Adv. Appl. Prob. **35,** 641–664 (2003)
Printed in Northern Ireland

MODELING GROWTH STOCKS VIA BIRTH–DEATH PROCESSES

S. C. KOU,* *Harvard University*
S. G. KOU,** *Columbia University*

Abstract

The inability to predict the future growth rates and earnings of growth stocks (such as biotechnology and internet stocks) leads to the high volatility of share prices and difficulty in applying the traditional valuation methods. This paper attempts to demonstrate that the high volatility of share prices can nevertheless be used in building a model that leads to a particular cross-sectional size distribution. The model focuses on both *transient* and *steady-state* behavior of the market capitalization of the stock, which in turn is modeled as a birth–death process. Numerical illustrations of the cross-sectional size distribution are also presented.

Keywords: Biotechnology and internet stocks; asset pricing; convergence rate; volatility; power-type distribution; Zipf's law; Pareto distribution; regression

AMS 2000 Subject Classification: Primary 60H30; 91B70

1. Introduction

1.1. Growth stocks

Issuing stocks is arguably the most important way for growth companies to finance their projects, and in turn helps transfer new ideas into products and services for the society. Although the content of growth stocks may change over time (perhaps consisting of railroad and utility stocks in the early 1900s, and biotechnology and internet stocks in 2000), studying the general properties of growth stocks is essential for understanding financial markets and economic growth.

However, uncertainty is manifest for growth stocks. For example, as demonstrated in the recent market from 1999 to 2002, (a) growth stocks tend to have low or even negative earnings; (b) the volatility of growth stocks is high (both their daily appreciation and depreciation rates are high); (c) it is difficult to predict the future growth rates and earnings. Consequently, it poses a great challenge to derive a meaningful mathematical model within the classical valuation framework, such as the net-present-value method (which relies on current earnings and the prediction of future earnings).

Indeed, since it appears that the only thing that we are sure about growth stocks is their uncertainty, we may wonder whether there is much more to say about them. The current paper attempts to illustrate that a mathematical model for growth stocks can, nevertheless, be built via birth–death processes, mainly by utilizing the *high volatility* of their share prices.

One motivation of the current study comes from a report on internet stocks in the *Wall Street Journal* (27 December 1999): researchers at Credit Suisse First Boston observed that

Received 2 December 2002; revision received 17 February 2003.
* Postal address: Department of Statistics, Science Center 603, Harvard University, Cambridge, MA 02138, USA.
** Postal address: Department of IEOR, 312 Mudd Building, Columbia University, New York, NY 10027, USA.
Email address: sk75@columbia.edu

'there is literally a mathematical relationship between the ranking of the (internet) stock and its capitalization'. (This observation is summarized later in a research report by Mauboussin and Schay (2000).) In the article, it is suggested that a linear downward pattern emerges when the market capitalizations of internet stocks are plotted against their associated ranks on a log–log scale, with rank one being the largest market capitalization. The same article, more interestingly, also reported that this phenomenon does not seem to hold for nongrowth stocks. The article challenges people to investigate whether such a phenomenon happens simply by chance or if there is certain mechanism behind it.

The model proposed in the current paper sheds light on this phenomenon. More interestingly, the model leads us to discover and explain another new empirical observation, which has a better goodness of fit; see (5.8), Remark 5.1, and Section 6. Roughly speaking, our result suggests that if the market capitalization of stocks is modeled as a birth–death process, then for stocks with high volatility (such as biotechnology and internet stocks) an almost (but not exactly) linear curve will appear, on the log–log scale, if the market capitalizations (normalized by the market capitalization of the largest stock within the group) are plotted against their relative ranks; see (5.8). Meanwhile for nongrowth stocks the model implies that such a phenomenon should not be expected, primarily because of the slow convergence of the birth–death process to its steady-state distribution due to their low volatility. Furthermore, the model applies not only to internet stocks (on which the *Wall Street Journal* article focuses) but to growth stocks (e.g. biotechnology stocks) in general.

1.2. Background of cross-sectional size distributions

Studying the stochastic relationships between some values of interest and their relative ranks within a group, termed the (cross-sectional) 'size distribution', has a long history in probability, dating back at least to Pareto (1896), Yule (1924), (1944), Gibrat (1931), and Zipf (1949). For some more recent developments, see, for example, Woodroofe and Hill (1975), Chen (1980), and Mandelbrot (1997).

Starting from Simon (1955), economists began to use various stochastic processes to model cross-sectional size distributions in economics, for example the sizes of business firms (see e.g. Ijiri and Simon (1977), Lucas (1978), Steindl (1965), Simon and Bonini (1958), Axtell (2001)), income distribution (see e.g. Rutherford (1955), Mandelbrot (1960), Shorrocks (1975), Feenberg and Poterba (1993)), and city-size distribution (see e.g. Glaeser *et al.* (1995), Krugman (1996a), (1996b), Gabaix (1999)). However, most of the theory developed so far focuses on the steady-state size distribution and pays little attention to the *transient* behavior of size distributions.

The contribution of the current paper is twofold.

1. From a probabilistic viewpoint, we give a detailed analysis (see Section 4) of the transient behavior of the size distribution, which is not well addressed in the size-distribution literature. The analysis of the transient behavior is crucial to our study (see Section 5), as it explains why the size-distribution theory can be applied to growth stocks but not to nongrowth stocks.

2. From an applied point of view, we point out that the theory of the size distribution may have an interesting application in studying growth stocks (see Section 5), which is difficult for traditional methods, such as the net-present-value approach.

The paper is organized as follows. Section 2 proposes the basic model, while Sections 3 and 4 analyze both the transient and steady-state properties of the model. The model is then applied in Section 5 to derive the size distribution of growth stocks, and to explain why the method can be used for growth stocks but not for nongrowth stocks. Numerical illustrations are presented in Section 6. The advantage and disadvantage of the model are discussed in the last section. Some proofs are deferred to Appendix A.

2. The model

In modeling growth stocks, instead of working on the price of a growth stock, it makes more sense to study the market capitalization, defined as the product of the total number of outstanding shares and the market price of the stock, because growth stocks tend to have frequent stock splits, which immediately makes the price drop significantly but has little effect on the market capitalization. We formulate the model as follows.

Consider at time t a growth stock with a total market capitalization $M(t)$. We postulate that

$$M(t) = \Theta(t)X(t), \tag{2.1}$$

where $\Theta(t)$ represents the overall economic and sector trend and $X(t)$ represents each individual variation within the sector. Hence, $\Theta(t)$ is the same for all firms within the same industry sector, and the individual variation term $X(t)$ varies for different firms within the sector.

The individual variation term $X(t)$ is modeled as a birth–death process: given that $X(t)$ is in state i, the instantaneous changes are $i \mapsto i+1$, with rate $i\lambda + g$ for $i \geq 0$, and $i \mapsto i-1$, with rate $i\mu + h$ for $i \geq 1$, where the parameters are such that $\lambda, \mu > 0$, $g > 0$, $h \geq 0$, $\lambda < \mu$. The unit of $X(t)$ could be, for example, millions or billions of dollars.

Under the standard notation, $X(t)$ is a birth–death process with the birth rate λ_i and the death rate μ_i satisfying

$$\begin{aligned} &\lambda_i = i\lambda + g, \qquad && \mu_i = i\mu + h, \qquad && i \geq 1, \\ &\lambda_0 = g, && \mu_0 = 0, && \end{aligned} \tag{2.2}$$

and the infinitesimal generator of $X(t)$ is given by the infinite matrix

$$\begin{pmatrix} -g & g & 0 & 0 & \cdots \\ \mu + h & -\lambda - \mu - g - h & \lambda + g & 0 & \cdots \\ 0 & 2\mu + h & -2\lambda - 2\mu - g - h & 2\lambda + g & \cdots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix}.$$

In the model, the state 0 only signifies that the size of $X(t)$ is below a certain minimal level. It does not imply, for example, that the company goes bankrupt.

The two parameters λ and μ represent the instantaneous appreciation and depreciation rates of $X(t)$ due to market fluctuation; the model assumes that they influence $X(t)$ proportionally to the current value. In general, because of the difficulty of predicting the instantaneous upward and downward price movements (partly thanks to the efficient market hypothesis), for both growth stocks and nongrowth stocks λ and μ must be quite close, $\lambda/\mu \approx 1$; in addition, for growth stocks, both λ and μ must be large, because of their high volatility. These two requirements will become Assumptions 5.3 and 5.4 in Section 5.2. The requirement that $\lambda < \mu$ is postulated here to ensure that the birth–death process $X(t)$ has a steady-state distribution.

The parameter $g > 0$ models the rate of increase in $X(t)$ due to nonmarket factors, such as the effect of additional shares being issued through public offerings or the effect of warranties

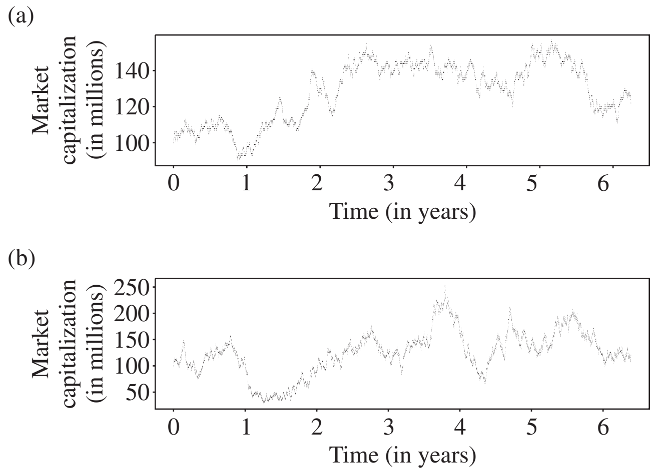

FIGURE 1: Two sample paths of the birth–death process with (a) $X(0) = 100$, $\lambda = 2.49$, $\mu = 2.51$, $g = 1$, $h = 0$; and with (b) $X(0) = 100$, $\lambda = 24.9$, $\mu = 25.1$, $g = 10$, $h = 0$.

on the stock being exercised (resulting in new shares being issued). The parameter h attempts to capture the rate of decrease in $X(t)$ due to nonmarket factors, such as the effect of dividend payments. For most growth stocks, $h \approx 0$, as no dividends are paid.

Remark 2.1. Although the individual stock's variation $X(t)$ is assumed to have a steady-state distribution, the overall economic (and sector) trend $\Theta(t)$ can have a positive drift.

Remark 2.2. In general, neither $X(t)$ nor $\Theta(t)$ is likely to be observed directly in the market. Instead, only $M(t)$ is directly observed in the market.

Remark 2.3. The model of $X(t)$ proposed here is a variation and a generalization of the models proposed in Simon (1955) and Shorrocks (1975) to study business and income sizes, etc. The key difference here is that we provide a detailed analysis of both *transient* and *steady-state* behavior, not just the steady-state analysis. The transient analysis not only presents some mathematical challenges (see Section 4), but also is essential to understand why the theory of the size distribution is useful for growth stocks but not for nongrowth stocks (see Section 5).

Figure 1 provides an illustration of the model by showing the sample paths of two realizations of the birth–death process $X(t)$ in (2.2) for about 6.5 years. In Figure 1(a), the instantaneous jump rates, λ and μ, are small, while in Figure 1(b), λ and μ are large. The sample paths suggest two points:

1. For reasonably large λ and μ, the jumps of the birth–death processes are almost unnoticeable, and the overall sample paths fit in well with our intuition of market fluctuation.

2. Although $\lambda < \mu$, the sample paths of $X(t)$ may still have some strong upward movements if λ is close to μ; for example, in Figure 1(b), $X(t)$ increases from about 20 to about 250 (more than 12 times) within a short period (about 2.5 years).

3. Preliminary results

3.1. Properties of the steady-state distribution

The steady-state measure of a birth–death process is given by

$$\pi_0 = 1, \qquad \pi_n := \frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\cdots\mu_n}, \quad n = 1, 2, \dots .$$

Normalizing $\{\pi_n\}$ provides the steady-state distribution of the birth–death process:

$$\lim_{t\to\infty} \mathrm{P}(X(t) = n) = \frac{\pi_n}{S}, \qquad S := \sum_{n=0}^{\infty} \pi_n$$

(see Lemma 3.1 below for the finiteness of S under the setting of (2.2)). In our case,

$$\pi_n = \left(\frac{\lambda}{\mu}\right)^n \frac{(g/\lambda)(1+g/\lambda)(2+g/\lambda)\cdots((n-1)+g/\lambda)}{(1+h/\mu)(2+h/\mu)\cdots(n+h/\mu)}, \qquad n \geq 1.$$

Using the gamma function, it can be succinctly expressed as

$$\pi_n = \frac{\Gamma(1+h/\mu)}{\Gamma(g/\lambda)}\left(\frac{\lambda}{\mu}\right)^n \frac{\Gamma(n+g/\lambda)}{\Gamma(n+1+h/\mu)}, \qquad n \geq 0. \tag{3.1}$$

Lemma 3.1. (Steady-state properties of $X(t)$.) (i) *The birth–death process $X(t)$ in (2.2) is positive recurrent; i.e. it will visit every state $\{0, 1, 2, \dots\}$ with probability 1, and the expected visiting time of any state is finite.*

(ii) *As $n \to \infty$,*

$$\pi_n \cong \frac{\Gamma(1+h/\mu)}{\Gamma(g/\lambda)}\left(\frac{\lambda}{\mu}\right)^n n^{g/\lambda - h/\mu - 1}. \tag{3.2}$$

Here and throughout this paper, $a \cong b$ means that $a/b \to 1$ asymptotically. This asymptotic order, in particular, implies that $S = \sum_{n=0}^{\infty} \pi_n$ is finite.

(iii) *The moment-generating function of the steady-state distribution is given by*

$$\eta(\theta) := \sum_{n=0}^{\infty} \frac{\mathrm{e}^{\theta n}\pi_n}{S} = \frac{F(g/\lambda, 1; 1+h/\mu; (\lambda/\mu)\mathrm{e}^{\theta})}{F(g/\lambda, 1; 1+h/\mu; \lambda/\mu)}, \tag{3.3}$$

where $F(a, b; c; z)$ is the hypergeometric function (see Abramowitz and Stegun (1972, p. 556)):

$$F(a, b; c; z) := \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{n=0}^{\infty} \frac{\Gamma(a+n)\Gamma(b+n)}{\Gamma(c+n)} \frac{z^n}{n!}.$$

In particular, the mean and the second moment of the steady-state distribution are

$$m_1 := \eta'(0) = \frac{1}{S}\frac{g}{\mu+h} F\left(1 + \frac{g}{\lambda}, 2; 2 + \frac{h}{\mu}; \frac{\lambda}{\mu}\right)$$

and

$$\begin{aligned} m_2 &:= \eta''(0) \\ &= \frac{1}{S}\frac{g}{\mu+h}\left\{F\left(1+\frac{g}{\lambda}, 2; 2+\frac{h}{\mu}; \frac{\lambda}{\mu}\right) + 2\frac{\lambda+g}{2\mu+h} F\left(2+\frac{g}{\lambda}, 3; 3+\frac{h}{\mu}; \frac{\lambda}{\mu}\right)\right\}. \end{aligned}$$

(iv) *Let the tail probability of the steady-state distribution be*

$$F(n) := \lim_{t\to\infty} \mathrm{P}(X(t) \geq n)$$
$$= \sum_{k=n}^{\infty} \frac{\pi_k}{S}.$$

Then, as $n \to \infty$,

$$F(n) \cong \frac{1}{S}\frac{\Gamma(1+h/\mu)}{\Gamma(g/\lambda)}\left(1-\frac{\lambda}{\mu}\right)^{-1}\left(\frac{\lambda}{\mu}\right)^{n} n^{g/\lambda-h/\mu-1}. \tag{3.4}$$

Although some of the results in Lemma 3.1 may be known, perhaps for $h = 0$, to the best of the authors' knowledge the representations in terms of the hypergeometric functions have not been given previously; the proof of Lemma 3.1 is deferred to Appendix A.

A careful inspection of the result of Lemma 3.1 reveals that, instead of the original parameters, only the three ratios λ/μ, h/μ, and g/λ determine the steady-state distribution. Thus, the steady-state properties only reflect the relative magnitude of the parameters λ, μ, g, and h, rather than the absolute magnitude. (This contrasts with the realizations of the birth–death process, such as in Figure 1, in which the dynamic behavior of the sample path does depend on the absolute magnitude of λ, μ, g, and h.)

3.2. Transient mean and variance

Lemma 3.1 only provides steady-state properties of $X(t)$, which might be relevant if the birth–death process $X(t)$ has been run for a long time; i.e. the stock has been traded in the market for a long period. However, it is quite possible that the parameters λ, μ, g, and h may have changed during the period, thus altering the steady-state distribution. Therefore, practically the steady-state properties are relevant *only if the convergence from the transient states to the steady states is fast enough*, i.e. if the convergence can be observed in a timely fashion.

There are several ways to judge the convergence speed. In this section we shall focus on the mean and variance of the transient distribution, which can lead to a measure of the convergence rate; see Section 5.1. A more accurate measure (which is of course more difficult to study) is the convergence rate for the transition probabilities, which attempts to capture the convergence of the whole distribution rather than just the first two moments; this will be analyzed in the next section.

Denote the transition probability at time t by

$$p_{i,j}(t) := \mathrm{P}(X(t) = j \mid X(0) = i),$$

the transient expectation at time t by

$$m_1(t) := \mathrm{E}\, X(t) = \sum_{j=0}^{\infty} j p_{i,j}(t),$$

and the second moment by

$$m_2(t) := \mathrm{E}\, X^2(t) = \sum_{j=0}^{\infty} j^2 p_{i,j}(t).$$

Lemma 3.2. (Transient mean and variance.) *Suppose that the birth–death process starts from $X(0) = i$. The first moment $m_1(t)$ at time t satisfies the following differential equation:*

$$m_1'(t) = (\lambda - \mu)m_1(t) + g + h(1 - p_{i,0}(t)),$$

whose solution is given by

$$m_1(t) = i\mathrm{e}^{(\lambda-\mu)t} + \frac{g}{\mu - \lambda}[1 - \mathrm{e}^{(\lambda-\mu)t}] + h\int_0^t \mathrm{e}^{(\mu-\lambda)(s-t)}(1 - p_{i,0}(s))\,\mathrm{d}s.$$

The second moment $m_2(t)$ satisfies the differential equation

$$m_2'(t) = 2(\lambda - \mu)m_2(t) + (\lambda + \mu + 2g - 2h)m_1(t) + g + h(1 - p_{i,0}(t)),$$

with solution given by

$$\begin{aligned} m_2(t) = {} & i^2\mathrm{e}^{2(\lambda-\mu)t} + \frac{g}{2(\mu - \lambda)}[1 - \mathrm{e}^{2(\lambda-\mu)t}] + h\int_0^t \mathrm{e}^{2(\mu-\lambda)(s-t)}(1 - p_{i,0}(s))\,\mathrm{d}s \\ & + (\lambda + \mu + 2g - 2h)\int_0^t \mathrm{e}^{2(\mu-\lambda)(s-t)}m_1(s)\,\mathrm{d}s. \end{aligned}$$

The proof of Lemma 3.2 is given in Appendix A. We note that, for the special case of $h = 0$, Karlin and McGregor (1958) derived a differential equation for $p_{i,j}(t)$, and solved it by using orthogonal polynomials.

4. The transient behavior of the model

As mentioned above, most of the literature on size distributions focuses on the steady-state properties, and, except for some numerical examples (for example, Shorrocks (1975) demonstrated through numerical calculation that, if the convergence rate is not large enough, it may take 15 to 181 years for some birth–death processes to reach steady states), the theoretical properties of the transient behavior are hardly addressed in the literature. In this sense, this section constitutes the main technical contribution of the current paper to the size-distribution literature.

The speed of convergence of a birth–death process can be measured by the decay parameter (see Kijima (1997)), which is defined by

$$\gamma := \sup\biggl\{\alpha \ge 0 : p_{i,j}(t) - \biggl(\frac{\pi_j}{S}\biggr) = O(\mathrm{e}^{-\alpha t}) \text{ for all } i, j \ge 1\biggr\},$$

where, as before, $p_{i,j}(t)$ is the transition probability at time t and π_j/S is the steady-state probability. For further background on the speed of convergence and rate of exponential ergodicity, see, for example, Meyn and Tweedie (1993).

Note that the decay parameter γ affects the convergence in an exponential way. In other words, a small difference in γ can have a remarkable effect on the speed of convergence, which in turn suggests that the steady-state analysis of the size distribution in our model based on the birth–death process is only relevant when the decay parameter is large. In addition, since the infinitesimal generator of the birth–death process is an infinite-state matrix, the analysis of the convergence rate is different from that for finite-state Markov chains.

Theorem 4.1. (The decay parameter.) *For the birth–death process in the model, if $h = 0$, then the decay parameter γ is equal to $\mu - \lambda$; otherwise, if $h > 0$, then*

$$\mu - \lambda \le \gamma < \mu - \lambda + h\left[1 - \min\left(\frac{\lambda}{\mu}, \frac{\lambda + g}{\mu + h}\right)\right].$$

The derivation of Theorem 4.1 is the main technical contribution of the current paper to the study of the size distribution. To prove Theorem 4.1, we start from the following lemma (Lemma 5.14 in Kijima (1997)).

Lemma 4.1. *There exists a sequence $\{k_i\}$ such that $k_0 = \infty$, $k_i > 0$ for all $i \ge 1$, and $\lambda_i + \mu_{i+1} - \lambda_i \mu_i / k_i - k_{i+1} = y$ for all $i \ge 0$ for some constant y if and only if $y \le \gamma$.*

To derive the upper bound for the decay parameter γ, we need the following lemma.

Lemma 4.2. *For any constant $c > 0$, consider the sequence $\{k_i\}$ defined by*

$$\begin{aligned} k_1 &= \lambda + g + h - c, \qquad c > 0,\ h \ge 0, \\ k_{i+1} &= \lambda_{i+1} + \mu_i - \frac{\lambda_i \mu_i}{k_i} - c, \qquad i \ge 1. \end{aligned} \tag{4.1}$$

Let $l_i := k_i - \lambda_i$, $i \ge 1$, so l_i has the following recurrence relation:

$$\begin{aligned} l_1 &= h - c, \qquad c > 0,\ h \ge 0, \\ l_{i+1} &= \frac{l_i}{\lambda_i + l_i}\mu_i - c, \qquad i \ge 1. \end{aligned} \tag{4.2}$$

Then $k_i > 0$ for all $i \ge 1$ if and only if $l_i > 0$ for all $i \ge 1$.

Proof. Suppose that $l_i > 0$ for all $i \ge 1$. Then immediately $k_i = \lambda_i + l_i > \lambda_i > 0$ for every i. We next prove the reversed statement by contradiction.

Suppose that $k_i > 0$ for all $i \ge 1$ and $l_m \le 0$ for some m. The recurrence relation of l_i gives $l_{m+1} \le -c$ and $l_{i+1} \le l_i \mu_i / k_i$ for all $i \ge 1$. Therefore,

$$l_{m+2} \le l_{m+1} \frac{\mu_{m+1}}{k_{m+1}} \le -c \frac{\mu_{m+1}}{k_{m+1}}.$$

In general, by induction, for any $M > m + 1$,

$$l_M \le -c \prod_{j=m+1}^{M-1} \frac{\mu_j}{k_j} < 0,$$

which implies that $0 < k_j = l_j + \lambda_j < \lambda_j$ for all $j \ge m + 1$. Therefore, replacing k_j by λ_j in the above equation, we have

$$l_M \le -c \prod_{j=m+1}^{M-1} \frac{\mu_j}{k_j} \le -c \prod_{j=m+1}^{M-1} \frac{\mu_j}{\lambda_j} < 0.$$

Since $\mu_j / \lambda_j \to \mu/\lambda > 1$, the above inequality implies that $l_M \to -\infty$ exponentially fast. Thus, we must have $k_j = l_j + \lambda_j < 0$ for some $j \ge m + 1$, since $\lambda_j \to \infty$ only linearly fast. This contradicts the initial assumption.

Proof of Theorem 4.1. The proof consists of two steps.

Step 1. We want to prove the lower bound, i.e. $\gamma \geq \mu - \lambda$. To do this, consider the sequence $\{k_i\}$ defined by

$$k_0 = \infty,$$
$$\mu - \lambda = \lambda_i + \mu_{i+1} - \frac{\lambda_i \mu_i}{k_i} - k_{i+1}, \qquad i \geq 0,$$

that is,

$$k_1 = \lambda_0 + \mu_1 - (\mu - \lambda) = \lambda + g + h,$$
$$k_{i+1} = (\lambda_{i+1} + \lambda) + (\mu_{i+1} - \mu) - \frac{\lambda_i \mu_i}{k_i} = \lambda_{i+1} + \mu_i - \frac{\lambda_i \mu_i}{k_i}, \qquad i \geq 1.$$

Let $l_i = k_i - \lambda_i$ for $i \geq 1$. Then

$$l_1 = h,$$
$$l_{i+1} = \frac{l_i}{\lambda_i + l_i} \mu_i, \qquad i \geq 1.$$

It is easy to see that $l_i \geq 0$ for all $i \geq 1$, which says that $k_i = \lambda_i + l_i > l_i \geq 0$ for every $i \geq 1$. By Lemma 4.1, we must have $\gamma \geq \mu - \lambda$.

Step 2. We now prove the upper bound for γ. First, note that the recurrence relationship of l_i in (4.2) implies that, for any number $d \geq 0$,

$$l_{i+1} > d \geq 0 \quad \text{if and only if} \quad l_i > (c + d) \frac{\lambda_i}{\mu_i - c - d}.$$

In particular, for any number $d \geq 0$,

$$\text{if } l_{i+1} > d \geq 0, \quad \text{then } l_i > (c + d) \frac{\lambda_i}{\mu_i} = c \frac{\lambda_i}{\mu_i} + d \frac{\lambda_i}{\mu_i}.$$

Using it once again, we know that, if $l_{i+1} > d \geq 0$, then

$$l_{i-1} > c \frac{\lambda_{i-1}}{\mu_{i-1}} + \left(c \frac{\lambda_{i-1}}{\mu_{i-1}} + d \frac{\lambda_{i-1}}{\mu_{i-1}} \right) \frac{\lambda_i}{\mu_i} = c \frac{\lambda_{i-1}}{\mu_{i-1}} + c \frac{\lambda_{i-1}}{\mu_{i-1}} \frac{\lambda_i}{\mu_i} + d \frac{\lambda_{i-1}}{\mu_{i-1}} \frac{\lambda_i}{\mu_i}.$$

In general, simple induction gives that, if $l_{i+1} > d \geq 0$, then, for any $j \leq i$,

$$l_j > c \left\{ \frac{\lambda_j}{\mu_j} + \frac{\lambda_j}{\mu_j} \frac{\lambda_{j+1}}{\mu_{j+1}} + \cdots + \frac{\lambda_j}{\mu_j} \frac{\lambda_{j+1}}{\mu_{j+1}} \cdots \frac{\lambda_i}{\mu_i} \right\} + d \frac{\lambda_j}{\mu_j} \frac{\lambda_{j+1}}{\mu_{j+1}} \cdots \frac{\lambda_i}{\mu_i}. \tag{4.3}$$

Next, suppose that $\gamma = \mu - \lambda + c$, with $c > 0$. Then Lemma 4.1 implies that there exists a sequence $\{k_i\}$ such that $k_0 = \infty$, $k_i > 0$ for all $i \geq 1$, and

$$\mu - \lambda + c = \lambda_i + \mu_{i+1} - \frac{\lambda_i \mu_i}{k_i} - k_{i+1}, \qquad i = 0, 1, 2, \ldots.$$

According to Lemma 4.2, $k_i > 0$ for all i implies that $l_j > 0$ for all j. Thus, we can set $d = 0$ in (4.3). Letting $j = 1$ and $i \to \infty$ and using the fact that

$$\frac{\lambda_i}{\mu_i} = \frac{i\lambda + g}{i\mu + h} \geq \rho := \min\left(\frac{\lambda}{\mu}, \frac{\lambda + g}{\mu + h}\right) \quad \text{for any } i \geq 1$$

yields that $l_1 > c\{\rho + \rho^2 + \cdots\} = c\rho/(1-\rho)$. But, by the definition (4.2), $l_1 = h - c$. Therefore, $h - c > c\rho/(1-\rho)$, which implies that

$$c < h(1-\rho) \quad \text{for all } h \geq 0. \tag{4.4}$$

Finally, consider the two cases $h = 0$ and $h > 0$. If $h = 0$, then (4.4) leads to a contradiction as c is assumed to be positive. Thus, when $h = 0$, γ must be equal to $\mu - \lambda$. If $h > 0$, then (4.4) yields that

$$\gamma = \mu - \lambda + c < \mu - \lambda + h(1-\rho) = \mu - \lambda + h\left[1 - \min\left(\frac{\lambda}{\mu}, \frac{\lambda + g}{\mu + h}\right)\right],$$

from which the conclusion follows.

5. The size distribution for growth stocks

In this section we shall apply the results on both the steady-state and the transient behavior of the model to study the size distribution of growth stocks. Since, for most growth stocks, there is no dividend payment, we shall assume from this section on that

$$h = 0. \tag{5.1}$$

5.1. Basic transient and steady-state properties for $h = 0$

Under the assumption (5.1), Lemma 3.1 implies that the steady-state measure for $X(t)$ is

$$\pi_n = \frac{1}{\Gamma(g/\lambda)} \left(\frac{\lambda}{\mu}\right)^n \frac{\Gamma(n + g/\lambda)}{n!},$$

and

$$F(n) = \sum_{k=n}^{\infty} \frac{\pi_k}{S} = \frac{\pi_n F(n + g/\lambda, 1; n+1; \lambda/\mu)}{S}, \qquad n \geq 0.$$

In addition,

$$S = \sum_{k=0}^{\infty} \pi_k = F\left(\frac{g}{\lambda}, 1; 1; \frac{\lambda}{\mu}\right) = \left(1 - \frac{\lambda}{\mu}\right)^{-g/\lambda},$$

thanks to the following property of the hypergeometric function: $F(a, b; b; z) = (1-z)^{-a}$. This, together with (3.4), yields that

$$F(n) = \lim_{t\to\infty} \mathrm{P}(X(t) \geq n) \cong \frac{1}{\Gamma(g/\lambda)} \left(1 - \frac{\lambda}{\mu}\right)^{g/\lambda - 1} \left(\frac{\lambda}{\mu}\right)^n n^{g/\lambda - 1}. \tag{5.2}$$

By (3.3), the moment-generating function of the steady-state distribution, under $h = 0$, is

$$\eta(\theta) = \left(\frac{\mu - \lambda e^\theta}{\mu - \lambda}\right)^{-g/\lambda}.$$

Thus, for the steady-state distribution, the first two moments are

$$\begin{aligned} m_1 &= \eta'(0) = \frac{g}{\mu - \lambda}, \\ m_2 &= \eta''(0) = \frac{g(\mu + g)}{(\mu - \lambda)^2}, \end{aligned} \tag{5.3}$$

and the variance is

$$m_2 - m_1^2 = \frac{\mu g}{(\mu - \lambda)^2}.$$

For the properties of the transient behavior, first note that, by Theorem 4.1, the decay parameter, which measures the speed of convergence to the steady state in an exponential way, is given by

$$\gamma = \mu - \lambda.$$

Secondly, by Lemma 3.2,

$$\begin{aligned} m_1(t) &= i\mathrm{e}^{(\lambda-\mu)t} + \frac{g}{\mu - \lambda}[1 - \mathrm{e}^{(\lambda-\mu)t}], \\ m_2(t) &= i^2\mathrm{e}^{2(\lambda-\mu)t} + i\frac{\lambda + \mu + 2g}{\lambda - \mu}(\mathrm{e}^{2(\lambda-\mu)t} - \mathrm{e}^{(\lambda-\mu)t}) \\ &\quad + \frac{g}{2(\mu - \lambda)}[1 - \mathrm{e}^{2(\lambda-\mu)t}] + \frac{g(\lambda + \mu + 2g)}{2(\mu - \lambda)^2}(1 - \mathrm{e}^{(\lambda-\mu)t})^2. \end{aligned}$$

The exponents in $m_1(t)$ and $m_2(t)$ are all related to $\lambda - \mu$, which also points out, from a different viewpoint, that $\mu - \lambda$ should affect the speed of convergence in an exponential way. In addition, it is easily seen that

$$\begin{aligned} \lim_{t\to\infty} m_1(t) &= \frac{g}{\mu - \lambda} = m_1, \\ \lim_{t\to\infty} m_2(t) &= \frac{g(\mu + g)}{(\mu - \lambda)^2} = m_2. \end{aligned}$$

5.2. The size distribution

Consider N growth firms within a particular sector, whose market capitalizations are governed by the model (2.1) (here N is an unknown quantity). Suppose that, among these N firms, we observe the K largest (in terms of their market capitalization). Denote the market capitalization of the K observed stocks by $M_i(t)$, $1 \le i \le K$. Since all these K firms are from the same sector, we have

$$M_i(t) = \Theta(t)X_i(t), \qquad 1 \le i \le K,$$

where $\Theta(t)$, the overall economic and sector trend, is the same for all K stocks, but the individual variation terms $X_i(t)$ are different.

Now suppose that we rank the market capitalizations such that $M_{(1)}(t) > M_{(2)}(t) > \cdots > M_{(K)}(t)$, where $M_{(1)}(t)$ denotes the largest firm, $M_{(2)}(t)$ the second largest firm, etc. Then we have

$$\log M_{(j)}(t) = \log \Theta(t) + \log(X_{(j)}(t)), \tag{5.4}$$

where $X_{(j)}(t)$ are the ranked values of $X_j(t)$, $1 \le j \le K$. Since the first term Θ in (5.4) is common for all firms, the plot of $\log M_{(j)}(t)$ versus $\log j$ and the plot of $\log X_{(j)}(t)$ versus $\log j$ will display similar patterns. Therefore, we first focus on the relationship between $\log X_{(j)}(t)$ and $\log j$.

As $X_{(1)}(t), X_{(2)}(t), \dots, X_{(K)}(t)$ are the ordered realizations of $X(t)$, the empirical tail distribution $\tilde{F}(x)$ (the empirical version of F) evaluated at $X_{(i)}(t)$ is simply $\tilde{F}(X_{(i)}(t)) = i/N$, $i = 1, \dots, K$. Next we make two assumptions.

Assumption 5.1. *The birth–death process has reached the steady state.*

Assumption 5.2. *For each stock included in the group, the market capitalization is large; that is, even $X_{(K)}(t)$ is large.*

According to (5.2), in the steady state, for large capitalization n,

$$\log F(n) \cong \log\left\{\frac{1}{\Gamma(g/\lambda)}\left(1-\frac{\lambda}{\mu}\right)^{g/\lambda-1}\right\} + n\log\left(\frac{\lambda}{\mu}\right) - \left(1-\frac{g}{\lambda}\right)\log(n).$$

Therefore, under Assumptions 5.1 and 5.2 empirically with $n = X_{(i)}$, we shall expect that

$$\begin{aligned}\log \tilde{F}(X_{(i)}(t)) &= \log\left(\frac{i}{N}\right)\\ &\approx \log\left\{\frac{1}{\Gamma(g/\lambda)}\left(1-\frac{\lambda}{\mu}\right)^{g/\lambda-1}\right\} + X_{(i)}(t)\log\left(\frac{\lambda}{\mu}\right) - \left(1-\frac{g}{\lambda}\right)\log(X_{(i)}(t)).\end{aligned}$$

Rearranging the terms above yields that

$$\log X_{(i)}(t) \approx C - \frac{1}{1-g/\lambda}\log i + \frac{1}{1-g/\lambda}X_{(i)}(t)\log\left(\frac{\lambda}{\mu}\right), \qquad 1 \le i \le K, \tag{5.5}$$

where the constant term

$$C = \frac{1}{1-g/\lambda}\log\left\{\frac{1}{\Gamma(g/\lambda)}\left(1-\frac{\lambda}{\mu}\right)^{g/\lambda-1}\right\} + \frac{\log(N)}{1-g/\lambda}. \tag{5.6}$$

Since N is unknown, C is essentially a free parameter.

Equation (5.5) provides a link between the ordered values of $X(t)$ and their relative ranks. However, since it involves a nuisance parameter C, a better equation can be obtained by eliminating C first, as is typical in many standard statistical procedures. To do this, observe that when $i = 1$ we have

$$\log X_{(1)} \approx C - \frac{1}{1-g/\lambda}\log 1 + \frac{1}{1-g/\lambda}X_{(1)}\log\left(\frac{\lambda}{\mu}\right). \tag{5.7}$$

Taking the difference between (5.5) and (5.7) cancels out the nuisance constant C and gives

$$\log\frac{X_{(i)}(t)}{X_{(1)}(t)} \approx -\frac{1}{1-g/\lambda}\log i + \frac{1}{1-g/\lambda}(X_{(i)}(t) - X_{(1)}(t))\log\left(\frac{\lambda}{\mu}\right), \qquad 1 \le i \le K.$$

Now substituting it into (5.4) we have, for $1 \le i \le K$,

$$\log\frac{M_{(i)}(t)}{M_{(1)}(t)} \approx -\frac{1}{1-g/\lambda}\log i + \frac{1}{1-g/\lambda}(M_{(i)}(t) - M_{(1)}(t))\log\left[\left(\frac{\lambda}{\mu}\right)^{1/\Theta(t)}\right]. \tag{5.8}$$

Equation (5.8) is the key cross-sectional empirical implication in this paper. It will lead to an 'almost' linear curve if the third term in (5.8) is small. This can be achieved if we make two more assumptions.

Assumption 5.3. *For both growth stocks and traditional stocks,* $\lambda/\mu \approx 1$.

Assumption 5.4. *For growth stocks, both* λ *and* μ *must be large.*

Assumption 5.3 is postulated because, generally, it is hard to predict instantaneous upward and downward price movements for both growth stocks and nongrowth stocks; thus, λ and μ must be quite close. Assumption 5.4 reflects the high volatility of growth stocks. Indeed, Kerins *et al.* (2001) show empirically that the volatility of internet stocks may be at least five times as high as that of traditional stocks. With Assumptions 5.3 and 5.4, we have the following lemma.

Lemma 5.1. *Suppose that* $\xi := \lambda/\mu \to 1$ *and* $g/\lambda \to 0$. *Then, in the steady state, the second term on the right-hand side of (5.8) goes to zero in* $\mathcal{L}_1$. *In other words, under these conditions, (5.8) is asymptotically a linear relationship.*

Proof. It is enough to prove that

$$\frac{1}{1-g/\lambda}\,\mathrm{E}(X_{(i)}(t)-X_{(1)}(t))\log\biggl(\frac{\lambda}{\mu}\biggr)\to 0, \qquad 1\le i\le K.$$

Since $0 \le X_{(i)}(t) \le X_{(1)}(t)$, we only need to show that

$$\frac{1}{1-g/\lambda}\,\mathrm{E}[X_{(1)}(t)]\log\biggl(\frac{\lambda}{\mu}\biggr)\to 0.$$

Noting that $\mathrm{E}[X_{(1)}(t)] \le N\,\mathrm{E}[X(t)]$, the problem is further reduced to showing that

$$\frac{1}{1-g/\lambda}\,\mathrm{E}[X(t)]\log\biggl(\frac{\lambda}{\mu}\biggr)\to 0.$$

In the steady state, using (5.3), this becomes

$$\frac{1}{1-g/\lambda}\,\frac{g}{\mu}\,\frac{\log(\xi)}{1-\xi}\to 0,$$

which is true because $\log(\xi)/(1-\xi) \to -1$ as $\xi \to 1$.

As we shall see in Section 6, g/λ indeed tends to be small in the numerical examples, ranging from 0.08 to 0.36. It is worth mentioning here that, even if g/λ is not small, the cross-sectional equation (5.8) still holds.

In addition, note also that Assumption 5.4 implies that the decay parameter $\gamma = \mu - \lambda$ (which affects the convergence in an exponential way) may also be large for growth stocks, thus leading to a fast convergence to the steady-state distribution and justifying Assumption 5.1.

With Assumptions 5.1–5.4, we note that the cross-sectional equation (5.8) postulates a relationship between the logarithm of the normalized (by the largest value) market caps and the logarithm of the ranks of 'large-cap' growth stocks (those satisfying Assumption 5.2). The term 'large-cap' here is used in a loose sense, and should not be confused with similar words used in stock exchanges; more precisely, it means that the market capitalization is large enough so that the asymptotic result (5.2) holds.

Equation (5.8) also implies that the same cross-sectional phenomenon should hold not only for large-cap internet stocks (as reported in the *Wall Street Journal* article) but also for other large-cap growth stocks, such as large-cap biotechnology stocks, with large λ and μ.

Remark 5.1. It may be worthwhile to point out the connection between the new cross-sectional implication (5.8) and the empirical puzzle reported in the *Wall Street Journal*, which amounts to

$$\log M_{(i)}(t) \approx a(t) + b(t) \log i,$$

where $a(t)$ and $b(t)$ do not depend on the index i. The new cross-sectional implication (5.8) appears to have a better fit to the data, as judged by the high R^2 statistic (see Section 6). There are two reasons for the improvement: (i) we eliminate the nuisance parameter by using the relative market caps $M_{(i)}/M_{(1)}$; (ii) we do not require the third term in (5.8) to be zero or small when we fit the model in Section 6, thus leading to a better fitting.

5.3. Why the model does not apply to nongrowth stocks

There are at least two reasons why the cross-sectional relationship in (5.8) between the logarithm of the market capitalization and the logarithm of the ranks should not be expected for nongrowth stocks. First, the birth–death process model may not be valid for nongrowth stocks. Secondly, even if the model is valid for nongrowth stocks, in order to empirically observe a phenomenon such as that implied by (5.8), several conditions must be satisfied, as (5.8) is based on the steady-state distribution.

Condition 5.1. *In terms of time, the convergence from the transient states to the steady state must be fast enough. This in turn depends on the magnitude of the decay parameter γ; in other words, γ must be large.*

Condition 5.2. *In terms of market capitalization, X and, hence, M must be large enough, as required by the asymptotic results in (5.2) and (5.5).*

For large-cap growth stocks (thus satisfying Condition 5.2 above), by Assumption 5.4, both λ and μ are large. So, if $\mu - \lambda$ is large, then the decay parameter γ is also large, thus resulting in a fast convergence to the steady state.

For nongrowth stocks, the volatility parameters, which in our model are λ and μ, are generally not large. As a consequence, the decay parameter $\gamma = \mu - \lambda$ (which affects the convergence in an exponential way) cannot be large in general. In other words, although in the steady state plotting the logarithm of the market capitalization against the logarithm of the relative ranks may display a certain relationship, the relationship may not emerge at all within a reasonable amount of time, due to the slow convergence from the transient state to the steady state. Furthermore, if the convergence rate is slow, many factors can lead the process to depart from the original steady state, e.g. changing of λ and μ, etc.

5.4. Further remarks about the cross-sectional implication

As is common for many cross-sectional studies, (5.8) should only be viewed as an understanding of growth stocks as a whole rather than as a trading tool, because we did not provide dynamics of the relative ranks for growth stocks. However, cross-sectional implications may lead to some useful economic models. In fact, it can be shown that a dynamic equilibrium model can be built as a result of the model proposed in the current paper, and the model is also linked to endogenous stochastic growth theory in macroeconomics; the details of the macroeconomic justification of the current model, being too long to be included here, are given in Kou and Kou (2002).

The size distribution in (5.8) is quite different from that observed for the city-size distribution. In the city-size distribution, the exponent of the power law (i.e. the slope of regressing log-city-size on log-city-rank) is very close to -1 (see, for example, Krugman (1996a), Gabaix (1999)). But here the exponent $-1/(1 - g/\lambda)$ is less than -1, as we shall see in Section 6. Furthermore, as pointed out in Gabaix (1999) and Krugman (1996b, pp. 96–97), it could take a process too much time to converge to the steady-state distribution (which is the power law), if the volatility of city-growth rates is not large; this, consequently, poses a serious problem for using birth–death processes to model the city-size distribution. However, in our case the volatility of growth stocks tends to be much higher than that of nongrowth stocks and that of city sizes. Therefore, the growth stocks tend to converge to the steady state much faster, resulting in a clear pattern of the size distribution (as shown in Section 6). This also underlines the importance of studying the transient behavior of the size distribution.

The regression using (5.8) is robust against possible truncation errors, thanks to the fact that the relative ranks are used. For example, if there are 200 growth stocks in total and only the top 100 stocks with large market capitalization are included in the estimation, then (5.8) will not alter.

Equation (5.8) is almost scale invariant: if the unit of $X(t)$ scales up by a factor of A, then g/λ in (5.8) remains the same, while λ/μ becomes $(\lambda/\mu)^A$. However, since $\lambda/\mu \approx 1$, the difference between λ/μ and $(\lambda/\mu)^A$ is generally insignificant unless A is extremely large.

6. Numerical illustrations of the cross-sectional size distribution

To illustrate the cross-sectional size distribution for biotechnology stocks, we plot in Figure 2 the logarithm of their market capitalization relative to the largest biotechnology stock versus the logarithm of their ranks. In other words, $\log(M_{(i)}/M_{(1)})$ is plotted against $\log i$. This

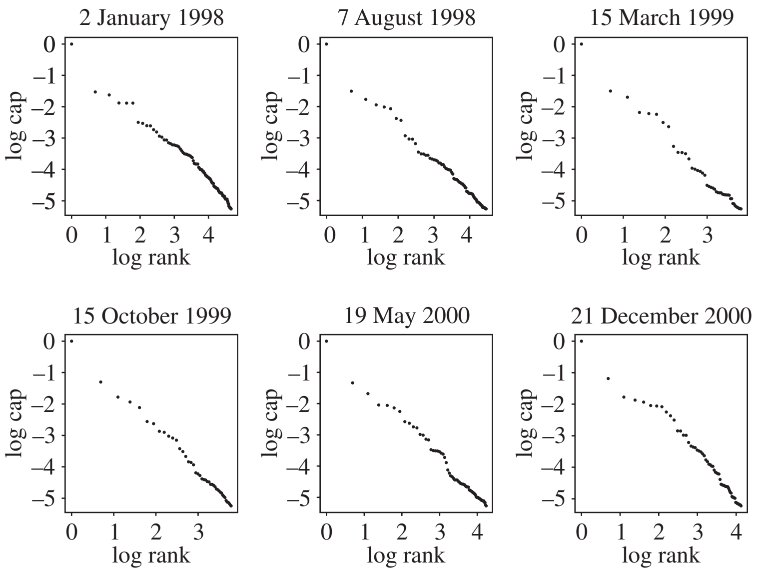

FIGURE 2: Size distribution of the biotechnology stocks.

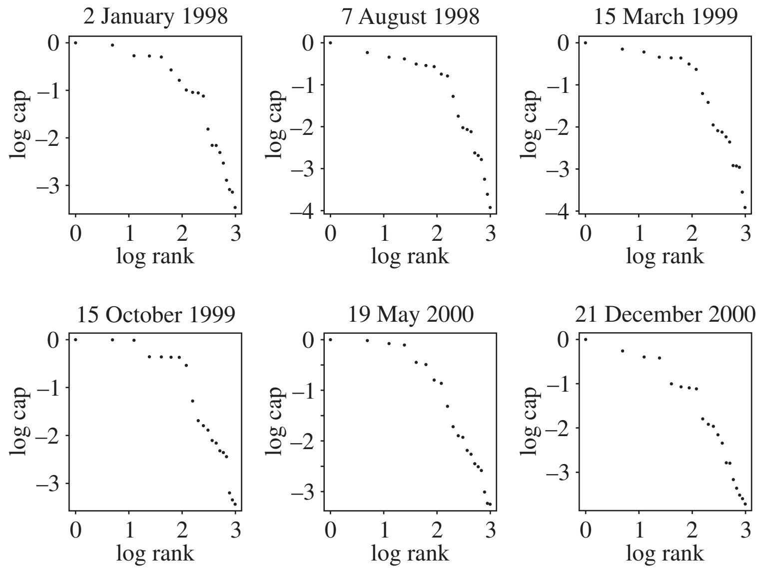

FIGURE 3: Plot for the 20 stocks of the Dow Jones Transportation Average.

can be viewed as choosing $M_{(1)}$ as the unit of measurement. The six graphs, which involve 139 biotechnology stocks, display results for 2 January 1998 and every 150 trading days thereafter. The 139 stocks include most of the stocks listed in the Nasdaq® biotechnology index and the Amex® biotechnology index (BTK). See Appendix B for a description of all the stocks used in the paper.

In each panel, the total market capitalizations of these 139 stocks are first computed by taking the product of the number of outstanding shares and the share price; then the stocks with market capitalization not smaller than 0.5% of that of the largest stock are plotted. The relationship (5.8) requires large market capitalization, and here 'large-cap' is taken (ad hoc) to be stocks having market capitalization at least as large as 0.5% of that of the largest stock. One advantage of categorizing 'largeness' relatively is that it automatically takes into account the fact that different groups of stocks could have different sizes (for example, even within growth stocks, internet stocks tend to be larger than biotechnology stocks).

It is worth noting that the six days shown in Figure 2 include days when the biotechnology stocks were performing well, as well as days when the biotechnology stocks were grounded heavily. Nevertheless, in all six graphs there is clearly an almost linear trend. In contrast, in Figures 3 and 4, for the same six trading days, the logarithm of the market capitalization of the 20 stocks of the Dow Jones Transportation Average and 88 saving and loan stocks relative to the largest one is plotted against the logarithm of their ranks. We include the saving and loan stocks because, based on the data of a single day, Mauboussin and Schay (2000) later (but not in the *Wall Street Journal*) stated that saving and loan stocks may show 'strong power-law characteristics' as well; so we want to investigate the issue.

Among the 20 transportation stocks, the smallest has a market capitalization about 2% of that of the largest. The 88 saving and loan stocks are described in Appendix B. The plot of the transportation stocks is far from linear. For the saving and loan stocks, based on the data of three years, from 1998 to 2000, it is fair to say that, although on some days there may be a linear

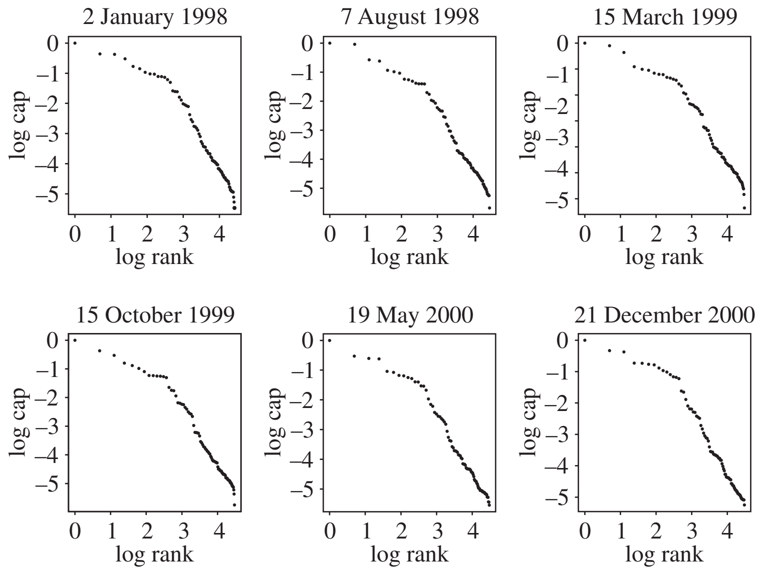

FIGURE 4: Plot for the saving and loan stocks.

pattern, the pattern disappears on other days and is not at all consistent. This is again expected from the model, since the convergence of nongrowth stocks to the steady-state distribution (governed by the decay parameter) is generally too slow to be observed in practice.

For the biotechnology stocks in Figure 2, the parameters can be estimated by fitting the model (5.8) to the data. More precisely, at any time t, the estimates $\widehat{g/\lambda}$ and $\hat{\beta}$ can be simply obtained by minimizing the squared errors for $\log(M_{(i)}/M_{(1)})$:

$$(\widehat{g/\lambda}, \hat{\beta}) = \arg\min_{(g/\lambda,\beta)} \sum_{i=1}^{K} \biggl[\log \frac{M_{(i)}}{M_{(1)}} - \biggl\{ -\frac{1}{1-g/\lambda} \log i + \frac{1}{1-g/\lambda} (M_{(i)} - M_{(1)}) \log \beta \biggr\} \biggr]^2, \tag{6.1}$$

with the constraints that $g/\lambda > 0$ and $0 < \beta < 1$, where $\beta = (\lambda/\mu)^{1/\Theta}$. Equation (6.1) is considered here mainly because (a) it is easy to implement; and (b) the focus here is an illustration of the cross-sectional size distribution, rather than an empirical test of the model.

We shall point out that there are other ways, such as likelihood-based methods, to estimate the parameters, which might be more efficient. The parameter $1 - g/\lambda$ is also related to the tail index for power-law distributions. In the above procedure we estimate g/λ by fitting a parametric model in (6.1). An alternative way to estimate g/λ is to use the nonparametric Hill estimator or its extension such as the Hill ratio plot; see, for example, Hill (1975), Lo (1986), Embrechts *et al.* (1997), Adler *et al.* (1998), Drees *et al.* (2000). However, our numerical experiment suggests that the Hill ratio plot may be too volatile to give a reliable estimator of g/λ, mainly because here the sample size (about 70 internet stocks and 139 biotechnology stocks) is too small for the nonparametric method to be effective. This seems to be consistent with the observation in the literature that a large sample size (in the magnitude of thousands

TABLE 1: The R^2 and estimated g/λ and β for biotechnology stocks.

	$\widehat{g/\lambda}$	$1-\hat{\beta}$	R^2
2 January 1998	0.080	1.38×10^{-9}	97.8%
7 August 1998	0.165	1.25×10^{-9}	98.2%
15 March 1999	0.295	1.06×10^{-9}	98.3%
15 October 1999	0.272	1.09×10^{-9}	99.2%
19 May 2000	0.197	1.20×10^{-9}	98.6%
21 December 2000	0.265	5.65×10^{-9}	97.5%

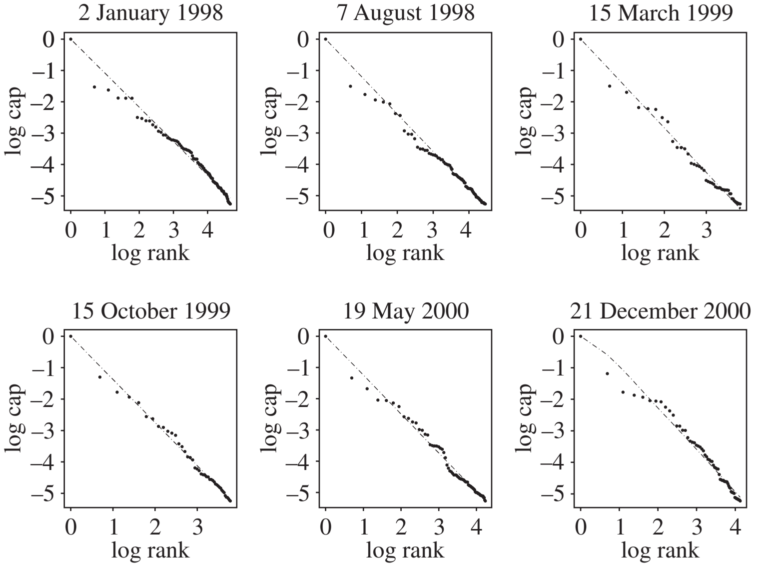

FIGURE 5: Empirical and estimated size distribution for biotechnology stocks.

or more) may be needed for the Hill ratio plot to produce reliable estimates; see, for example, Resnick (1997) and Heyde and Kou (2002).

Table 1 reports the estimates $\widehat{g/\lambda}$ and $\hat{\beta}$ as well as R^2, which measures the goodness-of-fit for all six graphs in Figure 2. Similar to the linear regression, here R^2 is simply defined as $1-$ (variance of the residuals)/(variance of the observed responses). Note that the $\widehat{g/\lambda}$ are all small. Furthermore, the $\hat{\beta}$ are all very close to 1; this hints that Θ might be quite large.

Using the estimated values of g/λ and β, the dashed lines in Figure 5 show the relationship between the log-market capitalization and the log-rank, as suggested by the model. They agree well with the empirical observation.

As a further illustration, Figure 6 shows the cross-sectional size distribution for internet stocks. The six graphs represent 4 January 1999 and every 100 trading days onward. In total, 70 internet stocks are involved. See Appendix B for details. The plot starts from 4 January 1999 because there were not many internet stocks before 1999. Again the expected pattern emerges. Table 2 reports the estimated parameters and R^2 for the internet stocks.

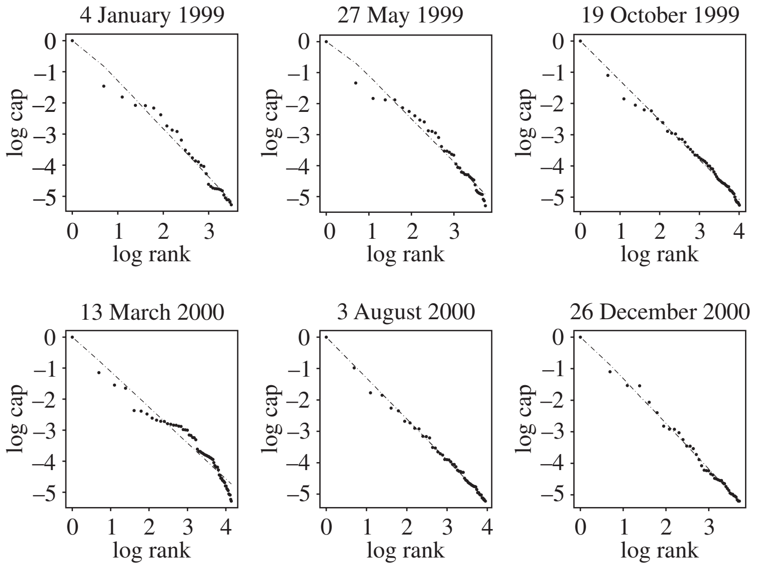

FIGURE 6: Empirical and estimated size distribution for internet stocks.

TABLE 2: The R^2 and estimated g/λ and β for internet stocks.

	$\widehat{g/\lambda}$	$1-\hat{\beta}$	R^2
4 January 1999	0.365	1.47×10^{-6}	97.3%
27 May 1999	0.298	1.60×10^{-6}	96.8%
19 October 1999	0.211	1.18×10^{-9}	99.0%
13 March 2000	0.135	1.12×10^{-7}	94.0%
3 August 2000	0.234	1.15×10^{-9}	99.5%
26 December 2000	0.315	4.51×10^{-7}	99.4%

We conclude this section by presenting a picture of the recent market. Figure 7 shows the size distribution of biotechnology and internet stocks as of 22 August 2001. The pattern for biotechnology and internet stocks expected by the model again emerges. Table 3 reports the estimated parameters and R^2. Note that the 'internet bubble' had burst by then; for example, the American stock exchange internet index (IIX) was 688.52 on 27 March 2000 and was only 141.21 on 22 August 2001. The fitting is good even under this severe market downturn.

Remark 6.1. Although the cross-sectional R^2 is encouraging for biotechnology and internet stocks, we should be very careful in interpreting the numerical results. First, the numerical results shown here only amount to an illustration of the cross-sectional size distribution, and they do not serve as an empirical test of the model. Secondly, the parameter estimate of g/λ seems to change, although perhaps slowly, over time. This may suggest some possible future research to extend the current model to include time-varying parameters (as in stochastic volatility models).

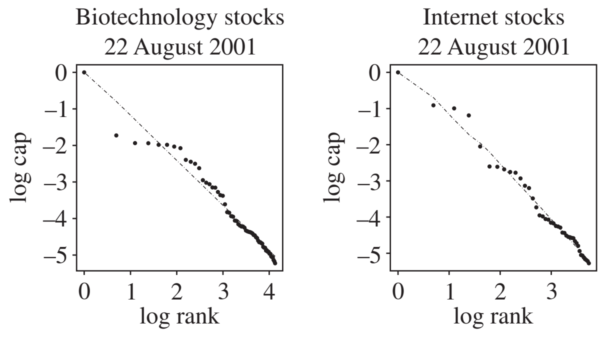

FIGURE 7: The size distribution for the recent market (22 August 2001).

TABLE 3: The estimated parameters and R^2 for the recent market (22 August 2001).

	$\widehat{g/\lambda}$	$1-\hat{\beta}$	R^2
Biotechnology stocks	0.192	9.23×10^{-7}	96.4%
Internet stocks	0.362	3.43×10^{-6}	98.5%

7. Discussion

By utilizing the high volatility of growth stocks, this paper proposes, based on both the transient and steady-state behavior of birth–death processes, a model for growth stocks, which are otherwise quite difficult to analyze using traditional valuation methods (partly because of the difficulties in predicting the future growth rate of the earnings).

The main contribution of the current paper is that it provides an understanding of the size distribution for growth stocks, by building a stochastic model. There are three useful properties of the model. First, the model leads to a cross-sectional equation (5.8) for growth stocks, including both biotechnology and internet stocks. Second, the cross-sectional model only uses regression and relative ranks, and is, thus, easy to implement. Third, the cross-sectional model remains valid irrespective to the market ups and downs, mainly because the model compares the relative value of a stock against the other stocks within its peer group.

There are several limitations of the model, which may be of interest for future research:

(a) A problem is the possible effect of merger and acquisition. For example, currently (as of 2001 and 2002) internet stocks have more merger and acquisition activity than biotechnology stocks. Thus, from this point of view, the current model is perhaps more suitable for biotechnology stocks than for internet stocks.

(b) The model only applies to growth stocks with a large enough market capitalization, i.e. large-cap growth stocks. It does not attempt to provide a solution to small-cap growth stocks.

(c) The model focuses on market capitalization, and does not take other possible factors (e.g. the outstanding debt of companies) into account. One intuitive explanation of why the fit is good without including the debt is that most growth companies may not use bonds as a major way of financing.

Because of these limitations, as a cautionary remark, the model is only intended to provide a quick and first-order, cross-sectional approximation to a difficult yet important problem: how to value volatile growth stocks without any earnings.

Appendix A. Proofs of Lemmas 3.1 and 3.2

A.1. Proof of Lemma 3.1

(i) To show that the birth–death process is positive recurrent, it is enough to check that

$$\sum_{n=0}^{\infty} \frac{1}{\lambda_n \pi_n} = \infty \quad \text{and} \quad \sum_{n=0}^{\infty} \pi_n < \infty;$$

see Kijima (1997, p. 245). The result follows as $\sum_{n=0}^{\infty} 1/\lambda_n \pi_n$ has the same order as

$$\sum_{n=0}^{\infty} \frac{1}{n\lambda + g} \left(\frac{\mu}{\lambda}\right)^n n^{1-g/\lambda+h/\mu} = \infty,$$

and $\sum_{n=0}^{\infty} \pi_n$ has the same order as

$$\sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n n^{g/\lambda-h/\mu-1} < \infty,$$

thanks to the assumption that $\mu > \lambda$.

(ii) The equation (3.2) follows from the fact that

$$\lim_{z\to\infty} z^{b-a} \frac{\Gamma(z+a)}{\Gamma(z+b)} = 1.$$

(iii) First we consider $\sum_{k=n}^{\infty} \pi_k$, which, according to (3.1), is

$$\alpha \sum_{k=n}^{\infty} \left(\frac{\lambda}{\mu}\right)^k \frac{\Gamma(k+g/\lambda)}{\Gamma(k+1+h/\mu)},$$

where $\alpha := \Gamma(1+h/\mu)/\Gamma(g/\lambda)$. The definition of the hypergeometric function yields that

$$\begin{aligned} \sum_{k=n}^{\infty} \pi_k &= \alpha \left(\frac{\lambda}{\mu}\right)^n \frac{\Gamma(n+g/\lambda)}{\Gamma(n+1+h/\mu)} F\left(n+\frac{g}{\lambda}, 1; n+1+\frac{h}{\mu}; \frac{\lambda}{\mu}\right) \\ &= \pi_n F\left(n+\frac{g}{\lambda}, 1; n+1+\frac{h}{\mu}; \frac{\lambda}{\mu}\right) \quad \text{for } n \geq 0. \end{aligned} \tag{A.1}$$

In particular, we obtain that

$$S = \sum_{n=0}^{\infty} \pi_n = \pi_0 F\left(\frac{g}{\lambda}, 1; 1+\frac{h}{\mu}; \frac{\lambda}{\mu}\right) = F\left(\frac{g}{\lambda}, 1; 1+\frac{h}{\mu}; \frac{\lambda}{\mu}\right).$$

The moment-generating function is given by

$$\begin{aligned}\eta(\theta) &= \sum_{n=0}^{\infty} \frac{\mathrm{e}^{\theta n}\pi_n}{S} \\ &= \frac{1}{S}\sum_{n=0}^{\infty} \alpha\left(\frac{\lambda}{\mu}\mathrm{e}^{\theta}\right)^n \frac{\Gamma(n+g/\lambda)}{\Gamma(n+1+h/\mu)} \\ &= \frac{F(g/\lambda, 1; 1+h/\mu; (\lambda/\mu)\mathrm{e}^{\theta})}{S} \\ &= \frac{F(g/\lambda, 1; 1+h/\mu; (\lambda/\mu)\mathrm{e}^{\theta})}{F(g/\lambda, 1; 1+h/\mu; \lambda/\mu)}.\end{aligned}$$

The results about the mean and the second moment follow easily via the following property of the hypergeometric function (see also Formula 15.2.1 of Abramowitz and Stegun (1972)):

$$\frac{\mathrm{d}}{\mathrm{d}z}F(a, b; c; z) = \frac{ab}{c}F(a+1, b+1; c+1; z).$$

(iv) By (A.1), $F(n) = \pi_n F(n + g/\lambda, 1; n + 1 + h/\mu; \lambda/\mu)/S$. So we only have to study the limiting behavior of $F(n + g/\lambda, 1; n + 1 + h/\mu; \lambda/\mu)$. Using Formula 15.3.5 of Abramowitz and Stegun (1972), $F(a, b; c; z) = (1 - z)^{-b}F(b, c - a; c; z/(z - 1))$, we have

$$F\left(n+\frac{g}{\lambda}, 1; n+1+\frac{h}{\mu}; \frac{\lambda}{\mu}\right) = \left(1-\frac{\lambda}{\mu}\right)^{-1} F\left(1, 1+\frac{h}{\mu}-\frac{g}{\lambda}; n+1+\frac{h}{\mu}; \frac{\lambda}{\lambda-\mu}\right).$$

But from the definition of the hypergeometric function, it is easily seen that $F(1, 1+h/\mu-g/\lambda; n+1+h/\mu; \lambda/(\lambda-\mu)) \to 1$ (see Section 2.3.2 of Erdélyi *et al.* (1953)). Therefore, we obtain

$$F(n) \cong \frac{\alpha}{S}\left(1-\frac{\lambda}{\mu}\right)^{-1}\left(\frac{\lambda}{\mu}\right)^n n^{g/\lambda-h/\mu-1},$$

which completes the proof of Lemma 3.1.

A.2. Proof of Lemma 3.2

We start from the forward Kolmogorov equations of a birth–death process (see Karlin and Taylor (1975, p. 136)):

$$\begin{aligned}p_{i,0}'(t) &= -\lambda_0 p_{i,0}(t) + \mu_1 p_{i,1}(t), \\ p_{i,j}'(t) &= \lambda_{j-1}p_{i,j-1}(t) - (\lambda_j + \mu_j)p_{i,j}(t) + \mu_{j+1}p_{i,j+1}(t), \qquad j \ge 1,\end{aligned}$$

which in our case become

$$\begin{aligned}p_{i,0}'(t) &= -gp_{i,0}(t) + (\mu + h)p_{i,1}(t), \\ p_{i,j}'(t) &= (\lambda(j-1)+g)p_{i,j-1}(t) - ((\lambda+\mu)j+g+h)p_{i,j}(t) + (\mu(j+1)+h)p_{i,j+1}(t), \\ & \qquad j \ge 1.\end{aligned}$$

Multiplying the jth equation by j and taking a sum yields $m_1'(t) = (\lambda - \mu)m_1(t) + g + h(1 - p_{i,0}(t))$, with the initial condition $m_1(0) = i$. Similarly, multiplying the jth equation by j^2 and summing leads to $m_2'(t) = 2(\lambda-\mu)m_2(t)+(\lambda+\mu+2g-2h)m_1(t)+g+h(1-p_{i,0}(t))$, with $m_2(0) = i^2$, from which the result follows.

Appendix B. Description of the stocks used in the numerical illustration

Except for the stocks (e.g. non-US stocks) that are not included in the Center for Research in Security Prices (CRSP) historical database and the stocks that no longer exist because of merger or bankruptcy, we use all the biotechnology stocks included in the Nasdaq biotechnology index (IXBT) and the Amex biotechnology index (BTK); all the internet stocks included in the Amex internet index (IIX), the Dow Jones composite internet index (DJINET), the Street.com internet index (DOT), the Amex Internet Infrastructure HOLDRS (IIH), the Amex B2B Internet HOLDRS (BHH), and the Amex Internet HOLDRS (HHH); and all the saving and loan stocks included in the Philadelphia exchange bank index (BKX), the S&P bank index (BIX), the regional bank HOLDRS (RKH), and the Nasdaq Financial-100 index (IXF). A detailed list of all the stocks can be obtained from the authors.

Acknowledgements

We are very grateful to an anonymous referee for many helpful comments and to Morris Cohen at the University of Pennsylvania for pointing out the article in the *Wall Street Journal* during his talk. We also want to thank many people who offered insights into this work including Mark Broadie, Martin Haugh, Yuji Ijiri, Paul Glasserman, David Weinstein, Hui Wang, and Willard Zangwill. A four-page summary of a preliminary version of the paper appeared in *Risk* magazine (December 2001). This research is supported in part by NSF grants DMS-0204674, DMS-0074637, and DMI-0216979.

References

Abramowitz, M. and Stegun, I. A. (eds) (1972). *Handbook of Mathematical Functions*. National Bureau of Standards, Washington, DC.

Adler, R. J., Feldman, R. E. and Taqqu, M. S. (eds) (1998). *A Practical Guide to Heavy Tails*. Birkhäuser, Boston, MA.

Axtell, R. L. (2001). Zipf distribution of U.S. firm sizes. *Science* **293,** 1818–1820.

Chen, W. C. (1980). On the weak form of Zipf's law. *J. Appl. Prob.* **17,** 611–622.

Drees, H., de Haan, L. and Resnick, S. (2000). How to make a Hill plot. *Ann. Statist.* **28,** 254–274.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events*. Springer, Berlin.

Erdélyi, A. *et al.* (1953). *High Transcendental Functions* (Bateman Manuscript Project), Vol. 1. McGraw-Hill, New York.

Feenberg, D. R. and Poterba, J. M. (1993). Income inequality and the incomes of very high-income taxpayers: evidence from tax returns. In *Tax Policy and the Economy*, Vol. 7, ed. J. M. Poterba, MIT Press, Cambridge, MA, pp. 145–177.

Gabaix, X. (1999). Zipf's law for cities: an explanation. *Quart. J. Econom.* **154,** 739–767.

Gibrat, R. (1931). *Les Inégalités Économiques*. Recueil Sirey, Paris.

Glaeser, E., Scheinkman, J. and Shleifer, A. (1995). Economic growth in a cross-section of cities. *J. Monetary Econom.* **36,** 117–143.

Heyde, C. C. and Kou, S. G. (2002). On the controversy over tailweight of distributions. Preprint, Columbia University.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3,** 1163–1174.

Ijiri, Y. and Simon, H. A. (1977). *Skew Distributions and the Sizes of Business Firms*. North-Holland, Amsterdam.

Karlin, S. and McGregor, J. (1958). Linear growth, birth and death processes. *J. Math. Mech.* **7,** 643–662.

Karlin, S. and Taylor, H. M. (1975). *A First Course in Stochastic Processes*, 2nd edn. Academic Press, New York.

Kerins, F., Smith, J. K. and Smith, R. (2001). New venture opportunity cost of capital and financial contracting. Preprint, Washington State University.

Kijima, M. (1997). *Markov Processes for Stochastic Modeling*. Chapman and Hall, London.

Kou, S. C. and Kou, S. G. (2002). A tale of two growths: modeling stochastic endogenous growth and growth stocks. Preprint, Harvard University and Columbia University.

Krugman, P. (1996a). Confronting the urban mystery. *J. Japanese Internat. Econom.* **10,** 399–418.

Krugman, P. (1996b). *The Self-Organizing Economy*. Blackwell, Cambridge, MA.

Lo, G. S. (1986). Asymptotic behavior of Hill's estimator and applications. *J. Appl. Prob.* **23,** 922–936.

LUCAS, R. (1978). On the size distribution of business firms. *Bell J. Econom.* **9,** 508–523.

MANDELBROT, B. B. (1960). The Pareto–Lévy law and the distribution of income. *Internat. Econom. Rev.* **1,** 79–106.

MANDELBROT, B. B. (1997). *Fractals and Scaling in Finance.* Springer, New York.

MAUBOUSSIN, M. J. AND SCHAY, A. (2000). Still powerful: the internet's hidden order. Equity Res. Rep., Credit Suisse First Boston Corporation.

MEYN, S. P. AND TWEEDIE, R. (1993). *Markov Chains and Stochastic Stability*. Springer, London.

PARETO, V. (1896). *Cours d'Économie Politique*. Rouge, Lausanne.

RESNICK, S. (1997). Heavy tail modelling and teletraffic data. *Ann. Statist.* **25,** 1805–1869.

RUTHERFORD, R. (1955). Income distributions: a new model. *Econometrica* **23,** 425–440.

SHORROCKS, A. F. (1975). On stochastic models of size distributions. *Rev. Econom. Studies* **42,** 631–641.

SIMON, H. A. (1955). On a class of skew distribution functions. *Biometrika* **52,** 425–440.

SIMON, H. A. AND BONINI, C. P. (1958). The size distribution of business firms. *Amer. Econom. Rev.* **48,** 607–617.

STEINDL, J. (1965). *Random Processes and the Growth of Firms*. Hafner, New York.

WOODROOFE, M. AND HILL, B. M. (1975). On Zipf's law. *J. Appl. Prob.* **12,** 425–434.

YULE, G. U. (1924). A mathematical theory of evolution, based on the conclusions of Dr. J. R. Willis. *Phil. Trans. R. Soc. London B* **213,** 21–83.

YULE, G. U. (1944). *The Statistical Study of Literary Vocabulary*. Cambridge University Press.

ZIPF, G. (1949). *Human Behavior and the Principle of Least Effort*. Addison-Wesley, Cambridge, MA.