THE UNIVERSITY *of York*

Discussion Papers in Economics

No. 2005/34

A Theoretical Foundation for Understanding Firm Size Distributions and Gibrat's Law

by

Christopher A Laincz and Ana Sofia Domingues Rodrigues

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

August, 2005

A Theoretical Foundation for Understanding Firm Size Distributions and Gibrat's Law*

Christopher A. Laincz[1] and Ana Sofia Domingues Rodrigues[2]

Abstract

This paper presents a dynamic model of the firm size distribution. Empirical studies of the firm size distribution often compare the moments to a log-normal distribution as implied by Gibrat's Law and note important deviations. Thus, the first, and basic questions we ask are how well does the dynamic industry model reproduce Gibrat's Law and how well does it match the deviations uncovered in the literature. We show that the model reproduces these results when testing the simulated output using the techniques of the empirical literature. We then use the model to study how structural parameters affect the firm size distribution. We find that, among other things, fixed and sunk costs increase both the mean and variance of the firm size distribution while generally decreasing the skewness and kurtosis. The rate of growth in an industry also raises the mean and variance, but has non-monotonic effects on the higher moments.

JEL Codes: L11, L13

Keywords: Firm Size Distribution, Gibrat's Law, R&D

1 Drexel University, claincz@drexel.edu (corresponding author)

2 University of York, asdr101@york.ac.uk

---

*We would like to thank Dong Chen, Huw Dixon, Ulrich Doraszelski, Patrick Paul Walsh, seminar participants at the Federal Reserve Bank of Philadelphia, the University of York (UK), Drexel University, the 2004 Conference of the Society for Computational Economics held at the University of Amsterdam, 2004 meeting of the SEA, the 2004 Conference on "Innovation, Entrepreneurship and Growth" held at the Royal Institute of Technology, Stockholm, and the 2005 IIOC conference in Atlanta for their helpful comments. Ana Rodrigues would also like to acknowledge the generous support of Fundação para a Ciência e a Tecnologia do Ministério da Ciência e do Ensino Superior de Portugal.

1 Introduction

Studies on the firm size distribution and Gibrat's Law to date have been the province of empiricists. We can write down various reduced form models, as in McCloughan (1995), to reproduce many of the statistical facts surrounding the firm size distribution and Gibrat's Law of Proportionate Effect which states that the growth rate of a firm is independent of its size. However, little of the empirical work has been guided by a formal structural model. In Caves' (1998) survey on the recent empirical findings in industrial organization, he states, "Although the importance of these facts for economic behavior and performance is manifest, their development has not been theory-driven." This paper seeks to take a step towards filling this gap.

We employ an extension of the Ericson and Pakes (1995) model of a dynamic industry that allows for firm growth developed by Laincz (2004a). By varying key priors, the simulations demonstrate potential sources for the various, and sometimes conflicting, results on Gibrat's Law uncovered in the empirical literature. We demonstrate that the model matches empirical findings on Gibrat's Law.

A more recent literature uncovers significant cross-industry variation in the higher moments of the firm size distribution. Machado and Mata (2000) find that industry characteristics such as technological orientation and capital-intensity are significantly related to the skewness. Lotti and Santerelli (2004) show how the distribution of a new cohorts differs across different industries and over time. Audretsch et al. (2004) present evidence suggesting that the firm size distribution of the service industry differs from manufacturing. We use the model to develop theoretical reasoning for many of these findings, however, our analysis also emphasizes that some variables have strong non-monotonic effects on the moments of the firm size distribution suggesting caution in generalizing empirical results based on linear specifications.

After briefly reviewing the lengthy empirical literature on Gibrat's Law and its relationship to the firm size distribution in the next section, section 3 presents the basic model. In section 4 we compare the results of a baseline simulation to the empirical literature on the firm size distribution and Gibrat's Law. Section 5 then documents how varying key structural parameters alters the firm size distribution. Section 6 summarizes the results.

2 Gibrat's Law and Empirical Findings

Following the seminal works of Hart and Prais (1956) and Ijiri and Simon (1964), the industrial organization literature devoted much energy into exploring the statistical regularity known as Gibrat's Law as it applies to the firm size distribution. Figure 1 shows the size distribution of enterprises for the U.S. in 2001. Notably, the distribution is significantly skewed to the right with the large peak for the smallest size class with non-zero employment. The following simple statistical process generates almost the same distribution. Let x_i be the size of firm i, then growth from one period to the next is represented as:

$$x_i(t) = x_i^{\beta}(t-1)\exp\left[u_i(t)\right], \, \beta > 0 \tag{1}$$

where $u_i(t) \backsim iid\ N(\mu, \sigma^2)$. Defining $y_i(t) = \ln x_i(t)$, then:

$$y_i(t) = \beta y_i(t-1) + u_i(t). \tag{2}$$

When $\beta = 1$ we have Gibrat's Law wherein the growth rate of a firm is independent of its size and the process yields a log-normal distribution of firm sizes. Empirical work on the firm size distribution finds that this characterization is a close, but imperfect proxy for the data. The earliest work on Gibrat's Law only had data available for large firms. Hart and Prais (1956), for example, included only firms listed on the London Stock Exchange between 1885 and 1950. They found that Gibrat's Law provided a good statistical approximation for the distribution. Simon and Bonini (1958) found similar results for large US firms.

More recently, Hart and Oulton (1996) compare the implications of (1) to a large sample of firms measured by employees, net sales, and net assets. They find that the distribution has a long right tail, with skewness coefficient estimates ranging from 0.19 to 0.75, and leptokurtic with values from 4.58 to 6.20. However, they argue the deviations should not be compared with the extreme of matching the log-normal distribution exactly and that the close approximation justifies the use of Gibrat's Law in empirical work.

Our task is rather different. We are specifically interested in the deviations themselves. We want to construct a sensible model of optimizing firm behavior that can both approximate the distribution and provide us with a tool to understand the deviations and, moreover, cross-industry differences. Before turning to the model, we look at the literature that explicitly rejects the strong form of Gibrat's Law where β is exactly one.

Mansfield (1962) was perhaps the first to explicitly deal with the problems that entry and exit present for the interpretation of Gibrat's Law. Specifically, since exiting firms effectively have a growth rate of -100%, does Gibrat's Law hold for all firms, only the survivors, or for firms exceeding a size threshold such as minimum efficient scale? Of the three, he found that the latter interpretation fit his data the best using a χ^2 test on the lognormality of the distributions for each of his industries in each time period. In growth size regressions, Mansfield found that in the entire sample of survivors, firms grow less than proportionally, i.e. $\beta < 1$. However, analyzing large firms only, he found that the mean growth rate is independent of size, i.e. $\beta = 1$. He still concluded that Gibrat's law does not hold for any of the versions considered due to the fact that, even for the case of larger firms only, the variance of growth rates decreases with size.

Subsequent empirical analysis largely confirmed Mansfield's initial foray into the subject. Using more advanced econometric techniques to deal with heteroscedasticity and sample selection bias, Hall (1987) and Evans (1987) found that Gibrat's Law generally holds for large firms, but not for the entire population. They uncover a negative relationship between size and growth. Dunne and Hughes (1994) also find that while size evolves proportionally for medium and large firms, small firms' growth rates have higher variance and tend to decrease with size.

Another set of growth regression studies focused on the persistence of deviations of firm size from the mean, which would imply biased estimates for β. Singh and Whittington (1975) and Kumar (1985) found evidence for serial correlation in the growth rates of firms supporting the variant of Gibrat's Law proposed in Ijiri and Simon (1964). Kumar (1985) confirms the previous findings rejecting the strong form of Gibrat's Law, by showing that the earlier conclusions were robust to correcting for autocorrelation in the growth rates.

One of the problems that has plagued this literature, particularly the early work by Hart and Prais (1956) and Simon and Bonini (1958), has been data without a balanced representation of small firms. Dunne and Hughes (1994) and Hart and Oulton's (1996) work tries to address the problem by using a database with broad representation of small firms. They use this database to test for differences in growth rates among firms of different size classes and find the differences to be significant in contrast to Gibrat's Law. In the analysis of our model, we find the same differences and we also note that how small firms are counted matters when analyzing the firm size distribution itself.

A newer literature focusses on cross-industry variation. Santarelli and Lotti (2004) look at the evolution of the size distribution of new firms in four industries. Over a period of five years most of the distributions approach the log normal distribution, however, they find that the more technologically oriented industries achieved the lognormal faster. Audretsch et al. (2004) find evidence that services may exhibit different distributional properties than manufacturing, the main focus of the empirical literature to date. Looking at the Dutch hospitality sector they find that growth is independent of size, whereas the majority of studies focusing on manufacturing find the negative growth-size relation discussed above. Machado and Mata (2000) use quantile regressions to examine the effect of industry characteristics on different portions of the distribution for Portuguese data. While their results are mixed for some characteristics and distribution measures, they find that the impact of industry characteristics on skewness is the most stable over time. Both technology measures and the rate of growth in an industry reduce the skewness of the distribution, while turbulence increases it.

However, all of the results of the previous paragraph lack a solid theoretical base for their findings. It is this gap in the literature we seek to fill by proposing a fully dynamic model of optimizing firms that generates the distributional characteristics found in the empirical literature.

3 The Model

To capture the forces that affect firm size distribution in a structural model, we apply a variant of the Ericson and Pakes (1995) model described in Laincz (2004a). The modification allows for continually falling marginal costs through process R&D such that we can discuss both firm and industry growth rates. That enables us to perform analogous growth-size tests on the resulting simulated data.

We specify an industry with a finite set of imperfect substitutes such that one of the common drawbacks of the Ericson-Pakes framework does not apply. Because the state space for a single industry can be very large, it limits the total number of firms that the computational algorithm can handle, often to no more than about 10 firms. In order to generate a cross-sectional distribution with a reasonable number of observations, our industry is characterized by a finite set of imperfect substitutes, but each good is produced by a Cournot oligopoly. We solve for the dynamics associated with each substitute separately treating them as highly disaggregated goods and then aggregate across the varieties. We think of each good as being defined at a 7-digit level in the SIC or NAIC codes, for example, and the aggregation taking place at a less detailed industry level such as 4 or 5 digits.[1]

By specifying independent products within the same broader market, we bridge the literature between the earlier stochastic models and the more recent literature devoted to strategic interaction. The older literature presumed that a market contained a series of isolated opportunities and assigned exogenous probabilities that these opportunities would be undertaken by either in-

[1] The approach is similar to Sutton (1998), pgs. 19-20. However, our use of the term "submarkets" differs from his and accords more with his notion of "subindustry" (see pages 297-298).

cumbents or new entrants. As Sutton (1997) states, the assumption is "crude," however, ". . . most conventionally defined industries exhibit both some strategic interdependence within submarkets, and some degree of independence across submarkets."[2] Our characterization allows for strategic interdependence within each product market, but independence across products within the industry.

3.1 *The Industry*

We characterize the industry as producing intermediate goods sold into a perfectly competitive final goods sector. Firms producing the intermediate goods choose quantity produced, investment in R&D, and whether to exit or not, if they are currently active in the product market, or enter if they are not currently active. The dynamic equilibrium is a Markov Perfect Nash Equilibrium which imposes that decisions are functions only of the current state which is the current market structure. The basic timing of the model begins with incumbent firms first choosing whether to exit or not. The remaining incumbent firms then compete in a Cournot fashion in the product market and determine their optimal levels of investment in R&D to lower future costs which follows a stochastic process. Potential entrants then compare their opportunity cost of remaining outside the industry to the expected value of entering in the next period. These potential entrants draw on a public stock of knowledge which increases overtime through spillovers according to another stochastic process. At the end of the period, R&D outcomes and the public stock of knowledge for the next period are determined by the results of the stochastic processes.

3.1.1 The Product Markets

Demand for intermediate goods comes from a perfectly competitive final goods sector with a CRS production function.[3] Output in period t, Y_t, of the final goods sector is given by the production

[2] Sutton (1997), p. 49.

[3] We could analagously think of (3) as the utility function for a consumer and apply the framework to imperfectly competitive final goods producers.

function:

$$Y_t = k_{1t}^{\theta 1} \cdot k_{2t}^{\theta 2} \cdots k_{Mt}^{\theta M}, \text{ where } \sum_{m=1}^{M} \theta_m = 1. \tag{3}$$

Each k_{mt} is the input from subsector m, where m denotes the products within the industry. Within each subsector, multiple firms engage in Cournot competition providing a homogenous good to gain market share. Normalizing the price of the final good to unity, the demand for each intermediate good k_{mt} is given by:

$$k_{mt} = \frac{\theta_m Y_t}{P_{mt}}. \tag{4}$$

Firms producing intermediate goods at any given time have a technology for production of intermediate goods where the marginal costs are constant although they vary across firms. All firms are assumed to face constant fixed costs which do not vary either with time or across firms. Each intermediate goods firm, $n \in [1, N_m]$, in industry m faces the following optimization problem for choosing quantity:

$$\max_{q_{jm}} \pi_{jm} = P_m \left(\theta_m Y, \sum_{n=1}^{N_m} q_{nm} \right) q_{jm} - MC_{jm} q_{jm} - f \tag{5}$$

where market size, $\theta_m Y$, and total quantity, $\sum_{n=1}^{N_m} q_{nm}$, determine the price of the intermediate good, P_m. q_{jm} is the quantity output of firm j producing product m, MC_{jm} are the marginal costs for firm j, and f is fixed costs. The implicit production function is linear in the input good with a coefficient equal to the inverse of the marginal cost.

We focus on one submarket to illustrate the model in the discussion that follows. Let N_m^* be the number of firms producing $q_{nm} > 0$. The Cournot-Nash equilibrium outcomes yield the profits for firm j as

$$\pi_{jm}^* = \max \left\{ \begin{array}{c} -f, \\ \theta_m Y \frac{\left(\sum_{n=1}^{N_m^*} MC_{mn} - (N_m^* - 1) MC_{jm} \right)^2}{\left(\sum_{n=1}^{N_m^*} MC_{mn} \right)^2} - f \end{array} \right\}. \tag{6}$$

Firms choose to produce if:

$$\left(\sum_{n=1}^{N_m^*} MC_{mn} - (N_m^* - 1)MC_{jm}\right) \geq 0. \tag{7}$$

Equation (7) simply states that a firm will choose not to produce if its marginal costs are too high relative to its competitors. The choice of produce or not to produce (also exit or not exit) is assumed made prior to quantity decisions and all firms know the decisions of their rivals to ensure uniqueness of the solution.

The Cobb-Douglas specification generates a Cournot solution for the intermediate goods firms in which profits are homogenous of degree zero in the vector of marginal costs across firms. Thus, a proportional change in the vector of marginal costs leaves profits the same despite falling marginal costs through process innovation (described below). Moreover, it allows for *continuously* declining marginal costs as opposed to the Ericson-Pakes framework where marginal costs are restricted to take on values in a finite set. The reason is that for any given vector of marginal costs, once the policy functions specifying R&D expenditures, entry, and exit are determined, these decisions will not vary provided the vector of marginal costs changes proportionally. Hence, policy functions for a finite subset of possible vectors of marginal costs are sufficient to characterize the long-run equilibrium as marginal costs continuously decline with process innovation.

However, the functional form of the demand system does create a problem in the case of an intermediate goods industry containing a monopolist. Because the price elasticity of market demand is unity, the monopolist's solution is not well defined. We assume that there is a minimum scale level of operations for a monopoly.[4] Let $\underline{q}$ be the minimum amount that a monopoly must produce in order to engage in the market. The assumption has two effects. First, it immediately defines a solution for the monopoly problem with a positive level of output while still providing

[4] There are other assumptions that could be made here instead, but do not significantly affect the results. For example, it would be more natural to think of the minimum scale assumption applying to all firms whether or not there is a monopoly. This assumption, while more plausible, only complicates the Cournot-Nash solution by changing the corner solution for output from 0 to $\underline{q}$ for affected firms. Moreover, Dixit-Stiglitz technology is a viable alternative that yields the same homogeneity of degree zero property, but it does not create a poorly defined monopoly problem as in Laincz (2004b). That extension introduces a more complicated problem to solve without adding much in the way of additional insights for the present inquiry.

the monopoly with incentives to invest in order to lower its costs. Furthermore, provided $\underline{q}$ is sufficiently small, there remain strong incentives for firms to strive to become monopolists. The minimum scale chosen for the simulations of the next section, while small enough to generate large monopoly profits, is such that in equilibrium firms always have sufficiently strong incentives to remain in the market or enter the market when the number of firms is small. Those incentives are discussed in the next subsection. Given that true monopolies without regulatory protection are exceedingly rare, the focus on markets where the probability of a monopoly emerging is quite small seems realistic and appropriate for the questions at hand regarding the distribution of firms.

3.1.2 Evolution of Market Structure

The number of firms operating in each product market and their relative levels of marginal cost determine the market structure at any point in time. The market structure evolves through process R&D which lowers a firm's marginal cost when R&D is successful. We track the level of marginal costs by accounting for the number of innovations that each firm j in market m has available at time t and denote it as i_{jmt}. The mapping from innovations to marginal costs is:

$$MC_{jmt} = \frac{1}{Z}\exp(-\eta i_{jmt}). \tag{8}$$

Marginal costs fall at the rate η with each additional innovation. Z is a scale parameter on costs which we use below to calibrate the model to match the mean employment level of firms observed in the data. Z captures the unit labor costs of firms relative to the price of the final output good. Firms with a greater number of innovations enjoy a cost advantage over rivals. The cost advantage generates higher profits and the motive for engaging in process R&D.

The total number of innovations accessible by firm j is the sum of publicly available innovations for product m, labelled I_{mt}, and each firm's private innovations, ip_{jmt},

$$i_{jmt} = I_{mt} + ip_{jmt}. \tag{9}$$

Private innovations of incumbents diffuse to the public stock at a constant rate, δ. Thus, I_{mt}

increments by one with probability δ in every period. We interpret δ as the strength of lead-time, secrecy, and patent protection within the industry. However, for incumbent firms an increment in the stock of public innovations also means a reduction in the stock of private innovations. Thus, in the absence of successful R&D in any period (discussed below), diffusion of an innovation to the public stock leaves the total stock of innovations for an incumbent unchanged. In section 5, we explore how the diffusion rate, δ, alters the observed firm size distribution.

The constantly growing public stock of innovations allows potential entrants to remain viable.[5] Completely new firms in a particular market do not have to invest to learn all of the innovations that have taken place in an industry since the beginning of time. Rather, we assume that most innovations are in the form of readily available public knowledge, while more recent innovations are held privately by incumbent firms. Existing firms have access to all of the publicly available technological innovations and have discovered some new ones through process R&D which is temporarily private information. It is through this process of knowledge diffusion that industries are prevented from becoming permanently monopolized.[6]

We assume that new firms generally enter at relatively lower efficiency levels than incumbents to capture the fact that hazard rates of exit decline with the age of the firm (See Dunne, Robertson, and Samuelson, 1988). Specifically, new firms will enter, on average, with fewer private innovations than incumbents:

$$0 < ip^{EN} < \widetilde{ip} = \int \left(\sum_{m=1}^{M} \sum_{n=1}^{N_m} \frac{1}{N_{mt}} ip_{nmt} \right) dt, \tag{10}$$

where ip^{EN} represents the number of private innovations of a new entrant. If new firms entered

[5] If all information was permanently private, a leading firm could innovate a sufficient number of times such that the cost to a new firm of acquiring enough innovations to generate positive profits would make entry costs prohibitively high.

[6] In the specification presented here, there are no spillovers between active firms which contrasts with the empirical evidence (e.g. Griliches, 1998). The spillover from private to the public stock of knowledge is necessary for continual growth because it enables new firms to enter at levels competitive with incumbents. The model can be adjusted to account for diffusion between incumbent firms. Doing so would enable analysis of the role of secrecy and lead time and how they interact with market structure. Overall, we do not believe it would change the main results presented in the next section, but we do believe it is worthy of exploration in future work.

at higher levels than incumbents, then incumbents would be more likely to die than entrants producing the counterfactual result that incumbents have a higher hazard rate of exit than new firms. The left side of the inequality implies that new firms are bringing some new ideas while the right-side, $\widetilde{ip}$, is the equilibrium average (over the long-run) number of private innovations held by incumbent firms. This assumption then creates the possibility that new firms can immediately establish themselves as the new leader if incumbents repeatedly fail to innovate.[7]

The stock of private innovations held by incumbent firm j increases through successful R&D. The role of R&D is given by:

$$ip_{jm,t+1} = ip_{jmt} + \upsilon_{jmt} \tag{11}$$

where:

$$\Pr(\upsilon_{jmt}) = \left\{ \begin{array}{l} \frac{ax_{jmt}}{1+ax_{jmt}}, \text{ for } \upsilon_{jmt} = 1 \\ \\ \frac{1}{1+ax_{jmt}} \text{ for } \upsilon_{jmt} = 0 \end{array} \right\}. \tag{12}$$

x_{jmt} is the level of R&D undertaken by a firm at time t. Note that the function does not vary with firm size, i.e. large firms do not possess an inherent advantage in successfully conducting R&D. We do not need to assume advantages owing to size to generate R&D spending distributions that match the highly skewed distributions in the data (see Laincz 2004a). This assumption is consistent with the arguments of Cohen and Klepper (1992) among others that there are no differences in the productivity of research investment owing to firm size.

The parameter a governs the productivity of R&D and is interpreted as measuring the technological opportunity and basic state of science. We assume this to be constant across firms and product markets. Clearly, the level of R&D productivity will be an important parameter for variation in the firm size distribution. Higher levels of a generate greater potential for any one

[7] This outcome occurs only rarely. Most of the time new firms will enter with a small market share relative to existing incumbents.

firm to extend its technological advantage and generate greater variance in firm sizes. We explore the relationship between technological opportunity and the firm size distribution in section 5.

The combination of the two stochastic variables, R&D and diffusion, in conjunction with the solution to the dynamic equilibrium results in an upper bound on how much of a lead firms will actually gain over potentially new firms in equilibrium. Because returns to investment are decreasing when marginal costs are relatively low, firms will enter a "coasting" state and choose not to invest because the gains eventually become outweighed by the costs.[8]

This specification for the evolution of marginal costs and innovation has several notable features. First, it is the relative marginal costs that matter to firms' profits as shown in (6); the absolute level of the marginal costs (or total stock of innovations) is irrelevant to the decisions of a firm. Second, in contrast to Ericson-Pakes, the spillover process does *not* change the marginal costs of active firms, but it does lower the costs of potential entrants because the stock of publicly available innovations continually grows. This feature allows for potential entrants to remain within striking distance of the incumbents. Hence, the contribution of private innovations to the public stock is an externality that benefits the pool of potential entrants.

3.1.3 Dynamic Equilibrium

Let s_{nm} be the number of firms with ip private innovations producing product m and define the vector $s_m = [s_{nm}]$ which describes the market structure at any point in time. There are two types of firms facing different problems: incumbents and potential entrants. Incumbents are either producing for the market or choosing to exit. Their problem is characterized by comparing the expected net present value of investment in R&D against a positive liquidation value given by ϕ. Potential entrants compare an outside alternative, ψ, against the net present value of entering minus sunk costs of establishing production facilities denoted by χ. Both ϕ and χ are assumed constant across time and equal across firms.

[8] See Ericson and Pakes (1995) for a discussion of the coasting states.

An incumbent's intertemporal decision can be described by the following Bellman equation where time subscripts are replaced with a prime indicating a future value and all others are current:

$$V^{I}_{jm}(ip_{jm}, s_m) = \max\left\{\phi,\ \pi(i_{jm}, s_m) - cx_{jm} + \left(\frac{1}{1+r}\right) E\left[V^{I}(ip'_{jm}, s'_m | ip_{jm}, s_m)\right]\right\} \qquad (13)$$

where the I superscript refers to the value of an incumbent. If the firm chooses to exit it receives the liquidation value ϕ, otherwise the firm receives current period profits minus its investment level in R&D, x_{jm} at a cost of c per unit plus the discounted expected value conditional on future market structure. The future market structure depends on the firm's current number of private innovations and the current market structure. $1/(1+r)$ is the common discount factor facing all firms. The expectation sign reflects the fact that the firm is assigning probability weights via the transition matrix of the market structure moving from its current state to all possible states. These include the probability of a spillover, the probability the firm itself will be successful in R&D, the probabilities of other firms being successful, and the probabilities of entry and exit.

A potential entrepreneur may enter a submarket, incur sunk entry costs, and establish production and R&D facilities. Production and sales do not begin until the following period. The Bellman equation resembles that for incumbents with few changes:

$$V^{EN}_{jm}(ip^{EN}, s_m) = \max\left\{\psi,\ \ -\chi - cx_{jm} + \left(\frac{1}{1+r}\right) EV^{I}(ip'_{jm}, s'_m | ip^{EN}, s_m)\right\} \qquad (14)$$

where the EN superscript refers to entrants and the future value corresponds to that of being an incumbent in the next period. ψ measures the opportunity cost of entering and χ represents the sunk entry costs. By endogenizing entry and exit, we can observe how turnover rates respond to changes in structural parameters as we observe changes in the firm size distribution in the analytical section of the paper.

The investment strategy of firms derives from the first order conditions on the above. Let $C_1(ip'_{jm} + 1, s'_m)$ denote the expected value of the firm conditional on successful innovation and $C_2(ip'_{jm}, s'_m)$ the expected value if it fails to innovate. We can then rewrite the Bellman equation for incumbents as:

$$V^I_{jm}(ip_{jm}, s_m) = \max \left\{ \begin{array}{c} \phi, \ \pi(i_{jm}, s_m) - cx_{jm} + \\ \left(\frac{1}{1+r}\right) \left[\frac{ax_{jm}}{1+ax_{jm}} C_1(ip'_{jm} + 1, s'_m) + \frac{1}{1+ax_{jm}} C_2(ip'_{jm}, s'_m)\right] \end{array} \right\} \tag{15}$$

From this, the first-order condition yields the following policy function:

$$x_{jm}(ip_{jm}, s_m) = \max \left\{ 0, \frac{-1 + \sqrt{\frac{a(C_1 - C_2)}{(1+r)c}}}{a} \right\}. \tag{16}$$

The firm chooses the value maximizing level of R&D investment subject to a non-negativity constraint on R&D expenditure. Investment in R&D rises with the expected marginal gain in value, $C_1 - C_2$, and falls with the discount rate, r, and the cost of investment, c. The productivity of R&D, captured by a, has offsetting effects. As a rises it increases the probability of successful R&D and an incentive to increase investment. However, the higher the level of a the lower the marginal product for any given level of investment which lowers R&D investment.

Overall, the industry exhibits growth in total output and thus, Y in equation (1), grows over time while the innovations constantly reduce the cost of inputs. This continually growing industry can exhibit a great deal of change over time in terms of the identities of firms, their relative sizes, and the degree of entry and exit. The model provides us with the ability to generate a long-run firm size distribution based on the ergodic distribution of the model *and* the ability to examine the growth-size relationship at the individual firm level. The numerical algorithm uses value function iteration to solve for the space of values given by all possible combinations of firms and private innovations. We extract the policy functions including R&D expenditure as well as the entry and exit decisions. From the solutions, we can simulate our product markets and industry for comparison with the results found in the empirical literature. We now turn to that analysis.

4 Firm Size Distribution

4.1 Simulation

Table 1 presents the baseline parameterization of the model. We set the discount rate to 1/1.08 as an approximation of the average cost of capital for firms following Ericson and Pakes (1995).

The rate of technological spillovers, δ, is set to 0.7 such that knowledge enters the public pool roughly one-and-a-half years after discovery. This fits with the empirical estimates of Mansfield (1981) on imitation time. Cost of a unit of R&D spending is set to one unit of the final good. The liquidation value and outside opportunity cost are chosen to be small to prevent them from dominating the incentives firms face. The liquidation value is about 7.5% of the average firm value, while the opportunity cost is roughly 15% of average firm value. We set both fixed and sunk costs equal to the outside opportunity cost.

The parameters a and η interact to determine the incentives for investment in R&D and ultimately the growth rate of the industry as measured by the rate of cost reduction. These parameters are set to 3 and 10% respectively. The latter implies that successful R&D will reduce marginal costs by 10%, but the former governs the incentives to engage in R&D such that we find at the mean level of R&D investment, the expected rate of cost reduction is just over 2%, which is approximately the industry growth rate. Both of these two parameters, plus the rate of knowledge diffusion, fixed and sunk costs will be allowed to vary in the following section to analyze their relationship to the moments of the firm size distribution.

The state space constraints we use have a maximum of six firms per submarket and each firm can hold up to 30 private innovations. To ensure that the state space boundaries do not drive the results we choose our demand parameters such that when six symmetric firms are in a submarket they are making negative profits. For the maximum number of private innovations, we checked in the simulations whether any firm attempted to obtain more private innovations than exist in the state space and made adjustments accordingly which led to our choice of 30.

Because we had no priors on how to vary the submarket sizes, we choose to use a simple, transparent linear function as follows:

$$\theta_m = \theta_1 + (m-1)\, b \tag{17}$$

where θ_1 is the market share of the smallest submarket and each submarket increments by b.[9] Upon simulating the model, we use the state space constraints to determine θ_1 and determine m by matching the model to the empirical results on the firm size distribution. b then follows from $\sum_{m=1}^{M} \theta_m = 1$.

The choices on the market size parameters, θ_1 and b, were determined as follows. We started with 10 submarkets, $m = 10$, where the smallest market share was determined by the lowest level of the market size that still produced positive levels of investment in R&D. At the tenth submarket firms began to invest at the upper bound of the state space. Therefore the increments in market share per submarket were determined to be 0.0318. For the analysis below, we then choose the number of submarkets, m, to analyze by matching the general shape of the log-normal distribution which closely, but not perfectly, resembles the empirical firm size distribution across industries. We found that if there are too few markets, the distribution is skewed left instead of right. Thus, for very narrowly defined markets with only one or two submarkets, the model generates a high frequency of average sized firms and a small number of tiny firms. On the other hand, as the number of submarkets expanded the model generated a bimodal distribution, in accord with some of the findings in Bottazi et. al. (2003b) for some industries. For the general log-normal distribution, we found that specifying five submarkets, $m = 5$, was the closest match to the results reported in Hart and Oulton (1996) discussed in the next subsection. As a further check on the validity of the results, beyond matching the general pattern of the firm size distribution, we then conduct cross-sectional regressions to see if the growth-size relations match the empirical literature in section 4.3.[10]

[9] This specification is similar to Sutton (1998) Chapter 13 where he uses a geometric distribution of market sizes. We also considered using a random process, or possibly demand shocks, but opted for the simple linear function on account of its transparency.

[10] Clearly, it would be more appealing to have the submarket sizes determined endogenously. This additional feature could perhaps be accomplished by specifying a Dixit-Stiglitz demand function. However, it would still require additional assumptions on how firms interact across sectors in terms of both price-taking (or not) behavior as well as specifying how innovations in one sector affect the other. The additional complications introduced would detract from the main task of this paper, which is to understand how the overall firm size distribution changes with underlying structural parameters. Moreover, because our model captures strategic interaction within each submarket, it is likely that most forms of strategic interaction across submarkets would be of second order

4.2 Distribution Results

In our first comparison of the model with the data, we compute the ergodic distribution by simulating the model.[11] From the distribution found in the simulations, we weight the observed outcomes by their probability of occurrence to generate the ergodic distributions for various size measures.[12]

Table 2 shows the results of the baseline parameterization compared with the statistics found in Hart and Oulton (1996) who use subsets (50 to 80 thousand companies) of a large UK database that includes very small firms in the sample. We calibrate the cost parameter, Z, to match the mean log size of employment reported in Hart and Oulton (1996). Because their data set has a good representation of small firms we felt that it was the most appropriate for comparison with the model. They find that the distribution of the natural log of various size measures (employment, sales, and net assets) exhibit positive skewness (long right-tails) and peaked (leptokurtic) distributions relative to the log normal distribution. We report analogous measures based on our model.[13] Sales are computed by extracting the quantities and prices while we use firm values, V^I, for net assets. All values below are reported in natural logs.

Figures 2-4 show the distributions in levels (a) and in logs (b). The distributions in levels, for all the three size proxies considered, exhibit long right tails, especially for the net assets distribution. The size range accumulating the higher probability mass lies to the left of the mean size in all the distributions. The distributions for both the log of sales and log of net assets are roughly bell shaped, but exhibit thicker tails and higher peaks than the standard normal.

importance. Our assumption of no strategic interaction across submarkets fits with the arguments of Sutton (1997), mentioned earlier, for blending strategic interaction with the independent opportunities assumed in the older literature on Gibrat's Law.

[11] The simulation runs the model for one million periods. In order to avoid any bias caused by the specification of the initial market structure, we simulate it first for 10,000 periods and find the modal market structure. The main simulation then uses the modal market structure as its starting point.

[12] It is important to note that the comparison here with the data is not direct. We take advantage of the fact that through simulations we can generate the probability distribution of the market structures. Empirical studies use a cross-section of firms at a point in time (we turn to this analysis later) while the ergodic distribution shows the probabilities of a market structure occurring at a point in time. That is, the ergodic distribution is generated as a time series, but it reveals what the *expected* cross-section would look like.

[13] All employment calculations add one in levels to represent the manager which we view as part of fixed costs.

The distribution of the log of employment exhibits less variance, but shows some skewness and leptokurtosis.

There are two noticeable differences between the model and the data. First, the standard deviations of the size measures are considerably smaller. This discrepancy is not surprising since the model is designed for a particular industry whereas the empirical estimates cover a large range of industries which would generate greater size variation.

Of greater concern is the slight negative skewness in the natural log of employment generated by the model versus the positive skewness observed in the data. Upon careful examination of the results, it turns out that the negative skewness is being driven by a tiny fraction of extremely small firms. These are firms with less than one employee which constitute about 0.1% of all firms and only 0.0025% of total employment. Those firms are to the right of the vertical line in Figures 2a and 2b. If we eliminate them from the distribution and recalibrate Z, the skewness in employment goes slightly positive and the negative skewness in sales is cut in half as shown in Table 3. Moreover, the high kurtosis value in sales comes down considerably and is much more in line with the data. If we drop more small firms, less than 5 employees (0.02% of total employment), the skewness for employment rises to approximately 0.41. In fact, we found that we can match the Hart and Oulton skewness figure almost exactly if we eliminate all firms of less than 10 employees (0.14% of total employment).

The negative skewness for sales remains even after eliminating the small firms, although the skewness value for sales reported by Hart and Oulton, while positive, is the smallest of the three. This result of our model is being driven by the strong implications of Cournot oligopoly pricing with homogenous goods in each submarket. For example, when multiple firms produce large quantities, and hence have substantial employment, the direct competition between them drives the price down significantly. What we find is that the model often generates 3 or 4 firms in a given submarket with marginal costs that are very close. Although quantities are reasonably high for these firms, the low level of the price accounts for the reduced skewness when comparing employment and sales.[14] Overall, the ergodic distribution of the model reasonably matches the observed data in terms of deviations from a log-normal distribution

When we turn to the growth-size relationship in the next subsection, we extract a balanced panel which eliminates exiting firms. These exiting firms include these extremely small firms that generate the negative skewness in employment. Thus, in our summary statistics on the balanced panel below, the skewness measures increase significantly. These results suggest an interesting hypothesis. First, although the skewness is generally smaller than that observed in the data, it is important to bear in mind that data sets rarely include the full population of small firms. Second, the model accounts, in some sense, for part-time workers while data typically do not. These differences may be relevant empirically for testing distributions against the log-normal distribution. For example, if data collected round workers upward it would imply an underweighting of the left-side of the distribution which would bias skewness upwards.

4.3 Cross-Sectional Growth-Size Properties

We examine the growth-size relationships of the simulated model and compare them with the empirical literature on Gibrat's Law. To extract cross-sectional data comparable to that used in the empirical literature, we simulate the model five times for 5,000 periods each and extract the final periods from each run.[15] That provides us with simulated panel data to test the growth-size relationship. We ran these simulations ten times to check the robustness of these results.

The average number of total firms observed in each simulation was 138.8 (range of 132 to 143) with an average of 43.4 new entries over two periods and 25.9 exits.[16] We eliminate all the new entrants, who do not produce in their initial period and all firms that exit to generate a balanced panel for analysis. Table 4 shows the average distributional characteristics for size measures

[14] Note that the statistics we report exclude monopolies altogether so as not to be affected by the minimum quantity assumption. In the baseline, monopolies account for less than 0.000001% of all observations.

[15] To prevent the variance of the size of the firms from being dominated by the overall growth process, we shut down the increments to the public stock of knowledge except for the periods we extracted for analysis.

[16] Note that entries and exits would match almost exactly if we included those firms that exited in the previous period.

across these simulations of the balanced panel of firms for the initial period. The measures are similar to those shown in Tables 2 and 3, but note the substantial increases in the skewness values when small exiting firms are eliminated.[17]

Table 5 provides the results of the regressions on each of the ten simulations of the following form:

$$y_t = \beta y_{t-1} + \epsilon_t$$

where y_t is the log of the various size measures. Of interest in terms of Gibrat's Law is whether the coefficient is significantly different from unity. Hall's (1987) estimates for β as applied to employment across three different samples were consistently 0.99 and significantly less than one. Evans (1987) found values for β that range between 0.93 and 0.97 for employment. The model here also generates a coefficient less than one, below Hall's estimates, but in accord with those of Evans'.[18] The last columns report the percentage of times the null hypothesis of $\beta = 1$ was rejected at the 10% level, followed by the percentage of times it was rejected at the 5%, and 1% levels, respectively.

We use the median size values to split our sample into large and small firms. Empirical evidence suggests that Gibrat's Law works better for the large firms. Our results show a similar pattern. The estimated β's are consistently closer to one for the large firm sample for each size measure and in every simulated sample. In fact, for the employment size measure we cannot reject $\beta = 1$ nine out of ten times and then only at the 10% level.

For the three size measures we also tested for the equality of the coefficients between the large and small firm subsets reported in the last rows of Table 5. For employment we reject equality in all cases at the 5% level or better and 80% of the time at the 1% level. For sales, the differences

[17] The negative skewness in sales persists, but becomes even smaller in absolute value than when we eliminated the smallest firms outright in the preceding subsection.

[18] We report the results using robust standard errors, but even without using them the results are hardly changed. The R-squared's are exceedingly high typically between 0.95 and 0.99. However, since there are only two simple stochastic processes in the model and nothing akin to demand shocks, it is not surprising in the least that past size is a good predictor of size in the short-run.

weaken somewhat and we reject equality 60% of the time all of those t-statistics at the 5% level or better. Equality of values is rejected in nearly all of the subsamples. It is worth emphasizing, however, that in all subsamples, the estimated beta for large firms was greater than that of small firms for all three size measures. Given the small sample size we draw, the large number of significant rejections of $\beta_{LARGE} = \beta_{SMALL}$, indicates the model matches the empirical findings.

In order to test for serial correlation we reduce our sample to those firms surviving in three consecutive periods for a balanced panel. The previous tests only had 69.5 observations on average and after one more period of eliminating exiting firms to retain a balanced panel, we were left on average with 46.9 firms. The test specification is similar to Kumar (1985) where growth is the dependent variable (instead of the log of size):

$$\left(\frac{y_t}{y_{t-1}}\right) = (1+\beta)\, y_{t-1} + \gamma \left(\frac{y_{t-1}}{y_{t-2}}\right) + \epsilon_t.$$

Persistence in growth will show up as a positive value for γ. We find that the coefficients for β and γ are significant at the 5% level for the majority of the ten samples. Their average estimated values are 0.9817 and 0.577 respectively. γ is positive and significant in all but one of the samples at the 1% level.

The positive and significant value of γ indicates serial correlation which comes as no surprise given the design of the model. There are several contributing factors to serial correlation in our model. First, successful firms seek to build on and protect any technological advantage and thus invest more heavily than small firms. In addition, a growing firm pushes rival firms closer to the exit threshold. Thus, the growing firm will get a subsequent additional increase in market share with the increase in the likelihood of rivals' exit. These processes of firm dynamics effectively embed serial correlation in error terms that do not control for innovative behavior and expected future changes in market share conditional on them. The results suggest that serial correlation should weaken in empirical studies if appropriate controls for own and rival R&D expenditure and innovations are included. We leave this hypothesis for future empirical work.[19]

Finally, we look at the variance in growth rates across firm sizes. A number of the studies find that the variance of the growth rate is larger for small firms (e.g. Dunne and Hughes, 1994). Again, we separate our simulated samples by the median size. Table 6 shows the average standard deviations in growth rates across the ten samples and for large and small firms according to the three size measures. By all three measures the variance in the growth rate of the small firms is larger than that of the large firms and across all ten samples. The final columns report the percentage of rejections based on the F-statistic for the variance ratio test for equality of the standard deviations. We reject equality at the 1% level based on the employment and sales measures in seven out of ten samples, but in only half the samples for net assets. The latter also has the highest level of the standard deviation. Overall, the results are encouraging in the sense that, again the model replicates empirical findings.

To summarize the section, we find that the model is able to replicate empirical studies of Gibrat's Law in two ways. First, it can generate a firm size distribution with the higher moments deviating from the log-normal distribution in the same direction as actual distributions. Secondly, in the cross-section the model generates a negative firm size-growth relationship, decreasing variance in the growth rate with firm size, and serial correlation, all found in the data. Based on the above we feel reasonably confident in using the model to understand how underlying structural parameters affect the overall firm size distribution.

[19] One note on the magnitude of serial correlation is required here. Our estimate of γ is larger than that found in either Singh and Whittington (1975) or Kumar (1985) who find values of approximately 0.3 and 0.l2 respectively. The distinguishing feature is in the difference in time periods. Those authors use a much longer time frame, 10 to 12 years, compared with our simulated data which corresponds to roughly three years based on the user cost of capital we specify. Because we know that the model will predict serial correlation that declines over time due to the Markov perfect nature of the equilibrium, we do not pursue that issue any further here. Suffice it to say, that the model does generate serial correlation in the errors when using the basic regression model found in the growth-size regressions related to Gibrat's Law. See Pakes and Ericson (1998) for the empirical implications of the Markov Perfect feature embedded in the model.

5 Variation in the Firm Size Distribution

The previous section established that the model reasonably matches the data in terms of the firm size distribution and in its cross-sectional empirics. Now we ask how the moments of the firm size distribution change with underlying structural parameters suggested in the literature. Specifically we vary the following parameters: sunk costs, fixed costs, productivity of R&D, rate of spillovers, and the rate of decline in marginal costs.[20] The goal of this section is to generate a set of hypotheses that can be examined empirically. We do not carry out that examination in this paper, but view the contribution of this analysis as setting an empirical research agenda on the firm size distribution guided by theory. How the model fares when taken to the data should provide insights for improvement in the model itself and a deeper understanding of the empirical work. All of the analysis below shows the distribution including all levels of employment, i.e. all firms are included no matter how small.[21] [22]

5.1 Fixed Costs

We start with fixed and sunk costs. In the baseline fixed costs were set to 0.2 and we allow that to vary from 0 to 0.25, when translated via the unit cost of labor, the range of the fixed costs then go from about 12% (for the smallest non-zero value, 0.025) to just over 50% of total costs of production excluding R&D costs for the mean sized firm in the sample.

We find that *increases in the level of fixed costs lead to a larger mean size of firms, but lower variance, skewness and kurtosis of the firm size distribution.* Figure 5a shows how the first four

[20] We do not vary the outside alternative parameter, ψ, because it enters in much the same way as sunk costs, and we do not vary the liquidation value because the parameter must be constrained to be less than $X_e/(1+r)$ such that firms cannot enter, produce nothing, and exit with a net gain. We also do not vary the discount rate because even if the discount factor varied uniformly across firms, such variations basically mean interest rate variations and those variations are typically short-run fluctuations rather than long-run characteristics inherent to an industry that shape the firm size distribution.

[21] We also examined the behavior of the first four moments when eliminating small firms as in the previous section, but found no qualitative differences. The only notable difference was that as we eliminated small firms from those of less than 1 employee, to less than 5, to less than 10, the effects became even more pronounced. By that we mean that the percentage changes in any moment of the firm size distribution were larger when eliminating small firms, but the direction of the effects was stable.

[22] We also analyzed the changes in the distributions of both sales and net assets, but we do not report those results here. Qualitatively they are very similar to the effects on the size distributions by employment.

moments of the firm size distribution change relative to the baseline while Figure 5b shows the same measures when we look at the firm size distribution in natural logs. The x-axis shows the level of fixed costs and the y-axis shows the percentage change from the baseline. Figure 5c shows the baseline distribution in levels against the low and high value of fixed costs.[23] In the latter, Figure 5c, low levels of fixed costs are associated with a greater mass of the distribution at lower levels of employment, but also with a longer right-tail and, hence, a higher skewness. The high levels of the fixed cost exhibit greater mass further to the right and there is a small, but noticeable, second mode emerging to the right of the peak.

Figure 5d shows the same distribution but in natural logs and we see similar changes. The mean size rises as the mass of small firms shrinks while the mass of larger firms grows. The variance falls as the distribution becomes more leptokurtic (in logs) as firms become more concentrated around the mean size. The skewness increases relative to the baseline at lower levels of the fixed costs, but then declines at the higher levels. The initial increase stems from an increase in the frequency of firms above the mean size creating more mass on the right-side of the distribution. The decrease follows from the flattening of the left-side of the distribution as smaller firms become more dispersed in their scale of operations.

When we set fixed costs to zero, the mean firm size is more than 11% below the mean size of firms in our baseline. As fixed costs rise the mean size of firms also rises with a more rapid increase at higher levels. Although the pattern exhibits some non-monotonicity, the variance generally falls with increases in fixed costs, while both skewness and kurtosis decline.

Intuitively, in the model higher fixed costs make it more likely that small firms will choose not to produce as in equation (7) and increase the likelihood of exit because future profit values are smaller for the same level of output. Thus, by reducing the fraction of small firms in the sample the mean size increases. Moreover, with small firms more likely to exit, higher fixed costs create

[23] In Figures 9 through 12 for the graphs showing the shape of the firm size distribution, we standardize the x-axis to maximums of 160 and 6, in levels and logs. However, in many cases the maximum sized firm exceeds those values. We choose to standardize the x-axis to facilitate comparison in the regions showing the bulk of the mass and how the parameters alter the distribution.

greater incentives to innovate for incumbent firms to distance themselves from the exit threshold which further increases the mean. The variance, however, declines and is related to the decrease in kurtosis. With a reduced fraction of small firms, the frequency of firms near the mean size increases, but there is also an increase in the mass of firms to the right of the peak reflecting higher R&D investment among incumbents which decreases the kurtosis. In natural logs we see more or less the same pattern, however, the higher moments behave differently. Skewness displays an inverted-U shaped pattern, while the kurtosis displays a generally increasing pattern. The reason is that the fattening of the tails in levels is primarily just to the right of the primary mode such that in logs the effect blends with the original mode and the tails remain relatively flat.

Skewness, in levels, falls because the region of small firms becomes smaller while the frequency of mid-sized firms increases. This effect flattens and lengthens the tail on the left-side which reduces the skewness. Thus, fixed costs act in a way that is fairly straightforward by making it more difficult for very small firms to survive, essentially requiring a minimum scale of operations to produce non-negative profits and survive. These results are consistent with the findings of Machado and Mata (2000) who use a Box-Cox quantile regression model to characterize the effect of covariates on the firm size distribution of Portuguese firms. They find that minimum efficient scale had a consistently positive impact on the size of firms, a negative effect on the skewness, and an ambiguous effect on kurtosis.[24]

5.2 Sunk Costs

Figures 6a, 6b, 6c, and 6d show the results from varying the sunk costs of entry. The range here starts from 0.1 such that $\left(\frac{1}{1+r}\right) X_e > \psi$ continues to hold. The upper bound here is much higher than for fixed costs to capture industries for which sunk entry costs will take, in expectation, significant time to recover. These values can be understood as a ratio to the value of the mean sized incumbent firm. The range is from 10% to 35% and equals approximately 21% at the

[24] Machado and Mata (2000) do not report measures of variance.

baseline. The vast majority of firms that enter the market do not recover their full sunk costs. However, those that survive and grow ultimately reap substantial rewards. At low levels of sunk costs, a firm does not need to survive for a long period of time before making entry optimal. However, at high levels of the sunk costs, firms require a substantial likelihood of sustained success to induce entry.

At low levels of the sunk costs we see little change in the mean size of firms, but a negative effect on all the higher moments. In fact, we find that *industries characterized by lower levels of sunk and/or fixed costs will more nearly match the log-normal distribution and the strong form of Gibrat's Law.* In Figure 9c its clear that these changes are fairly small when comparing the shape of the baseline distribution to the low end for sunk costs. However, once sunk costs reach 0.25 (or approximately 26% of the value of the mean sized firm), the mean size of firms rises rapidly, while the variance increases though somewhat non-monotonically. The entry barrier discourages new firms reducing the mass of small firms. Markets become more concentrated with fewer firms, but of greater average size. Thus, *industries with high levels of sunk entry costs will exhibit greater average size of firms, higher variance in the size, but a flatter distribution potentially with multiple modes.*

Higher sunk costs have offsetting effects for incumbents which increases the variance but continues to reduce skewness and kurtosis. With smaller firms less likely to enter and pose a threat to incumbents, firms have less competition reducing the benefits of engaging in R&D which flattens the far end of the right-tail. However, at the same time, among incumbents, because the sunk costs help extend the expected lifetime of any one firm, competition in terms of R&D intensifies. Thus, once a firm does enter it has strong incentives to try to develop a technological lead over its rivals. This incentive leads to an increase in the mass of firms in the mid-sized range. Once a sufficient cost advantage has been established the first effect comes to dominate and discourages firms from establishing an even larger technological lead because the threat of entry has been reduced.

The model thus suggests that industries with large sunk costs should have a larger mean size, greater variance and a flatter distribution overall. The flatter distribution and Figures 6c and 6d suggest that bi- or multi-model distributions are quite likely for industries characterized by high sunk entry costs. The sunk entry costs protect incumbents such that once a firm reaches a sufficiently large size, it seeks to maintain that size by investing in R&D to maintain its advantage but with less incentive to increase that advantage.

Audretsch et. al. (2004) argue that the service sector is more likely to approximate the strong form of Gibrat's Law and therefore the lognormal distribution because the link between firm size and survival rates is weaker in industries with lower sunk costs and where capital intensity and scale economies play less of a role.[25] In our analysis, we find that to be true particularly for the fixed costs which imply a higher requirement for scale in order for net profits to be non-negative. We also find that the distribution more closely approximates the log-normal distribution at the low end of the sunk costs.

5.3 Rate of Cost Reduction

The rate of cost reduction is captured by the parameter η. More specifically, η, which first appears in (8), is the percentage decline in marginal costs of production conditional on a successful innovation. Thus increases in η will translate into a faster industry growth rate for the same level of investment as measured by output of the final good. We think of η as a key parameter in governing the rate of technological progress which in the context of the model is the rate at which costs fall.

The parameter ranges from 0.07 to 0.20 with our baseline value set to 0.1 (10%), but to make better sense out of this specification, we convert it to the expected cost reduction at the mean level of investment. At the low end, few firms are actually engaged in R&D and thus the mean expected rate of cost reduction is only 0.32%, an anemically growing industry with little

[25] Audretsch et al (2004) study the Dutch hospitality sector.

innovation. However, at the upper bound, there is a fair amount of R&D and the mean rate rises to 12.13%.[26]

Looking at Figures 7a-7d, we see that *increases in the rate of technological progress lead to an increase in both the mean and variance of the firm size distribution.* The higher levels of cost reduction lead to greater incentives to engage in R&D and capture market share from rivals which leads to increases in both of the first two moments. At the same time, *variation in the rate of technological progress has non-monotonic effects on both the skewness and the kurtosis.* Both exhibit convexity as η rises. Skewness falls initially because at low levels of cost reduction, there is a very small percentage of extremely large firms. These are firms that established a technological advantage and raced ahead to cement their leading position. With increases in the rate of cost reduction and, therefore, greater incentives for small and medium sized firms to use R&D, the scope for business stealing rises. As a result the industry becomes more competitive leading to more firms and more competition with fewer truly giant firms which reduces the overall skewness.

As the rate of cost reduction increases, around $\eta = 0.15$ or a mean expected cost reduction of 7.1%, the skewness begins to increase as a larger mass of firms emerges in the mid-sized range as can been seen in Figures 7c and 7d. Kurtosis undergoes a similar change. In fact, the whole distribution almost completely flattens out at our extremely high range. This effect occurs because the range of relative marginal costs throughout the incumbent firms increases along with the strong incentives to engage in R&D to defend existing market share as well as capture market share from rivals. Thus, rapid growth should lead to a high variance and a flatter distribution. This leads to the hypothesis that *industries with high rates of technological progress are more likely than those with low rates to exhibit multi-modal distributions.*

Machado and Mata (2000) also measure empirically the marginal effects of industry growth

[26] Jorgenson and Stiroh (2000) report industry growth rates for highly aggregated industries with the fastest growing industry, eletronic and electric equipment, growing at an annual rate of 5.457% from 1958 to 1996. That would suggest an upper bound for η of approximately 0.14 or 0.15. However, since that growth rate is for an industry at roughly the 2-digit SIC level, it therefore averages across more detailed sectors. Thus, we examine the effects for even faster rates of growth.

rates on the firm size distribution in their paper. They find that firms in faster growing industries have a higher mean, but more rapid growth reduces the skewness. For kurtosis they also find a negative effect, but it is not statistically significant.

5.4 Technological Opportunity

The productivity of R&D which we think of as the technological opportunity facing an industry is captured by a in equation (12). It is related to the rate of cost reduction, η, in the sense that those two parameters jointly affect the equilibrium rate of cost reduction. a governs the incentives to engage in R&D and η defines the gains of success in terms of cost reduction. Higher levels of a imply a higher probability of success for any given level of R&D expenditure. However, the marginal impact of an increase in R&D expenditure falls with higher levels of a. Moreover, the solution for the optimal level of investment based on the first-order condition of the value function shown in (16) implies that changes in a will have countervailing effects. Thus, as a prior, we expect to find non-monotonic behavior as a varies.

The range of a that we used went from 2 to 4, centered around the baseline value of 3. At both the lower and upper limits the computational algorithm began to generate extreme results. At the lower level, we found that virtually no firms were investing in R&D while at the upper bound firms began to exceed the limitations of the state space. To provide some economic interpretation of these values, a firm investing at the average level from the baseline, would expect success in R&D with a probability of 20.7% and thus an expected cost reduction by the following period of 2.07%. At the lower bound of a, 2, those values fall to less than 10% and under 1% while at the upper bound they are slightly less than double the baseline.

Figures 8a through 8d show the effect of varying a. *Both the mean and variance of firm sizes rise with productivity of R&D.* In natural logs the pattern is similar, but there is some concavity at the higher levels of a with respect to the variance. *Increases in the productivity of R&D have non-monotonic effects on both the skewness and the kurtosis.* Skewness and kurtosis both exhibit concavity, which contrasts with the results for the rate of cost reduction. The difference between these two parameters appears to come from the left-side of the distribution and their effects on smaller firms. Changes in the productivity parameter affect the right-tail of the distribution in much the same way as an increases in the rate of cost reduction. In both we observe a steady increase in the mass of firms to the right of the peak while the peak itself shrinks which eventually lowers the skewness and the kurtosis. In natural logs the pattern is similar with, again, another mode emerging on the right.

At very low levels for a, we found that the right-tail was much shorter and thinner than for the higher values. This follows from the much lower productivity in R&D which stunts the incentive to engage in R&D and the mean size of firms is considerably smaller as a result. Thus, at the lowest levels of a, as R&D expenditures yield greater returns with the higher marginal product, larger firms emerge and stretch the right-tail initially leading to increases in skewness and the kurtosis. As a rises beyond 2.5, more firms engage in R&D leading to the increase in the variance and hence a flatter distribution overall with less skewness.

Of the structural parameters we investigate, this parameter is almost certainly the most difficult to capture empirically. However, we do wish to emphasize the strong non-monotonicity in this variable and in the rate of cost reduction. The main conclusion we draw here is that empirically we should not expect proxies for either a or η to have clear monotonic effects on the higher moments of the distribution and caution should be exercised in generalizing results found in studies of the firm size distribution for a selected group of industries.

5.5 Rate of Spillovers

In the model, the parameter δ governs the rate at which the public stock of knowledge grows. The faster it grows the easier it is for entrepreneurs drawing on the public stock to enter the industry and challenge incumbents for market share. If $\delta = 0$, it would imply that no knowledge enters the public stock and over time no entrant would be able to challenge existing incumbents. At the other extreme, if $\delta = 1$ then all new innovations enter the public stock in the following period which would be similar to Klepper (2002) where all R&D is costlessly imitated.[27]

Mansfield (1981) reports imitation times that range from about 6 months to nearly three years. Thus, we allow δ to vary from 0.3 to 0.9 which generates an expected lifetime for a single innovation to remain private from just over one year to more than three years. Low values of δ can be interpreted as pertaining to an industry where incumbent firms possess strong advantages through secrecy, patent protection, and/or lead time to implement their innovations.

Figures 9a through 9d show the results which are quite striking. Changes in the rate of spillovers generate an enormous impact on both the mean and variance. *Industries with stronger patent protection (secrecy, or lead time) will have a higher mean and variance in the size of firms.* For example, at the low end of $\delta = 0.3$ the mean firm size is nearly six times that of the baseline! Intuitively the stronger the protection for private innovations, the greater their value to any one firm. Therefore firms will accumulate a great number of private innovations and establish a large presence in the market making it difficult for any new entrant to mount a successful challenge. However existing incumbents will compete fiercely in the R&D arena which contributes to both the high mean and the large variance.

Looking at Figures 9c and 9d, we show the distributions when we move away from the baseline of $\delta = 0.7$ by ± 0.2. The changes, particularly as δ falls, are more substantial here than for other parameters examined. When δ is increased the peak mode becomes more pronounced with less variance in firm sizes. Because private knowledge passes quickly from any firm to the public stock, there is less ability and incentive for firms to engage heavily in R&D to separate themselves from rivals. Firms in an industry with a high rate of spillovers are facing an uphill battle on a slope that is nearly vertical.

At the low level of δ, the distribution has no obvious peak and shows great variation over the

[27] In Klepper (2002), he assumes randomly assigned R&D productivities which allows for survival of the more productive firms while generating high rates of exit during the product life cycle. Here we do not allow R&D productivity to vary by firm but allow the innovations to diffuse slowly which generates the advantage of size because large firms will hold more private innovations that smaller firms.

mid-sized range. For the same distribution, a small, but noticeable mode emerges to the far right (around 250 in levels and 5.6 in natural logs) which we do not see in other distributions. In fact the distribution generated by the model fails to resemble the empirical distributions. Turning to the higher moments there does not appear to be any straightforward effect and no discernible pattern. We draw no conclusions regarding the effects on skewness and kurtosis here other than to say they appear to be highly non-monotonic.

Based on the analysis here clearly the diffusion rate plays a critical role in shaping the firm size distribution. While δ represents the rate at which knowledge becomes available to new firms, it does not capture spillovers between incumbents. The extreme changes in the shape of the firm size distribution that follow from modifying δ at levels that are empirically plausible, suggest that our measure is simply too crude to capture all that secrecy and lead time entail. Extending the framework to account for spillovers, imitation costs, and absorptive capacity between active firms seems a highly fruitful avenue for further work.

6 Summary

Understanding the forces that generate differences in the firm size distribution enables us to identify the forces that generate more or less concentration across industries. This study provides a model for undertaking this task. We show that the model replicates the characteristics of the firm size distributions reported in the literature and reproduces the empirical growth-size relationships. The model generates a substantial list of empirical hypotheses for testing the effect of various structural parameters on the first four moments of firm size distributions. In addition the model suggests that serial correlation in firm growth should weaken in empirical studies if appropriate controls for R&D expenditure and innovations of existing and rival firms are included.

It is worth emphasizing that the model is quite flexible and can be adapted to serve as a baseline for analyzing particular industries by matching the parameters and the moments of an observed industry level firm size distribution. With that baseline, counterfactual experiments can be conducted and the effects of policies on the firm size distribution can be analyzed, such as subsidization of R&D or regulations that affect barriers to entry.

We note some missing elements in our framework that could be incorporated in future work. First, merger activity is one of the major concerns in the empirical literature (for example see Kumar, 1985, and Dunne and Hughes, 1994). Our model can incorporate mergers by combining it with Gowrisankaran's (1999) extension to mergers of the Ericson-Pakes framework. Second, the model here relies on stochastic R&D success and diffusion of knowledge to generate entry, growth, survival, and exit. More could be done to capture other risks that entrepreneurs face such as uncertainty of true costs as in Jovanovic (1982). That would enable an exploration of how the rise of venture capital and lowering of entry barriers, other than the sunk costs discussed here, affect the firm size distribution.

We view the work presented here as a step forward in the interplay between the theory and empirics of the firm size distribution. While the motivation for the theory comes from a host of empirical observations, the theory provides us with a list of hypotheses that can be examined empirically across industries. We would be surprised to find that all of the hypotheses generated apply to all industries and it is highly likely that the model may serve well for some industries but not for others. That probable outcome would lead to both further refinements of the model and, we hope, a better understanding of the forces that shape the firm size distributions across industries and their consequences.

References

Audretsch, D., L. Kompf, E. Santarelli, and A. Thurik, 2004, "Gibrat's Law: are the Services Different?" forthcoming in *Review of Industrial Organization.*

Bottazi, G. and A. Secchi, 2003a, "Sectoral Specificities in the Dynamics of US Manufacturing Firms." *LEM Working Paper* 2003/18, http://www.sssup.it/~LEM/.

Bottazi, G., E. Cefis, G. Dosi, and A. Secchi, 2003b, "Invariances and Diversities in the Evolution of Manufacturing Industries." *LEM Working Paper* 2003/21, http://www.sssup.it/~LEM/.

Caves, R., 1998, "Industrial Organization and New Findings on the Turnover and Mobility of Firms." *Journal of Economic Literature*, 36 (December, 1998), pp. 1947-1982.

Dunne, P. and A. Hughes, 1994, "Age, Size, Growth and Survival: UK Companies in the 1980s." *Journal of Industrial Economics*, 42 (2), pp. 115-140.

Ericson, R. and A. Pakes, 1995, "Markov-Perfect Industry Dynamics: A Framework for Empirical Work." *Review of Economic Studies*, 62, pp. 53-82.

Evans, D., 1987, "The Relationship Between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries." *Journal of Industrial Economics*, 35 (4): 567-581.

Gowrisankaran, G., 1999. A Dynamic Model of Endogenous Horizontal Mergers. *RAND Journal of Economics*, 30, pp. 56-83.

Hall, B., 1987, "The Relationship Between Firm Size and Firm Growth in the US Manufacturing Sector." *Journal of Industrial Economics*, 35 (4): 583-606.

Hart, P. and N. Oultan, 1996, "Growth and Size of Firms." *The Economic Journal*, 106 (438): 1242-1252.

Hart, P. and S. Prais, 1956, "The Analysis of Business Concentration: A Statistical Approach." *Journal of the Royal Statistical Society.* Series A, 119 (2): 150-191.

Ijiri, Y. and H. Simon, 1964, "Business Firm Growth and Size." *American Economic Review*, 54 (2), pp. 77-89.

Jorgenson, D.W. and K. Stiroh, 2000, "Industry-Level Productivity and Competitiveness Between Canada and the United States." *American Economic Review*, 90 (2), pp. 161-167.

Jovanovic, Boyan, 1982. "Selection and the Evolution of Industry." *Econometrica*, 50 (3), pp. 649-670.

Klepper, S., 2002. "Firm Survival and the Evolution of Oligopoly." *RAND Journal of Economics* 33 (1), pp. 37-61.

Kumar, M. (1985), "Growth, Acquisition Activity and Firm Size: Evidence from the United Kingdom." *Journal of Industrial Economics*, 33, pp. 327-338.

Laincz, C., 2004a, "Market Structure and Endogenous Productivity Growth: How do R&D Subsidies Affect Market Structure?" *Journal of Economic Dynamics and Control*, 29 (1-2), pp. 187-224.

Laincz, C., 2004b, "A Model of Growth with Dynamic Market Structure" Working Paper.

Laincz, C. and A. Rodrigues, 2005, “The Impact of Cost-Reducing R&D Spillovers on the Ergodic Distribution of Market Structures.” Working Paper.

Lotti, F. and E. Santarelli, 2004, “Industry Dynamics and the Distribution of Firm Sizes: A Nonparametric Approach.” *Southern Economic Journal* 70 (3), pp. 443-466.

Machado, J. and J. Mata, 2000, “Box-Cox Quantile Regression and the Distribution of Firm Sizes.” *Journal of Applied Econometrics*, 15, pp. 253-274.

Mansfield, E., 1962, “Entry, Gibrat’s Law, Innovation, and the Growth of Firms.” *American Economic Review*, 52 (5), pp. 1023-1051.

Mansfield, E., M. Schwartz, and S. Wagner, 1981, “Imitation Costs and Patents: An Empirical Study.” *The Economic Journal*, 91, pp. 907-918.

Mansfield, E., 1985. “How Rapidly Does New Industrial Technology Leak Out?” *Journal of Industrial Economics*, 34, 2, pp. 217-223.

McCloughan, P., 1995, “Simulation of Concentration Development from Modified Gibrat Growth-Entry-Exit Processes.” *Journal of Industrial Economics*, 43 (4), pp. 405-433.

Pakes, A. and R. Ericson, 1998. “Empirical Implications of Alternative Models of Firm Dynamics.” *Journal of Economic Theory*, 79, pp. 1-45.

Prais, S., 1976, *The Evolution of Giant Firms in Britain*, Cambridge, Cambridge University Press.

Simon, H. and C. Bonini (1958), “The Size Distribution of Business Firms.” *American Economic Review*, 48, pp. 607-617.

Singh, A. and G. Whittington, 1975, “The Size and Growth of Firms.” *Review of Economic Studies*, 42 (1): 15-26.

Sutton, J., 1997, “Gibrat’s Legacy.” *Journal of Economic Literature*, 35 (1), pp. 40-59.

Sutton, J., 1998, Technology and Market Structure. MIT Press: Cambridge, MA.

TABLE 1: Baseline parameter values for simulations.

Parameter	Symbol	Value
Discount Rate facing firms	$1/(1+r)$	1/1.08
Rate of Technological Spillover	δ	0.7
Productivity of R&D Investment	a	3
Sunk Entry Costs	X_e	0.2
Cost per unit of R&D Spending	c	1
Fixed Costs	f	0.2
Liquidation Value	ϕ	0.1
Outside Alternative Value	ψ	0.2
Rate of Decrease in Marginal Costs	η	0.10
Smallest Submarket Market Share	θ_1	0.136
Increments in Submarket Size	b	0.0318
Unit Cost of Labor	Z	131.12

TABLE 2: Distribution Statistics for Baseline (Natural Logs)

	Normal	H&O	Model	H&O	Model	H&O	Model
Measure	Distribution	Emp.	Emp.	Sales	Sales	Net Assets	Net Assets
Mean	-	3.1582	3.1582	7.2015	5.1321	5.5539	4.9010
Std. Dev.	-	1.5197	0.2803	1.6628	0.3697	1.9635	2.1468
Skewness	0	0.7487	-0.1114	0.1932	-1.0220	0.4366	0.7825
Kurtosis	3	4.5794	4.7265	6.1876	11.7373	4.835	2.9123

Note: H&O refers to the results reported in Hart and Oulton (1996).

TABLE 3: Distribution Statistics for Baseline Excluding Firms with <1 Employee (Natural Logs)

	Normal	H&O	Model	H&O	Model	H&O	Model
Measure	Distribution	Emp.	Emp.	Sales	Sales	Net Assets	Net Assets
Mean	-	3.1582	3.1582	5.1825	5.1321	5.5539	4.8469
Std. Dev.	-	1.5197	0.2718	1.6628	0.3516	1.9635	2.1499
Skewness	0	0.7487	0.0777	0.1932	-0.5251	0.4366	0.77770
Kurtosis	3	4.5794	3.9254	6.1876	5.6798	4.835	2.9031

TABLE 4: Average Summary Statistics in Natural Logs

Size Measure	Employment	Sales	Net Assets
Observations	69.5	69.5	69.5
Average	3.521	5.027	5.189
Std. Dev.	0.521	0.493	1.410
Skewness	0.867	-0.402	0.717
Kurtosis	4.183	3.819	2.714

TABLE 5: Regression Results

Firms	Size Measure	Average Coefficient	$H_0 : \beta = 1$ Rejection Rate 10% level	5%	1%
All	Employment	0.971	100%	90%	80%
	Sales	0.966	100%	100%	100%
	Net Assets	0.944	100%	100%	100%
Large	Employment	0.998	10%	0%	0%
	Sales	0.979	100%	90%	80%
	Net Assets	0.970	70%	70%	50%
Small	Employment	0.929	100%	100%	90%
	Sales	0.947	100%	100%	90%
	Net Assets	0.883	100%	100%	100%
		Average Difference	$H_0 : \beta_{LARGE} = \beta_{SMALL}$ Rejection Rate 10% level	5%	1%
Equality	Employment	0.069	100%	100%	80%
	Sales	0.019	60%	60%	20%
	Net Assets	0.061	90%	90%	70%

TABLE 6: Tests of Standard Deviation of Growth Rates by Size Class

Size Measure	Total	Large	Small	$H_0 : \sigma_1 = \sigma_2$ Rejection Rate 10% level	5%	1%
Employment	0.301	0.159	0.361	80%	80%	70%
Sales	0.321	0.195	0.401	70%	70%	70%
Net Assets	0.542	0.438	0.584	50%	50%	30%

Figure 1: US Firm Size Distribution by Employment, 2002
Source: Statistics of US Business (SUSB), Small Business Administration

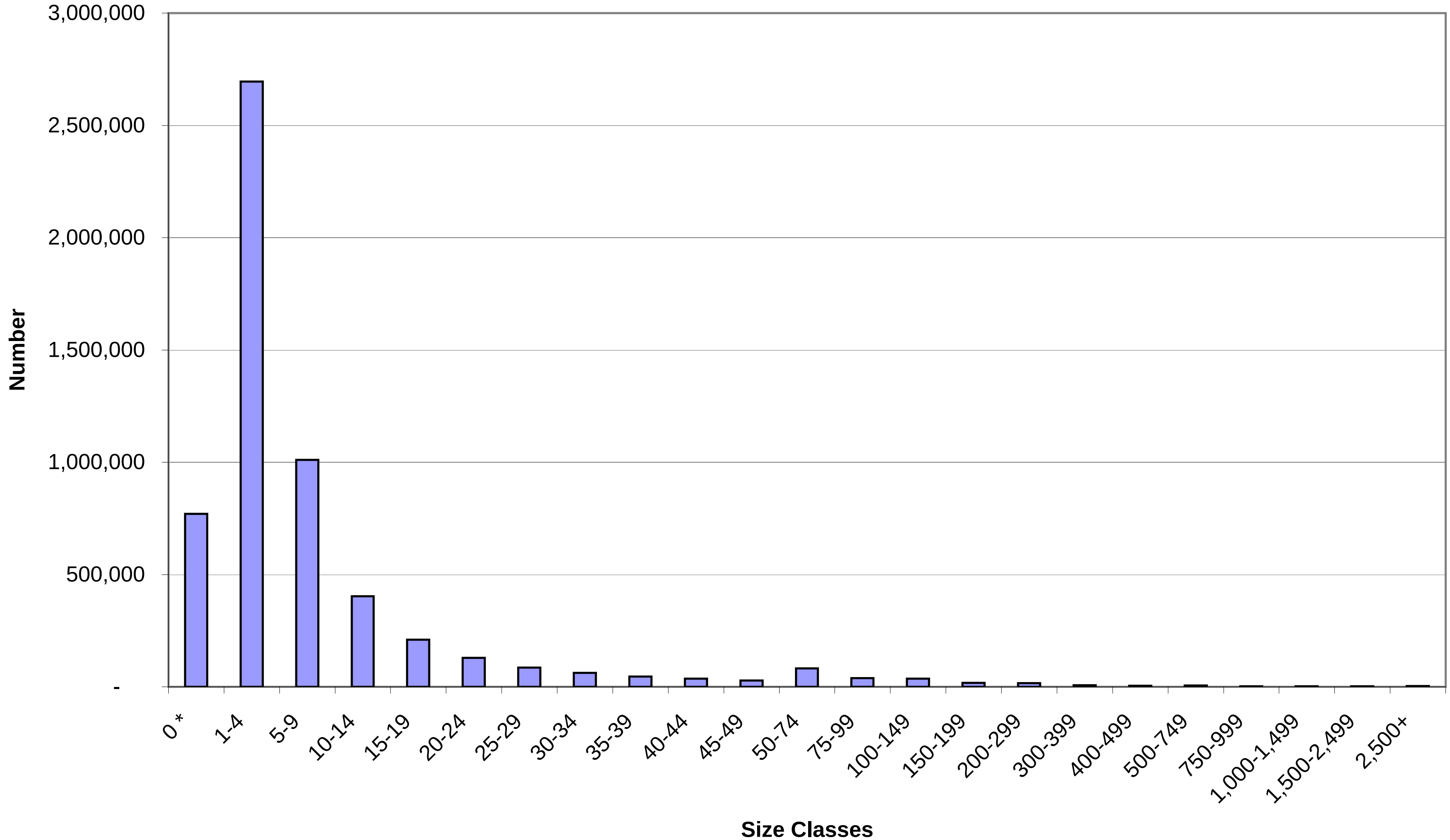

Figure 2a: Employment Distribution

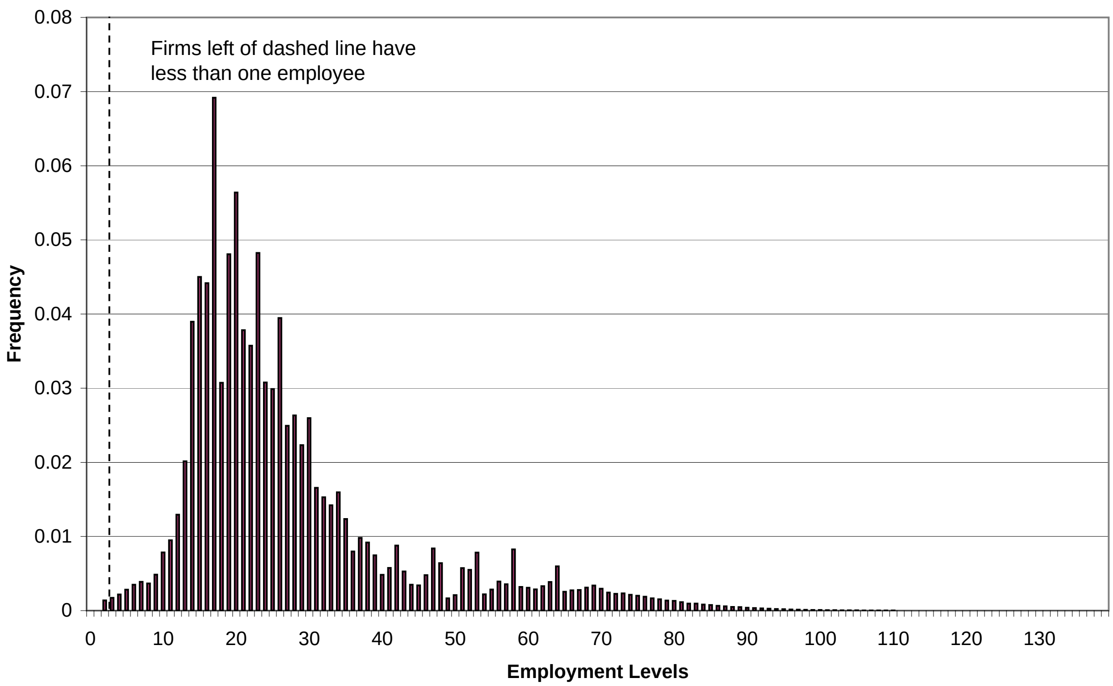

Fig 2b: Ln of Employment Distribution

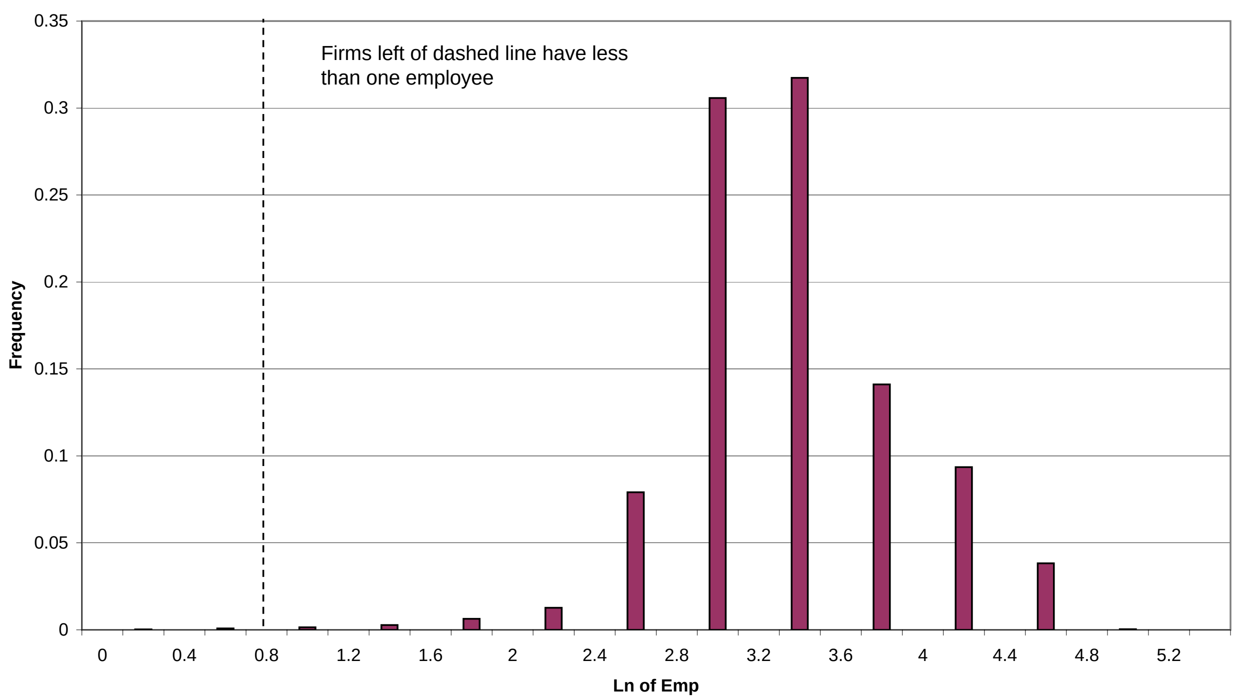

Fig 3a: Values Distribution

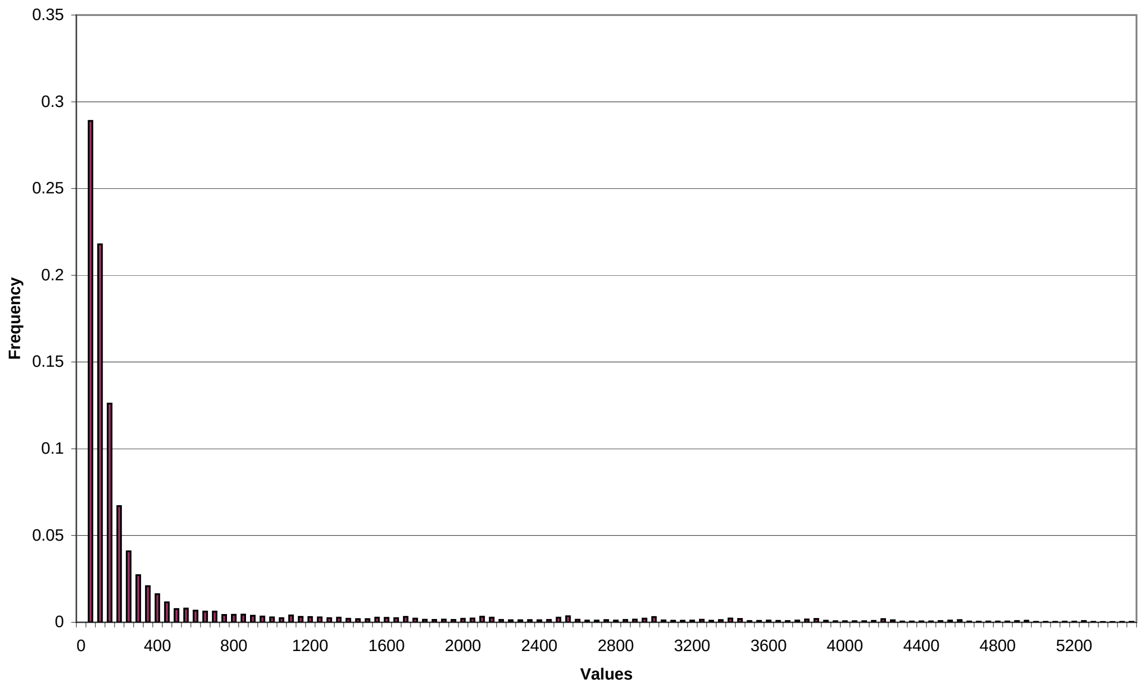

Fig 3b: Ln of Values Distribution

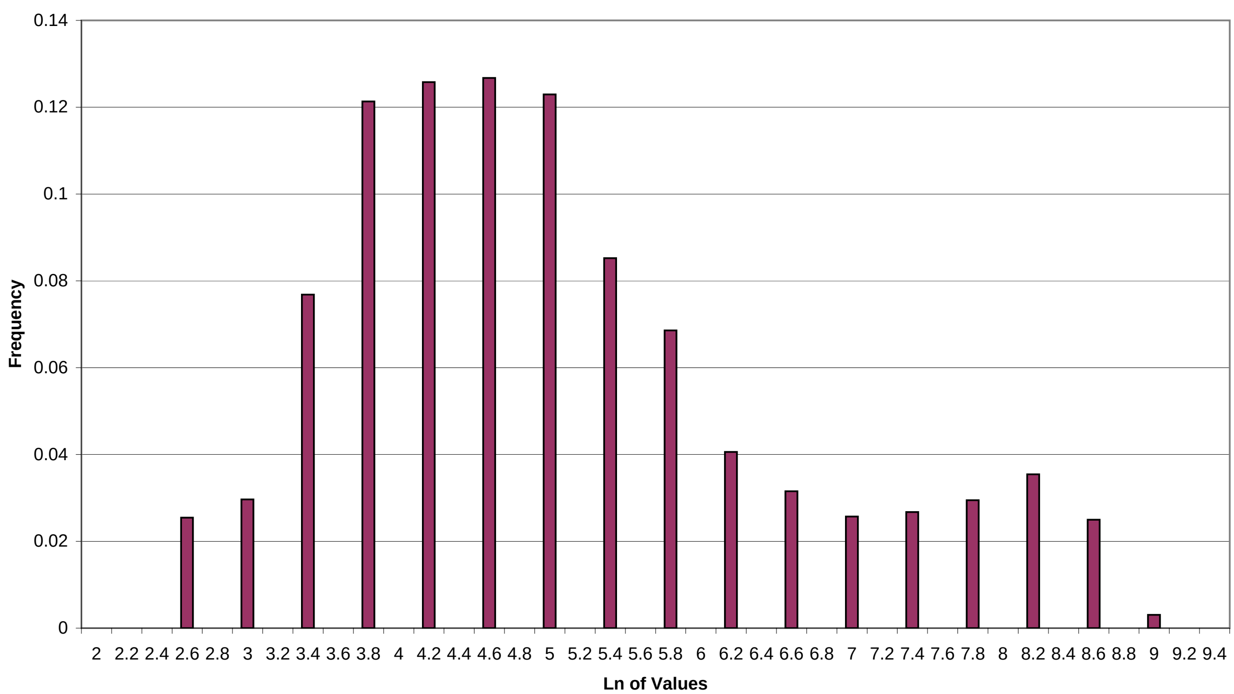

Fig 4a: Sales Distribution

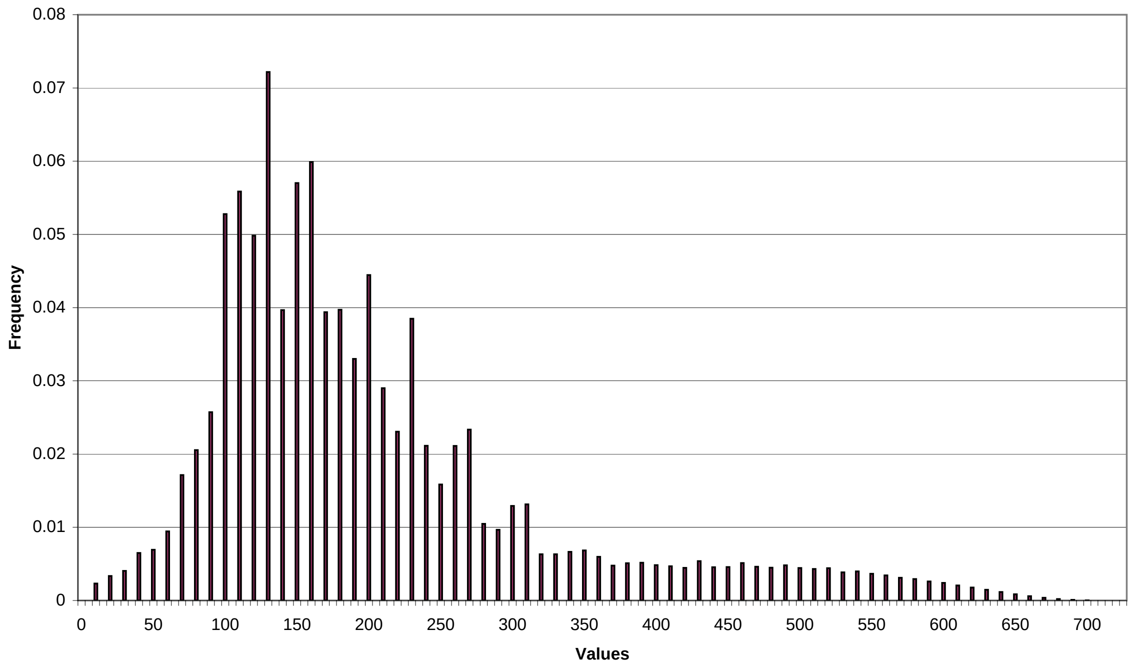

Fig 4b: Ln of Sales Distribution

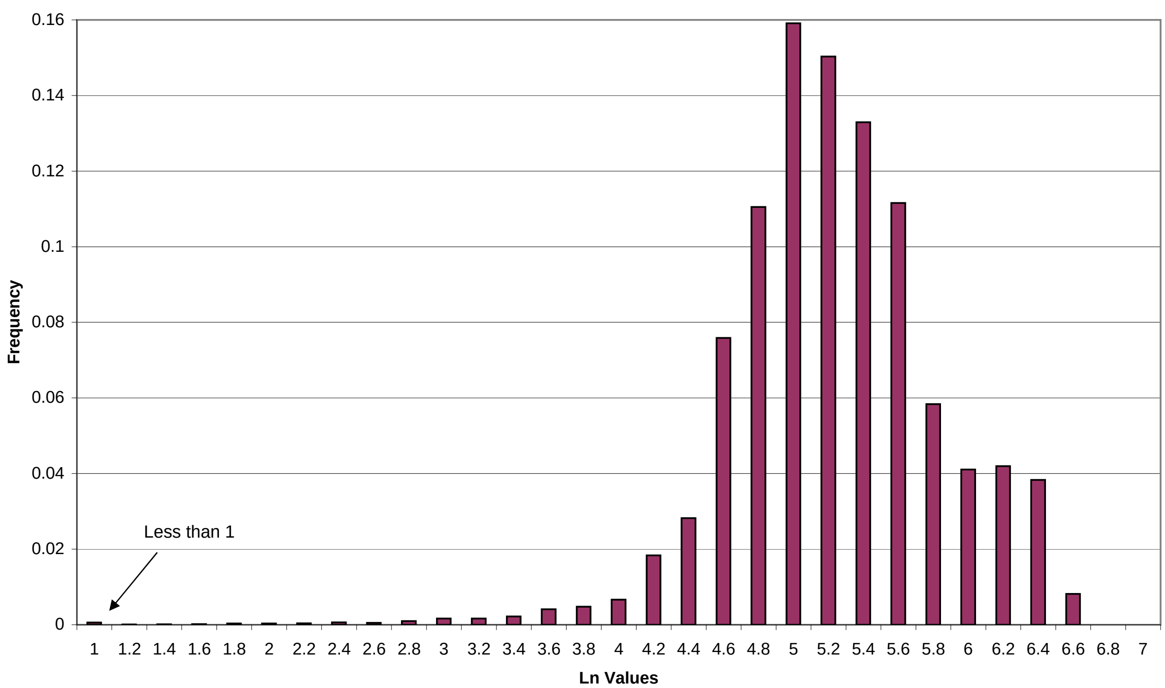

FIGURE 5a: Fixed Cost Effects (Levels)

Percentage Deviations from Baseline Values of Firm Size Distribution by Employment

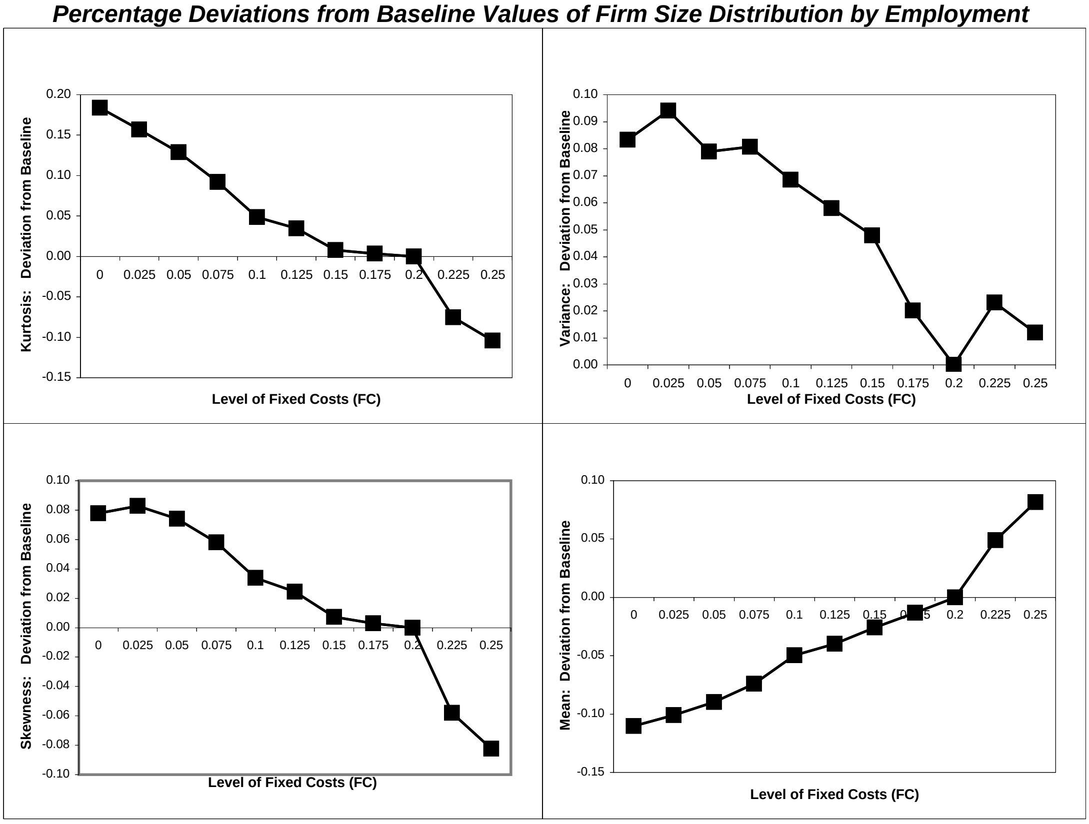

FIGURE 5b: Fixed Cost Effects (Logs)

Percentage Deviations from Baseline Values of Firm Size Distribution by Employment

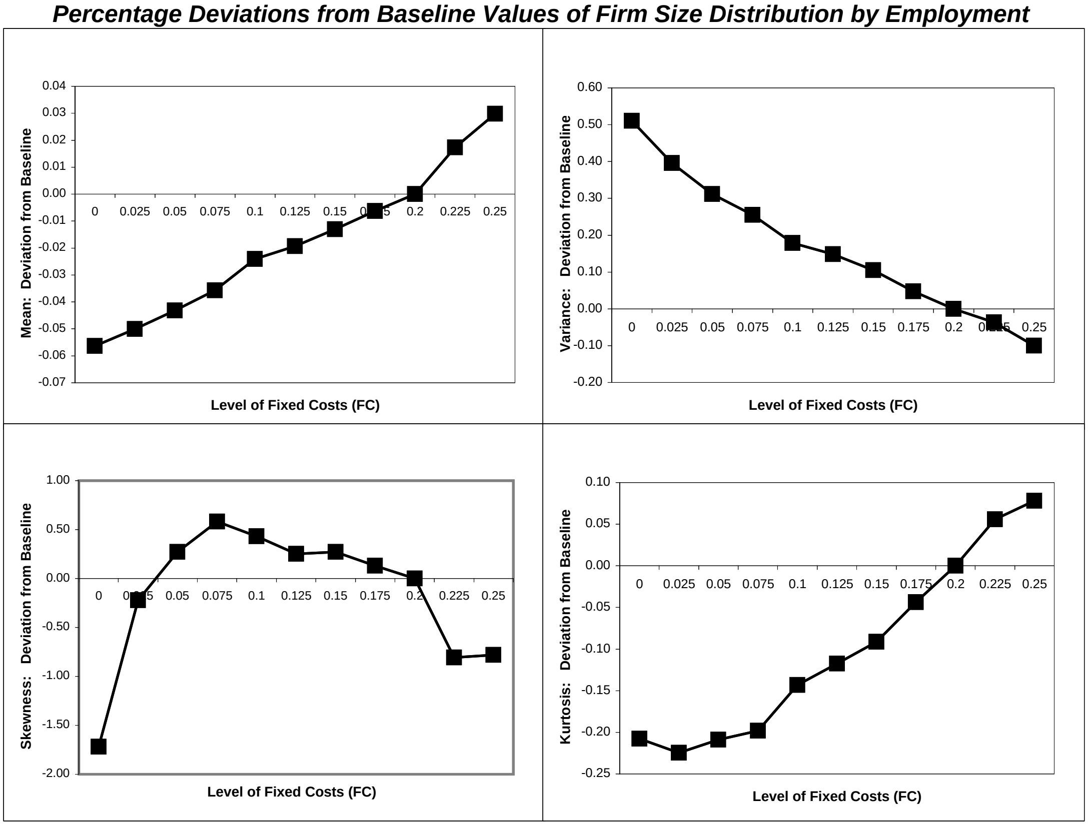

FIGURE 5c: Effects of Fixed Costs on Firm Size Distribution (Levels)

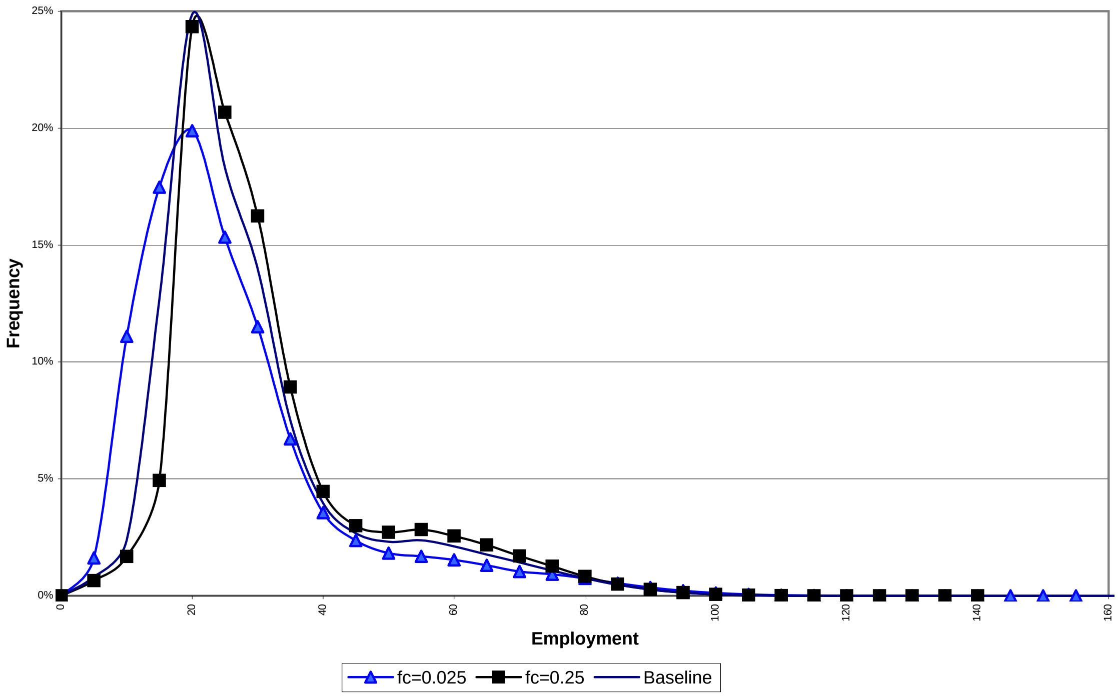

FIGURE 5d: Effects of Fixed Costs on Firm Size Distribution (Logs)

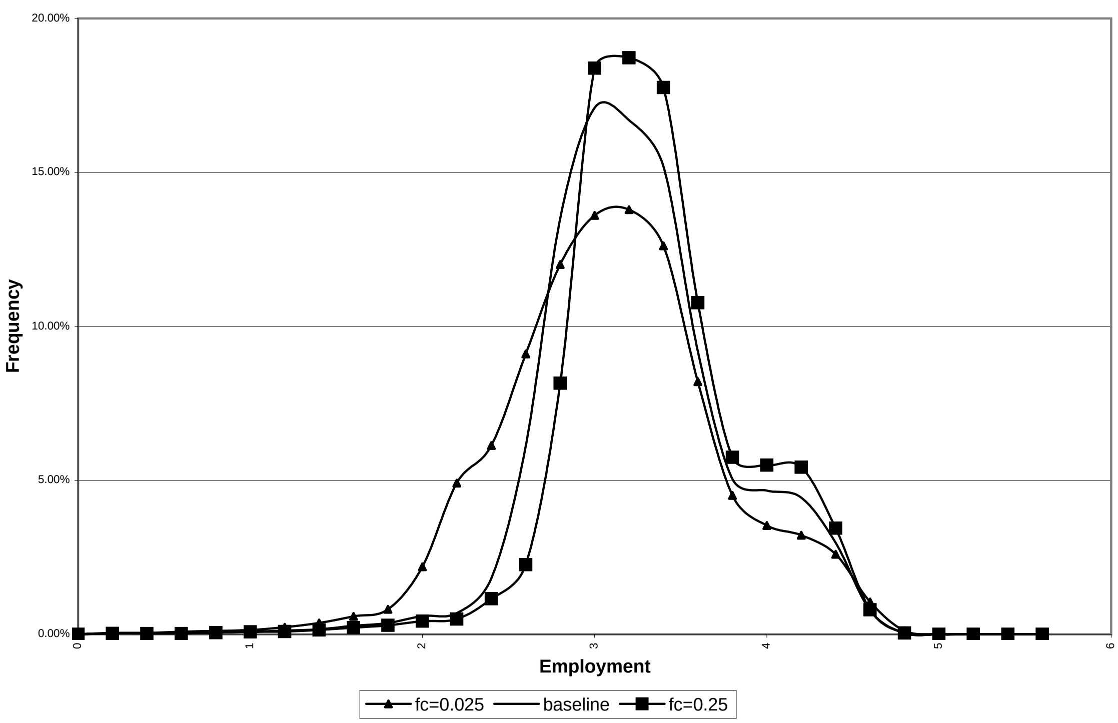

FIGURE 6a: Sunk Cost Effects (Levels)

Percentage Deviations from Baseline Values of Firm Size Distribution by Employment

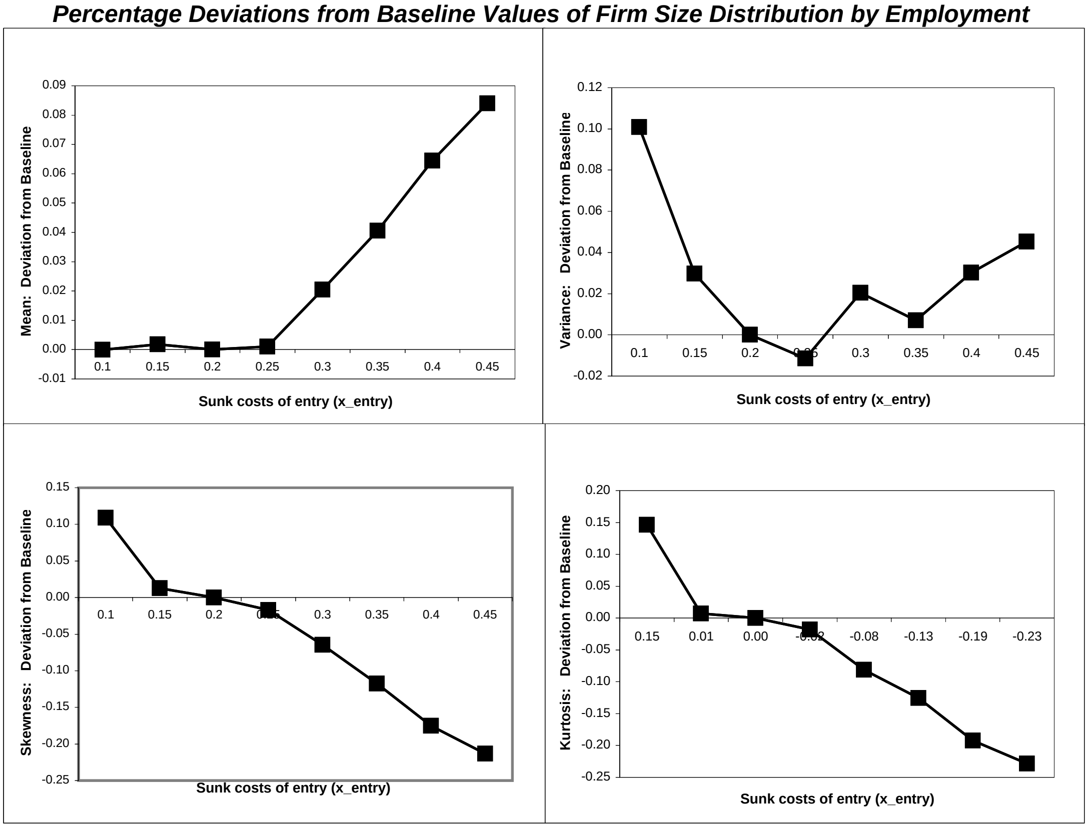

FIGURE 6b: Sunk Cost Effects (Logs)

Percentage Deviations from Baseline Values of Firm Size Distribution by Employment

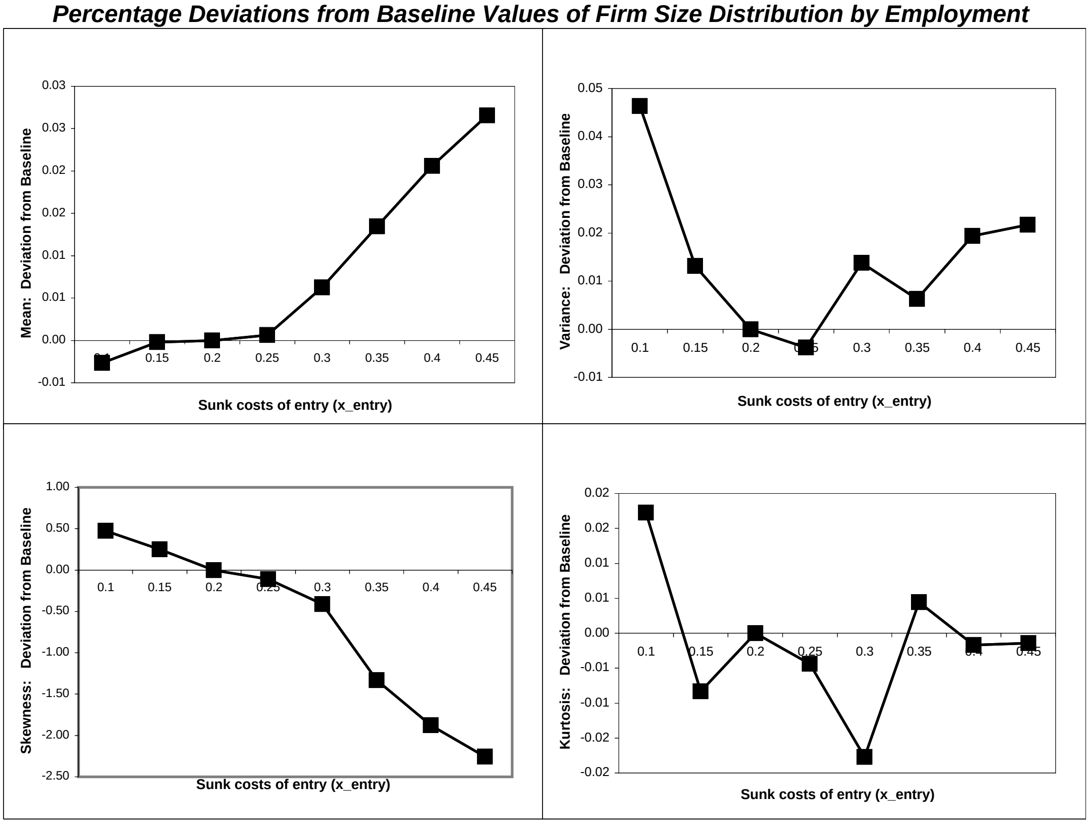

FIGURE 6c: Effects of Sunk Costs on Firm Size Distribution (Levels)

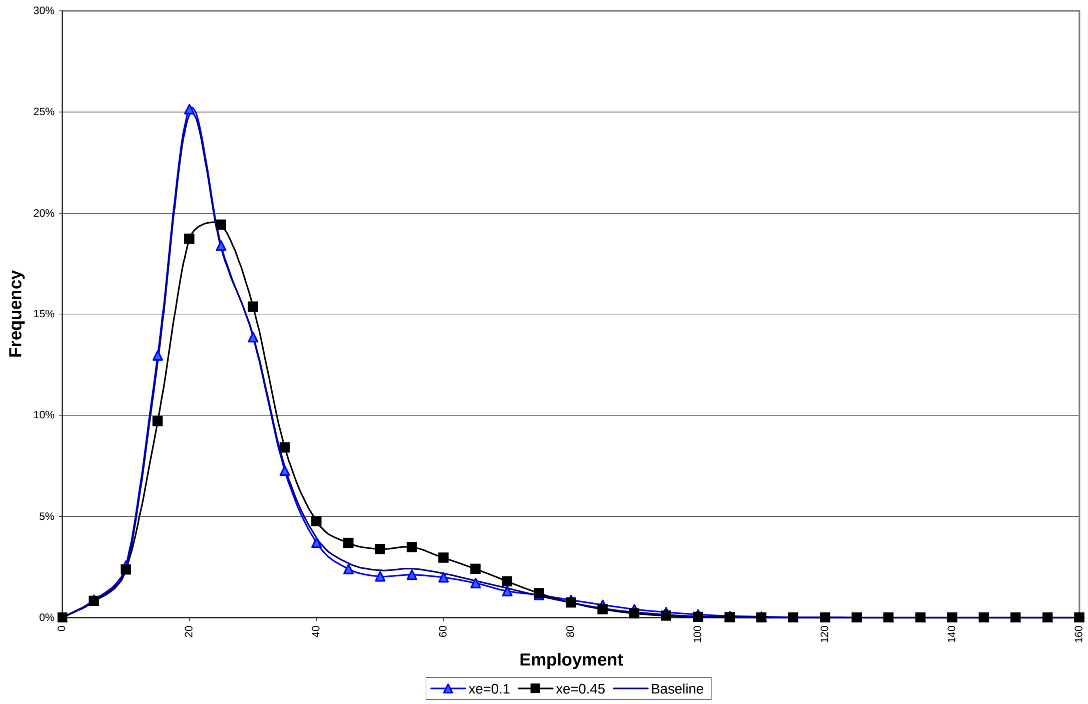

FIGURE 6d: Effects of Sunk Costs on Firm Size Distribution (Logs)

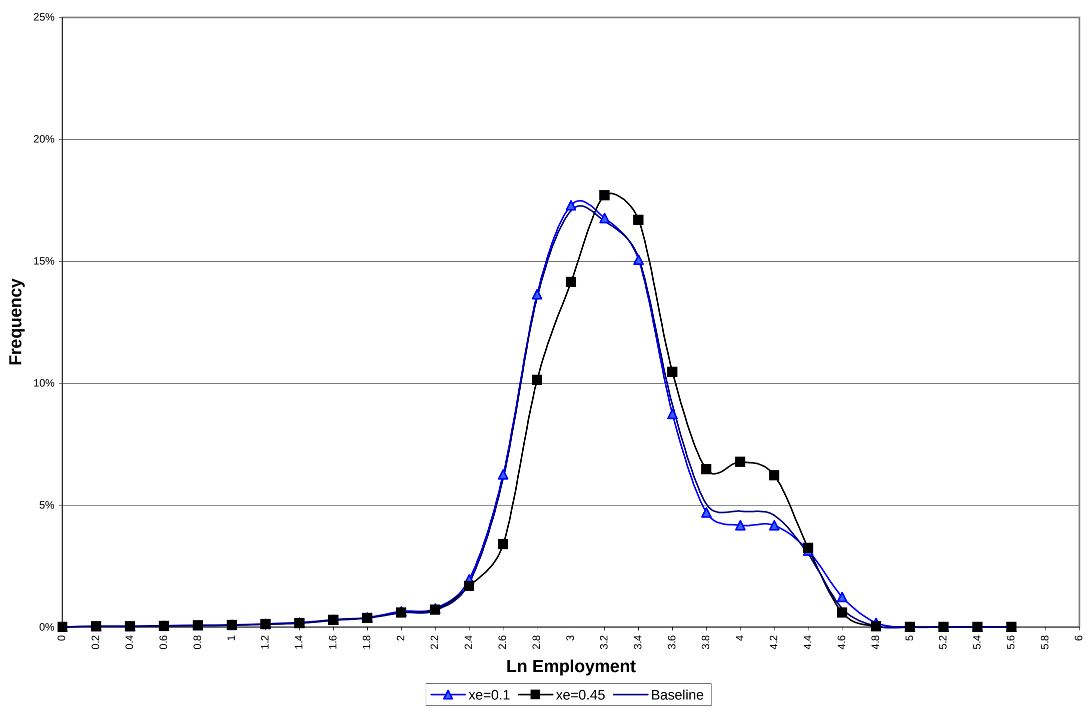

FIGURE 7a: The Rate of Cost Reduction Effects (Levels)

Percentage Deviations from Baseline Values of Firm Size Distribution by Employment

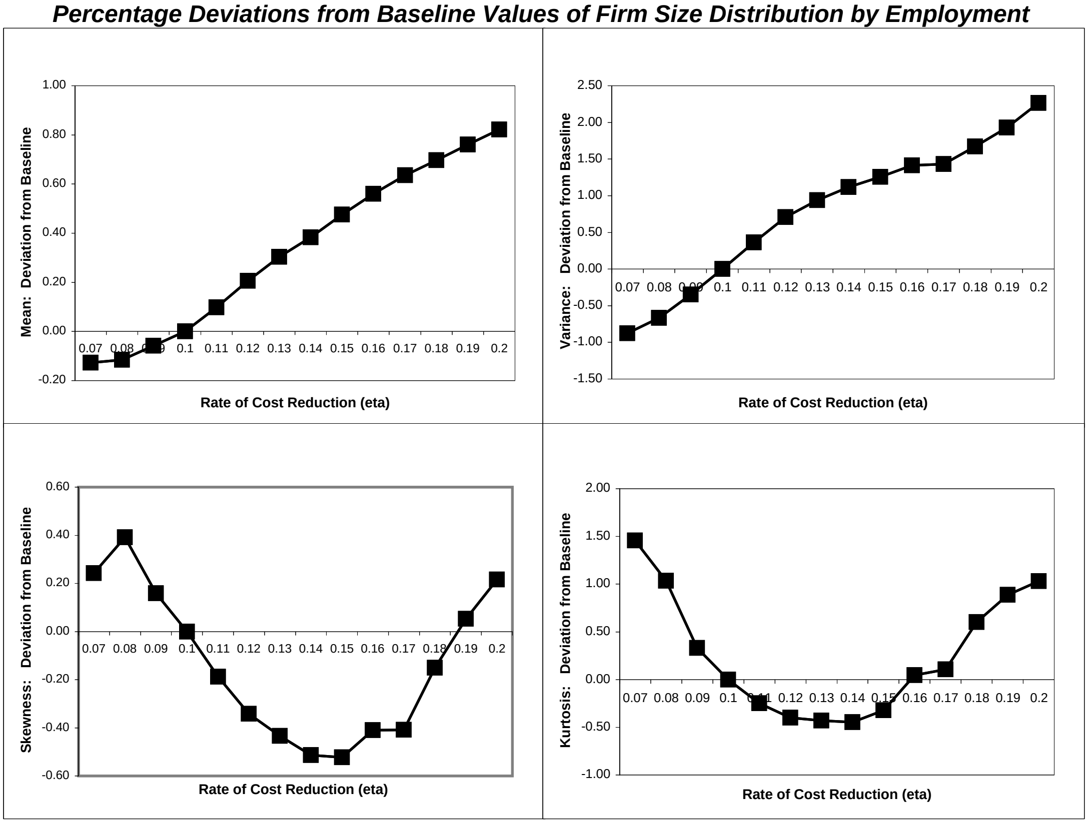

FIGURE 7b: The Rate of Cost Reduction Effects (Logs)

Percentage Deviations from Baseline Values of Firm Size Distribution by Employment

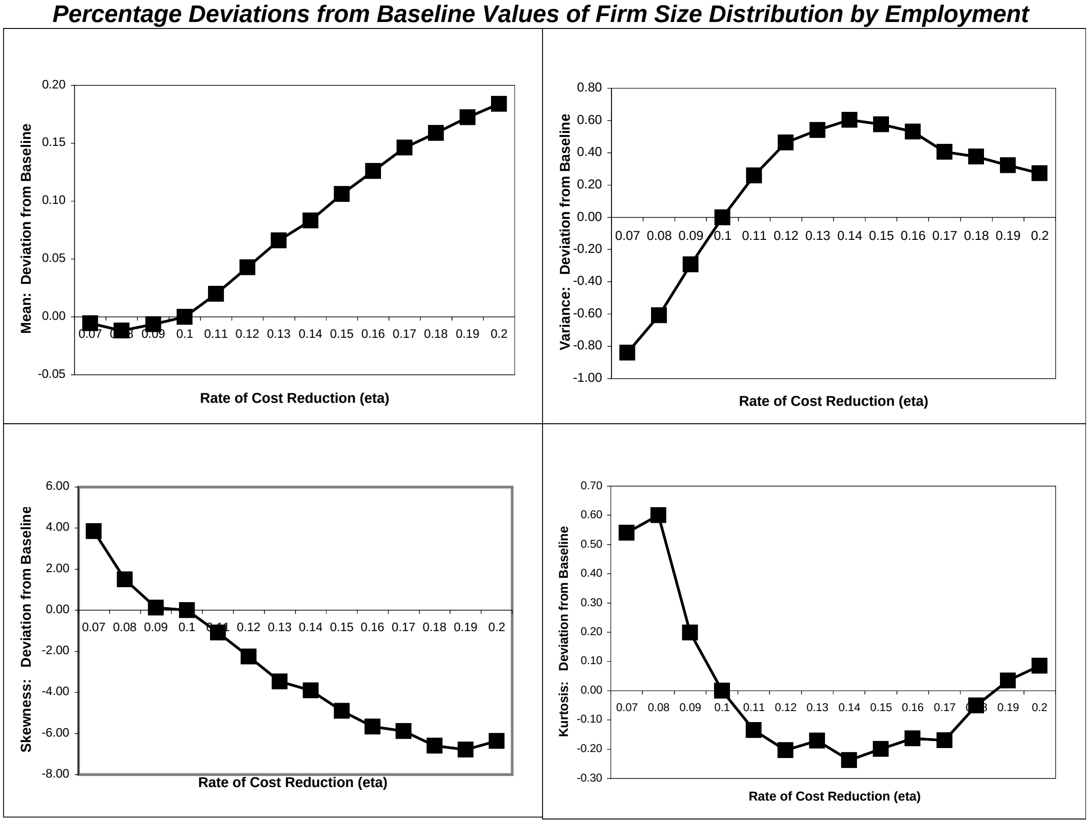

FIGURE 7c: Effects of the Rate of Cost Reduction on Firm Size Distribution (Levels)

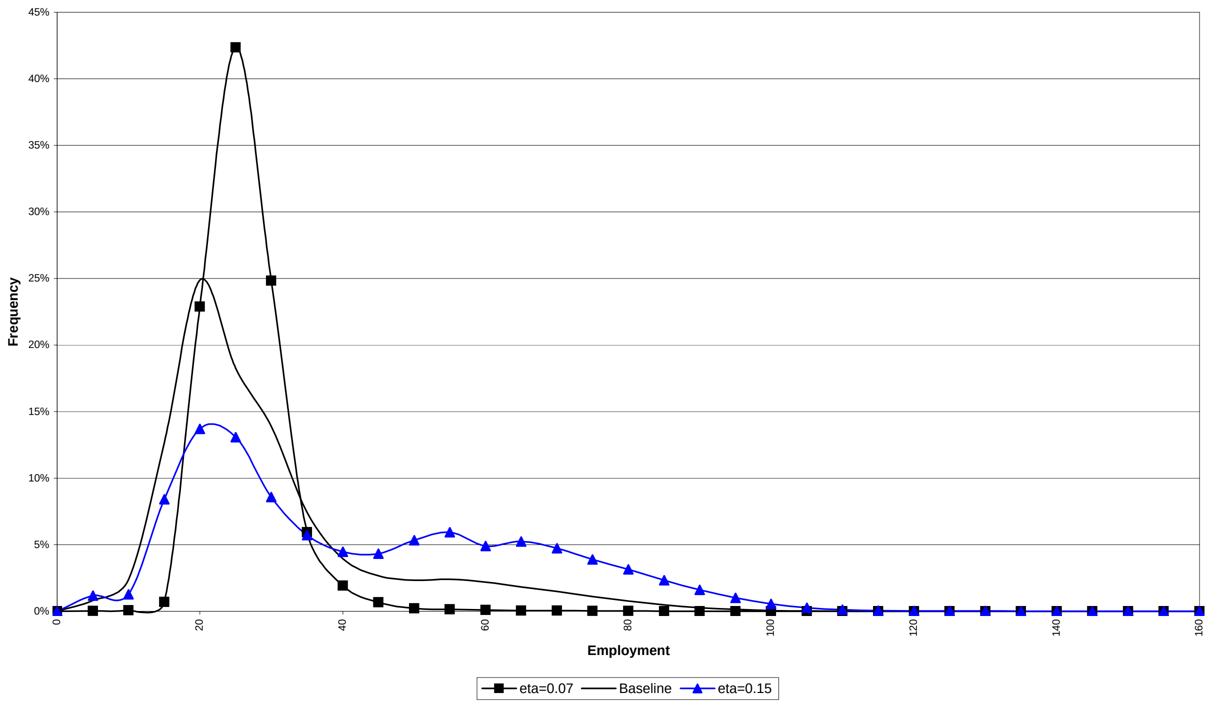

FIGURE 7d: Effects of the Rate of Cost Reduction on Firm Size Distribution (Logs)

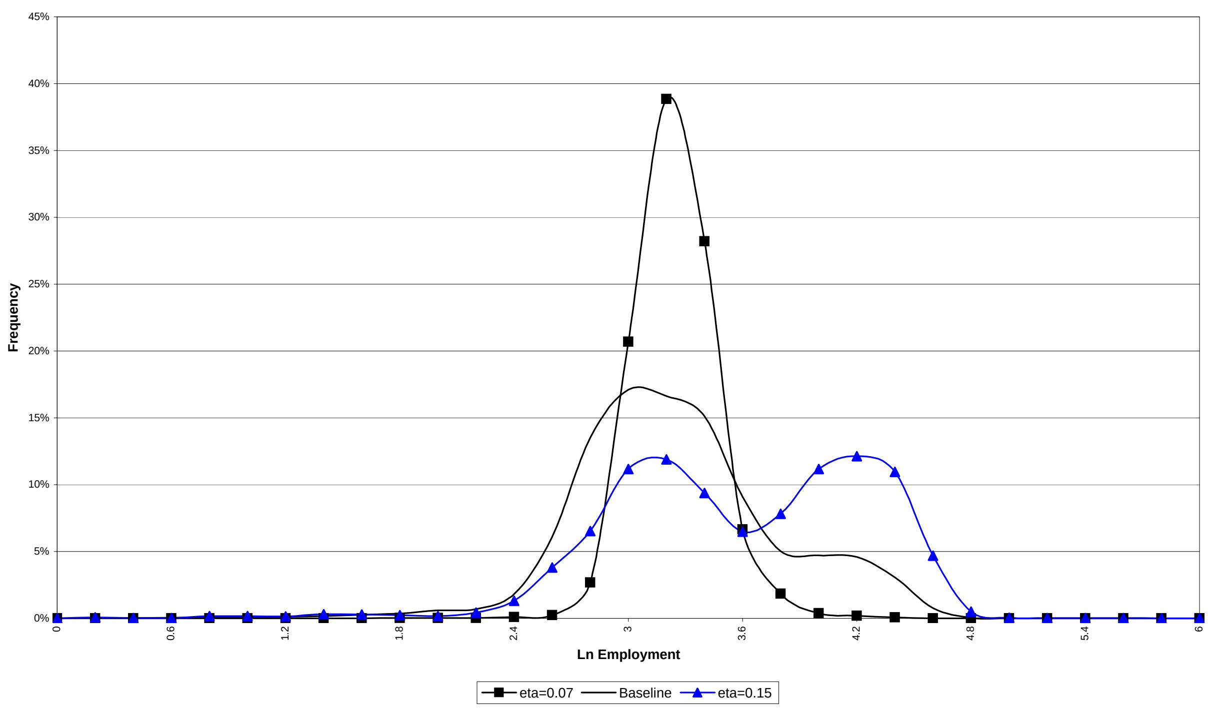

FIGURE 8a: Technological Opportunity Effects (Levels)

Percentage Deviations from Baseline Values of Firm Size Distribution by Employment

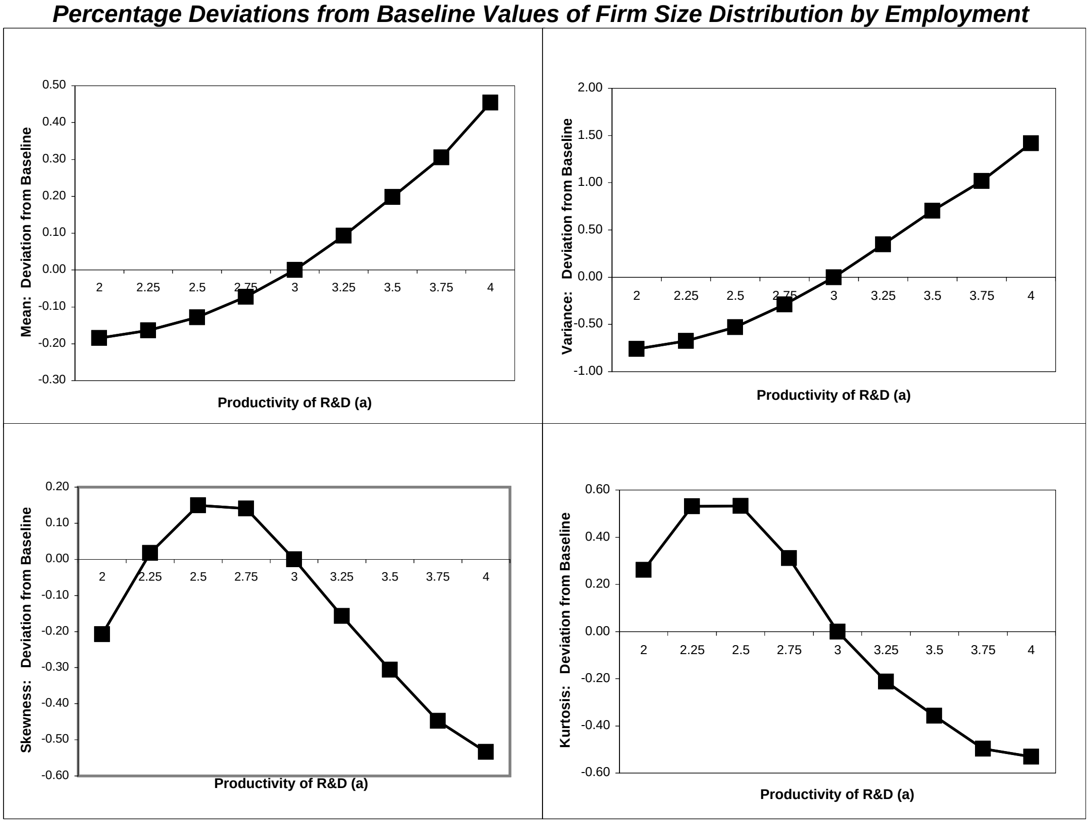

FIGURE 8b: Technological Opportunity Effects (Logs)

Percentage Deviations from Baseline Values of Firm Size Distribution by Employment

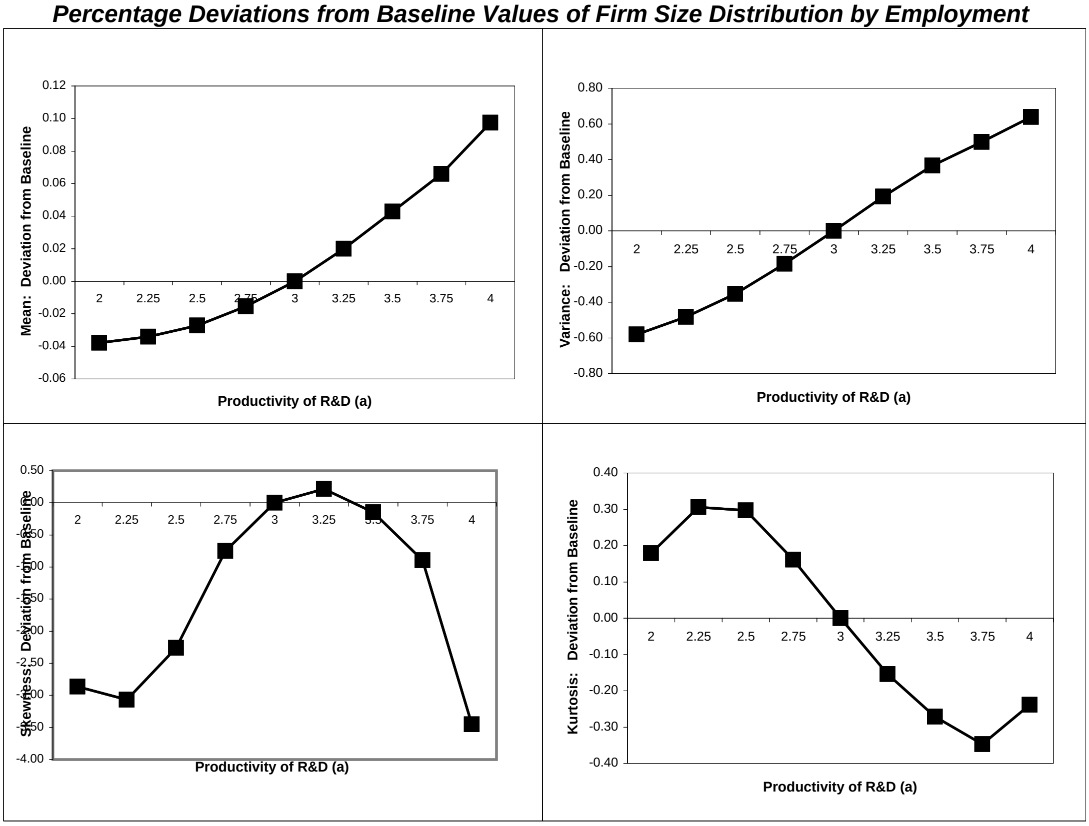

FIGURE 8c: Effects of the Productivity of R&D on Firm Size Distribution (Levels)

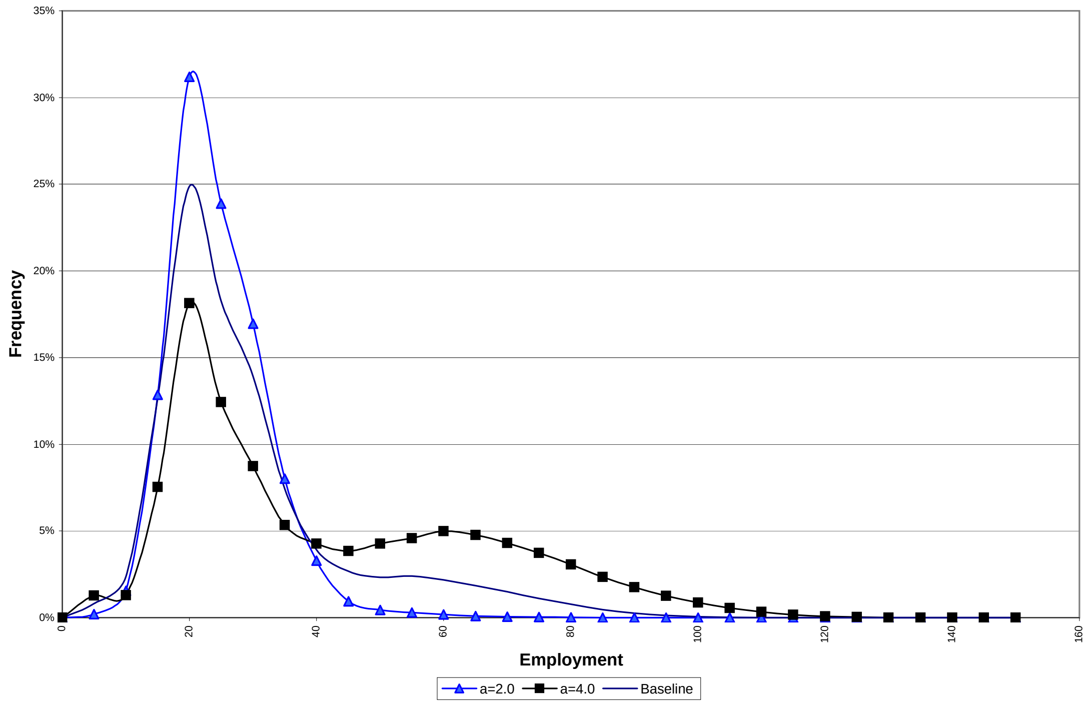

FIGURE 8d: Effects of the Productivity of R&D on Firm Size Distribution (Logs)

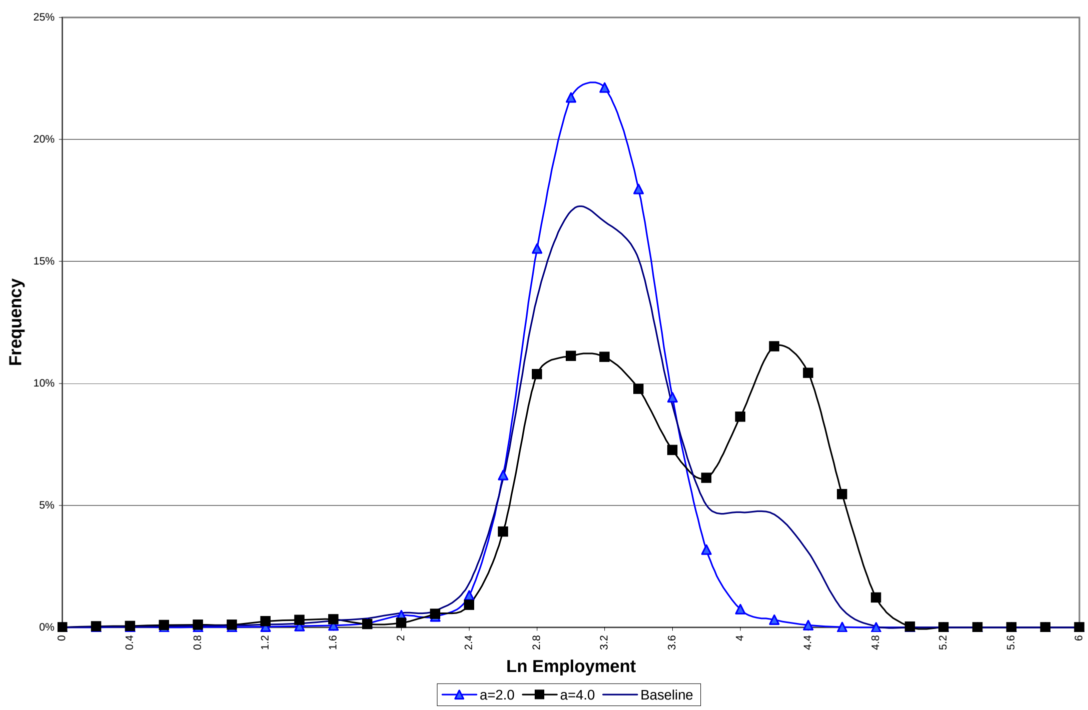

FIGURE 9a: The Rate of Technological Spillovers Effects (Levels)

Percentage Deviations from Baseline Values of Firm Size Distribution by Employment

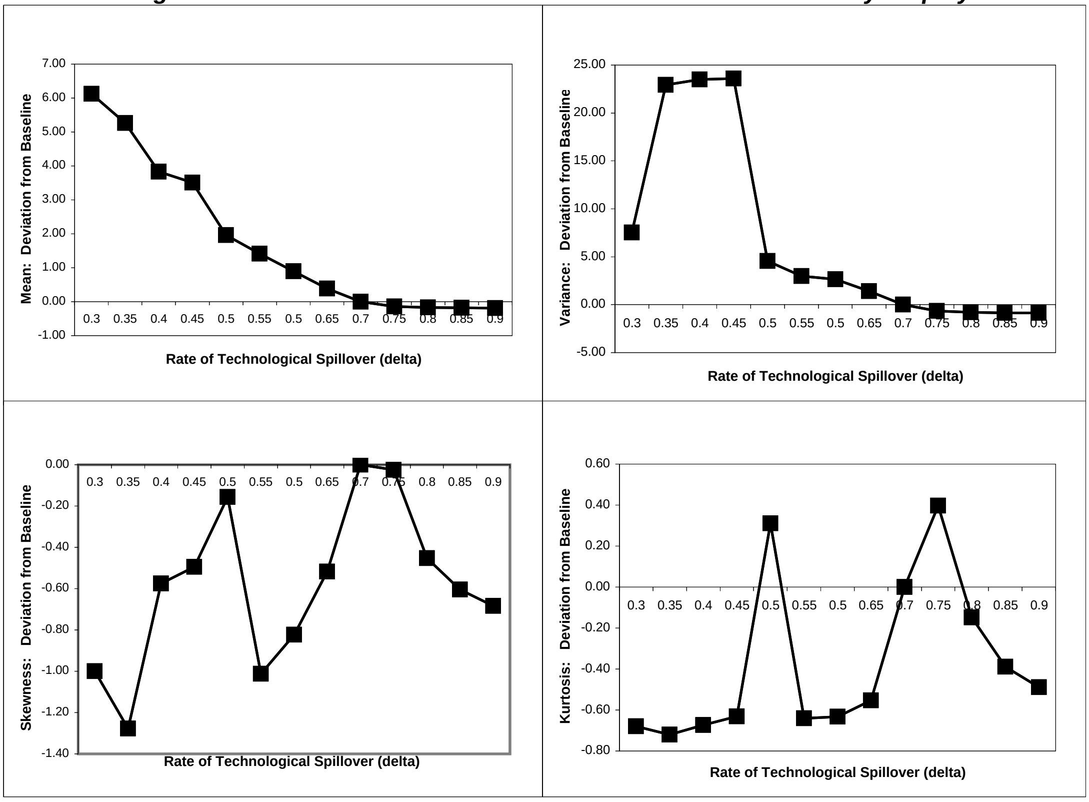

FIGURE 9b: The Rate of Technological Spillovers Effects (Logs)

Percentage Deviations from Baseline Values of Firm Size Distribution by Employment

FIGURE 9c: Effects of the Rate of Spillovers on Firm Size Distribution (Levels)

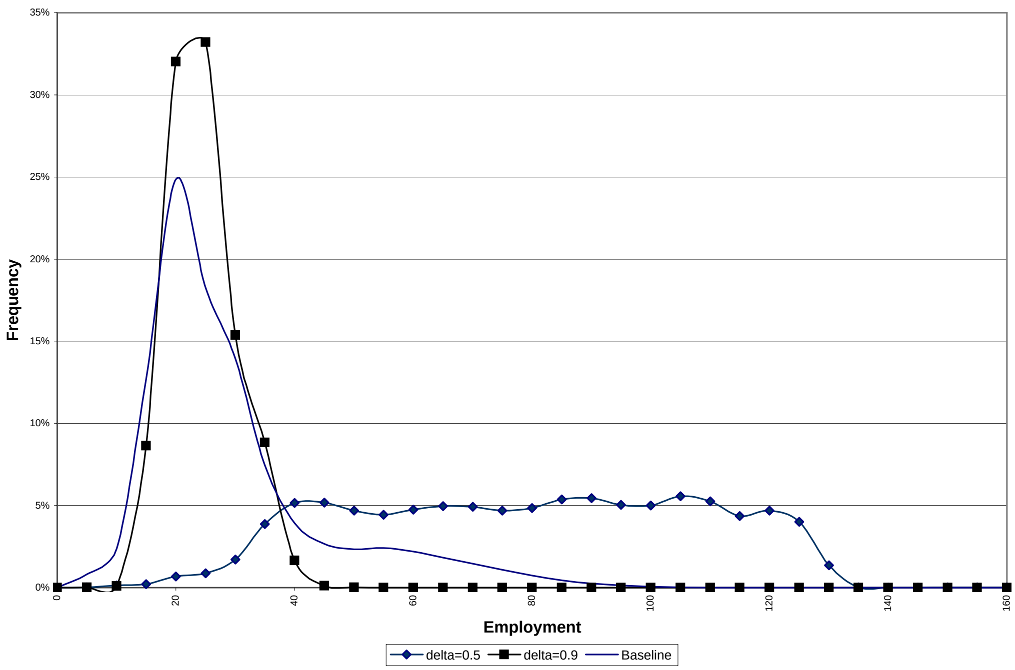

FIGURE 9d: Effects of the Rate of Spillovers on Firm Size Distribution (Logs)