

Understanding Gibrat's Law

with a Markov-Perfect Dynamic Industry Model*

Christopher A. Laincz¹ and Ana Sofia Domingues Rodrigues²

PRELIMINARY AND INCOMPLETE,

PLEASE DO NOT CITE

Abstract

Gibrat's Law of proportionate effect, as applied to firms, states that the growth rate of a firm is independent of its size. Empirical work on firm dynamics finds crucial deviations from Gibrat's Law such as smaller firms growing faster than larger firms (Evans, 1987, and Hall, 1987), a negative relationship between the variance of growth rates and size (Dunne and Hughes, 1994), and first-order positive autocorrelation in the growth rates (Kumar, 1995). Moreover, the degree of deviation from Gibrat's Law varies across industries. This paper contributes to our understanding of the forces that make Gibrat's Law a close, but imperfect approximation of firm size distributions and seeks to determine potential sources of cross-industry variation. Here, we employ an extension of the Ericson-Pakes (1995) theoretical framework that allows for firm growth developed by Laincz (2004a). By varying key parameters, the simulations demonstrate potential sources for the various, and sometimes conflicting, results on Gibrat's Law uncovered in the empirical literature.

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

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

*We would like to thank seminar participants at the University of York (UK) for their helpful comments on an earlier draft of this paper.

1 Introduction

Gibrat's Law of proportionate effect, as applied to firms, states that the growth rate of a firm is independent of its size. Empirical work on firm dynamics finds crucial deviations from Gibrat's Law such as smaller firms growing faster than larger firms (Evans, 1987 and Hall, 1987), a negative relationship between the variance of growth rates and size (Dunne and Hughes, 1994), and first-order positive autocorrelation in the growth rates (Kumar, 1995). Moreover, the type of deviation from Gibrat's Law varies across industries. This paper contributes to our understanding of the forces that make Gibrat's Law a close, but imperfect approximation of firm size distributions and seeks to determine potential sources of cross-industry variation. Here, we employ an extension of the Ericson-Pakes (1995) theoretical framework 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.

Studies on the firm size distribution and Gibrat's Law to date have been the province of empiricists. While we can write down various reduced form models, as in McCloughan (1995), to reproduce many of the statistical facts, little of the 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. 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 can approximate these results when testing the simulated output using the techniques of the empirical literature.

In addition, a more recent literature uncovers significant cross-industry variation in the higher moment 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 at least some segments of the service industry may fit Gibrat’s Law in its strong form. We explore how key structural parameters alter the higher moments of the firm size distribution to understand cross-industry differences.

After briefly reviewing the lengthy empirical literature on Gibrat’s Law in the next section, section 3 presents the basic model. In section 4 we compare the results of a baseline simulation to the basic predictions and deviations of Gibrat’s Law. Section 5, which is incomplete, will explore how varying key parameters alters the firm size distribution and compare these results with the literature that looks at a cross-section of industries. Section 6 summarizes the results and indicates avenues for future work.

2 Gibrat’s Law and Empirical Findings

Following on 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. Gibrat’s Law yields a log-normal distribution of firm sizes. Growth in the size of firm i from one period to the next is given as:

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

where $u_i(t) \sim iid N(\mu, \sigma^2)$. 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 initial size. 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.

If $\beta \neq 1$, then firm growth is not independent of its size. In McCloughan's simulations, he draws on Prais (1975) and uses 0.75, 0.95, and 1.05 as alternative values for β . Figures 1-4 display the results of simulating (2) to illustrate why this simple statistical process has such appeal.¹ The procedure starts 50 firms at equal sizes and runs for 50 periods. The process is run 20 times and the reported results are based on the average of the 20 simulations.

Figure 1 is the outcome of Gibrat's Law in its strong form with the mean growth rate set to zero. The distribution is roughly symmetric with a skewness coefficient just below zero of -0.035 and the index of kurtosis is 2.84 (0 and 3 for the normal distribution, respectively). When $\beta < 1$, firms revert to the average size for the industry. Figures 2 and 3 display this process. Notably the distributions become skewed to the right with coefficients equal to 0.11 and 0.17. The distributions' kurtosis is hardly changed in these simulations, 2.83 and 2.88. When $\beta > 1$ (Figure 4) firm sizes diverge and the skewness remains to the right, but at a smaller magnitude of 0.08 while the kurtosis rises to 3.06 indicating a distribution slightly more peaked (leptokurtic) than the normal distribution.

Comparing these results to empirical estimates based on a large sample of firms, Hart and Oulton (1996) measure firms by employees, net sales, and net assets. In all cases they find the distribution is skewed right, with 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 normal distribution exactly and that the close approximation justifies its use in empirical work.

Our task is rather different. We are specifically interested in the deviations themselves. We

¹ We follow McCloughan's (1995) simulation methodology. Our results nearly match his reported results in Table II, p. 413.

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.

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 while in the entire sample of firms and survivors only, firms grow less than proportionally. If one considers large firms only, the mean growth rate is independent of size. He still concludes that Gibrat's law does not hold for either of the versions considered due to the fact that even for the latter case the variance of growth rates decreases with size.

Subsequent empirical analysis largely confirmed Mansfield's initial foray into the subject. The studies were of two main types: testing the moments of the firm size distribution against lognormality and utilizing growth-size regressions. In the latter, using more advanced econometric techniques to deal with heteroscedasticity and sample selection bias, Hall (1987) and Evans (1987) find 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) expand the question looking at the size-growth relationship across small, medium, and large firms. They 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 implied 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). Hart and Oulton (1996) use a Galton-Markov process, such as proposed by Chesher (1979), and find that firm growth is a mean reverting process, and that, on average, small firms grow faster than large ones. 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 of growth rates.

We want to understand how key structural parameters affect the firm size distribution and their relationship with Gibrat's Law. Recent literature suggests significant cross-industry variation in the evolution of the firm size distribution. Santarelli and Lotti (2004) look at the evolution of the size distribution of new firms in bread, office machinery, radio and TV equipment, and footwear. Over a period of five years most of the distributions approach the log normal distribution, however, the more technologically oriented industries achieved the lognormal faster. Footwear, also converged to the log-normal relatively quickly and the authors attribute this to the agglomeration effects. Bread, however, remained bimodal and did not completely converge by the end of the time period analyzed.

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 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.

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 discussed 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, thinking of each as a highly disaggregated good and then aggregate across the varieties. Thus, one can think of each good as being defined at a 7-digit level, for example, and the aggregation taking place at a less detailed industry level such as 4 or 5 digits in the SIC or NAIC codes. The result of our approach yields industries averaging about 20 firms at any one time.

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 incumbents 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

² Sutton (1997), p. 49.

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.

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 of the final goods sector is given by the production function:

$$Y_t = x_{1t}^{\beta_1} \cdot x_{2t}^{\beta_2} \cdot \dots \cdot x_{Mt}^{\beta_M}, \text{ where } \sum_{m=1}^M \beta_m = 1. \quad (3)$$

Each x_{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. We allow the β 's to vary across substitutes to generate greater variation in market shares.⁴ The Cobb-Douglas framework also fixes the share that the final goods sector spends on each intermediate product. Normalizing the price of the output good to unity, the demand for each intermediate good x_{mt} is given by:

$$x_{mt} = \frac{\beta_m Y_t}{p_{mt}}. \quad (4)$$

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

⁴ An obvious extension to our framework would be to endogenize the shares across the goods. This could be accomplished by changing the final good production function to Dixit-Stiglitz 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.

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 between firms. All firms are assumed to face constant fixed costs which do not vary either with time or across firms. Each intermediate goods firm faces the following optimization problem for choosing quantity:

$$\max \pi_{km} = P_m \left(\beta_m Y, \sum_{n=1}^N q_{nm} \right) q_{km} - MC_{km} q_{km} - f \quad (5)$$

where P_m is the price of the intermediate good, q_{km} is the quantity output of firm k producing product m , MC_{km} are the marginal costs for firm k , 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. Price is a function of share of total demand for that subsector's good which is given by $\beta_m Y$. Due to the presence of fixed costs, some firms may choose not to produce. 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.

We focus on one submarket to illustrate the model in the discussion that follows. Let N_m^* be the number of firms producing $q_{mn} > 0$. The Cournot-Nash equilibrium outcomes yield the quantity produced by firm k in product market m as

$$q_{km}^* = \beta_m Y \frac{(N_m^* - 1) \left(\sum_{n=1}^{N_m^*} MC_{mn} - (N_m^* - 1) MC_{km} \right)}{\left(\sum_{n=1}^{N_m^*} MC_{mn} \right)^2}, \quad (6)$$

at industry price

$$P_m^* = \frac{1}{(N_m^* - 1)} \sum_{n=1}^{N_m^*} MC_{mn}, \quad (7)$$

and profits for firm k are given by

$$\pi_{km}^* = \max \left\{ \begin{array}{l} -f, \\ \beta_m Y \frac{\left(\sum_{n=1}^{N_m^*} MC_{mn} - (N_m^* - 1) MC_{km} \right)^2}{\left(\sum_{n=1}^{N_m^*} MC_{mn} \right)^2} - f \end{array} \right\}. \quad (8)$$

Firms choose to produce if:

$$\left(\sum_{n=1}^{N_m^*} MC_{mn} - (N_m^* - 1) MC_{mn} \right) \geq 0. \quad (9)$$

Equation (9) simply states that a firm will choose not to produce if their marginal costs are too high relative to its competitors.

The Cobb-Douglas specification simplifies solutions to the Cournot-Nash equilibrium tremendously. Specifically, it generates a Cournot solution for the intermediate goods firms in which profits are homogenous of degree zero in the vector of costs across firms. Thus, a proportional change in the vector of marginal costs leaves profits the same despite falling marginal costs. 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 is 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.⁵ 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 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, 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

⁵ 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, as noted in footnote 4, 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.

the market when the number of firms is small. Those incentives are discussed in the next section. 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.

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. A stochastic process governs success where the probability of innovation rises at a decreasing rate with the amount of investment.

We track the level of marginal costs by accounting for the number of innovations a firm has available at time t and denote it as i_{kmt} . The total number of innovations is the sum of the publicly available innovations for product m , labelled I_{mt} , and each firm's private innovations, ip_{kmt} ,

$$i_{kmt} = I_{mt} + ip_{kmt}. \tag{10}$$

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. In section 5, we explore how this diffusion rate alters the observed firm size distribution.

The constantly growing public stock of innovations allows potential entrants to remain viable. 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

A new entrant has access to all of the available public knowledge process innovations from the spillover process. In addition, entrants possess some private innovations developed earlier or during the construction phase of its manufacturing plant. 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:

$$0 < ip^{EN} < \tilde{ip}_m = \int \left(\sum_1^{N_m} \frac{1}{N_{mt}} ip_{nmt} \right) dt. \quad (11)$$

where ip^{EN} represents the number of private innovations of a new entrant. The left side implies that new firms are bringing some new ideas while the right-side, \tilde{ip}_m , is the equilibrium average (over the long-run) number of private innovations held by incumbent firms. If new firms entered 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. If new firms only entered with the public stock of knowledge, this would preclude them from ever being able to capture a large initial market share. What this allows for is the possibility that when incumbents fail to innovate, new firms have the opportunity to immediately establish themselves as the new leader. This possibility occurs only rarely. Most of the time new firms will enter with a small market share relative to existing incumbents.

The mapping from innovations to marginal costs is:

$$MC_{kmt} = \frac{1}{Z} \exp(-\eta i_{kmt}). \quad (12)$$

Marginal costs fall at the rate η with each additional innovation. Z is a scale parameter on costs

⁶ If all information were 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 prohibitively high.

which we use below to calibrate the model to match the mean size of firms measured by employees as reported in Hart and Oulton (1996). Z captures the unit labor costs of firms relative to the price of the final output good. Because all firms have access to the public stock of knowledge, what differentiates them in terms of marginal cost and market share is their private stocks of innovations. Firms with a greater number of privately held innovations enjoy a cost advantage over rivals. The cost advantage generates higher profits and the motive for engaging in process R&D.

This specification for the evolution of marginal costs and innovation has several notable features. First, process innovation will sustain growth. Marginal costs of production continually decline and go to zero as $t \rightarrow \infty$. Since it is the relative marginal costs that matter to firms' profits as shown in (8), the absolute level of the marginal costs (or total stock of innovations) is irrelevant to the decisions of a firm. For a given set of marginal costs across an industry, even when marginal costs are arbitrarily close to zero, firms will continue to acquire innovations to increase their profits by improving their *relative* position. 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.⁷

The stock of private innovations held by incumbent firm k increases through successful R&D.

The role of R&D is given by:

$$\dot{p}_{km,t+1} = \dot{p}_{kmt} + v_{kmt} \tag{13}$$

⁷ 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 not change the main results presented in the next section, but worthy of exploration in future work.

where:

$$\Pr(v_{kmt}) = \begin{cases} \frac{ax_{kmt}}{1+ax_{kmt}}, & \text{for } v_{kmt} = 1 \\ \frac{1}{1+ax_{kmt}} & \text{for } v_{kmt} = 0 \end{cases}. \quad (14)$$

x_{kmt} is the level of investment undertaken by a firm at time t . Note that the R&D production 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 products. 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 firm to extend its technological advantage and generate greater variance in firm sizes. We explore the consequences of higher and lower technological opportunity on 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

3.1.3 Dynamic Equilibrium

Let s_{nm} be the number of firms with i 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

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

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.

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:

$$V_{km}^I(ip_{km}, s_m) = \max \left\{ \phi, \pi(ip_{km}, s_m) - cx_{km} + \left(\frac{1}{1+r} \right) E [V^I(ip'_{km}, s'_m | x_m, ip_{km}, s_m)] \right\} \quad (15)$$

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_{km} 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, the current market structure, and the vector of investment $x_m = [x_{nm}]$ for all firms. $1/(1+r)$ 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 probability of all other firms being successful, and the probabilities of entry and exit. If the solution to the expected discounted profits is less than the liquidation value, ϕ , the firm chooses to exit the market and sell its capital. Accordingly, investment is zero for an exiting firm.

A potential entrepreneur may enter a product market and establish production and R&D facilities. If entry is optimal, the entrepreneur incurs a sunk entry cost. Production and sales do not begin until the following period. The Bellman equation resembles that for incumbents with

few changes:

$$V^{EN}(ip^{EN}, s_m) = \max \left\{ \psi, -\chi + \left(\frac{1}{1+r} \right) EV^I(ip'_{km}, s'_m | x_m, ip_{km}, s_m) \right\} \quad (16)$$

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.

The investment strategy of firms derives from the first order conditions on the above. Let $C_1(ip'_{km} + 1, s'_m)$ denote the expected value of the firm conditional on successful innovation and $C_2(ip'_{km}, s'_m)$ the expected value if it fails to innovate. We can then rewrite the Bellman equation for incumbents as:⁹

$$V_{km}^I(ip_{km}, s_m) = \max \left\{ \begin{array}{l} \phi, \pi(ip_{km}, s_m) - cx_{km} + \\ \left(\frac{1}{1+r} \right) \left[\frac{ax_{km}}{1+ax_{km}} C_1(ip'_{km} + 1, s'_m) + \frac{1}{1+ax_{km}} C_2(ip'_{km}, s'_m) \right] \end{array} \right\} \quad (17)$$

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

$$x_{km}(ip_{km}, s_m) = \max \left\{ 0, \frac{-1 + \sqrt{\frac{a(C_1 - C_2)}{(1+r)c}}}{a} \right\}. \quad (18)$$

Since investment cannot be negative, the firm chooses the value maximizing level of R&D investment subject to a non-negativity constraint. 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 itself rises it increases the probability of successful R&D and, hence, a higher value. That imparts an incentive to increase investment. However, the higher the level of a the less investment is required to achieve any given probability of success reducing the incentive to invest.

The numerical algorithm employed uses value function iteration to solve for the space of values given by all possible combinations of firms and private innovations. From the value function,

⁹ We only present the investment function for the incumbent to save on space.

the policy functions, including the entry and exit decisions, are fairly straightforward to extract. From the solutions, we can simulate our product markets and industry, comparing the results with the empirical literature.

4 Firm Size Distribution

4.1 The Ergodic Distribution

In our initial runs of the model, we specified five product markets with different levels of demand. The state space constraints we use have a maximum of six firms per product market 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 firms are in the market they are making negative profits. For the maximum number of private innovations, we check in the simulations whether any firm attempts to obtain more private innovations than exist in the state space and make adjustments accordingly.

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
Liquidation Value	ϕ	0.1
Outside Alternative Value	ψ	0.2
Rate of Decrease in Marginal Costs	η	0.05
Unit Cost of Labor	Z	127.65

The choice of parameters is as follows: We set the discount rate to $1/1.08$ as an approximation of the average cost of capital for firms. 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. Ex post, we find that the liquidation value is about 7.5% of the average firm value, while the opportunity cost is roughly 15% of average firm value. We calibrate the unit cost of production parameter, Z , to match the mean log size as measured by employment reported in Hart and Oulton (1996). The other parameters including the productivity of R&D investment, a , sunk costs of entry, X_e , and the rate of decrease in marginal costs, η are set to 3, 0.2, and 5% respectively and the values follow Laincz (2004a). These parameters, and the rate of knowledge diffusion, will be allowed to vary in the following section to determine their impact on the firm size distribution.

In our first comparison of the model with the data, we compute the ergodic distribution by simulating the model.¹⁰ 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. 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.

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 database

¹⁰ The simulation runs the model for 25,000 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.

that includes very small firms in the sample. 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 normal distribution. We report analogous measures based on our model. Sales are easily 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.

TABLE 2: Distribution Statistics¹¹

Measure	Normal	H&O	Model	H&O	Model	H&O	Model
	Distribution	Emp.	Emp.	Sales	Sales	Net Assets	Net Assets
Mean	-	3.1582	3.1582	7.2015	5.1638	5.5539	5.0285
Std. Dev.	-	1.5197	0.4865	1.6628	0.5052	1.9635	1.3010
Skewness	0	0.7487	0.3912	0.1932	0.1079	0.4366	0.8942
Kurtosis	3	4.5794	3.9685	6.1876	4.3851	4.835	3.1772

The model does reasonably well in matching Hart and Oulton’s data. However, 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. The normal distribution, which would result from Gibrat’s Law, would yield a skewness measure of 0 and kurtosis of 3. The skewness coefficients in both the data and the model indicate a long-right tail and the model produces a distribution more leptokurtic than the normal in line with the empirical results.

Figures 5-7 show the distributions in levels (a) and in logs (b). The distributions in levels, for all the three proxies considered, exhibit longer right tails, especially for the net assets distribution. The size range accumulating the higher probability mass lies significantly 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

¹¹ H&O refers to the results reported in Hart and Oulton (1996). Note that they report the kurtosis in the original paper after subtracting off three such that the normal distribution would generate a kurtosis measure of zero. Here we report the fourth moment without subtracting off three.

clearly not bell shaped, but rather appear to be slightly skewed to the right and exhibit thicker tails and higher peaks than the standard normal. The distribution of the log of employment exhibits less variance as indicated in Table 2, but shows some skewness and leptokurtosis.

4.2 Cross-Sectional Growth-Size Properties

While the above shows that the ergodic distribution of the models reasonably match the observed data in terms of deviations from Gibrat’s Law, we now turn to examining the growth-size relationships of the simulated model. 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.¹² That provides us with simulated panel data to test the growth-size relationship. Overall, the simulation produces 137 firms over the last two periods. Of those, 11 exited immediately, 22 new firms entered in the first period, 23 exited the following period, with 12 more entries in the subsequent period. The average entry and exit rates were 0.214 and 0.182 respectively. That leaves 69 firms that produced positive output in both periods. Table 3 provides sample statistics for the data in the initial period on mean size, variance in size, entry, and exit. The measures are largely the same as in Table 2, allow the kurtosis measure for employment is exceptionally high. If we were to simulate repeated times and take averages the number would undoubtedly come down, though part of the increase follows from removing new entrants and exiting firms to generate a balanced panel.

TABLE 3: Summary Statistics in Natural Logs

¹² 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.

Size Measure	Employment.	Sales	Net Assets
Observations	69	69	69
Average	3.558	5.098	5.066
Variance	0.238	0.141	1.485
Skewness	2.517	0.443	1.246
Kurtosis	12.190	3.543	4.925

Table 4 provides the results of the regressions of the following form:

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

where y_t is the log of the various size measures. We report the results using robust standard errors, but even without using them the results are hardly changed. The final column shows the t-statistic for the null hypothesis that $\beta = 1$.

TABLE 4: Regression Results

Firms	Size Measure	Obs.	Coefficient	Std. Error	R^2	$H_0 : \beta = 1$
All	Employment	69	0.969	0.008	0.99	14.28***
	Sales	69	0.964	0.006	0.99	37.00***
	Net Assets	69	0.933	0.014	0.98	21.94***
Large	Employment	36 ¹³	0.993	0.008	0.99	2.17
	Sales	35	0.968	0.006	0.99	24.36***
	Net Assets	35	0.944	0.016	0.99	11.83***
Small	Employment	33	0.937	0.013	0.99	23.26***
	Sales	34	0.956	0.011	0.99	15.14***
	Net Assets	34	0.911	0.027	0.97	11.24***

Note: * Significant at the 10% level; ** for the 5% level; and *** for the 1% level.

¹³ For employment, several firms were exactly at the median. We report the results with those firms included in the large category. Changing their designation to small does not alter the results.

Hall's (1987) estimates for β as applied to employment across three different samples were consistently 0.99 and significantly less than one.¹⁴ Evans (1987) on the other hand finds 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'. The R-squared's are exceedingly high. 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. Over more extended periods of time we would expect this relationship to weaken for the model because of the Markov Perfect nature of the equilibrium. Yet, the main result here is consistent with what the empirical literature finds in that there is a negative growth-size relationship.

In the bottom portion of the table we split our sample between large and small firms defined as above or equal to the median of the sample or below it. Empirically we have seen that Gibrat's Law works better for the large firms. Our results show a similar pattern. The estimated β 's are consistently higher for the large firm sample. In fact, for the employment size measure we cannot reject equality with one 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. For employment we reject equality at the 1% significance level, but we cannot reject equality for the other two measures.

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 observations and after one more period 23 exited, leaving only 46 remaining firms. The specification is similar to Kumar (1985) where growth is the dependent variable (instead of the log of size):

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

Persistence in the shocks will show up as a positive value for γ , while the coefficient on the log of the previous size corresponds to $1 + \beta$ in the previous regression. We find that the coefficients

¹⁴ In addition to rejecting $\beta = 1$ for employment, we also tested $\beta = 0.990$, the smallest value reported in Hall (Table II). We reject the null with a t-statistic of 7.29.

for β and γ , even for the small sample, are significant at the 5% level (though the former just barely). Their estimated values are -0.0157 (or 0.9843 for comparison with Table 4) and 1.257 for the latter.

The positive 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 rivals 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 decisions. These processes of firm dynamics effectively embed serial correlation in error terms that cannot account 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 R&D expenditure and innovations of existing and rival firms are included. We leave this hypothesis for future empirical work.

One note on the magnitude of serial correlation is required here. Our estimate of γ is an order of magnitude larger than that found in either Singh and Whittington (1975) or Kumar (1985) who find values of approximately 0.3 and 0.12 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 one year based on the user cost of capital we specify. Because we know that the model will predict serial correlation that declines over time, again 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.

Finally, we look at the variance in growth rates across firm sizes. Again, we separate our simulated sample by the median. Table 5 shows the standard deviations in growth rates for the total sample and large and small firms according to the three size measures. In all three cases the variance in the growth rate of the small firms is larger than that of the large firms. The final

column reports the F-statistic for the variance ratio test for equality of the standard deviations. We can reject equality at the 5% level based on the employment measure and at the 1% level for sales, but we fail to reject equality for measuring size by firm 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. Naturally, we could produce more simulations and increase the size of our samples to ensure we reject equality for the standard deviation in the growth rate of net assets, but that seems overkill to us at this point.

TABLE 5: Tests of Standard Deviation of Growth Rates by Size Class

Size Measure	Obs.	Total	Large	Small	$H_0 : \sigma_1 = \sigma_2$
Employment	36/33	0.251	0.192	0.288	2.25**
Sales	35/34	0.270	0.204	0.326	2.55***
Net Assets	35/34	0.630	0.596	0.671	1.27

Note: * Significant at the 10% level; ** for the 5% level; and *** for the 1% level.

To summarize the section, we find that the model is able to replicate the empirical studies 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.

5 Variation in the Firm Size Distribution

This section is in progress. 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 will vary the following parameters: sunk costs, fixed costs,

productivity of R&D, rate of spillovers, and the rate of decline in marginal costs. Of interest is whether the variations the model generates are in accord with cross-industry variation uncovered in the empirical literature.

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 can replicate the firm size distributions reported in the literature. In addition, it can be used to replicate the growth-size relationships of more recent interest.

The remaining work, in this preliminary and incomplete version, goes after the main question. We intend to examine how variations in critical parameters such as sunk costs of entry and technological opportunity affect the firm size distribution and compare those results with the empirical results on cross-industry differences.

We should also note some missing elements in our framework that could be incorporated in future work. We have already mentioned that the model could be extended to allow for strategic interaction across products in addition to the within product market strategic interaction. We doubt such an extension would significantly change the results, but that remains to be seen. Secondly, merger activity is one of the major concerns in the empirical literature on the firm size distribution (for example see Kumar, 1985, and Dunne and Hughes, 1994). Our model can be extended to handle mergers by combining it with Gowrisankaran's (1999) work on merger activity in the context of the Ericson-Pakes framework. Finally, most empirical work refers to the model of Jovanovic (1982) as a theoretical base. It features a passive learning process wherein each firm is uncertain about its true costs leading to a Bayesian learning behavior over time and endogenous exit. The model here relies on R&D success and diffusion of knowledge to generate entry, growth, survival, and exit. Clearly, there are other elements of uncertainty in the real world, especially

with regards to new firms. More could be done to capture the risks that entrepreneurs face such as uncertainty of true costs as in Jovanovic. That would allow us to explore how the rise of venture capital and lowering of entry barriers, other than the sunk costs discussed here, affect the firm size distribution and its evolution.

References

- Audretsch, D., L. Kompf, E. Santarelli, and A. Thurik, 2004, "Gibrat's Law: are the Services Different?" forthcoming in *Review of Industrial Organization*.
- 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. Oulton, 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.
- Jovanovic, Boyan, 1982. Selection and the Evolution of Industry. *Econometrica*, 50, 3, pp. 649-670.
- 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?" Forthcoming *Journal of Economic Dynamics and Control*.
- Laincz, C., 2004b, "A Model of Growth with Dynamic Market Structure" 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.
- 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.

Figure 1: Simulation of Gibrat's Law
Beta = 1

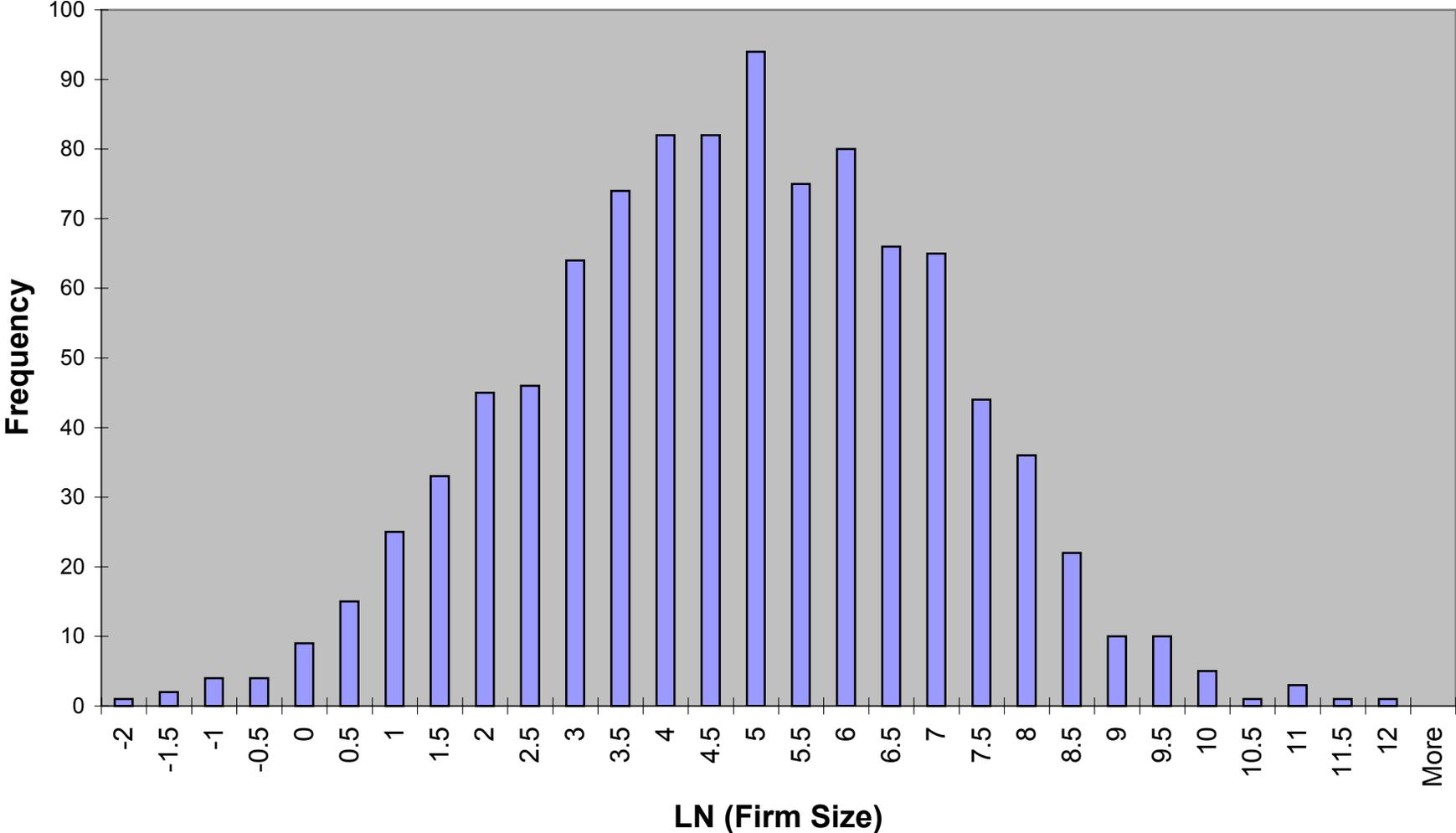


Figure 2: Strong Reversion to the Mean
Beta = 0.75

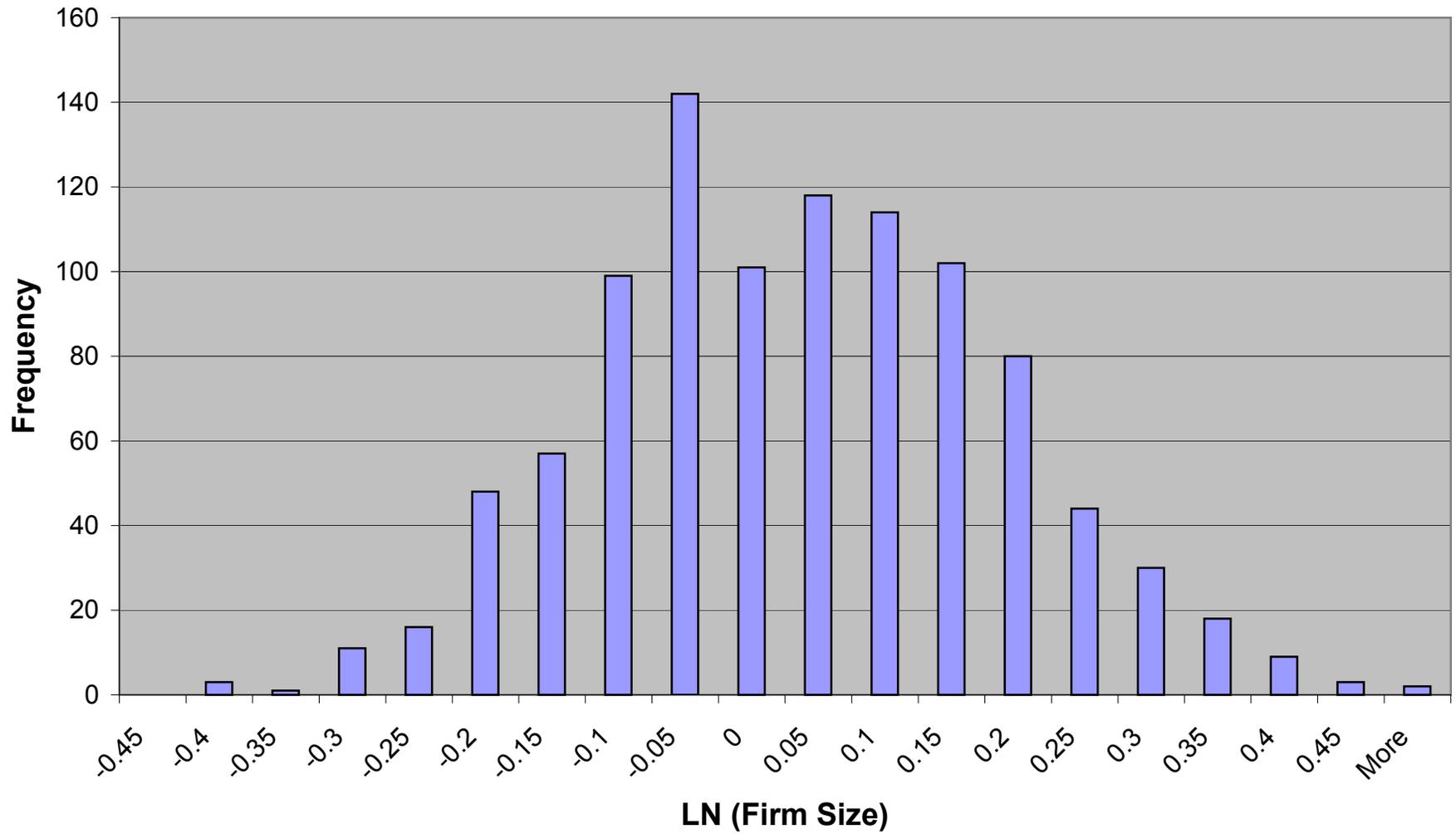


Figure 3: Weak Reversion to the Mean
Beta = 0.95

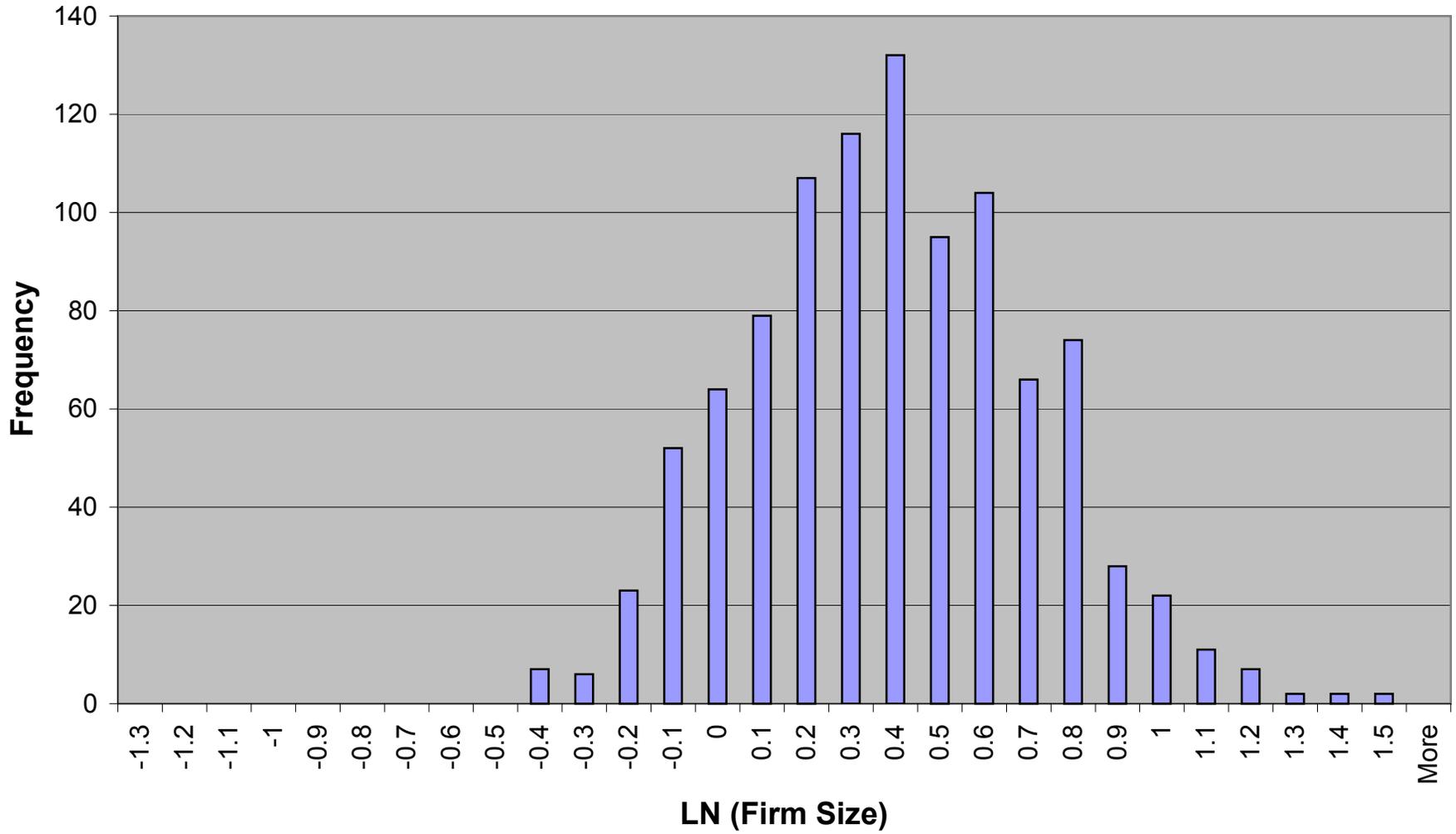


Figure 4: Diversion from the Mean
Beta = 1.05

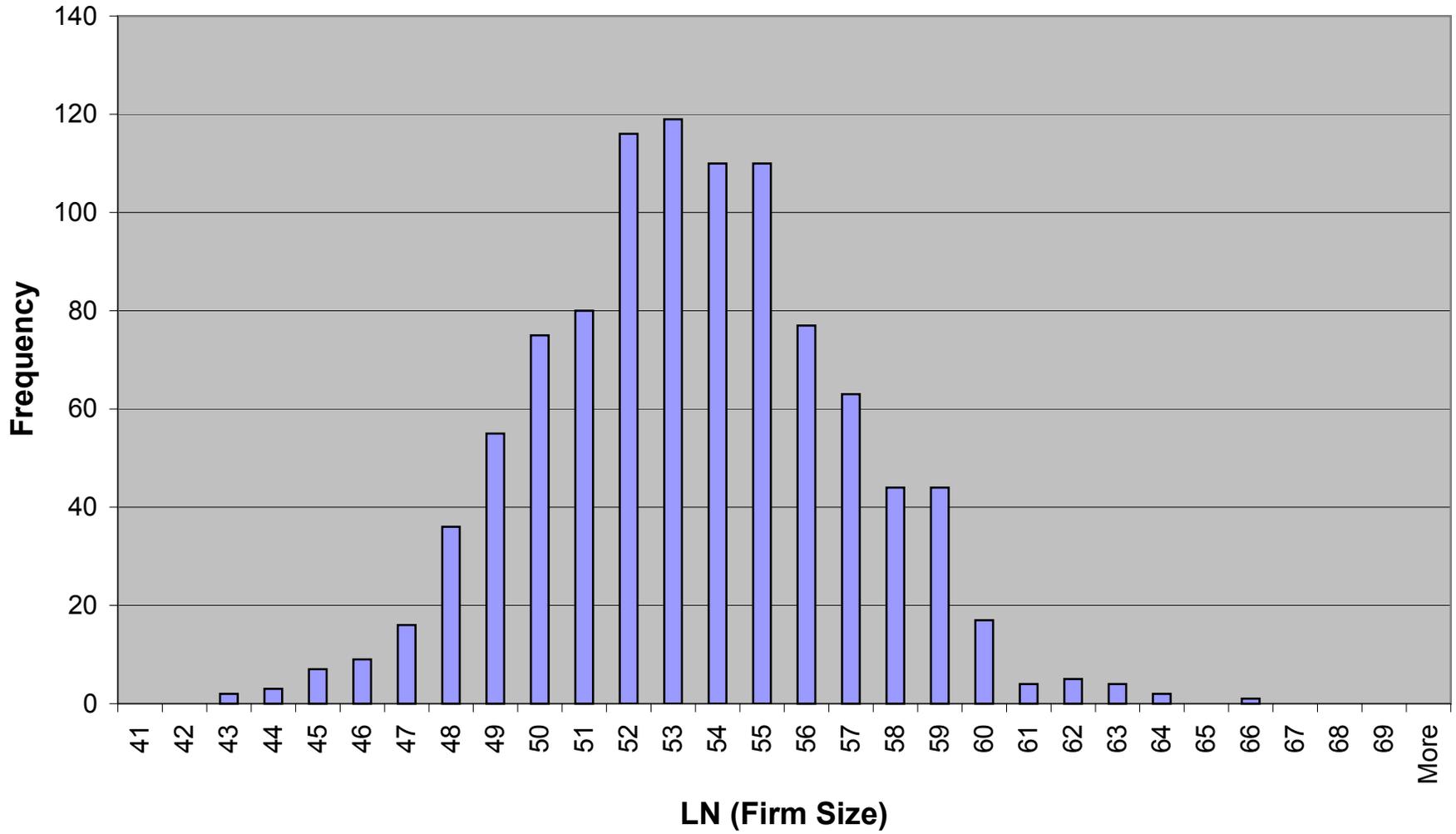


Figure 5a: Employment Distribution

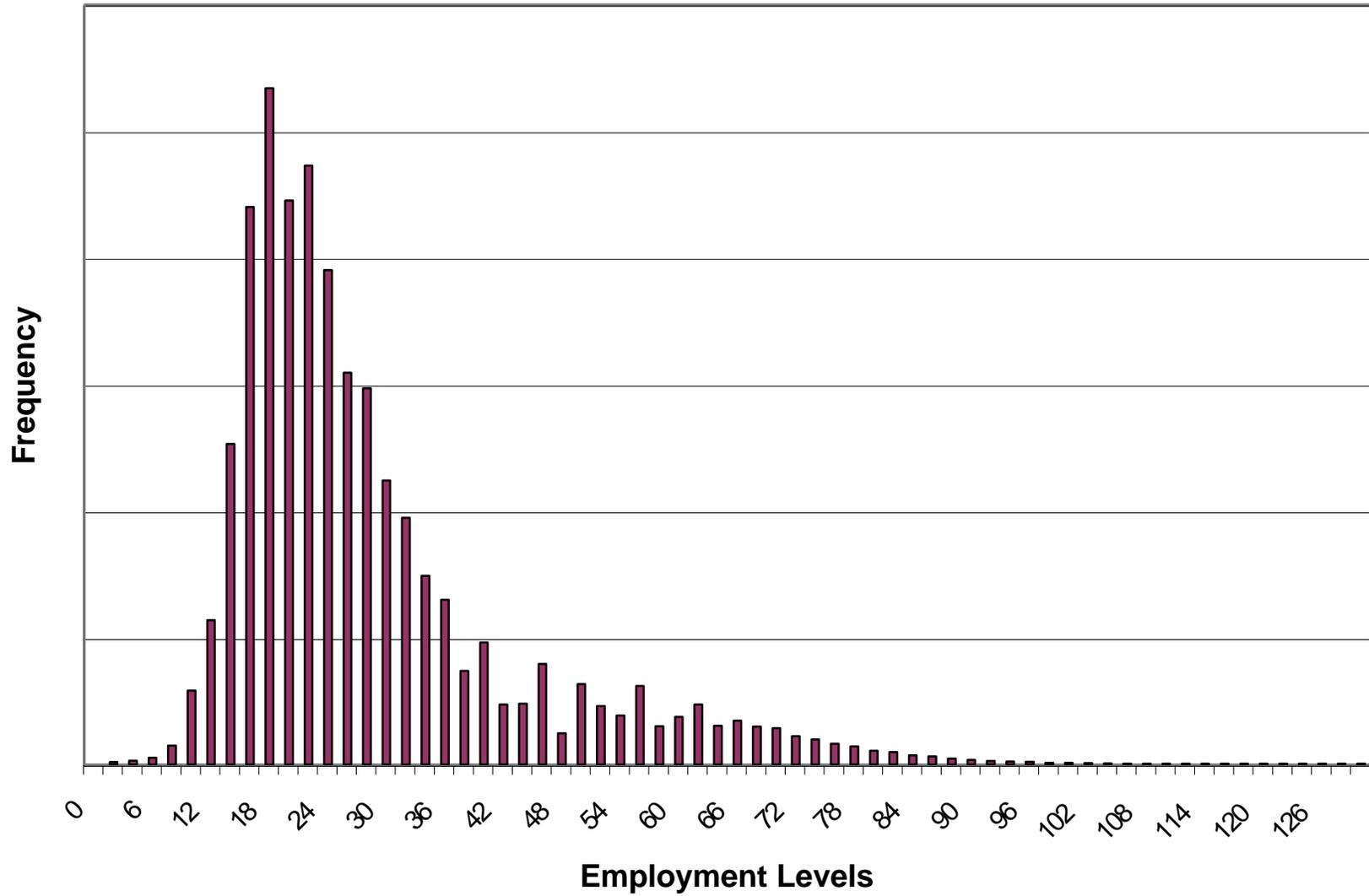


Figure 5b: Log Employment Distribution

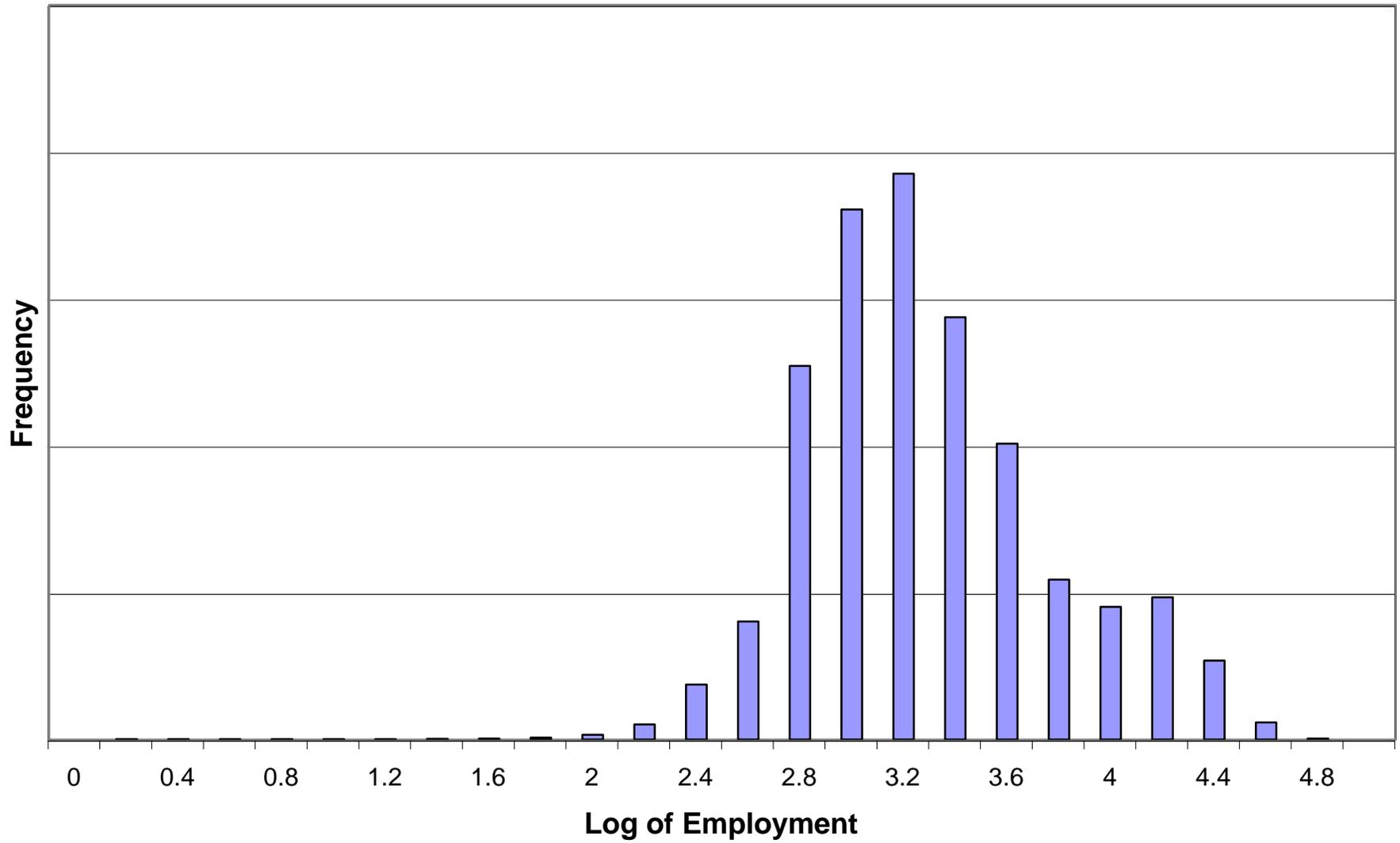


Figure 6a: Sales Distribution

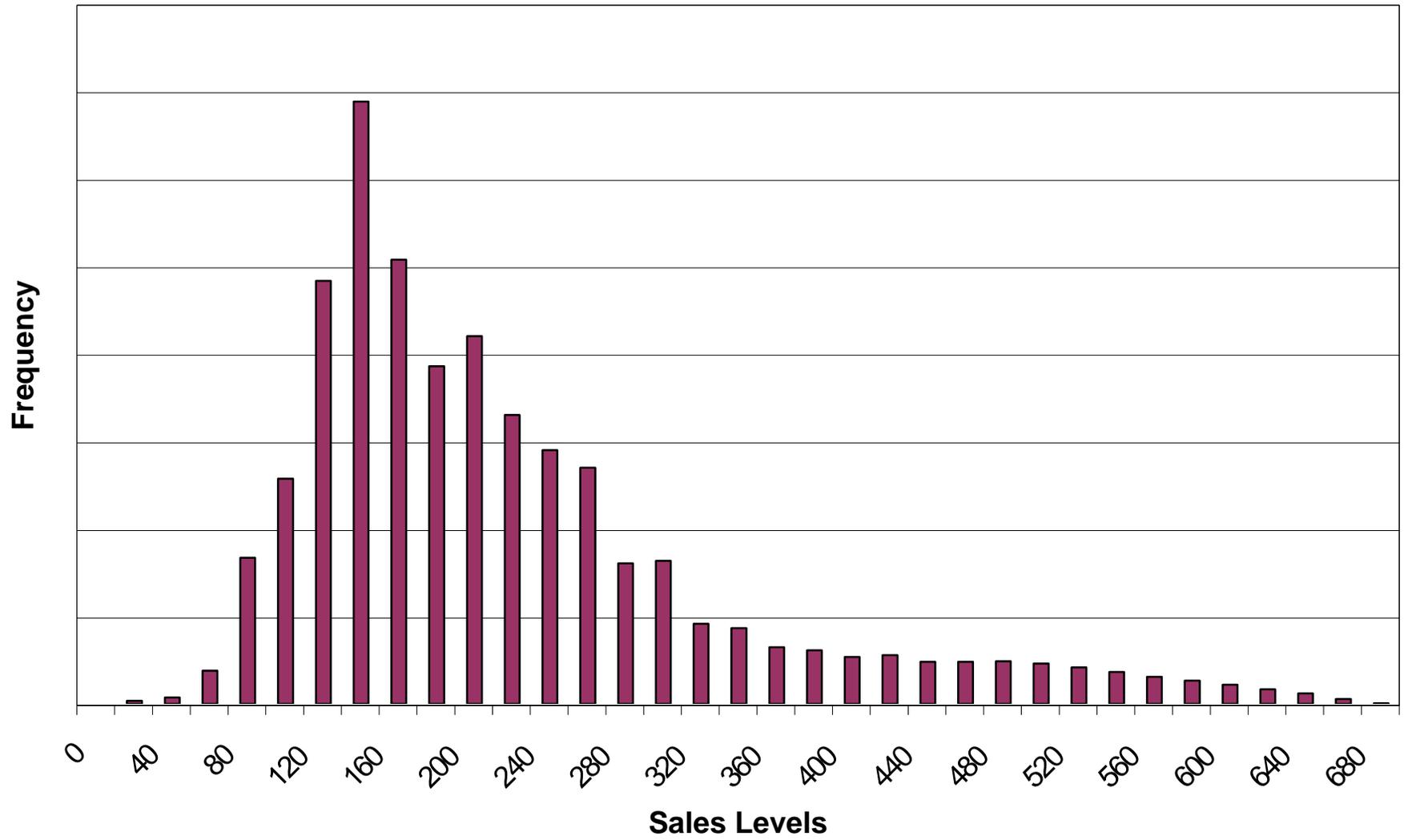


Figure 6b: Log Sales Distribution

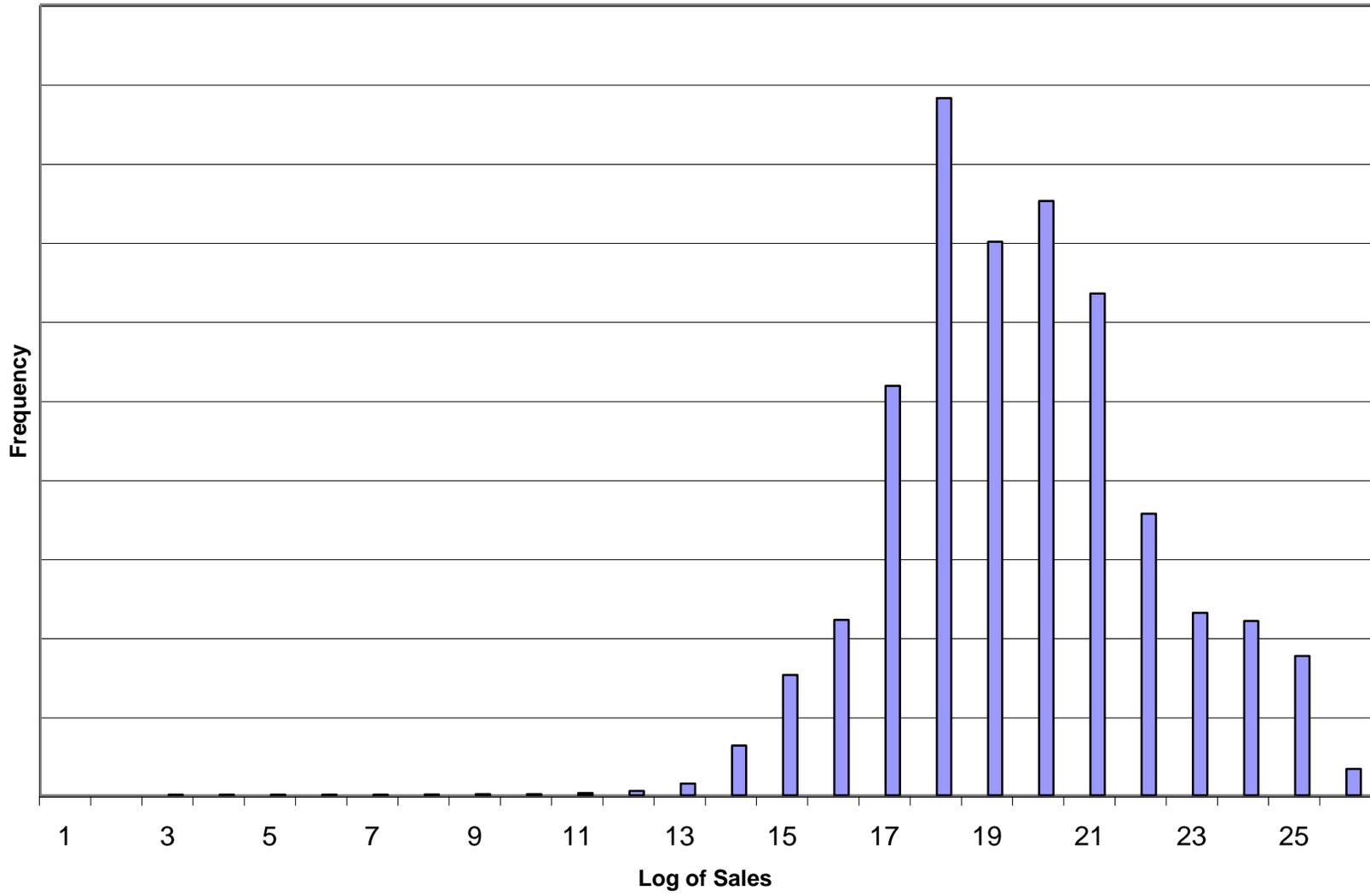


Figure 7a: Values Distribution

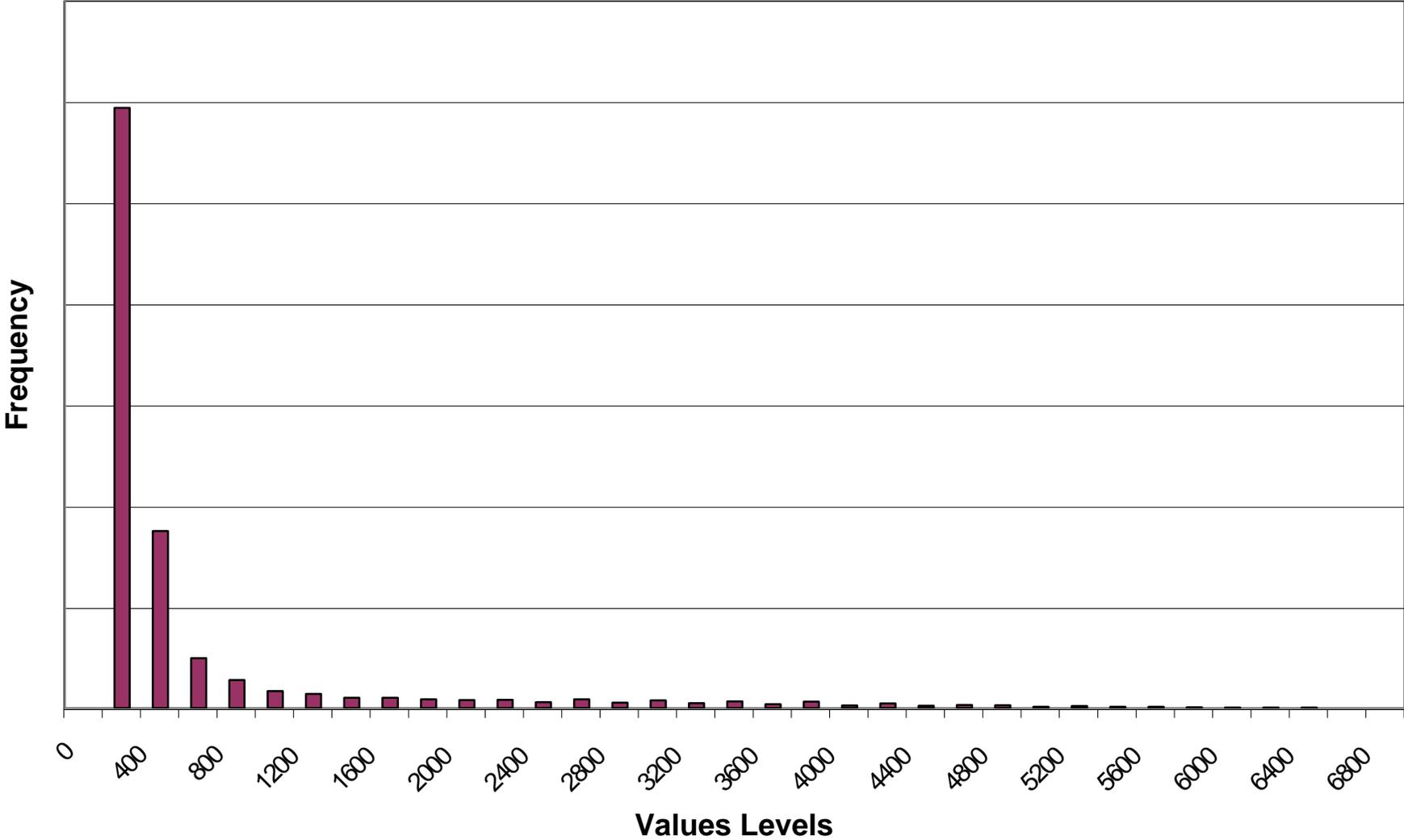


Figure 7b: Log Values Distribution

