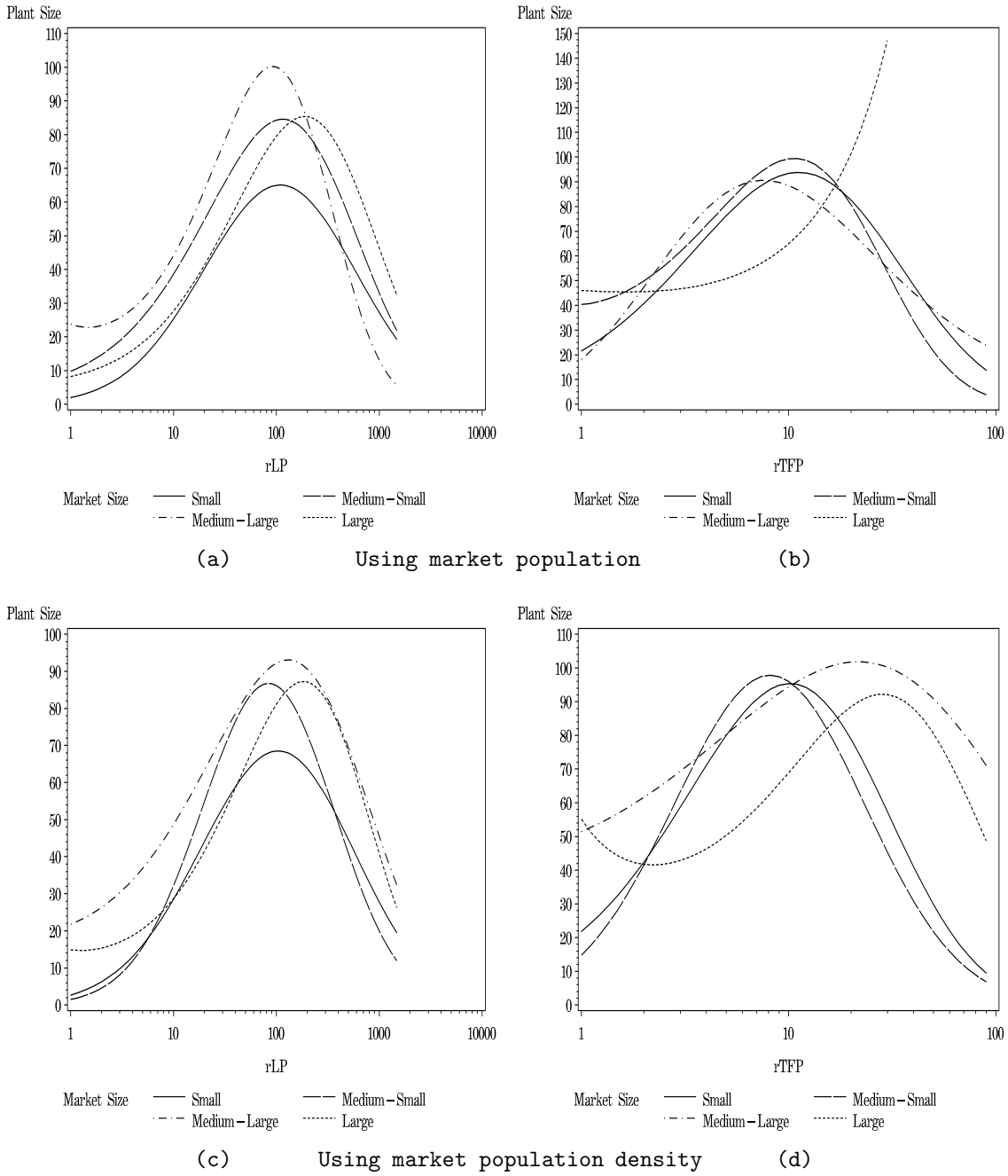


Figure 9: Estimated relationship between employment and productivity using (a) rLP and market population, (b) rTFP and market population, (c) rLP and market population density, and (d) rTFP and market population density.

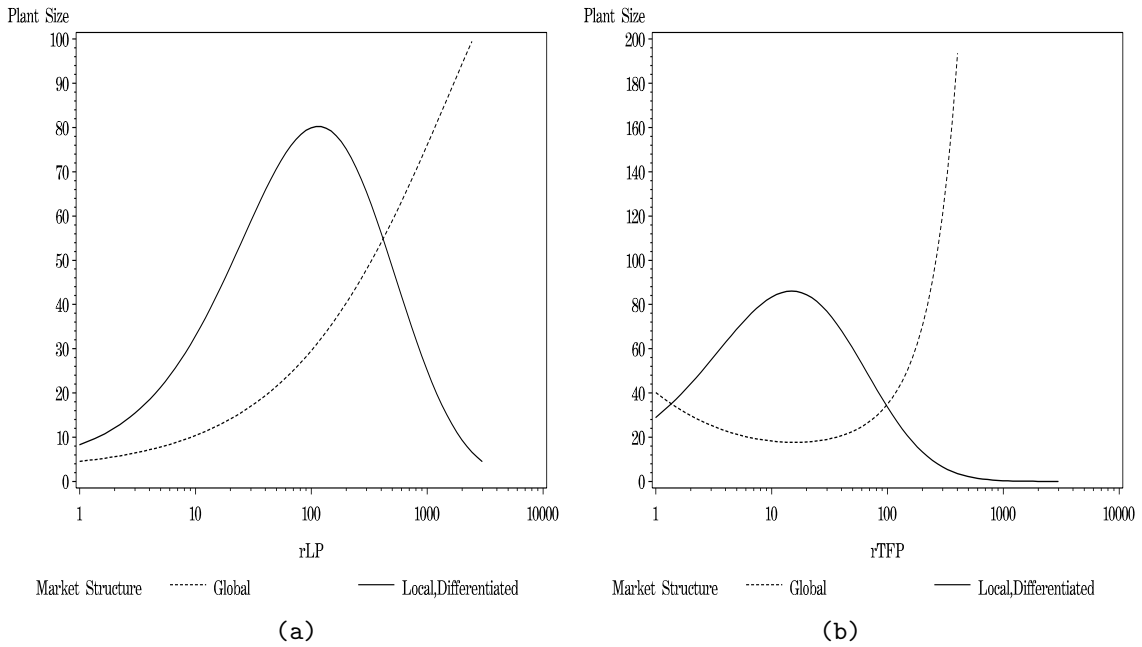


Figure 10: Comparative plots of the estimated size-productivity relation for local market differentiated products industries (Solid line) and the global industries (dashed line), using (a) rLP and (b) rTFP.

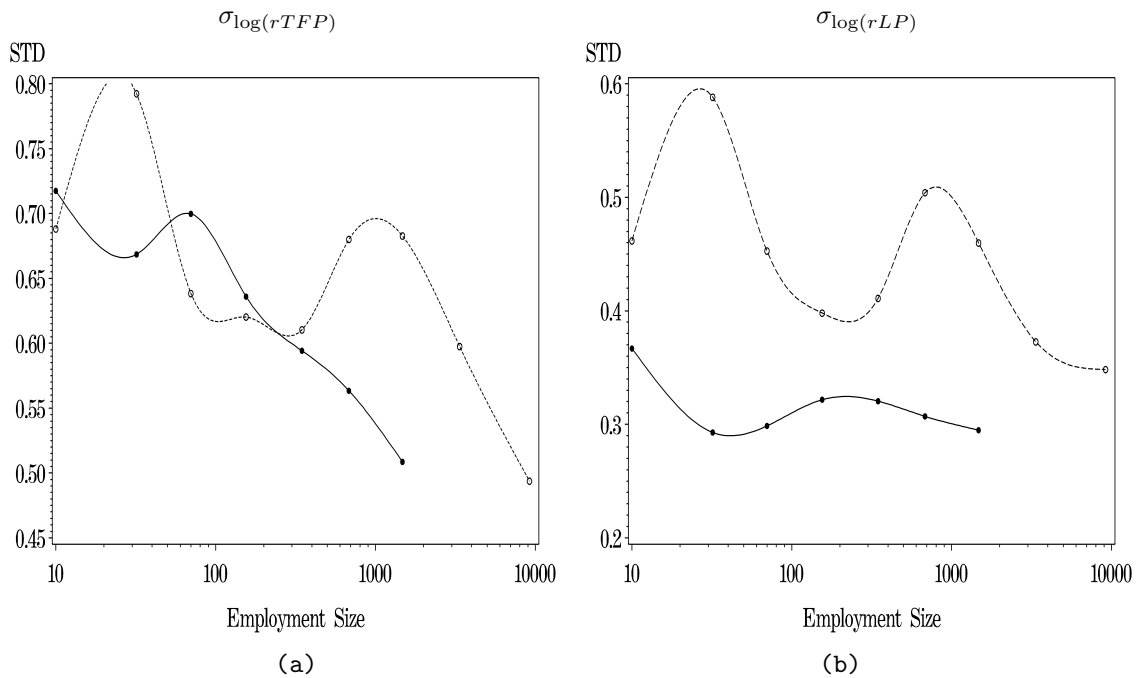


Figure 11: Behavior of productivity dispersion by employment for two groups of industry: localized market differentiated product industries (solid line) and globalized market industries (dashed line).

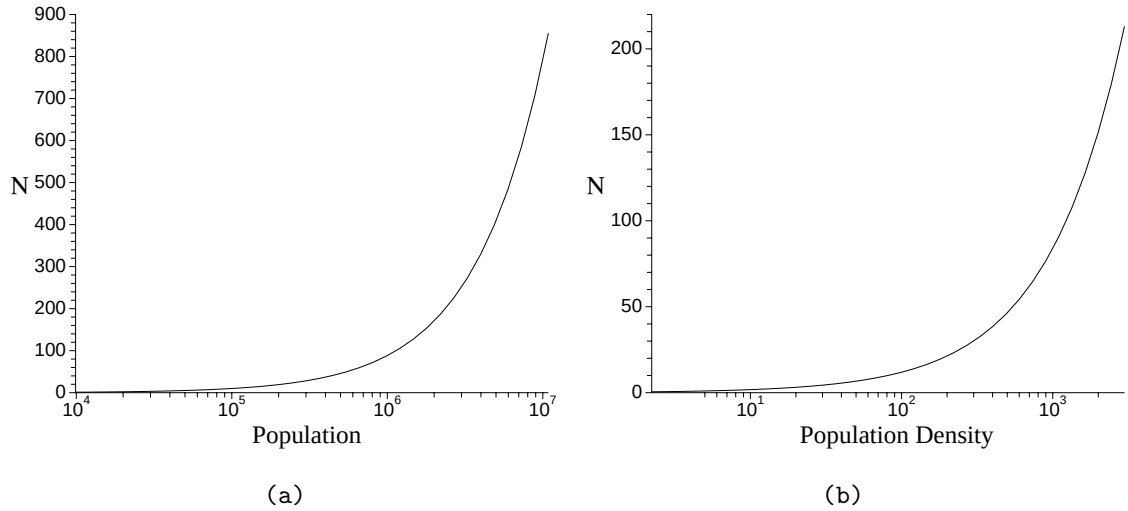


Figure 12: Number of manufacturing plants as a function of market size measured in (a) total population (b) population density.

returns production, where the production function intrinsically favors a large number of smaller operators. In the data, each plant can be associated with a market size through its geographical link. Then the sum of sample weights for plants belonging to each market provides an estimate of the number of plants operating there. A penalized least squares method is again used to produce a smooth semi-parametric relation between the number of plants and market size in the data. The relation is of the form

$$n_m = v(\log(L_m)) + \zeta_m. \tag{5.4}$$

Here m indexes each market (metro- or micropolitan area), and n_m is the weighted number of plants operating in market m, in log levels. L is either the population or the population density of each market. The value of penalty parameter λ is set to 100 in my preferred specification, and results are shown in Figure 12. The pictures suggest that larger markets are host to a larger number of plants, as expected. As shown in the simulation results below, the theoretical model is able to create a similar relation for the special case of constant returns to scale.

6 Numerical Simulation

The first-order condition of (3.7) simplifies to a linear equation when $\nu = 0.5$ or $\nu = 1$. In these two special cases closed form solutions for output and input size are possible. Each of these solutions constitutes an extreme case of the model behavior, with the behavior of the model for any ν between 0.5 and 1 being enveloped by them[12].

Existing evidence suggests that some industries produce under close to constant returns (Basu & Fernald 1997, Syverson 2004). When I estimate Syverson's model on my selected set of industries in

[12]This is a direct result of Proposition 1.

Description	SIC	$\hat{\nu}$	Std.Err.
Natural, Processed, and Imitation Cheese	2022	0.979	(0.006)
Dry, Condensed, and Evaporated Dairy Products	2023	0.981	(0.014)
Bread and Other Bakery Products	2051	0.997	(0.005)
Cookies and Crackers	2052	1.006	(0.009)
Frozen Bakery Products	2053	0.961	(0.016)
Soft Drinks and Carbonated Waters	2086	0.944	(0.008)
Concrete Products	3272	0.883	(0.008)
All		0.951	(0.003)

Table 4: Estimated returns to scale for sample industries.

Table 2, the resulting values for returns to scale are in the range 0.88 to 0.99 (Table 4). The last row in the table pools plants in all industries, purging the industry effect by including proper industry dummies. The results are consistent with a near constant returns to scale technology; consequently, I treat the case $\nu = 1$ with special attention, while deferring the analysis with $\nu = 0.5$ to Appendix C.

Also, as in the previous section, the size of plants is measured using employment sizes.

6.1 Estimation Methodology

Simulating the theoretical model entails estimating a set of parameters $\Lambda = \{\alpha, \eta, \gamma, \phi_M, w, f_E, f\}$ that minimizes the weighted squared error between the data provided moments and the simulated moments from the model. Section 5 provided two sets of data moments that can be of use in the estimation: (1) the relationship between employment and productivity, and (2) the relationship between number of plants and the market size.

Before describing the estimation method, it is useful to note that the profit function (3.6) can be written in the following form[13]

$$\pi_j = \left(\frac{f}{(1-\nu)q_*} - \frac{\gamma}{L}\frac{2\nu - 1}{1-\nu} q_* - \frac{\gamma}{L} q_j \right) q_j - w \frac{q_j^{1/\nu}}{\phi_j^{1/\nu}} - f. \tag{6.1}$$

From (3.9) and (6.1), q_* depends only on ϕ_* and the reduced set of parameters $\Lambda_0 = \{\gamma, \phi_M, w, f_E, f\}$. It then follows from (3.8) that ϕ_* is implicitly a function of Λ_0. Since all plant performance measures depend only on ϕ_* and Λ_0, it is easier to first estimate the reduced set Λ_0 using a weighted nonlinear least squares method, and then use the estimated parameters to pin point α and η [14]. However, the dependency of ϕ_* on Λ_0 creates an identification problem: I need to know ϕ_* to estimate Λ_0, but to compute ϕ_* the parameter set Λ_0 must be known. At the same time, the unavailability of data on physical output productivity makes separate identification of some parameters impossible when ϕ_* is not known. These facts, together, make it impractical to use standard nonlinear least squares methods to estimate the parameters.

[13] Please refer to the Appendix A for details of how to derive this relation.

[14] α and η enter the performance measures in a certain form easily replaceable by (A.13) from Appendix A.

Instead, I use a recursive method of simulated moments described in the following algorithm.

Algorithm 1

1. An initial Λ_0 is assigned,

2. Using Λ_0, ϕ_* is computed and simulated moments are produced as a function of revenue productivity θ (not ϕ).

3. Using a search method, a new parameter set Λ_0 is found that reduces the sum of square errors between the empirical and simulated moments[15].

4. Steps 2 and 3 are repeated until the change in parameter set falls below an acceptable tolerance.

The form of the weighted nonlinear least squares problem is

$$\min_{\Lambda_0} \quad (L^{data} - \hat{L}(\Lambda_0))W_0(L^{data} - \hat{L}(\Lambda_0))', \tag{6.2}$$

where L^{data} is the vector of estimated employment moments from Section 5.2 stacked for four market size classes. I use the relationship estimated for rLP data in Figure 9(a), as the estimates are more stable than those for rTFP. $\hat{L}$ is the vector of corresponding simulated moments generated by the model when using Λ_0. The moments are simulated using revenue productivities for each market class separately and plugging the average market size in that class as L into the model. In this way, the dimension of vectors L^{data} and $\hat{L}$ is the same. W_0 is a weighting matrix that governs the importance of different moments in setting the parameters. I used a diagonal weighting matrix initially. However, the computed moments on the market class "Large" seemed out of sync with other classes. To reduce estimation error, the weights on these moments were set to 0, while setting equal weights for the rest. In my data, employment size is estimated for 74 revenue productivity points in each market class. Therefore, 222 points are used for the estimation altogether.

In the second stage, I take the estimated Λ_0 from (6.2) and estimate the parameters α and η, using nonlinear least squares to minimize the following error function

$$\min_{\alpha,\eta} \quad (N^{data} - \hat{N}(\Lambda))W_1(N^{data} - \hat{N}(\Lambda)) + \lambda_1\hat{N}(\Lambda)I[\hat{N}(\Lambda) < 0]\hat{N}(\Lambda)'. \tag{6.3}$$

Here, N^{data} is the vector of the number of plants operating for each market size class and has 36 point estimates in my data. I use data from Figure 12(a), with population as the measure of market size. $\hat{N}$ is the simulated number using the complete set of parameters Λ, where Λ_0 are the estimated values from stage 1 and are fixed. $\hat{N}$ is estimated for the market size classes from the data, so that N^{data} and $\hat{N}$ are

[15] I use a pattern search with trust region adjustments to perform this search. Due to presence of implicit and complicated functions, finding analytical gradients and Hessians proved to be non-trivial. Gradient methods using numerical gradient computations with BFGS adjustments also got stalled. Alternatively, pattern search is completely insensitive to such irregularities.

forced to have the same dimension. W_1 is the weighting matrix, and I use an identity matrix to weight all estimates uniformly. The extra term in (6.3) is a penalty term that forces the simulated vector $\hat{N}$ to have non-negative values. λ_1 is the penalty parameter and $I[\,]$ is the diagonal matrix of indicator functions. By imposing a large penalty parameter λ_1, I make sure that the estimated values for α and η will not result in a negative simulated number of plants for any market. In my exercise, I set the value of λ_1 to 100.

6.2 The Constant Returns Case

With a CRTS production function ($\nu = 1$) a closed form solution to (3.7) can be found which is

$$q_j = \frac{L}{2\gamma}\left(\frac{\alpha\gamma + \eta N\bar{p}}{\gamma + \eta N} - \frac{w}{\phi_j}\right). \tag{6.4}$$

Using (6.4), the optimal profit can be computed and used in (3.9) to compute the cutoff productivity. The computed value for cutoff productivity is

$$\phi_* = \frac{w}{\frac{\alpha\gamma+\eta N\bar{p}}{\gamma+\eta N} - \sqrt{2\frac{\gamma f}{L}}} \tag{6.5}$$

A feature of the above cutoff productivity is that it summarizes the effects of endogenous variables N and $\bar{p}$ on the plant behavior. In fact, all plant's performance measures can be expressed as a function of ϕ_* and model parameters in the following way,

$$\begin{aligned}
q_j &= \frac{Lw}{2\gamma}\left(\frac{1}{\phi_*} - \frac{1}{\phi_j} + K\right), && (6.6)\\
p_j &= \frac{w}{2}\left(\frac{1}{\phi_*} + \frac{1}{\phi_j} + K\right), && (6.7)\\
l_j &= \frac{Lw}{2\gamma\phi_j}\left(\frac{1}{\phi_*} - \frac{1}{\phi_j} + K\right), && (6.8)\\
\pi_j &= \frac{Lw^2}{4\gamma}\left(\frac{1}{\phi_*} - \frac{1}{\phi_j} + K\right)^2 - f, && (6.9)\\
N &= \frac{2\gamma}{\eta w}\left(\frac{\alpha - \frac{w}{\phi_*} - wK}{\frac{1}{\phi_*} - \overline{\left(\frac{1}{\phi_j}\right)} + \frac{1}{2}K}\right), && (6.10)\\
K &= \frac{2}{w}\sqrt{\frac{\gamma f}{L}}.
\end{aligned}$$

with l_j being the employment size. Also, using the definition (3.5), the revenue productivity θ_j can be expressed as

$$\theta_j = \frac{w}{2}\left(1 + \phi_j\left(\frac{1}{\phi_*} + K\right)\right). \tag{6.11}$$

Obviously, θ is a function of the input price w and the market elasticity of demand embodied in ϕ_* and K, as well as the efficiency of production ϕ. The analysis of Section 3.3 together with the definition

of K show that larger L or smaller γ decrease the coefficient multiplying ϕ_j in (6.11). That, in turn, causes revenue productivity to under-represent the efficiency of production, especially in large markets. The dispersion of revenue productivity is actually affected by the scaling effects of w, ϕ_* and K, as well as by ϕ_* cutting the distribution from below. Since the under-representation affects larger markets more seriously, the dispersion of revenue productivity should fall faster with employment than its physical productivity counter part.

Upon entry, plants draw their random productivity from a variant of the Pareto distribution whose cumulative distribution function is

$$G(\phi) = \frac{\log(1+\phi)}{\log(1+\phi_M)}. \tag{6.12}$$

This distribution has a number of advantages for my analysis. As in the data, it implies a low probability of high productivity draws. Also, this functional form reduces the computational burden and improves the convergence of solutions. With this productivity distribution, the free entry equilibrium condition (3.8) can be solved for ϕ_* as a function of L, w, γ, f, and f_E, as elaborated in Section 6.1.

Using (6.8) and (6.10) to generate simulated relationships between size and productivity and between market size and the number of plants, the model parameters are estimated as follows:

α	η	γ	ϕ_M	w	f_E	f
1.835	0.222	4.059	8573.44	33.822	2431.4	658.94

The simulated performance of the model with the above parameters is illustrated in Figure 13; while simulated values for a selection of market sizes are listed in Table 5. A bell-shaped relationship between employment and both revenue and physical output productivity is demonstrated in both plots (a) and (b). As expected, in larger markets, plants can get larger and are more productive on average. Variations of the cutoff productivity and the measure of operating plants in the market are illustrated in plots (c) and (d), and both behave exactly as estimated in the data. It is also interesting to note that both ϕ_* and N grow at a slower rate than L [16], although by the analysis of Section 3.3 I know that ϕ_* will eventually hit an upper bound. I approximate these two relationships by functional forms $a_1 L^{b_1}$ and $a_2 L^{b_2}$. Applying a simple regression model to the simulated data, I estimate

$$\phi_* = 0.393 L^{0.418}, \quad N = 0.00026 L^{0.946}.$$

As a result, ϕ_* grows at a rate approximately half as fast as L.

Due to the bell-shaped relationship between employment and productivity, the correlation of employment and productivity is not necessarily positive as in other economic models. Table 5 shows negative correlations between employment and productivity for small markets, which increase toward positive values as markets get larger. This is reasonable, since larger markets give the more productive plants the

[16] Notice that the horizontal axis is in log space.

L	ϕ_*	N	$corr(\phi, l)$
10,000	18.013	0	-0.286
43,939	34.350	9.9	-0.248
193,069	64.708	56.9	-0.191
848,342	74.943	179.3	-0.155
3,727,593	218.699	313.4	0.035
10,000,000	321.942	494.8	0.155

Table 5: Cutoff productivity, variety measure, and size-productivity correlation by market size.

chance to be larger. It is also clear from the linear relation of (6.11) between ϕ and θ that the correlation of employment size l and revenue productivity θ within a market is identical to the correlation using physical productivity ϕ.

The simulated productivity dispersion curves are numerically obtained by performing a Monte Carlo simulation of productivity and market size draws. Market sizes are drawn from a uniform distribution in the range 10,000 to 10,000,000. Draws of ϕ are drawn from the distribution (6.12). In my simulation, I assume that each plant chooses its productivity and market size independently upon entry, making it trivial to write the joint probability density. 100,000 random draws of productivity and market size are taken from the resulting joint distribution and productivity dispersions are computed for both ϕ and θ and for size classes spaced in logs. The curves are illustrated in Figure 14. As expected, productivity dispersion falls with employment, slowly in the case of physical productivity and faster with revenue productivity.

To complete the results, productivity dispersions are also plotted against output size (Figure 15). These plots show an increasing trend in productivity dispersion as output gets larger, closely resembling the empirical counterpart in Figure 1, especially that of rLP. Note that the model of Section 3 was neither composed nor calibrated to mimic the behavior of productivity dispersion with output size. The similarity results shows that this model can perform in a more general and versatile way.

So far the model of Section 3 has proven successful in producing significant productivity dispersion at each employment level, with the productivity dispersion falling as employment increases. However, the results to this point have relied on the assumption that the wage rate is constant within and across markets. The next section examines whether results are robust to allowing for wage variations.

6.3 The Effect of Wage Variations

A fixed wage within a market is justified in the absence of worker's skill heterogeneity when homogeneous workers are mobile within a market. In equilibrium, wages will level across producers to make workers indifferent between staying with their current employers or changing jobs. However, ruling out worker mobility across the markets creates different labor supply and demand curves in markets of different sizes. The most likely outcome is wages that vary by market size. The sensitivity of the model behavior is tested by assuming that wages follow a certain functional relation with market size. The functional form

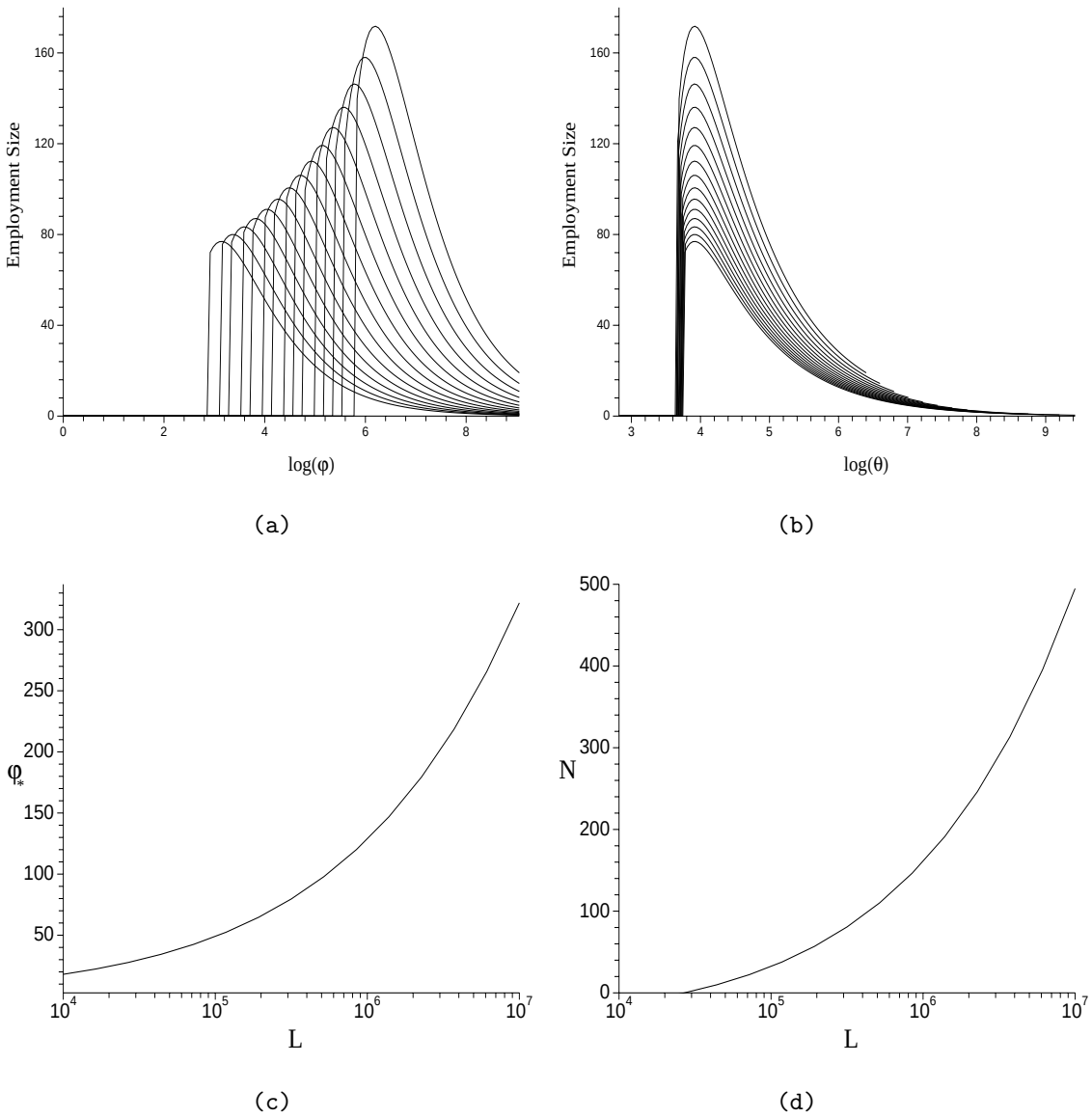
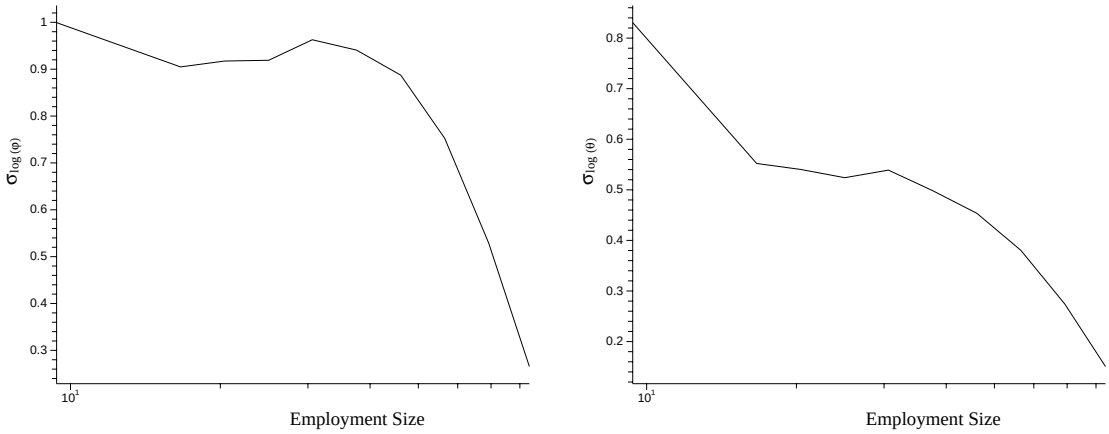
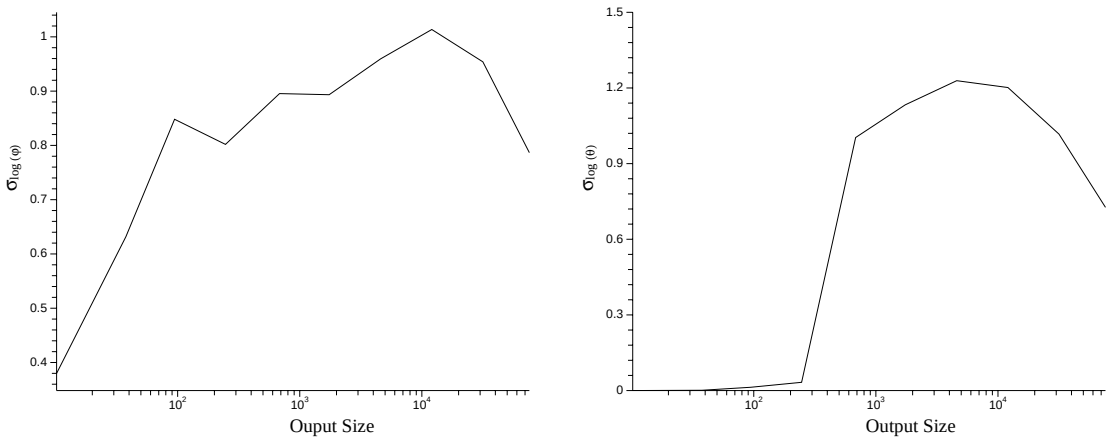


Figure 13: Plant size-productivity relation with a constant returns to scale technology with respect to (a) physical productivity ϕ and (b) revenue productivity θ. Variations of (c) cutoff productivity and (d) measure of product variety are also illustrated.



(a) (b)

Figure 14: Productivity dispersion by employment size using (a) productivity ϕ and (b) revenue productivity θ.



(a) (b)

Figure 15: Productivity dispersion by output size using (a) productivity ϕ and (b) revenue productivity θ.

assumed here is logarithmic and is written as

$$w = w_0(1 + \Delta_w \log(L)), \tag{6.13}$$

where w_0 is an offset wage and Δ_w is a non-negative variation factor. The effect of this wage variation on the performance of plants will be stronger in larger markets, where plants now face an increased cost of input. As markets get larger, their cutoff productivity is pushed higher because plants find it harder to be profitable. Also, in larger markets, operating plants will hire less labor in response to an increased wage. On the productivity dispersion plots, most of the change will happen at the smallest employment levels, as they are the plants being hit hardest by this increased labor cost. Since higher wages force more small plants to exit the market, productivity dispersion should fall at low employment levels.

In the data, the smallest market is Pecos, Texas with a population of 11,638, and the largest one is

New York, New York with a population of 18,747,320. The US Bureau of Labor Statistics (BLS) reports the mean annual wage of manufacturing production worker in Pecos and New York areas to be \$23860 and \$28999, respectively. This amounts to a roughly 21% wage difference. I will simulate the model with 5%, 15%, and 25% maximum wage variations. Note that from (6.13) the total percentage difference in wages across markets relates to Δ_w in the following way:

$$\Delta w\% \simeq 79\frac{\Delta_w}{1+\Delta_w}, \tag{6.14}$$

so I will simulate the model using Δ_w values of 0.07, 0.23, and 0.46. Figure 16 illustrates results for a CRTS production function. As expected, the cutoff productivity moves up as the wage variations get larger and markets get bigger. The number of product varieties in each market goes down as wage variations get larger, reflecting the fact that, especially in large markets, the high cost of labor causes fewer plants to operate. With higher cutoff productivities, productivity dispersion was predicted to fall for plants with smaller employment. With revenue productivities, plants are ordered not just based on their physical productivity, but by their elasticity of demand. Therefore, the impact of wage variation on productivity dispersion is harder to predict. Figure 16(c,d) show the dispersion for both ϕ and θ as wage variation increases. In summary, results under CRTS production function show very good amount of robustness even in presence of very large wage differences.

In the theoretical analysis, the fixed cost of operation f is also assumed constant within and across markets. I am not doing a separate analysis with varying fixed cost because the effect of varying fixed cost across markets will essentially be the same as the effect of varying wage (but not quite identical). Higher fixed costs in larger markets raise the cutoff productivity and drive more plants out of the market, causing productivity dispersion to change in a similar way to what is observed in Figure 16.

7 Conclusion

Current economic models typically produce a monotonic long-run relationship between the input size and productivity. While these models are useful for some purposes, they are unable to explain for example why plants with the same levels of employment are significantly dispersed in their productivities. In this paper, I examined a model of product differentiation with localized markets in which more productive plants cannot improve their demand by trading with other markets and are forced to choose their sizes according to what their demand size dictates. Smaller markets, especially, force the very productive plants to stay small. This generates a bell-shaped relationship between input size and productivity within each market, which in turn generates considerable productivity dispersion at any input level. Simulation results show that this model is capable of producing productivity dispersion patterns, exactly as observed in the data. My empirical work detects a bell-shaped relationship between employment and productivity within

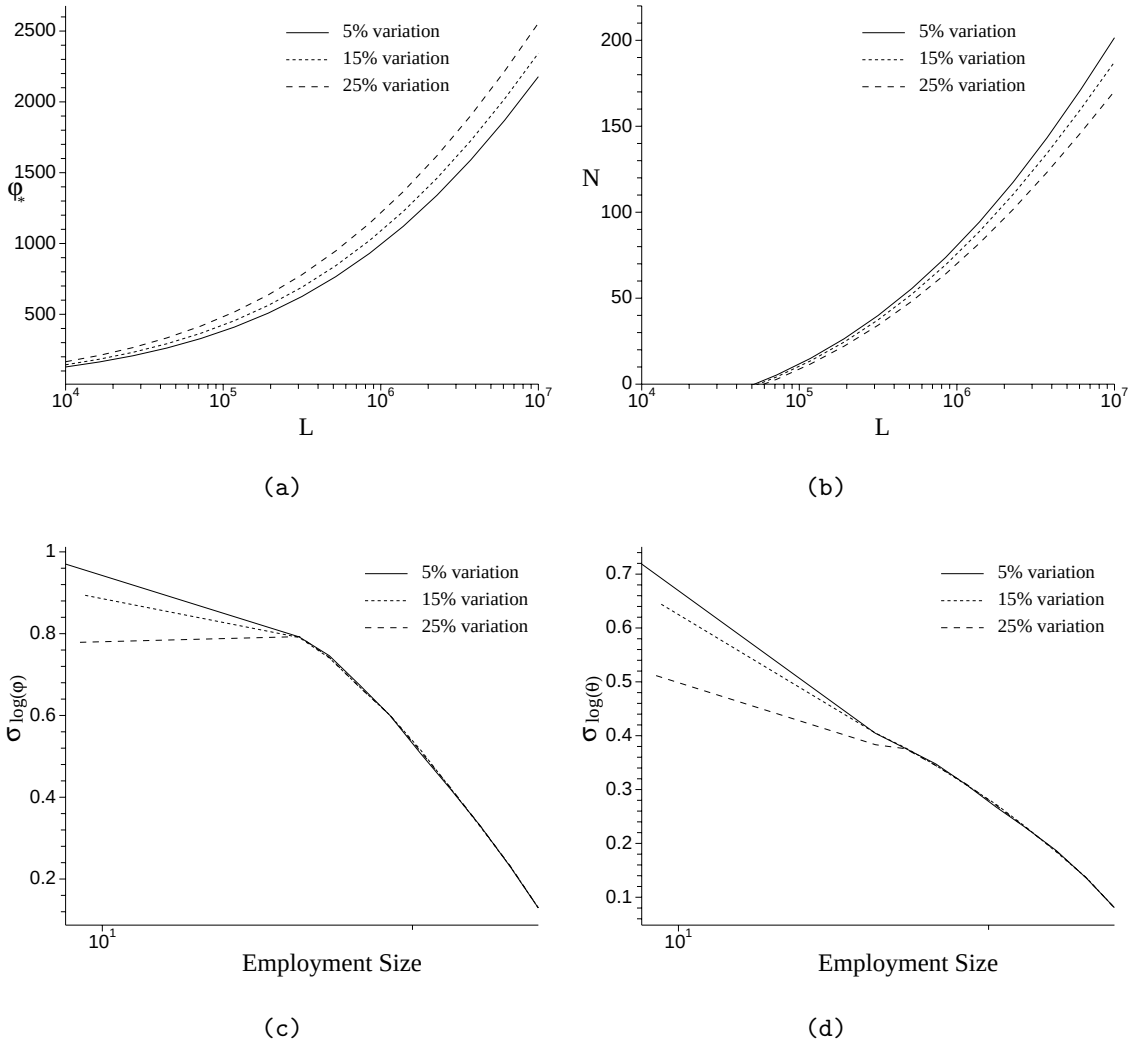


Figure 16: Effect of wage variation with constant returns to scale technology (a) cutoff productivity (b) measure of varieties (c) dispersion of ϕ by employment and (d) dispersion of θ by employment.

a subset of industries with localized markets. On the contrary, plants in industries with globalized markets do not face a demand constraint and their employment sizes increase monotonically with productivity. The theoretical results are shown to be fairly robust to wage variations across markets. Both revenue and physical productivity measures are used in simulations and results are shown to have the same qualitative implications.

References

Abel, A. B. & Eberly, J. C. (1996), 'Optimal investment with costly reversibility', *Review of Economic Studies* **63**, 581–593.

Acs, Z. J. & Audretsch, D. B. (1991), *Innovation and Technological Change*, The University of Michigan Press, Ann Arbor, chapter 4,6.

Bakhtiari, S. (2007), The behavior of productivity dispersion by plant size: A cross-industry study, Mimeo, University of Maryland.

Bartelsman, E. J. & Doms, M. (2000), 'Understanding productivity: Lessons from longitudinal microdata', *Journal of Economic Literature* **38**(3), 569–594.

Basu, S. & Fernald, J. G. (1997), 'Returns to scale in u.s. production: Estimates and implications', *Journal of Political Economy* **105**(2), 249–283.

Bertola, G. & Garibaldi, P. (2001), 'Wages and the size of firms in dynamic matching models', *Review of Economic Dynamics* **4**(2), 335–368.

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.

Brown, C. & Medoff, J. (1989), 'The employer size-wage effect', *Journal of Political Economy* **97**(5), 1027–1059.

Chew, W., Clark, K. B. & Bresnahan, T. (1990), *Measure for Manufacturing Excellence*, Harvard Business School Press, Boston, Massachusetts, chapter 5. edited by Robert S. Kaplan.

Chiang, H. (2004), Learning by doing, worker turnover and productivity dynamics, Mimeo, University of Maryland.

Churchill, N. C. & Lewis, V. L. (1983), 'The five stages of small business growth', *Harvard Business Review* pp. 30–50.

Davis, S. J. & Haltiwanger, J. (1991), 'Wage dispersion between and within u.s. manufacturing plants, 1963-1986', *Brookings Papers on Economic Activity, Microeconomics* **1991**, 115–200.

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.

Dunne, T., Foster, L., Haltiwanger, J. & Troske, K. (2000), Wage and productivity dispersion in u.s. manufacturing: The role of computer investment, Technical Report 7465, NBER Working Paper series.

Ericson, R. & Pakes, A. (1995), 'Markov-perfect industry dynamics: A framework for epirical work', *The Review of Economic Studies* **62**(1), 53–82.

Evans, D. S. (1987), 'The relationship between firm growth, size, and age: Estimates for 100 manufacturing industries', *The Journal of Industrial Economics* **35**(4), 567–581.

Fisher, F. M. (1993), *Aggregation*, 1st edn, MIT Press, Cambridge.

Foster, L., Haltiwanger, J. & Syverson, C. (2006), Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?, Mimeo, University of Maryland.

Gollop, F. M. & Monahan, J. L. (1991), 'A generalized index of diversification: Trends in u.s. manufacturing', *The Review of Economics and Statistics* **73**(2), 318–330.

Greiner, L. E. (1998), 'Evolution and revolution as organizations grow', *Harvard Business Review* pp. 55–67.

Hall, B. H. (1987), 'The relationship between firm size and firm growth in the us manufacturing sector', *The Journal of Industrial Economics* **35**(4), 583–606.

Haltiwanger, J. C., Lane, J. I. & Spletzer, J. (2000), Wages, productivity, and the dynamic interaction of businesses and workers, Technical Report 7994, NBER Working Paper series.

Haltiwanger, J., Jarmin, R. & Schank, T. (2003), Productivity, investment in ict and market experimentation: Micro evidence from germany and the u.s., Technical Report CES-WP-03-06, CES Discussion papers.

Hamermesh, D. S. (1995), 'Labour demand and the source of adjustment costs', *The Economic Journal* **105**(430), 620–634.

Jovanovic, B. (1982), 'Selection and the evolution of industry', *Econometrica* **50**(3), 649–670.

Katayama, H., Lu, S. & Tybout, J. R. (2003), Why plant-level productivity studies are often misleading, and an alternative approach to inference, Technical Report 9617, NBER Working Paper series.

McGuckin, R. H. & George A. Pascoe, J. (1988), The longitudinal research database (LRD): Status and research possibilities, Technical Report CES-88-2, CES Discussion papers.

Melitz, M. & Ottaviano, G. I. (2005), Market size, trade, and productivity, Mimeo, Harvard University.

Nelson, R. R. (1981), 'Research on productivity growth and productivity differences: Dead ends and new departures', *Journal of Economic Literature* **19**, 1029–1064.

Olley, G. S. & Pakes, A. (1996), 'The dynamics of productivity in the telecommunications equipment industry', *Econometrica* **64**(6), 1263–1297.

Syverson, C. (2003), Product substitutability and productivity dispersion, Technical Report 10049, NBER Working Paper series.

Syverson, C. (2004), 'Market structure and productivity: A concrete example', *Journal of Political Economy* **112**(6), 1181–1222.

Zadeh, L. A. (1965), 'Fuzzy sets', *Information and Control* **8**, 338–353.

Zimmermann, H.-J. (1996), *Fuzzy Set Theory and Its Applications*, 3rd edn, Kluwer Academic Publishers.

A Technical Appendix

Proof of Proposition 1

For $\nu = 1$ the proof is trivial. Let $0 < \nu < 1$. Testing (3.7) for two extreme values $q_j = 0$ and $q_j \to \infty$ and with the continuity of $\partial\pi_j/\partial q_j$, at least one crossing point is found in the range $q_j > 0$. Moreover, the second derivative of the profit function is

$$\frac{\partial^2 \pi_j}{\partial q_j^2} = -\frac{2\gamma}{L} - \frac{(1-\nu)w}{\nu^2 \phi_j^{1/\nu}} q_j^{1/\nu - 2}, \tag{A.1}$$

which is always negative for any $q_j > 0$. Since two maxima cannot appear next to each other without any local minimum in between them, then, there is only one positive solution to (3.7).

To show that the solution can never be negative, let $q_j < 0$ be the solution to (3.7). We notice that a negative solution can always be written in complex form as $q_j = qe^{\imath\pi}$, where $q > 0$ and q is real. Replacing this in (3.7) results in a left hand side with nonzero imaginary part for any $\nu < 1$. Having a real right hand side, this contradicts the fact that q_j is a solution. ♢

Proof of Proposition 2

Since prices must be non-negative, it follows from (3.3) that output size is bounded above within a certain market. Now if we let $\phi_j \to \infty$ in (3.7) and having $\bar{p} \leq \alpha$ (and therefore $(\alpha\gamma + \eta N\bar{p})/(\gamma + \eta N) \leq \alpha$), then q_j will converge to $L\alpha/2\gamma$. ♢

Proof of Lemma 1

From (3.7) the optimal price for each plant can be written as

$$p_j = \frac{\gamma}{L} q_j + \frac{w}{\nu}\frac{x}{q}. \tag{A.2}$$

Therefore, using (A.2) and knowing that $q_j = \phi_j x_j^{\nu}$, the revenue productivity can be written as

$$\theta_j = \frac{p_j q_j}{x_j} = \frac{\gamma}{L}\phi_j^{1/\nu} q_j^{2-1/\nu} + \frac{w}{\nu}. \tag{A.3}$$

Taking derivatives with respect to ϕ results in

$$\frac{d\theta_j}{d\phi_j} = \frac{\gamma}{L\nu}\phi_j^{\frac{1}{\nu}-1} q_j^{2-\frac{1}{\nu}} + \frac{\gamma(2\nu-1)}{L\nu}\phi_j^{\frac{1}{\nu}} q_j^{1-\frac{1}{\nu}} \frac{dq_j}{d\phi_j} > 0, \tag{A.4}$$

and the above result follows because of (3.10). ♢

Proof of Proposition 4

First, I show that the endogenous term $\frac{\alpha\gamma+\eta N\bar{p}}{\gamma+\eta N}$ can never converge to zero. If so, then the only possible case is when $N \to \infty$ and $\bar{p} \to 0$. But it means that $p_j \to 0, \forall j$. In turn, (3.3) implies that $q_j \to 0, \forall j$. But this means that all plants will exit the market, driving N to zero. This contradicts the original assumption that $N \to \infty$. Hence, $0 < \frac{\alpha\gamma+\eta N\bar{p}}{\gamma+\eta N} \leq \alpha < \infty$.

To complete the proof, two cases must be treated separately.

Case 1, $\nu = 1$: Then using (3.7) and knowing that $q_j = \phi_j x_j$, I can write

$$x_j = \frac{L}{2\gamma}\left(\frac{\alpha\gamma + \eta N\bar{p}}{\gamma + \eta N}\phi_j - w\right). \tag{A.5}$$

Then since the term $\frac{\alpha\gamma+\eta N\bar{p}}{\gamma+\eta N}$ is always positive, size of plants with productivities above a certain cutoff productivity will go to infinity.

Case 2, $\nu < 1$: Using (3.7) and with boundedness of the right-hand side, it is clear that q_j cannot grow faster than L. Then as $L \to \infty$, (3.7) converges to

$$a_0 + \frac{w}{\nu\phi_j^{1/\nu}} q_j^{1/\nu - 1} = A_0, \tag{A.6}$$

where a_0 is a positive constant if L and q_j grow at the same rate, and zero if q_j grows slower than L. A_0 is the limit value of $\frac{\alpha\gamma+\eta N\bar{p}}{\gamma+\eta N}$ and non-negative. Note that (A.6) requires that $q_j \geq 0$ and $a_0 \leq A_0$. Using $q_j = \phi_j x_j^{\nu}$ results in

$$x_j = \left(\frac{(A_0 - a_0)\nu}{w}\phi_j\right)^{\frac{1}{1-\nu}}. \tag{A.7}$$

The case where $a_0 = A_0$ (q_j and L grow at the same rate) can be immediately rejected here as it implies that $q_j \to 0$, and that contradicts the fact that $a_0 > 0$. Therefore, x_j will be exponentially increasing in ϕ_j.

From Lemma 1, it also follows that he limit relationship between input size and revenue productivity is a monotonic one, and that completes the proof. ♢

Algebraic Steps to (3.14)**:** Let's define

$$A = \frac{\alpha\gamma + \eta N\bar{p}}{\gamma + \eta N}. \tag{A.8}$$

At this point A is an endogenous variable that will facilitate further algebra. The cutoff conditions will be

$$\frac{2\gamma}{L}q_* + \frac{w}{\nu\phi_*^{1/\nu}} q_*^{(1-\nu)/\nu} = A \tag{A.9}$$

$$\left(A - \frac{\gamma}{L}q_*\right) q_* - \frac{w}{\phi_*^{1/\nu}} q_*^{1/\nu} = f \tag{A.10}$$

Eliminating A between (A.9) and (A.10), and substituting $q_* = \phi_* x_*^{\nu}$, gives

$$\frac{\gamma}{L}q_*^2 + \frac{(1-\nu)w}{\nu}x_* = f. \tag{A.11}$$

Equation (A.11) can be rewritten in the following way

$$q_*\left(\frac{2\nu - 1}{1-\nu}\frac{\gamma}{L}q_* + \frac{2\gamma}{L}q_* + \frac{w}{\nu\phi_*^{1/\nu}} q_*^{(1-\nu)/\nu}\right) = \frac{f}{1-\nu}. \tag{A.12}$$

Looking at (A.12), it is easy to recognize and replace the term from (A.9). Thus, with some simple algebra, (A.12) yields

$$A = \frac{f}{(1-\nu)q_*} - \frac{2\nu - 1}{1-\nu}\frac{\gamma}{L}q_*. \tag{A.13}$$

By replacing A in the profit function, a plant's profit at an optimum can be expressed as below which is a function of q_* only:

$$\pi_j = \left(\frac{f}{(1-\nu)q_*} - \frac{\gamma}{L}\frac{2\nu-1}{1-\nu}q_* - \frac{\gamma}{L}q_j \right) q_j - w\frac{q_j^{1/\nu}}{\phi_j^{1/\nu}} - f. \tag{A.14}$$

To find $\partial\phi_*/\partial L$, I need to find $\partial q_*/\partial L$ first. To find the derivatives, note that a change in market size affects q_* both directly and indirectly, through ϕ_*. With this in mind, taking partial derivatives of (A.11) with respect to L results in

$$\frac{\partial q_*}{\partial L} = \frac{\frac{\gamma}{L^2}q_*^2 + \frac{(1-\nu)w}{\nu^2\phi_*}x_*\frac{\partial\phi_*}{\partial L}}{\frac{2\nu-1}{\nu}\frac{\gamma}{L}q_* + \frac{f}{\nu q_*}}. \tag{A.15}$$

Next, insert (A.14) into (3.8) and take partial derivatives with respect to L. Replacing $\partial q_*/\partial L$ from (A.15) leads to (3.14). ♢

Proof of Proposition 5: Both the nominator and the denominator in the (3.14) will be unambiguously positive in the light of (3.10) and if $\nu \geq 0.5$. Therefore, it immediately follows that $\partial\phi_*/\partial L > 0$. Having a fixed maximum productivity, the higher the cutoff productivity goes, the smaller the productivity dispersion will become. ♢

B Globalized Market Industries

SIC	Description
2034	Dried Fruits, Vegetables, Jams, Jellies
2371	Fur Goods
2386	Leather and Sheep-lined Clothing
2387	Apparel Belts
3021	Rubber and Plastic Footwear
3111	Leather Tanning and Finishing
3131	Boot and Shoe Cut Stock and Findings
3142	House Slippers
3143	Men's Footwear
3144	Women's Footwear
3151	Leather Gloves and Mittens
3161	Luggage
3171	Women's Handbags and Purses
3172	Personal Leather Goods
3671	Electron Tubes
3672	Printed Circuit Boards
3674	Semiconductor and Related Devices
3675	Electronic Capacitors
3676	Electronics Resistors
3677	Electronic Coils, Transformers and Inductors
3678	Electronic Connectors
3691	Storage Batteries
3692	Primary Batteries, Dry and Wet
3711	Motor Vehicles and Passenger Cars
3721	Aircraft
3724	Aircraft Engines and Engine Parts
3728	Aircraft Parts and Auxiliary Equipment
3761	Guided Missiles and Space Vehicles
3764	Guided Missile and Space Vehicle Propulsion Units
3841	Surgical and Medical Instruments
3842	Orthopedic, Prosthetic, and Surgical Appliances
3843	Dental Equipment and Supplies
3844	X-Ray Apparatus and tubes
3845	Electromedical and Electrotherapeutic Apparatus
3851	Ophthalmic Goods
3942	Dolls and Stuffed Toys
3944	Games, Toys, and Children Vehicles

Table 6: The list of industries with DISTANCE$\geq$ 500 miles.

C Model Behavior Under Decreasing Returns Production

The first-order condition (3.7) can also be analytically solved when $\nu = 0.5$. This analysis is particularly useful, because Proposition 1 suggests that the behavior of the model should change continuously as ν moves from 0.5 to 1.

With $\nu = 0.5$ the solution to (3.7) is

$$q_j = \frac{\alpha\gamma + \eta N\bar{p}}{\gamma + \eta N}\frac{L\phi_j^2}{2(\gamma\phi_j^2 + Lw)}, \tag{C.1}$$

which will result in the following profit function

$$\pi_j = \frac{1}{2}\frac{\alpha\gamma + \eta N\bar{p}}{\gamma + \eta N}q_j - f. \tag{C.2}$$

Again, all the plant performance measures can be expressed as a function of this cutoff productivity :

$$q_j = \frac{AL\phi_j^2}{2(\gamma\phi_j^2 + Lw)}, \tag{C.3}$$

$$p_j = \frac{A}{2}\left(1 + \frac{Lw}{\gamma\phi_j^2 + Lw}\right), \tag{C.4}$$

$$l_j = \frac{A^2L^2\phi_j^2}{4(\gamma\phi_j^2 + Lw)^2}, \tag{C.5}$$

$$\pi_j = \frac{fLw(\phi_j^2 - \phi_*^2)}{\phi_*^2(\gamma\phi_j^2 + Lw)}, \tag{C.6}$$

$$N = \frac{2\gamma(\alpha - A)}{\eta A(1 - \bar{K})}, \tag{C.7}$$

$$A^2 = \frac{4f(\gamma\phi_*^2 + Lw)}{L\phi_*^2}, \quad K = \frac{Lw}{\gamma\phi_j^2 + Lw},$$

where l is the employment size. Using the definition (3.5), the revenue productivity θ_j can be expressed as

$$\theta_j = \frac{\gamma}{L}\phi_j^2 + 2w. \tag{C.8}$$

Looking at (C.8), it is clear that higher elasticity of demand results in an under-representation of productivities in revenue term. Therefore, using revenue productivities will result in steeper productivity dispersion curves.

Using (C.5) and (C.7) to generate simulated relationships between size and productivity and between market size and the number of plants, model parameters are estimated as follows:

α	η	γ	ϕ_M	w	f_E	f
6.566	1.825	22.601	37368.7	17.516	2903.66	255.01

To test the reliability of the estimates, it is especially noted that the estimated w is lower than that of Section 6.2 and the estimated ϕ_M is actually higher. This is consistent with a decreasing returns production where being larger is a disadvantage. Here the model compensates for that effect by estimating lower input costs and expanding the range of possible productivities.

The model's behavior is simulated and illustrated in Figure 17. Table 7 lists market performance measures for selected market sizes. Again, approximate functional forms $a_1L^{b_1}$ and $a_2L^{b_2}$ are used to examine the rates by

L	ϕ_*	N	$corr(\phi, l)$	$corr(\theta, l)$
10,000	19.592	0	-0.273	-0.190
43,939	38.484	24.8	-0.263	-0.191
193,069	74.835	76.6	-0.239	-0.189
848,342	143.581	176.4	-0.189	-0.182
3,727,593	270.278	363.8	-0.085	-0.152
10,000,000	405.833	568.0	0.033	-0.102

Table 7: Cutoff productivity, variety measure, and size-productivity correlation by market size.

which ϕ_* and N respond to the market size. Applying simple linear regression to the simulated data, I estimate

$$\phi_* = 0.349L^{0.440}, \quad N = 0.004L^{0.764}.$$

To generate productivity dispersion, a Monte Carlo simulation is performed here by drawing 100,000 random samples from the same distribution in Section 6.2 productivity dispersion is computed for each employment class. The dispersions of ϕ and θ by employment are shown in (Figure 18). Figure 19 shows productivity dispersions computed by output size. The findings here are again consistent with the model prediction and resembling the results of Section 6.2.

The sensitivity of the model to different degrees of wage variation is also tested for %5, %15, and %25 maximum wage variation and results are illustrated in Figure 20. Similar to the results of Section 6.3, very good amount of robustness to wage variation is achieved here too.

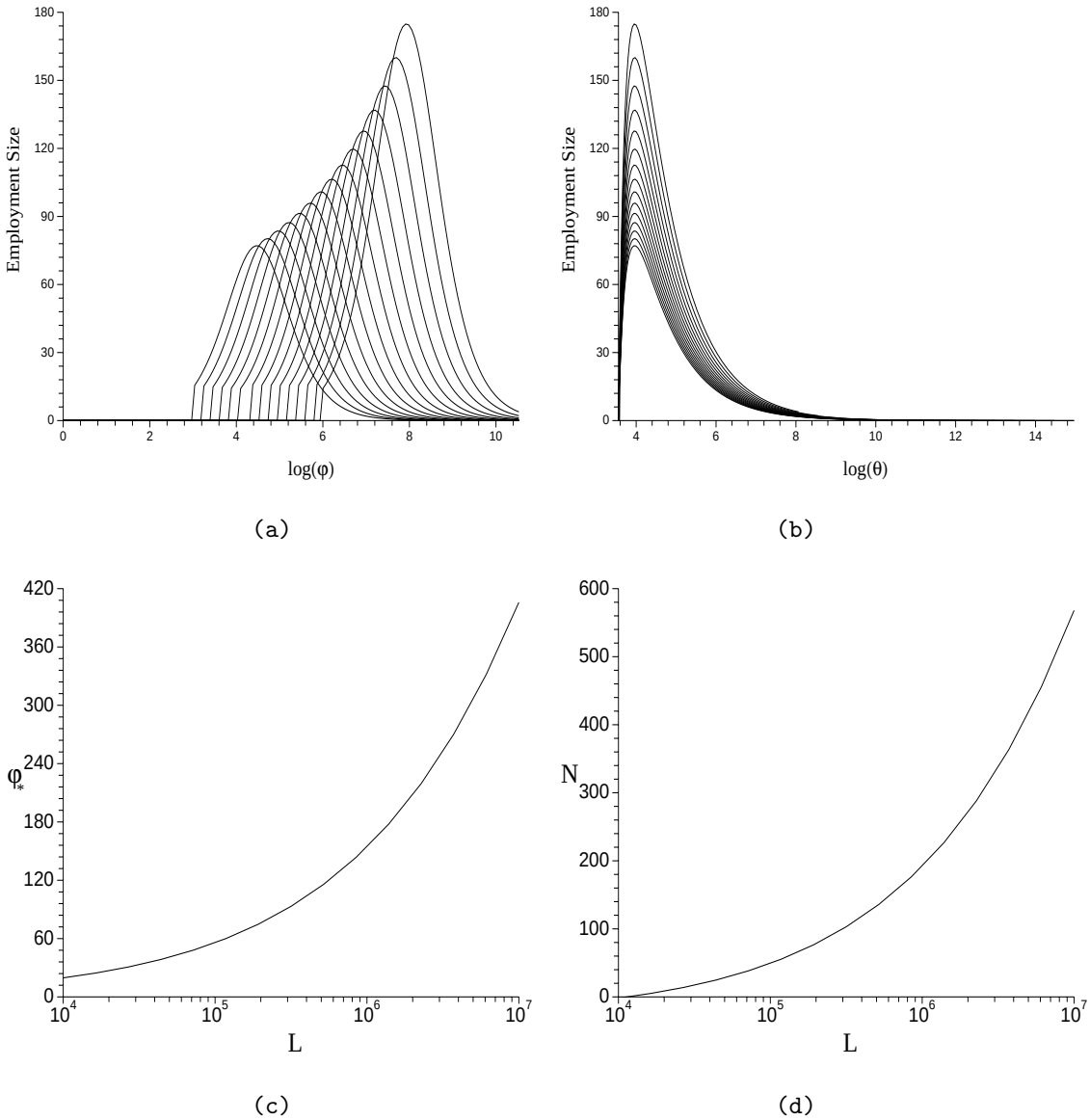
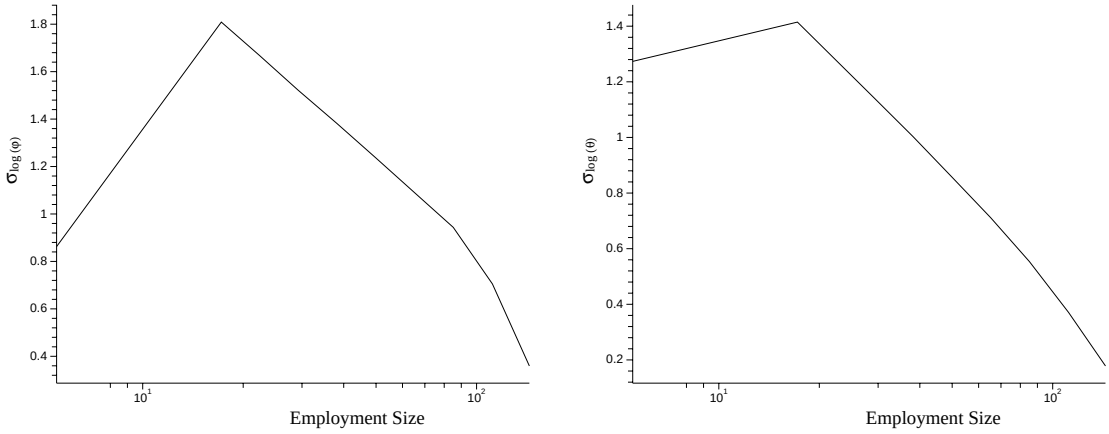
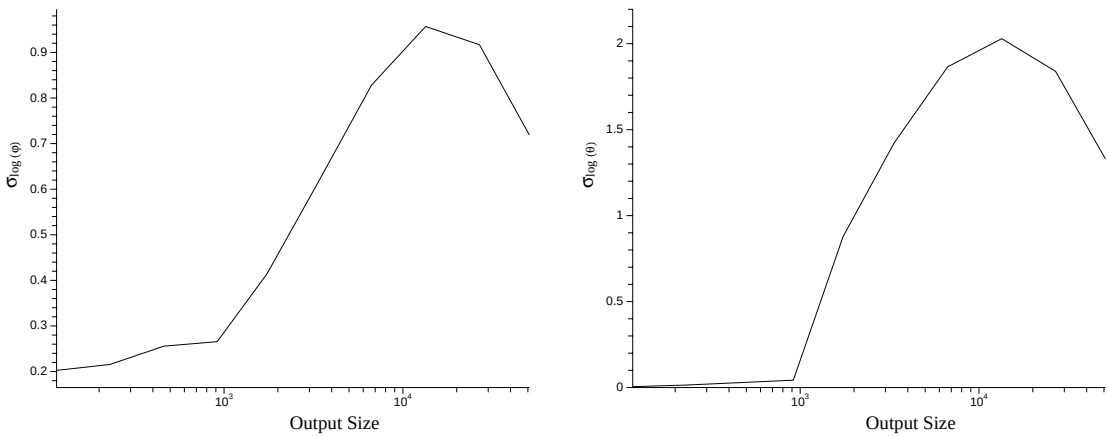


Figure 17: Plant size-productivity relation with a decreasing returns to scale technology with respect to (a) productivity ϕ and (b) revenue productivity θ. Variations of (c) cutoff productivity and (d) measure of product variety are also illustrated.



(a) (b)

Figure 18: Productivity dispersion by employment using (a) productivity ϕ and (b) revenue productivity θ.



(a) (b)

Figure 19: Productivity dispersion by output size using (a) productivity ϕ and (b) revenue productivity θ.

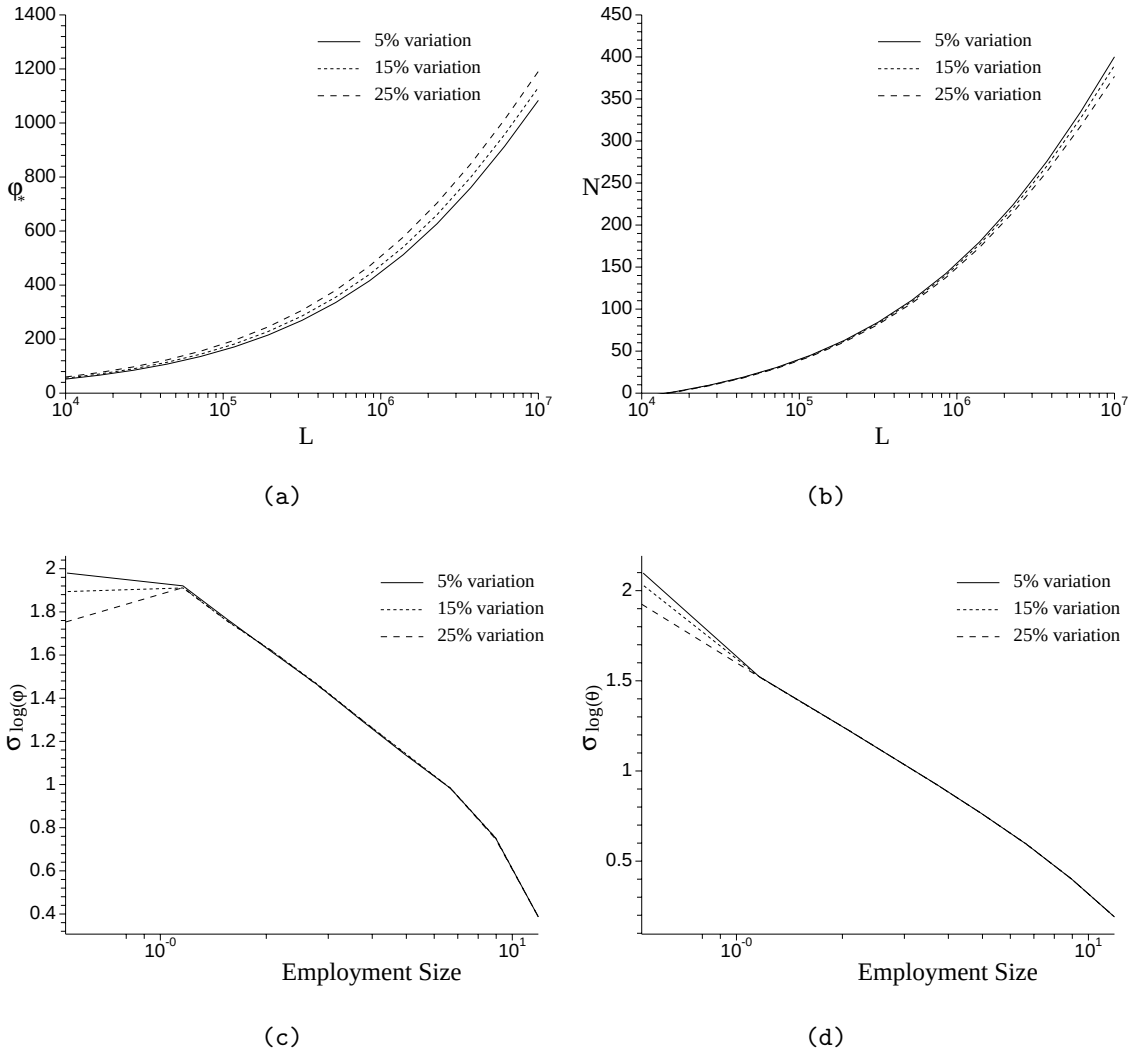


Figure 20: Effect of wage variation with decreasing returns to scale technology (a) cutoff productivity (b) measure of varieties (c) dispersion of ϕ by employment and (d) dispersion of θ by employment.