A Dynamic Model of Size Distribution of Firms Applied to U.S. Biotechnology and Trucking Industries

Fariba Hashemi

ABSTRACT. The present paper attempts to contribute to the existing literature on industry dynamics by proposing a tractable structure for the analysis of the dynamic process governing the size distribution of firms. An analytical model is proposed which describes the density of the cross-sectional distribution of firm size within an industry. The model is based on the theory of diffusion processes, and the method illustrates how information on the time-evolution of size distribution of firms over an extended period of time can be used to make inferences about an underlying process. An empirical application to the evolution of size distribution of population of firms in *(i)* the U.S. biotechnology industry, and *(ii)* the U.S. interstate for-hire trucking industry illustrates the applicability of the proposed model in industry studies.

1. Introduction

This paper attempts to contribute to the existing literature on industry dynamics by proposing a tractable structure for the analysis of the dynamic process governing the size distribution of firms. The analysis builds on earlier literature on firm size and the distribution of firm size within an industry,[1] and focuses on the *dynamics* of the *distribution of firm size*.

An analytical model is presented which describes the *density* of cross-sectional *distribution* of firm size within an industry. The model is based on the theory of diffusion processes, and the method illustrates how information on the time-evolution of the size distribution of firms over an extended period of time can be used to make inferences about an underlying process.

The study is motivated by the observation that the distribution of size of firms within an industry varies as a function of time. It would be reasonable to assume that there exists an equilibrium distribution of firm sizes with a certain mean and variance, towards which the ensemble of firms considered tend to converge. The present study makes an attempt to model the time evolution of the size distribution of firms towards this equilibrium.

The dynamics of the proposed model rely on two opposing forces: (*i*) a mean reversion process which is meant to describe supply and demand forces which tend to concentrate the distribution, and (*ii*) a counteracting random (or diffusion) process which is meant to describe the effect of firm's search and learning behavior, as well as bounded rationality on the size distribution of firms.

The proposed model is applied to the evolution of the size distribution of population of firms in (*i*) the U.S. biotechnology industry between the years of 1989–1996, and (*ii*) the U.S. interstate for-hire trucking industry between the years of 1977–1991. An empirical analysis estimates the parameters of the proposed analytical model using non-linear least squares procedure. The estimated results are discussed and the usefulness of the model in estimating (*i*) the final equilibrium distribution as well as (*ii*) the rate of progress toward this equilibrium is illustrated.

The paper is organized as follows. Section 2 develops and provides an analysis of the model. Section 3 reports some empirical results. Section 4 concludes and explores ways in which the model might fruitfully be elaborated.

Final version accepted on November 11, 2001

Swiss Federal Institute of Technology
Lausanne, Switzerland
E-mail: fariba.hashemi@epfl.ch

Small Business Economics **21**: 27–36, 2003.

2. Modeling the dynamics of firm size distribution

2.1. *Basic assumptions*

Consider an industry consisting of a constant number of firms with different sizes. This set of firms forms a distribution which evolves over time. It is assumed that there exists an equilibrium distribution of firm sizes with a certain unknown mean and variance. At any moment in time, a discrepancy exists between the current actual size distribution of firms and the equilibrium distribution. Firms find themselves in a dynamic adjustment process toward such an equilibrium, and the evolution of the size distribution reflects the convergence towards this equilibrium.

2.2. *Development of the model*

Let $n(s, t)$ be the measure of firms of size s at time t, where s measures some relevant aspect of size in logarithms. The total number of firms between s_1 and s_2 is

$$N(s_1, s_2, t) = \int_{s_1}^{s_2} n(s, t)\, \mathrm{d}s. \tag{1}$$

Set

$$N_0 = \int_{-\infty}^{\infty} n(s, t)\, \mathrm{d}s < \infty$$

where N_0 is the total number of firms which is constant, so firms are neither created nor destroyed and

$$\frac{\mathrm{d}N}{\mathrm{d}t}(s_1, s_2, t) = Q(s_1, t) - Q(s_2, t) \tag{2}$$

where $Q(s, t)$ is the flux of firms of size s at time t. Hence $Q(s_1, t) - Q(s_2, t)$ is a measure of the number of firms entering or leaving the size interval $[s_1, s_2]$ (A complete explanation of Q will follow in equations (8) to (13) below).

Let us assume that $N(s, t)$ is sufficiently regular, e.g. $N(s, .)$ is continuously differentiable on $(0, \infty)$ and $N(., t)$ is twice continuously differentiable on $(-\infty, \infty)$. It follows from (1) that

$$\frac{\mathrm{d}N}{\mathrm{d}t}(s_1, s_2, t) = \int_{s_1}^{s_2} \frac{\partial n}{\partial t}(s, t)\, \mathrm{d}s. \tag{3}$$

Thus we get

$$\int_{s_1}^{s_2} \frac{\partial n}{\partial t}(s, t)\, \mathrm{d}s + Q(s_2, t) - Q(s_1, t) = 0. \tag{4}$$

It follows from the fundamental theorem of calculus that

$$Q(s_2, t) - Q(s_1, t) = \int_{s_1}^{s_2} \frac{\partial Q}{\partial s}(s, t)\, \mathrm{d}s. \tag{5}$$

Thus

$$\int_{s_1}^{s_2} \left(\frac{\partial n}{\partial t} + \frac{\partial Q}{\partial s} \right) \mathrm{d}s = 0. \tag{6}$$

Since this equation is true for all s_1 and s_2 it follows

$$\frac{\partial n}{\partial t} + \frac{\partial Q}{\partial s} = 0.$$

So let

$$f(s, t) = \frac{n(s, t)}{N_0}$$

denote the probability density of firms. It then follows that

$$\frac{\partial f}{\partial t} + \frac{\partial q}{\partial s} = 0 \quad \text{where} \quad q = \frac{Q}{N_0}. \tag{7}$$

This is the basic conservation law and q is a flux of probability (which can be interpreted as the variation, or number of firms entering and number of firms exiting a size interval).

In this paper it is assumed that the flux is made of two different parts: a *drift* q_c and some *diffusion* (*dispersion*) q_d. *Drift* describes supply and demand forces at work, and *dispersion* describes random processes. Thus,

$$q = q_c + q_d. \tag{8}$$

More precisely, for the *drift flux*, it is assumed that there exists some equilibrium distribution of firm sizes with a certain mean and variance, towards which the ensemble of firms tend. i.e., a flux towards the equilibrium distribution. Let $u > 0$ be the mean of this distribution.

To illustrate, consider an industry in which average costs of producing an amount x of output are a non-increasing function of the size of the firm, for a given quality of output (each firm may have significant fixed costs, and marginal costs may essentially be constant). Furthermore, consumers prefer small firms for perceived higher quality of service. Under these assumptions, there would exist an equilibrium distribution of firm sizes with a certain mean and variance, determined by the tension between producer and consumer preferences regarding firm size. The equilibrium trades off productive efficiency against consumer preferences.

The term q_c measures the portion of the function f transported by the drift velocity

$$v(s) = \lambda(u - s) \tag{9}$$

where $\lambda > 0$ is a fixed parameter. Thus,

$$q_c = vf \tag{10}$$

i.e.

$$q_c(s, t) = \lambda(u - s)f(s, t). \tag{11}$$

For the *diffusion flux*, although an equilibrium distribution exists, this equilibrium is assumed uncertain from the point of view of the firm. Moreover, owing to individuals' imperfect information and limited computational capacity, bounded rationality is assumed (Nelson and Winter, 1982; Simon, 1955). Boundedly rational firms use various processes to arrive at the equilibrium distribution, and they may switch between processes over time. Firms follow no precise law to arrive at this optimum, they perform *local search* (Nelson and Winter, 1982), and learn by means of *imitation* and *trial and error* (Alchian, 1950). Organizational change is the product of searches for new practices in the neighborhood of an organization's existing routines (Cyert and March, 1963; Stuart and Podolny, 1996), within a setting in which patterns of behavior stabilize as formal structures and routines become institutionalized over time (Cyert and March, 1963). Moreover, long-established competencies stand in the way of adaptation to major changes in technological regimes (Abernathy and Utterback, 1978; Burgelman, 1994; Henderson and Clark, 1990; Tushman and Anderson, 1986). This search process generates randomness in the system.

Random effects therefore tend to cause a flux from regions of low probability to high probability. The simplest choice is Fick's law (Okubo, 1980):

$$q_d(s, t) = -\varepsilon \frac{\partial f}{\partial s}(s, t) \tag{12}$$

where $\varepsilon > 0$ is a constant diffusion parameter. Thus we have

$$q = \lambda(u - s)f - \varepsilon \frac{\partial f}{\partial s} \tag{13}$$

so

$$\frac{\partial f}{\partial t} + \frac{\partial}{\partial s}(\lambda(u - s)f) = \varepsilon \frac{\partial^2 f}{\partial s^2}. \tag{14}$$

This is our equation for the density of firm size distribution, where f measures density and s measures some relevant aspect of size in logarithms. The model is based on a drift-diffusion continuous process obtained from an Ornstein-Uhlenbeck one. The partial differential equation in this paper can also be regarded as the expectation of a classical linear stochastic differential equation.[2] This type of model has been widely used in Biomathematics (see for example Okubo, 1980; Ricciardi, 1977 and Taylor, 1984).

Initial conditions in time and boundary conditions with respect to s are needed to completely formulate the model. In order to allow an analytic solution to the model equation, I impose the initial condition:

$$f(s, 0) = f_0(s) = Ne^{\frac{-(s-u_0)^0}{2\sigma_0^2}} \tag{15}$$

on $s \in (-\infty, \infty)$, where N is the normalization constant, $N = 1/\sigma_0\sqrt{2\pi}$.

2.3. *Analysis of the model*

In order to describe the time-behavior of the proposed model, the model is solved for its analytic solution $f(s, t)$, describing the evolution of the distribution through time. There exists a unique analytic solution for this model. The expression representing the time-development of the distribution is:

$$\begin{aligned} f(s, t) &= Ne^{\lambda t}\sqrt{\frac{\alpha}{\alpha + \beta}}\, e^{-\frac{((s-u)e^{\lambda t}+u-u_0)^2}{2\sigma_0^2+\frac{2\varepsilon}{\lambda}(e^{2\lambda t}-1)}} \\ &= Ne^{\lambda t}\sqrt{\frac{\alpha}{\alpha + \beta}}\, e^{\frac{(s-u_t)^2}{2\sigma_t^2}} \end{aligned} \tag{16}$$

where

$$a = \frac{\sigma_0^2}{2}$$

$$\beta = \frac{\varepsilon}{2\lambda}(e^{2\lambda t} - 1)$$

$$u_t = E[f]_t = u(1 - e^{-\lambda t}) + u_0 e^{-\lambda t}$$

$$\sigma_t^2 = \sigma_0^2 e^{-2\lambda t} \frac{\varepsilon}{\lambda}(1 - e^{-2\lambda t})$$

u_0 represents the initial mean size, σ_0^2 represents the initial variance of the distribution, and N is the normalization constant: $N = 1/\sigma_0\sqrt{2\pi}$.

3. Empirical application

3.1. *Data*

The empirical analysis applies the model to the log size distribution of population of firms in two different industries. The first data comes from an industry which is at a relatively early stage of its development: the U.S. biotechnology industry between the years of 1989–1996. The data set describes a total of 250 biotechnology firms specializing in the development of human diagnostics and therapeutics. Observations were available annually.[3] Figure 1 provides a description of the distribution of firm size in the data, where our measure of firm size is the total number of employees. The vertical axis measures firm size in logarithms, and the horizontal axis measures time in years. The solid line represents the mean size of the industry and the dotted lines one standard deviation around the mean.

The second data describes a relatively mature industry: the U.S. interstate for-hire trucking industry between the years of 1977–1991. Observations were available annually,[4] and the sample is restricted to firms with annual revenues exceeding \$1 million.[5] Figure 2 provides a description of the evolution of the distribution of firm size in this data, where annual revenue is used as proxy for size. The vertical axis measures firm size in logarithms, and the horizontal axis measures time in years. Once again, the solid line represents the mean size of the industry and the dotted lines one standard deviation around the mean.

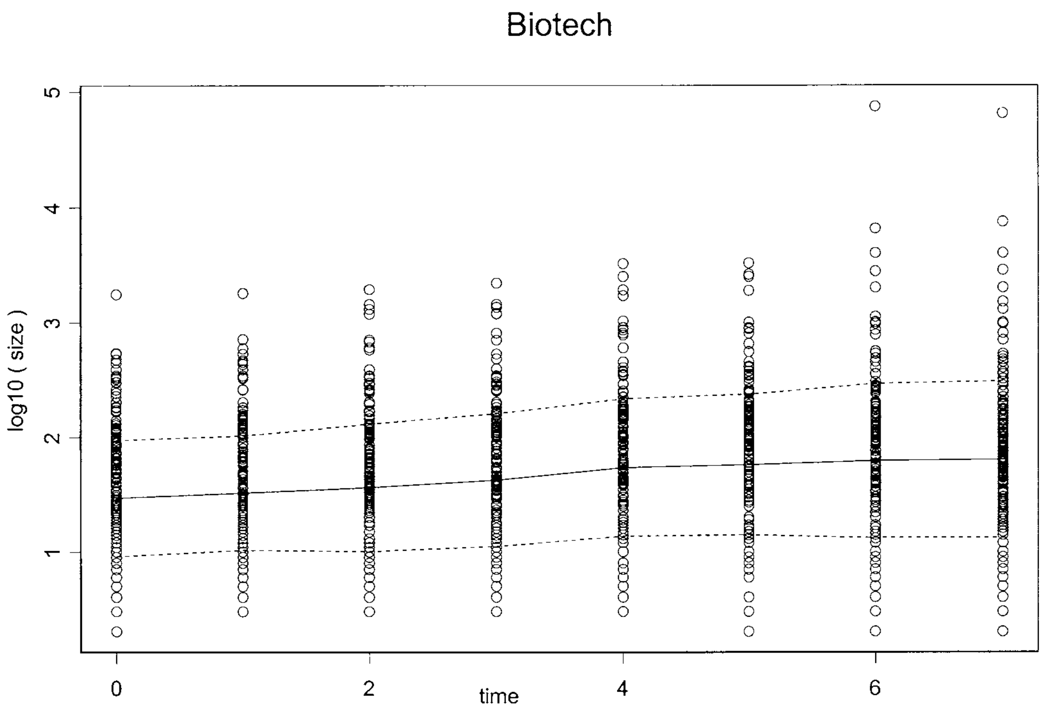

Figure 1.

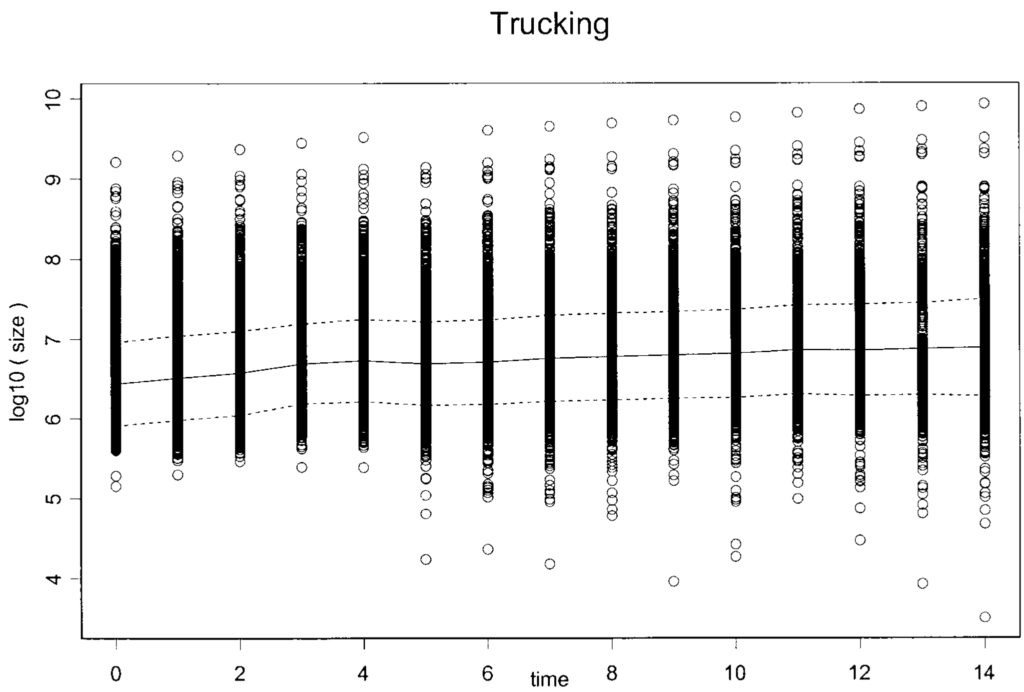

Figure 2.

3.2. *Method of estimation*

A second order partial differential equation model has been proposed to express the dynamics of firm size distribution. The model has five parameters: u_0, u, ε, σ_0^2, and λ. u_0 denotes the initial mean of the size distribution, and u denotes where the initial mean is heading. σ_0 is the standard deviation at time zero, ε represents the diffusion parameter, and λ determines the rate of convergence.

The expectation of the distribution shown in equation (16) is the first moment, expressed by:

$$u_t = u(1 - e^{-\lambda t}) + u_0 e^{-\lambda t}. \tag{17}$$

Moreover, using the expression for the second moment of the distribution, the variance of the distribution, σ_t^2, can be expressed by:

$$\sigma_t^2 = \sigma_0^2 e^{-2\lambda t} + \frac{\varepsilon}{\lambda}(1 - e^{-2\lambda t}). \tag{18}$$

To examine the behaviour of f in equation (16) as $t \to \infty$, one observes that $f(s, t) \to f_\infty(s)$ as $t \to \infty$, where

$$f(s, t) \to_{t\to\infty} N\sqrt{\frac{2a\lambda}{\varepsilon}}\, e^{\frac{(s-u)^2}{2\frac{\varepsilon}{\lambda}}} = f_\infty(s) \tag{19}$$

and that $\lim_{t\to\infty} E[f]_t = u$. Moreover, the diffusive limit, i.e., the limit as $t \to \infty$ of the variance is:[6]

$$\lim_{t\to\infty} \sigma_t^2 = \varepsilon/\lambda.$$

The model has been applied to the log size distribution of the two populations as a function of time, using the non-linear least square procedure in S-Plus, and using a two-steps procedure. First the values for u_0, u, and λ were estimated using equation (17):

$$u_t = u(1 - e^{-\lambda t}) + u_0 e^{-\lambda t}.$$

In the second step, the values for ε and σ_0 were computed using equation (18):

$$\sigma_t^2 = \sigma_0^2 e^{-2\lambda t} + \frac{\varepsilon}{\lambda}(1 - e^{-2\lambda t}).$$

3.3. *Estimation results and model checks*

Table I reports estimates for the five model parameters u_0, u, ε, σ_0, and λ, along with the standard errors and t-values for the biotechnology industry.

The value for the speed of convergence λ, is small but positive as expected. Furthermore, the

TABLE I
Parameter estimates for biotechnology industry

Parameter	Value	Std. error	T-value
λ	0.0946157	0.119884	0.78923
μ	2.22727	0.724686	3.07342
ε	0.0487127	0.0023842	20.4315
σ_0	0.471071	0.0216143	21.7945
μ_0	1.45201	0.0334421	43.4186

TABLE II
Parameter estimates for trucking industry

Parameter	Value	Std. error	T-value
λ	0.18	0.014	12.59
μ	6.91	0.015	466.87
ε	0.05	0.001	35.73
σ_0	0.50	0.0197	25.424
μ_0	6.44	0.008	810.25

diffusion parameter ε is small and positive, conforming to our theoretical predictions. The results predict that if we start with a normal distribution and let the model drive the distribution, the distribution variance will tend toward a constant ε/λ, and concentrated around a mean u.

Similarly, Table II reports estimates for the five model parameters u_0, u, ε, σ_0, and λ, along with the standard errors and t-values for the trucking industry.

Once again, the value for the speed of convergence λ, is small but positive as expected. Furthermore, the diffusion parameter ε is small and positive, conforming to our theoretical predictions.

Figure 3 graphically illustrates the mean of the firm size distribution in the biotech data (dotted line), superimposed on the mean of the size distribution as predicted by the model (bold solid line, +/– one standard deviation). The vertical axis on this panel measures the mean of the size distribution (in logarithms) and the horizontal axis measures time in years. Likewise, Figure 4 graphically illustrates the mean of the firm size distribution in the trucking data (dotted line), superimposed on the mean of the size distribution as predicted by the model (bold solid line, +/– one standard deviation). Once again, the vertical axis on this panel measures the mean of the size distribution (in logarithms) and the horizontal axis measures time in years.

Figures 5 and 6 graphically illustrate the standard deviation of the size distribution in the

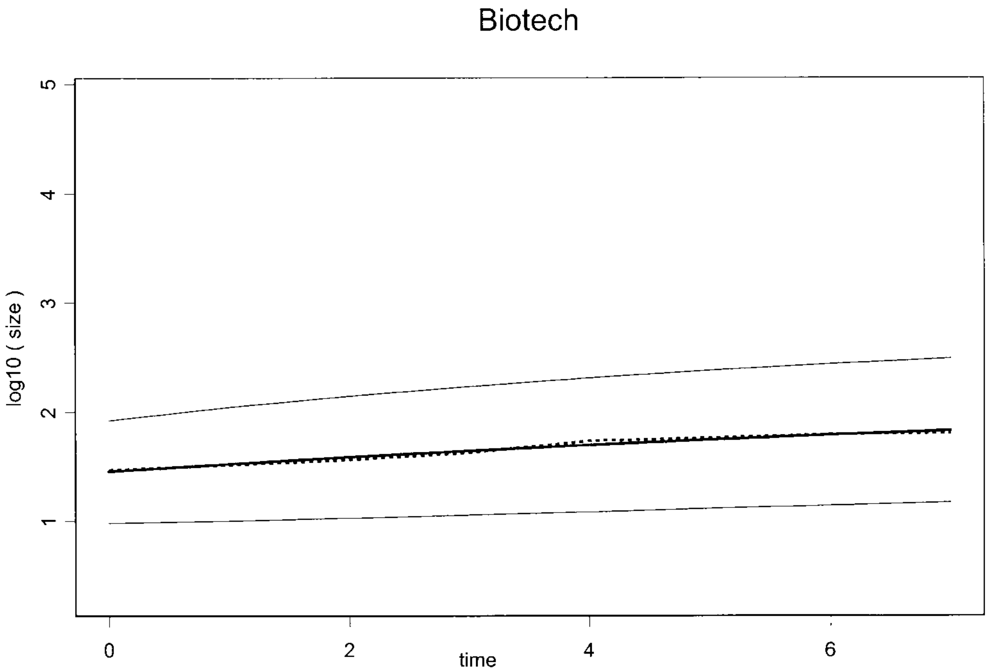

Figure 3.

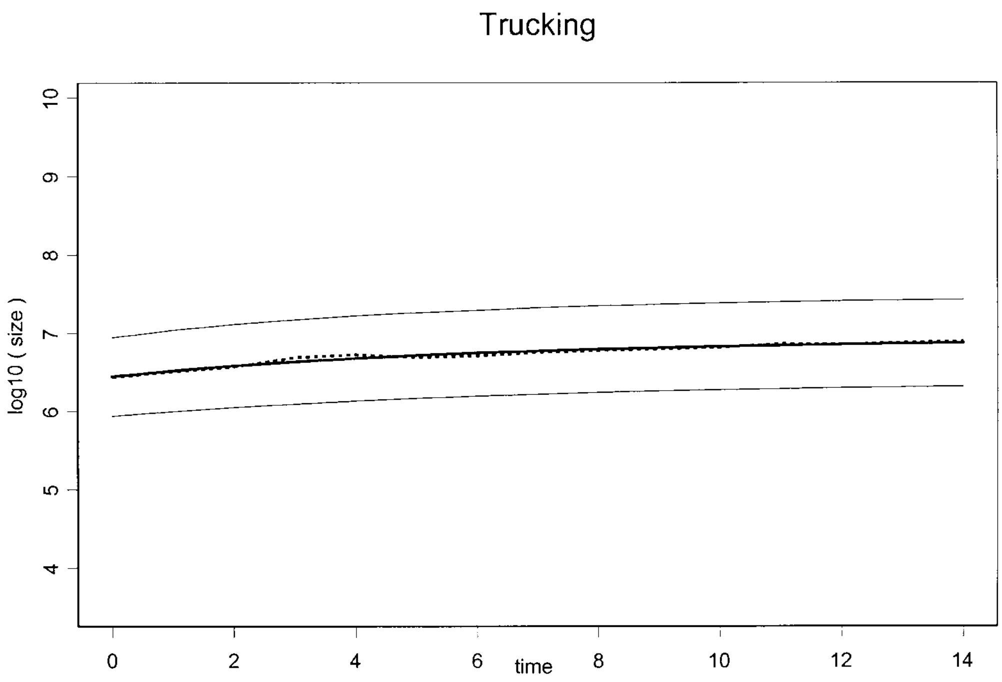

Figure 4.

biotech and trucking data respectively (dots), superimposed on the standard deviation of the respective distributions as predicted by the model (solid curves).

Figures 7 and 8 graphically illustrate the evolution of the firm size distribution (log-normals) over time for the biotech and trucking industries respectively, superimposed on histograms which describe the time evolution of the distribution of firm sizes in the data (for selected years). The solid curves in these figures illustrate the distribution of firm size as predicted by the model, and the dotted curves illustrate the distribution of firm size in the data. The vertical axes in these figures denote frequency, and the horizontal axes measure firm size in logarithms. The figures illustrates that the nice pattern which we see in the fitted log normals are being pulled out of a set of histograms whose shapes are irregular.

4. Concluding remarks and model limitations

This paper presents an analytical study of the evolution of the distribution of firm size in an industry. The dynamics of the model rely on two opposing forces: a mean reversion process which is meant to describe supply and demand forces tending to concentrate the distribution, and a counteracting diffusion process which is meant to describe the effects of a firm's search and learning behavior as well as bounded rationality on the size distribution of firms. The dynamic behavior of the model is analytically derived. An empirical application of the general structure to the log size distribution of populations in (*i*) the U.S. biotechnology industry, and (*ii*) the U.S. for-hire trucking industry is presented. The model parameters are estimated by non-linear least squares estimation techniques in S-Plus.

By considerations of analytic tractability, the model developed in this paper constitutes a considerable simplification. One limitation owes to the assumption regarding the initial condition. In Section 2, the initial size distribution of the population was approximated to be distributed normally about some average value u_0: $f(s, 0) = f_0(s)$ on $s \in [-\infty, \infty]$. This assumption, although limiting, was only chosen on grounds that it allowed an analytic solution to the diffusion model. It can be removed using numerical methods. A second limitation of the model developed in this paper, owes to the fact that it does not

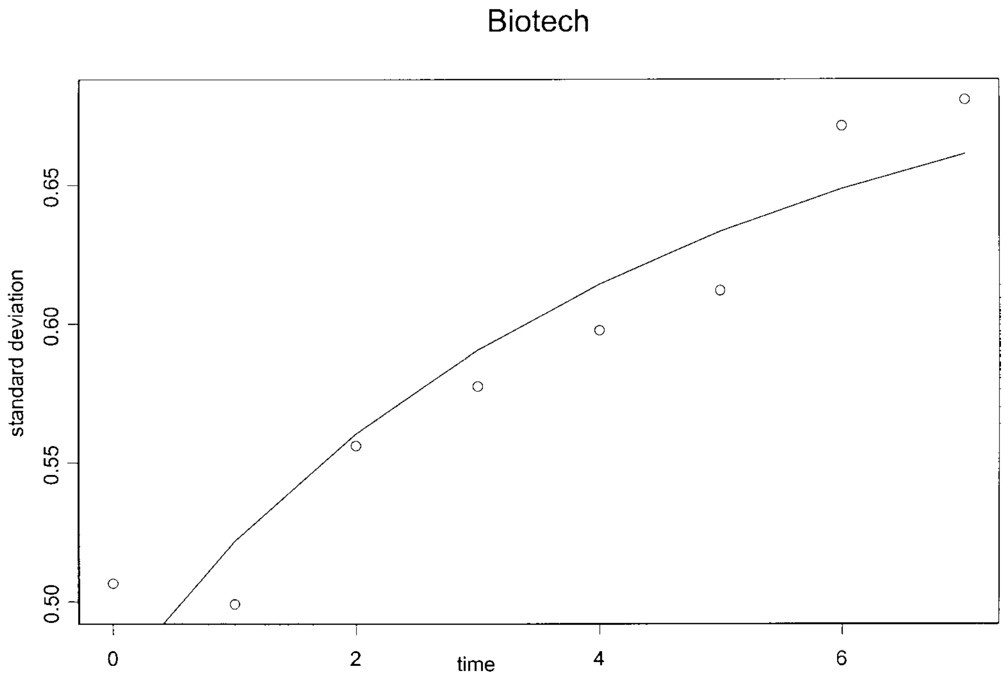

Figure 5.

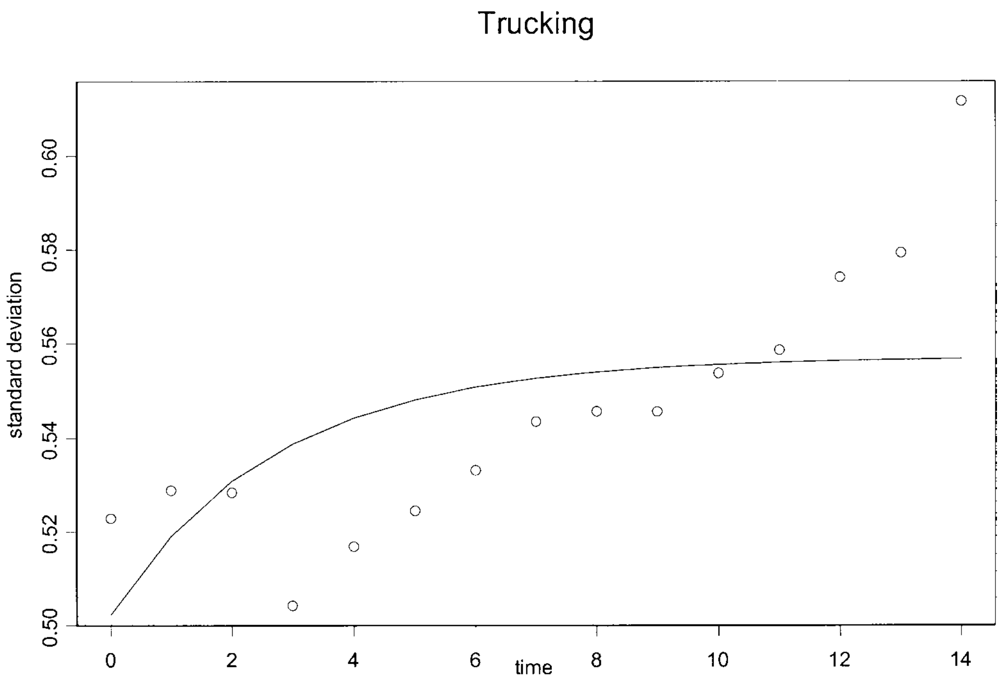

Figure 6.

allow entry into and exit out of the population (births and deaths). As a result, the analysis does not allow for the determination of the relative importance of entry/exit versus growth of existing

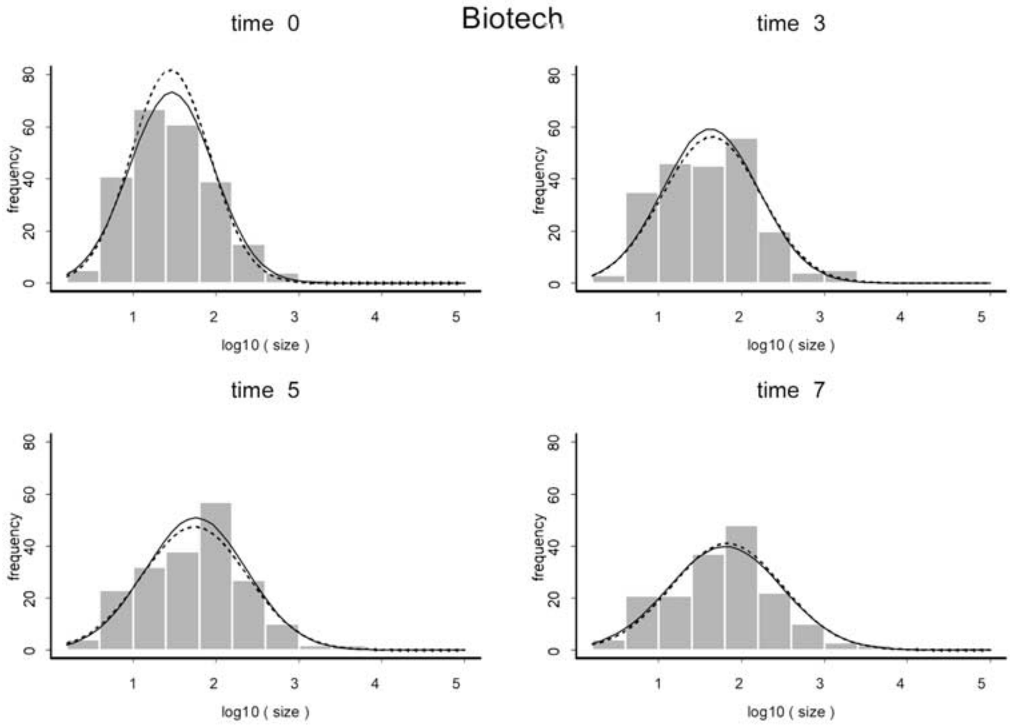

Figure 7.

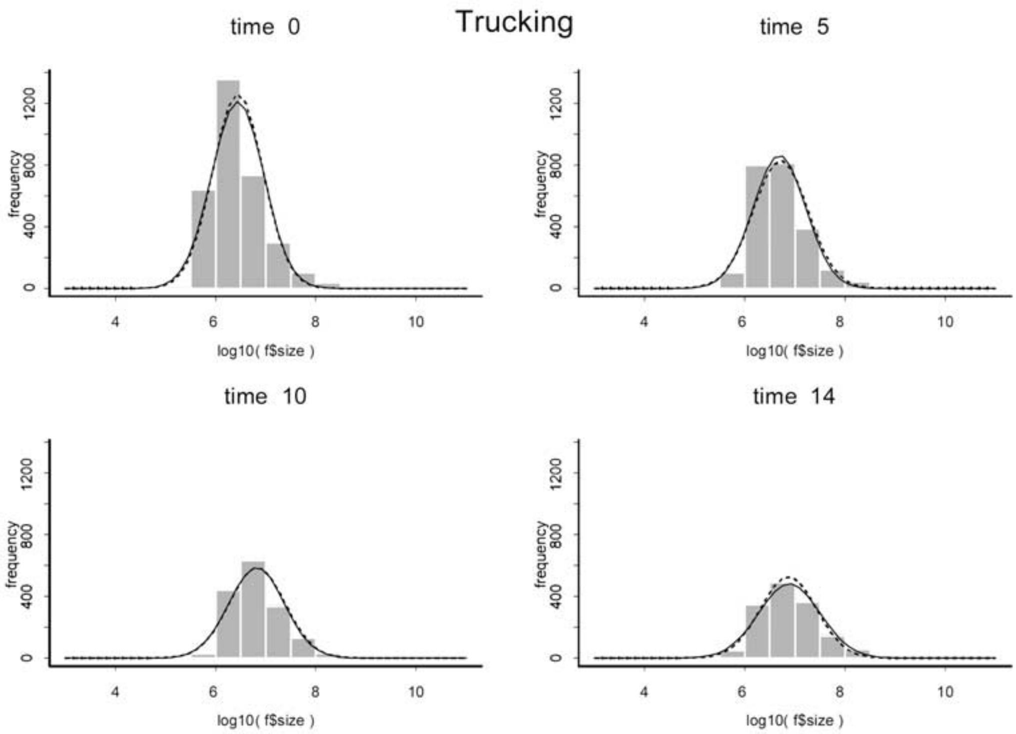

Figure 8.

firms, in bringing about the evolution in the size distribution of firms. Finally, a third limitation pertains to the functional form of the parameters. In this paper, the drift and diffusion parameters λ and ε, are assumed to be constant. One could assume these parameters to be a function of both time t and space s.

Generally, the diffusion model developed in this paper is capable of extrapolation to new and different situations. The model can be solved analytically for a variety of realistic assumptions pertaining to the rates of entry and exit, the diffusion parameter, and the speed of convergence. Moreover, the model can fruitfully be applied to analyze and compare industries in different sectors and different economies, to determine how a change of industry definition will influence the dynamics of the size distribution. Is there an equilibrium distribution of firm size in general and to what extent is it industry and/or country dependent. How does the environment (new technologies, changes in market demand, regulatory factors, investment rates etc. . . .) influence the model parameters and the results predicted by the model? These extensions would no doubt enrich the model and should prove insightful but only at the expense of considerable complexity. I hope this paper is sufficient to indicate that the endeavor shows promise.

Notes

1 These studies include Gibrat (1931); Hart and Prais (1956); Hashemi (2000); Hopenhayn (1992); Ijiri and Simon (1977); Jovanovic (1982); Simon and Bonini (1958); and Stanley et al. (1996), among others.

2 The equation in (14) is associated with the linear stochastic differential equation

$$\mathrm{d}X_t + \lambda(u - X_t)\,\mathrm{d}t = \sqrt{2\varepsilon}\,\mathrm{d}B_t$$

where B_t is the Brownian motion.

3 I am grateful to Toby Stuart for his generous provision of this data set.

4 I am grateful to Brian Silverman for his generous provision of this data set.

5 The U.S. interstate for-hire trucking industry was regulated extensively by the Interstate Commerce Commission from 1935 through 1980. For much of that time, the ICC required that all trucking firms within its regulatory jurisdiction (‘certified’ carriers) submit detailed annual reports – if the firms had annual revenues exceeding some cutoff figure ($1 million).

6 The process derived from the diffusion model evolves according to an Ornstein-Uhlenbeck, but with a transition, such that the mean tends to u, instead of 0 (Feller, 1966). The Ornstein-Uhlenbeck process is the most general normal stationary Markovian process with zero expectations. For $t > T$, the transition density from (T, s) to (t, y) is normal with expectation $e^{-\lambda(r-t)}s$ and variance $\sigma^2(1 - e^{-2\lambda(r-t)})$. As $t \to \infty$, the expectation tends to 0 and the variance to σ^2. The analytic solution derived for our diffusion equation is a normal distribution for all t. There is the $se^{\lambda t}$ factor; with a change of variables, it can be shown that the solution is normal with a constant multiplied by it.

References

Abernathy and Utterback, 1978, ‘Patterns of Industrial Innovation’, *Technology Review*, June/July, 41–47.

Alchian, 1950, ‘Uncertainty, Evolution, and Economic Theory’, *Journal of Political Economy* **58**, 211–221.

Burgelman, 1994, ‘Fading Memories: A Process Theory of Strategic Business Exit in Dynamic Environments’, *Administrative Science Quarterly* **39**, 24–56.

Cohen and Levinthal, 1990, ‘Absorptive Capacity: A New Perspective on Learning and Innovation’, *Administrative Science Quarterly* **35**, 128–152.

Cyert and March, 1963, *A Behavioral Theory of the Firm*, Englewood Cliffs, NJ: Prentice-Hall.

Hashemi, 2000, ‘An Evolutionary Model of Size Distribution of Firms’, *Journal of Evolutionary Economics* **10**, 507–521.

Henderson and Clark, 1990, ‘Architectural Innovation: The Reconfiguration of Existing Product Technology and the Failure of Established Firms’, *Administrative Science Quarterly* **35**, 9–31.

Ijiri and Simon, 1977, ‘Interpretations of Departures from the Pareto Curve Firm-size Distributions’, *Journal of Political Economy* **82**, 315–331.

Nelson and Winter, 1982, *An Evolutionary Theory of Economic Change*, Cambridge, MA: Belknap Press of Harvard University.

Okubo, 1980, *Diffusion and Ecological Problems: Mathematical Models*, Biomathematics Vol. 10, Springer-Verlag.

Ricciardi, 1977, *Diffusion Processes and Related Topics in Biology*, Lecture notes in Biomathematics, Springer-Verlag.

Simon, 1955, ‘A Behavioral Model of Rational Choice’, *Quarterly Journal of Economics* **69**.

Simon and Bonini, 1958, ‘The Size Distribution of Business Firms’, *American Economic Review* **48**, 607–617.

Stanley et al., 1996, ‘Scaling Behavior in the Growth of Companies’, *Nature* **379**, 804–806.

Stuart and Podolny, 1996, ‘Local Search and the Evolution of Technological Capabilities’, *Strategic Management Journal* **17**, 21–38.

Taylor, 1984, *Predation*, Population and Community Biology Series, Chapmann and Hall.

Tushman and Anderson, 1986, ‘Technological Discontinuities and Organizational Environments’, *Administrative Science Quarterly* **31**, 439–465.