A generalized preferential attachment model for business firms growth rates

II. Mathematical treatment

S.V. Buldyrev[1,a], F. Pammolli[2,3], M. Riccaboni[2,3], K. Yamasaki[4], D.-F. Fu[5], K. Matia[5], and H.E. Stanley[5]

[1] Department of Physics, Yeshiva University, 500 West 185th Street, New York, NY 10033, USA
[2] Faculty of Economics, University of Florence, Milan, Italy
[3] IMT Institute for Advanced Studies, via S. Micheletto 3, 55100 Lucca, Italy
[4] Tokyo University of Information Sciences, Chiba City 265-8501, Japan
[5] Center for Polymer Studies and Department of Physics, Boston University, Boston, MA 02215, USA

Abstract. We present a preferential attachment growth model to obtain the distribution $P(K)$ of number of units K in the classes which may represent business firms or other socio-economic entities. We found that $P(K)$ is described in its central part by a power law with an exponent $\varphi = 2 + b/(1-b)$ which depends on the probability of entry of new classes, b. In a particular problem of city population this distribution is equivalent to the well known Zipf law. In the absence of the new classes entry, the distribution $P(K)$ is exponential. Using analytical form of $P(K)$ and assuming proportional growth for units, we derive $P(g)$, the distribution of business firm growth rates. The model predicts that $P(g)$ has a Laplacian cusp in the central part and asymptotic power-law tails with an exponent $\zeta = 3$. We test the analytical expressions derived using heuristic arguments by simulations. The model might also explain the size-variance relationship of the firm growth rates.

PACS. 89.75.Fb Structures and organization in complex systems – 89.65.Gh Economics; econophysics, financial markets, business and management

1 Introduction

Here we introduce a mathematical framework that provides an unifying explanation for the growth of business firms based on the number and size distribution of their elementary constituent components [1–8]. Specifically we present a model of proportional growth in both the number of units and their size and we draw some general implications on the mechanisms which sustain business firm growth [4,9–13]. According to the model, the probability density function (PDF) of growth rates, $P(g)$ is Laplace [14] in the center [15] with power law tails [16,17] decaying as $g^{-\zeta}$ where $\zeta = 3$.

Two key sets of assumptions in the model are described in subsections A (the number of units K in a class grows in proportion to the existing number of units) and B (the size of each unit fluctuates in proportion to its size). Our goal is to first find $P(K)$, the probability distribution of the number of units in the classes at large t, and then find $P(g)$ using the convolution of $P(K)$ and the conditional distribution of the class growth rates $P(g|K)$, which for large K converges to a Gaussian.

[a] e-mail: buldyrev@yu.edu

2 Analytical results

2.1 The proportional growth of number of units

The first set of assumptions [18] is:

(A1) each class α consists of $K_\alpha(t)$ number of units. At time $t = 0$, there are $N(0)$ classes consisting of $n(0)$ total number of units. The initial average number of units in a class is thus $n(0)/N(0)$;

(A2) at each time step a new unit is created. Thus the number of units at time t is $n(t) = n(0) + t$;

(A3) with birth probability b, this new unit is assigned to a new class, so that the average number of classes at time t is $N(t) = N(0) + bt$;

(A4) with probability $1-b$, a new unit is assigned to an existing class α with probability $P_\alpha = (1-b)K_\alpha(t)/n(t)$, so $K_\alpha(t+1) = K_\alpha(t) + 1$.

This model can be generalized to the case when the units are born at any unit of time t' with probability μ, die with probability λ, and in addition a new class consisting of one unit can be created with probability b' [18]. This model can be reduced to the present model if one introduce time $t = t'(\mu - \lambda + b')$ and probability $b = b'/(\mu - \lambda + b')$.

Our goal is to find $P(K)$, the probability distribution of the number of units in the classes at large t. This model in two limiting cases (i) $b = 0$, $K_\alpha = 1$ ($\alpha = 1, 2 \ldots N(0)$) and (ii) $b \neq 0$, $N(0) = 1$, $n(0) = 1$ has exact analytical solutions $P(K) = N(0)/t(t/(t+N(0)))^K(1+O(1/t))$ [19,20] and $\lim_{t\to\infty} P(K) = (1+b')\Gamma(K)\Gamma(2+b')/\Gamma(K+2+b')$ [21] respectively, where $b' = b/(1-b)$. In general, an exact analytical solution of this problem cannot be presented in a simple close form. Accordingly, we seek for an approximate mean-field type [22] solution which can be expressed in simple integrals and even in elementary functions in some limiting cases. First we will present a known solution of the preferential attachment model in the absence of the influx of new classes [23]:

$$P_{\text{old}}(K) = \lambda^K \frac{1}{K(t) - 1}$$
$$\approx \frac{1}{K(t)} \exp(-K/K(t))[1 + O(t^{-1})], \quad (1)$$

where $\lambda = 1 - 1/K(t)$ and $K(t) = [n(0)+t]/N(0)$ is the average number of units in the old classes at time t. Note that the form of the distribution of units in the old classes remains unchanged even in the presence of the new classes, whose creation does not change the preferential attachment mechanism of the old classes and affects only the functional form of $K(t)$.

Now we will treat the problem in the presence of the influx of the new classes. Assume that at the beginning there are $N(0)$ classes with $n(0)$ units. Because at every time step, one unit is added to the system and a new class is added with probability b, at moment t there are

$$n(t) = n(0) + t \quad (2)$$

units and approximately

$$N(t) = N(0) + bt \quad (3)$$

classes, among which there are approximately bt new classes with n_{new} units and $N(0)$ old classes with n_{old} units, such that

$$n_{old} + n_{new} = n(0) + t. \quad (4)$$

Because of the preferential attachment assumption (A4), we can write, neglecting fluctuations [22] and assuming that t, n_{old}, and n_{new} are continuous variables:

$$\frac{dn_{new}}{dt} = b + (1-b)\frac{n_{new}}{n(0) + t}, \quad (5)$$

$$\frac{dn_{old}}{dt} = (1-b)\frac{n_{old}}{n(0) + t}. \quad (6)$$

Solving the second differential equation and taking into account initial condition $n_{old}(0) = n(0)$, we obtain $n_{old}(t) = (n(0)+t)^{1-b} n(0)^b$. Analogously, the number of units at time t in the classes existing at time t_0 is

$$n_e(t_0, t) = (n(0)+t)^{1-b}(n(0)+t_0)^b \quad (7)$$

where the subscript 'e' means "existing". Accordingly, the average number of units in old classes is

$$K(t) = \frac{n_{old}(t)}{N(0)} = \frac{(n(0)+t)^{1-b}}{N(0)} \, n(0)^b. \quad (8)$$

Thus according to equation (1), the distribution of units in the old classes is

$$P_{old}(K) \approx$$
$$\frac{N(0)}{(n(0)+t)^{1-b}n(0)^b} \, \exp\left(-\frac{K\,N(0)}{(n(0)+t)^{1-b}n(0)^b}\right), \quad (9)$$

and the contribution of the old classes to the distribution of all classes is

$$\tilde{P}_{old}(K) = P_{old}(K)N(0)/(N(0) + bt). \quad (10)$$

The number of units in the classes that appear at t_0 is $b\,dt$ and the number of these classes is $b\,dt$. Because the probability that a class captures a new unit is proportional to the number of units it has already gotten at time t, the number of units in the classes that appear at time t_0 is

$$n_{new}(t_0, t) = n_e(t_0, t)bdt/[n(0) + t_0]. \quad (11)$$

The average number of units in these classes is

$$K(t_0, t) = n_{new}(t_0, t)/b\,dt = (n(0)+t)^{1-b}/(n(0)+t_0)^{1-b}. \quad (12)$$

Assuming that the distribution of units in these classes is given by a continuous approximation (1) we have

$$P_{new}(K, t_0) \approx \frac{1}{K(t_0, t)} \, \exp\left(-K/K(t_0, t)\right). \quad (13)$$

Thus, their contribution to the total distribution is

$$\frac{b\,dt_0}{N(0) + b\,t} \frac{1}{K(t_0, t)} \, \exp\left(-K/K(t_0, t)\right).$$

The contribution of all new classes to the distribution $P(K)$ is

$$\tilde{P}_{new}(K) \approx \frac{b}{N(0) + b\,t} \int_0^t \frac{1}{K(t_0, t)} \, \exp\left(-K/K(t_0, t)\right) \, dt_0. \quad (14)$$

If we let $y = K/K(t_0, t)$ then $\tilde{P}_{new}(K) = P_{new}(K)bt/(N(0) + bt)$ where

$$P_{new}(K) \approx \frac{n(0)/t + 1}{1 - b} K^{\left(-\frac{1}{1-b}-1\right)} \int_{K'}^K e^{-y} y^{\frac{1}{1-b}} \, dy, \quad (15)$$

and the low limit of integration, K' is given by

$$K' = K\left(\frac{n(0)}{n(0)+t}\right)^{1-b}. \quad (16)$$

Finally the distribution of units in all classes is given by

$$P(K) = \frac{N(0)}{N(0) + bt} P_{old}(K) + \frac{bt}{N(0) + bt} P_{new}(K). \quad (17)$$

Now we investigate the asymptotic behavior of the distribution in equation (15) and show that it can be described by the Pareto power law tail with an exponential cut-off.

1. At fixed K when $t \to \infty$, we have $K' \to 0$, thus

$$P_{new}(K) = \frac{1}{1-b} K^{-\frac{1}{1-b}-1} \int_0^K e^{-y} y^{\frac{1}{1-b}} \, dy,$$

$$= \frac{1}{1-b} \Big[\Gamma\left(1 + \frac{1}{1-b}\right)$$

$$- \int_K^\infty e^{-y} y^{\frac{1}{1-b}} \, dy \Big] K^{-1-\frac{1}{1-b}}. \quad (18)$$

As $K \to \infty$, $P_{new}(K)$ converges to a finite value:

$$P_{new}(K) = K^{-1-\frac{1}{1-b}} \left(\frac{1}{1-b}\right) \Gamma\left(1 + \frac{1}{1-b}\right). \quad (19)$$

Thus for large $K \gg 1$, but such that $K' \ll 1$ or $K \ll [1 + t/n(0)]^{1-b}$, we have an approximate power-law behavior:

$$P_{new}(K) \sim K^{-\varphi}, \quad (20)$$

where $\varphi = 2 + b/(1-b) \geq 2$.

As $K \to 0$,

$$P_{new}(K) = \frac{1}{1-b} K^{\left(-\frac{1}{1-b}-1\right)} \frac{K^{\left(1+\frac{1}{1-b}\right)}}{1 + \frac{1}{1-b}} = \frac{1}{2-b}. \quad (21)$$

2. At fixed t when $K \to \infty$, we use the partial integration to evaluate the incomplete Γ function:

$$\int_x^\infty e^{-y} y^\alpha \, dy = -e^{-y} y^\alpha|_x^\infty$$

$$+ \alpha \int_x^\infty e^{-y} y^{\alpha-1} \, dy \approx e^{-x} x^\alpha.$$

Therefore, from equation (15) we obtain

$$\tilde{P}_{new}(K) \approx \frac{n(0) + t}{N(0) + bt} \frac{b}{1-b} K^{-\frac{1}{1-b}-1}$$

$$\times \int_{K\left(\frac{n(0)}{n(0)+t}\right)^{1-b}}^\infty e^{-y} y^{\frac{1}{1-b}} \, dy,$$

$$= \frac{n(0)}{N(0) + bt} \frac{b}{1-b} \frac{1}{K} \exp\left(-K \left(\frac{n(0)}{n(0)+t}\right)^{1-b}\right), \quad (22)$$

which always decays faster than equation (9) because $n(0) \geq N(0)$ and there is an additional factor K^{-1} in front of the exponential. Thus the behavior of the distribution of all classes is dominated for large K by the exponential decay of the distribution of units in the old classes.

Note that equations (9) and (15) are not exact solutions but continuous approximations which assume K
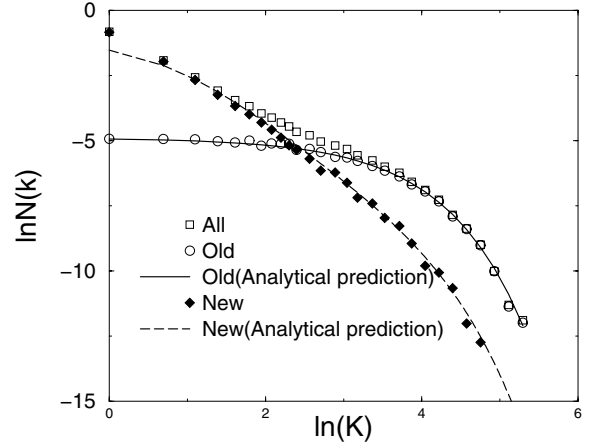

Fig. 1. Comparison of the distributions $P(K)$ for the new and old classes obtained by numerical simulations of the model with the predictions of equations (14) and (10) respectively. For large K the agreement is excellent. The discrepancy exists only for $\tilde{P}_{new}$ at small K, e.g. equation (14) significantly underestimates the $\tilde{P}_{new}(1)$ and $\tilde{P}_{new}(2)$.

is a real number. This approximation produces the most serious discrepancy for small K. To test this approximation, we perform numerical simulations of the model for $b = 0.1$, $N(0) = n(0) = 10\,000$ and $t = 400\,000$. The results are presented in Figure 1. While the agreement is excellent for large K, equation (15) significantly underestimates the value of $\tilde{P}_{new}(K)$ for $K = 1$ and $K = 2$. Note that in reality the power-law behavior of $\tilde{P}_{new}(K)$ extends into the region of very small K.

2.2 The proportional growth of sizes of units

The second set of assumptions of the model is:

(B1) At time t, each class α has $K_\alpha(t)$ units of size $\xi_i(t)$, $i = 1, 2, ... K_\alpha(t)$ where K_α and $\xi_i > 0$ are independent random variables taken from the distributions $P(K_\alpha)$ and $P_\xi(\xi_i)$ respectively. $P(K_\alpha)$ is defined by equation (17) and $P_\xi(\xi_i)$ is a given distribution with finite mean and standard deviation and $\ln \xi_i$ has finite mean $m_\xi = \langle \ln \xi_i \rangle$ and variance $V_\xi = \langle (\ln \xi_i)^2 \rangle - m_\xi^2$.

The size of a class is defined as $S_\alpha(t) \equiv \sum_{i=1}^{K_\alpha} \xi_i(t)$.

(B2) At time $t + 1$, the size of each unit is decreased or increased by a random factor $\eta_i(t) > 0$ so that

$$\xi_i(t + 1) = \xi_i(t) \eta_i(t), \quad (23)$$

where $\eta_i(t)$, the growth rate of unit i, is an independent random variable taken from a distribution $P_\eta(\eta_i)$, which has finite mean and standard deviation. We also assume that $\ln \eta_i$ has finite mean $m_\eta \equiv \langle \ln \eta_i \rangle$ and variance $V_\eta \equiv \langle (\ln \eta_i)^2 \rangle - m_\eta^2$.

Let us assume that due to the Gibrat process [24,25], both the size and growth of units (ξ_i and η_i respectively) are

distributed lognormally

$$p(\xi_i) = \frac{1}{\sqrt{2\pi V_\xi}} \frac{1}{\xi_i} \exp\left(-(\ln \xi_i - m_\xi)^2/2V_\xi\right), \quad (24)$$

$$p(\eta_i) = \frac{1}{\sqrt{2\pi V_\eta}} \frac{1}{\eta_i} \exp\left(-(\ln \eta_i - m_\eta)^2/2V_\eta\right). \quad (25)$$

If units grow according to a multiplicative process, the size of units $\xi_i' = \xi_i \eta_i$ is distributed lognormally with $V_{\xi'} = V_\xi + V_\eta$ and $m_{\xi'} = m_\xi + m_\eta$.

The nth moment of the variable x distributed lognormally is given by

$$\mu_x(n) = \int_0^\infty \frac{1}{\sqrt{2\pi V}} \frac{x^n}{x} dx \exp\left(-(\ln x - m_x)^2/2V_x\right)$$
$$= \exp\left(nm_x + n^2 V_x/2\right). \quad (26)$$

Thus, its mean is $\mu_x \equiv \mu_x(1) = \exp(m_x + V_x/2)$ and its variance is $\sigma_x^2 \equiv \mu_x(2) - \mu_x(1)^2 = \mu_x(1)^2 (\exp(V_x) - 1)$.

Let us now find the distribution of the growth rates of classes. The growth rate g of the class α is defined as

$$g \equiv \ln \frac{S_\alpha(t+1)}{S_\alpha(t)} = \ln \sum_{i=1}^{K_\alpha} \xi_i' - \ln \sum_{i=1}^{K_\alpha} \xi_i. \quad (27)$$

Here we neglect the influx of new units, so $K_\alpha = K_\alpha(t+1) = K_\alpha(t)$.

The resulting distribution of the growth rates of all classes is determined by

$$P(g) \equiv \sum_{K=1}^\infty P(K)P(g|K), \quad (28)$$

where $P(K)$ is the distribution of the number of units in the classes, computed in the previous stage of the model and $P(g|K)$ is the conditional distribution of growth rates of classes with given number of units determined by the distribution $P_\xi(\xi)$ and $P_\eta(\eta)$.

Now our goal is to find an analytical approximation for $P(g|K)$. According to the central limit theorem, the sum of K independent random variables with mean $\mu_\xi \equiv \mu_\xi(1)$ and finite variance σ_ξ^2 is

$$\sum_{i=1}^K \xi_i = K\mu_\xi + \sqrt{K}\nu_K, \quad (29)$$

where ν_K is the random variable with the distribution converging to Gaussian

$$\lim_{K\to\infty} P(\nu_K) \to \frac{1}{\sqrt{2\pi\sigma_\xi^2}} \exp\left(-\nu_K^2/2\sigma_\xi^2\right). \quad (30)$$

Accordingly, we can replace $\ln(\sum_{i=1}^K \xi_i)$ by its Tailor's expansion $\ln K + \ln \mu_\xi + \nu_K/(\mu_\xi \sqrt{K})$, neglecting the terms

of order K^{-1}. Because $\ln \mu_\eta = m_\eta + V_\eta/2$ and $\ln \mu_{\xi'} = \ln \mu_\xi + \ln \mu_\eta$ we have

$$g \equiv \ln S(t+1) - \ln S(t)$$
$$= \ln(K\mu_{\xi'}) + \frac{\nu_K'}{\sqrt{K}\mu_{\xi'}} - \ln(K\mu_\xi) - \frac{\nu_K}{\sqrt{K}\mu_\xi},$$
$$= m_\eta + \frac{V_\eta}{2} + \frac{\nu_K'\mu_\xi - \nu_K\mu_{\xi'}}{\sqrt{K}\mu_\xi\mu_{\xi'}}. \quad (31)$$

For large K the last term in equation (31) is the difference of two Gaussian variables and that is a Gaussian variable itself. Thus for large K, g converges to a Gaussian with the mean, $m = m_\eta + V_\eta/2$, and certain standard deviation which we must find.

In order to do this, we rewrite

$$\frac{\nu_K'}{\sqrt{K}\mu_{\xi'}} = \frac{\sum_{i=1}^K (\xi_i' - \mu_{\xi'})}{K\mu_{\xi'}},$$

and

$$\frac{\nu_K}{\sqrt{K}\mu_\xi} = \frac{\sum_{i=1}^K (\xi_i - \mu_\xi)}{K\mu_\xi}.$$

Thus

$$g = m_\eta + \frac{V_\eta}{2} + \frac{\sum_{i=1}^K \xi_i(\eta_i\mu_\xi - \mu_{\xi'})}{K\mu_\xi\mu_{\xi'}},$$
$$= m_\eta + \frac{V_\eta}{2} + \frac{\sum_{i=1}^K \xi_i(\eta_i - \mu_\eta)}{K\mu_{\xi'}}. \quad (32)$$

Since $\mu_{\xi'} = \mu_\xi\mu_\eta$, the average of each term in the sum is $\mu_{\xi'} - \mu_\xi\mu_\eta = 0$. The variance of each term in the sum is $\langle(\xi_i\eta_i)^2\rangle - \langle 2\xi_i^2\eta_i\mu_\eta\rangle + \langle\xi_i^2\mu_\eta^2\rangle$ where $\xi_i\eta_i$, $\xi_i^2\eta_i$ and ξ_i^2 are all lognormal independent random variables. Particularly, $(\xi_i\eta_i)^2$ is lognormal with $V = 4V_\eta + 4V_\xi$ and $m = 2m_\eta + 2m_\xi$; $\xi_i^2\eta_i$ is lognormal with $V = 4V_\xi + V_\eta$ and $m = 2m_\xi + m_\eta$; ξ_i^2 is lognormal with $V = 4V_\xi$ and $m = 2m_\xi$. Using equation (26)

$$\langle(\xi_i\eta_i)^2\rangle = \exp(2m_\eta + 2m_\xi + 2V_\eta + 2V_\xi), \quad (33a)$$
$$\langle\xi_i^2\eta_i\rangle = \exp(m_\eta + 2m_\xi + 2V_\xi + V_\eta/2), \quad (33b)$$
$$\langle\xi_i^2\rangle = \exp(2m_\xi + 2V_\xi). \quad (33c)$$

Collecting all terms in equations (33a–33c) together and using equation (32) we can find the variance of g:

$$\sigma^2 = \frac{K \exp(2m_\xi + 2V_\xi + 2m_\eta + V_\eta)(\exp(V_\eta) - 1)}{K^2 \exp(2m_\xi + V_\xi + 2m_\eta + V_\eta)},$$
$$= \frac{1}{K} \exp(V_\xi) (\exp(V_\eta) - 1). \quad (34)$$

Therefore, for large K, g has a Gaussian distribution

$$P(g|K) = \frac{\sqrt{K}}{\sqrt{2\pi V}} \exp\left(-\frac{(g-m)^2 K}{2V}\right), \quad (35)$$

where
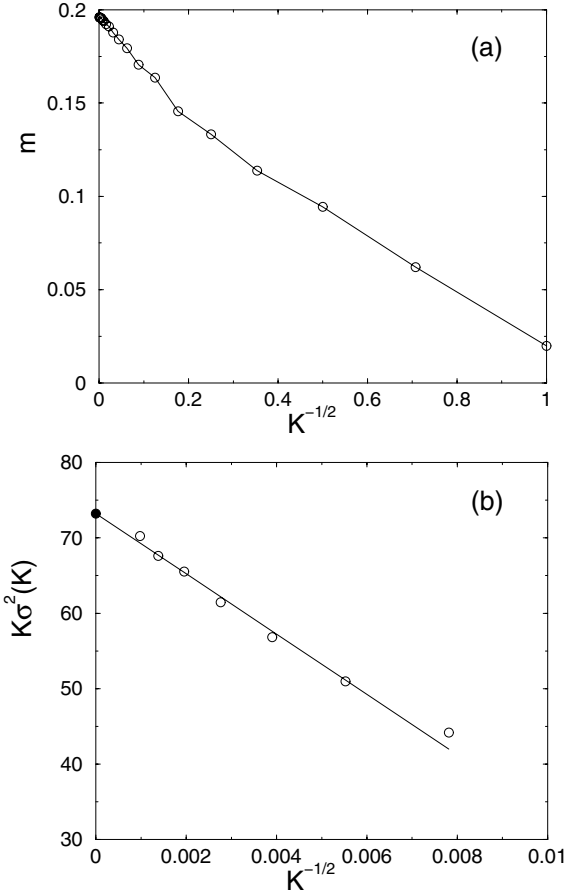
$$m = m_\eta + V_\eta/2 \quad (36)$$

Fig. 3. Crossover of the size-variance relationship $\sigma(K)$ from $K^{0.2}$ for small K to $K^{0.5}$ for large K. The parameters of the simulations are the same as in Figure 2.

Fig. 2. Convergence of the parameters of the simulated $P(g|K)$ to the values, which follow from the central limit theorem: (a) the mean $m(k)$ and (b) the normalized variance $K\sigma^2(K)$. In both cases the speed of convergence is $1/\sqrt{K}$ as can be seen from the straight line fits versus $1/\sqrt{K}$ with the intercepts equal to the analytical values $m = 0.196$ and $V = 73.24$, respectively. The parameters of the simulations $V_\xi = 5.13$ $m_\xi = 3.44$, $V_\eta = 0.36$, and $m_\eta = 0.016$ are taken from the empirical analysis of the pharmaceutical data base [26].

Fig. 4. Convergence of the shape of the distribution of $P(g|K)$ found in simulations to limiting Gaussian. One can see the developments of the tent-shape wings as K grows. The parameters of the simulations are the same as in Figure 2.

and

$$V \equiv K\sigma^2 = \exp(V_\xi)(\exp(V_\eta) - 1). \qquad (37)$$

Note, that the convergence of the sum of lognormals to the Gaussian given by equation (29) is a very slow process, achieving reasonable accuracy only for $K \gg \mu_\xi(2) \sim \exp(2V_\xi)$. For a pharmaceutical database [26], we have $V_\xi = 5.13$, $m_\xi = 3.44$, $V_\eta = 0.36$, and $m_\eta = 0.16$. Accordingly, we can expect convergence only when $K \gg 3 \times 10^4$. Figure 2 demonstrates the convergence of the normalized variance $K\sigma^2(K)$ and mean $m(K)$ of g to the theoretical limits given by equations (36) and (37) respectively: $V = 73.24$ and $m = 0.196$. In both cases, the discrepancy between the limiting values and the actual values decreases as $1/\sqrt{K}$. Interestingly, equation (35) predicts $\sigma(K) \sim K^{-\beta}$, where $\beta = 1/2$. This value is much larger than the empirical value $\beta \approx 0.2$ observed for the size-variance relationships of various socio-economic enti-
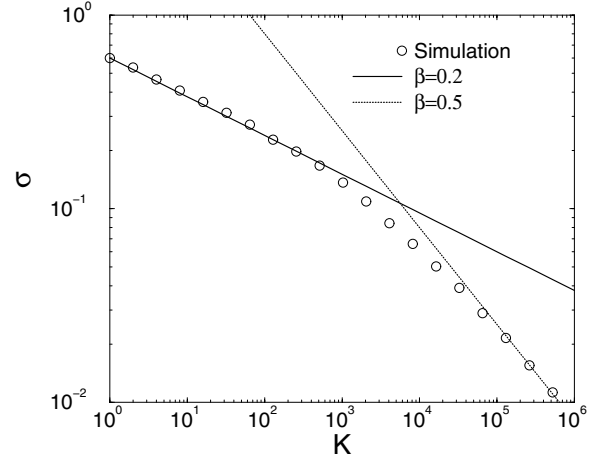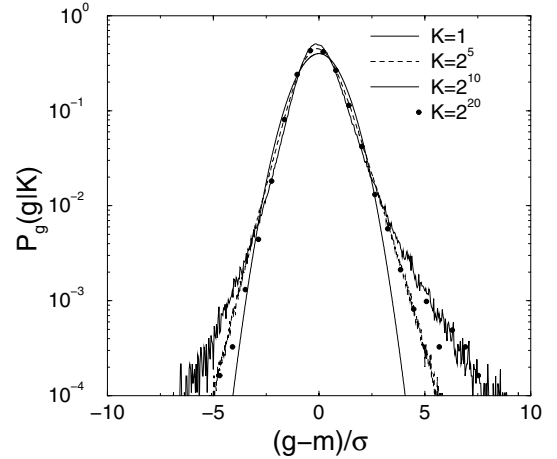
ties [1,2,15,27]. However, the slow convergence of $V(K)K$ suggests that for quite a wide range of $K < 1000$, $\sigma(K) \sim K^{-0.2}$ and only at $K > 10^4$ there is a crossover to the theoretical value $\beta = 0.5$, (Fig. 3). Finally, the simulated distribution of $P(g|K)$ has tent-shape wings which develop as K increases (Fig. 4). This feature of the model growth rates may explain the abundance of the tent-shaped wings of the growth rates of various systems in nature. The most drastic discrepancy between the Gaussian shape and the simulated distribution $P(g|K)$ can be seen when $K \approx 1000$ and than it starts to decrease slowly, and remains visible even for $K = 10^6$.

Nevertheless, in order to obtain close form approximations for the growth rate, we will use the Gaussian approximation (35) for $P(g|K)$. The distribution of the growth rate of the old classes can be found by equation (28). In order to find a close form approximation, we replace the summation in equation (28) by integration and replace
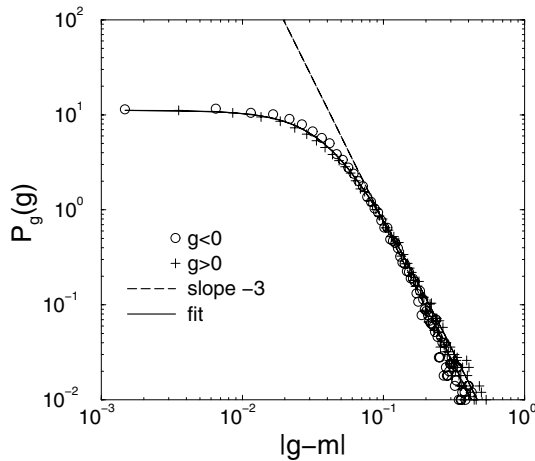
Fig. 5. Comparison of the prediction of equation (38) with the distribution $P(g)$ of the growth rates of the classes simulated for the exponential distribution of the number of units in a class $P(K) = 1/\langle K \rangle \exp(-K/\langle K \rangle)$ with $\langle K \rangle = 2^{15}$. The parameters of the simulation are the same as in Figure 2. The fitting parameter $V = 33$ in equation (38) gives the best agreement with the simulation results. One can see a very good convergence to the inverse cubic law for the wings.

the distributions $P(K)$ by equation (9) and $P(g|K)$ by equation (35). Assuming $m = 0$, we have

$$P_{old}(g) \approx \frac{1}{\sqrt{2\pi V}} \int_0^\infty \frac{1}{K(t)}$$
$$\times \exp\left(\frac{-K}{K(t)}\right) \exp\left(-\frac{g^2 K}{2V}\right) \sqrt{K}\, dK,$$
$$= \frac{\sqrt{K(t)}}{2\sqrt{2V}} \left(1 + \frac{K(t)}{2V} g^2\right)^{-\frac{3}{2}}, \qquad (38)$$

where $K(t)$ is the average number of units in the old classes (see Eq. (8)). This distribution decays as $1/g^3$ and thus does not have a finite variance. In spite of drastic assumptions that we make, equation (38) correctly predicts the shape of the convolution $P_{old}(g)$. Figure 5 shows the comparison of the simulation of the growth rates in the system with the exponential distribution of units $P(K)$ with $K(t) = 2^{15}$ and the same empirical parameters of the unit size and growth distributions as before. The parameter of the analytical distribution characterizing its width (variance does not exist), must be taken $V = 33$ which is much smaller than the analytical prediction $V = 73.23$. This is not surprising, since for $K = 2^{15}$ $K\sigma^2(K) = 50$ (see Fig. 2b). Moreover, since we are dealing with the average $\sigma^2(K)K$ for $K < 2^{15}$, we can expect $V < 50$. Nevertheless the nature of the power-law wings decaying as $1/g^3$ is reproduced very well.

For the new classes, when $t \to \infty$ the distribution of number of units is approximated by

$$P_{\text{new}}(K) \approx \frac{1}{1-b} K^{-1-\frac{1}{1-b}} \int_0^K y^{\frac{1}{1-b}} e^{-y}\, dy. \qquad (39)$$

Again replacing summation in equation (28) in the text by integration and $P(g|K)$ by equation (35) and after the

switching the order of integration we have:

$$P_{new}(g) \approx \frac{1}{1-b} \frac{1}{\sqrt{2\pi V}} \int_0^\infty \exp(-y)\, y^{\frac{1}{1-b}}\, dy$$
$$\times \int_y^\infty \exp(-g^2 K/2V)\, K^{(-\frac{1}{2}-\frac{1}{1-b})}\, dK. \quad (40)$$

As $g \to \infty$, we can evaluate the second integral in equation (40) by partial integration:

$$P_{new}(g) \approx \frac{1}{1-b} \int_0^\infty \frac{1}{\sqrt{2\pi V}} \frac{2V}{g^2}\, y^{-\frac{1}{1-b}-\frac{1}{2}}\, y^{\frac{1}{1-b}}\ \exp(-y)$$
$$\times \exp(-y\, g^2/2V)\, dy,$$
$$= \frac{1}{1-b} \frac{1}{\sqrt{2\pi V}} \frac{2V}{g^2}\, \frac{1}{\sqrt{g^2/2V+1}}\ \sqrt{\pi} \sim \frac{1}{g^3}. \quad (41)$$

We can compute the first derivative of the distribution (40) by differentiating the integrand in the second integral with respect to g. The second integral converges as $y \to 0$, and we find the behavior of the derivative for $g \to 0$ by the substitution $x = Kg^2/(2V)$. As $g \to 0$, the derivative behaves as $g\, g^{2[-(3/2)+1/(1-b)]} \sim g^{2b/(1-b)}$, which means that the function itself behaves as $C_2 - C_1|g|^{2b/(1-b)+1}$, where C_2 and C_1 are positive constants. For small b this behavior is similar to the behavior of a Laplace distribution with variance V: $\exp(-\sqrt{2}|g|/\sqrt{V})/\sqrt{2V} = 1/\sqrt{2V} - |g|/V$.

When $b \to 0$, equation (40) can be expressed in elementary functions:

$$P_{new}(g)|_{b\to 0} \approx \frac{1}{\sqrt{2\pi V}} \int_0^\infty K^{-3/2} \exp(-K\, g^2/2\, V)\, dK$$
$$\times \int_0^K \exp(-y)y\, dy,$$
$$\approx \frac{1}{\sqrt{2\, V}} \left(-\frac{1}{\sqrt{1+g^2/2\, V}}\right.$$
$$\left. +\frac{2}{|g|/\sqrt{2\, V} + \sqrt{g^2/2\, V + 1}}\right).$$

Simplifying we find the main result:

$$P_{\text{new}}(g)|_{b\to 0} \approx \frac{2V}{\sqrt{g^2 + 2V}\,(|g| + \sqrt{g^2 + 2V})^2}, \qquad (42)$$

which behaves for $g \to 0$ as $1/\sqrt{2V} - |g|/V$ and for $g \to \infty$ as $V/(2g^3)$. Thus the distribution is well approximated by a Laplace distribution in the body with power-law tails. Because of the discrete nature of the distribution of the number of units, when $g \gg \sqrt{2V}$ the behavior for $g \to \infty$ is dominated by $\exp(-g^2/2V)$.

In Figure 6a we compare the distributions given by equation (38), the mean field approximation equation (40) for $b = 0.1$ and equation (42) for $b \to 0$. We find that all three distributions have very similar tent shape behavior in the central part. In Figure 6b we also compare the distribution equation (42) with its asymptotic behaviors for $g \to 0$ (Laplace cusp) and $g \to \infty$ (power law), and find the crossover region between these two regimes.
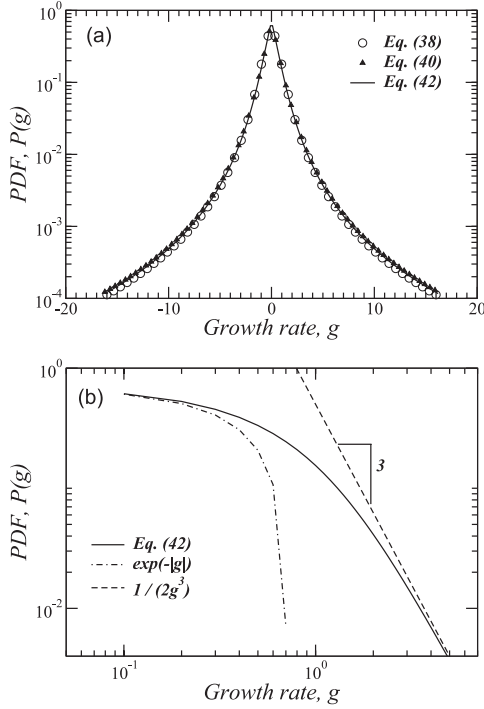
Fig. 6. (a) Comparison of three different approximations for the growth rate PDF, $P(g)$, given by equation (38), mean field approximation equation (40) for $b = 0.1$ and equation (42). Each $P(g)$ shows similar tent shape behavior in the central part. We see there is little difference between the three cases, $b = 0$ (no entry), $b = 0.1$ (with entry) and the mean field approximation. This means that entry of new classes ($b > 0$) does not perceptibly change the shape of $P(g)$. Note that we use $K(t)/V_g = 2.16$ for equation (38) and $V_g = 1$ for equation (42). (b) The crossover of $P(g)$ given by equation (42) between the Laplace distribution in the center and power law in the tails. For small g, $P(g)$ follows a Laplace distribution $P(g) \sim \exp(-|g|)$, and for large g, $P(g)$ asymptotically follows an inverse cubic power law $P(g) \sim g^{-3}$.

3 Conclusions

The analytical solution of this model can be obtained only for certain limiting cases but a numerical solution can be easily computed for any set of assumptions. We investigate the model numerically and analytically and find:

(1) In the presence of the influx of new classes ($b > 0$), the distribution of units converges for $t \to \infty$ to a power law $P(K) \sim K^{-\varphi}$, $\varphi = 2 + b/(1-b) \geq 2$. Note that this behavior of the power-law probability density function leads to the power law rank-order distribution where rank of a class R is related to the number of its units K as

$$R = N(t) \int_K^\infty P(K)dk \sim K^{-\varphi+1}. \qquad (43)$$

Thus $K \sim R^{-\zeta}$, where $\zeta = 1/(\varphi-1) = 1-b \leq 1$, which leads in the limit $b \to 0$ to the celebrated Zipf's law [28] for cities populations, $K \sim 1/R$. Note that

this equation can be derived for our model using elementary considerations. Indeed, due to proportional growth the rank of a class, R, is proportional to the time of its creation t_0. The number of units $n(t_0)$ existing at time t_0 is also proportional to t_0 and thus also proportional to R. According to the proportional growth, the ratio of the number of units in this class to the number of units in the classes existed at time t_0 is constant: $K(t_0, t)/n_e(t_0, t) = 1/n(t_0)$. If we assume that the amount of units in the classes, created after t_0 can be neglected since the influx of new classes b is small, we can approximate $n_e(t_0, t) \approx n(t) \sim t$. Thus for large t, $n_e(t_0, t)$ is independent of t_0 and hence $K(t_0, t) \sim 1/R$. If we do not neglect the influx of new classes, equation (7) gives $n_e(t_0, t) \sim t_0^b$, hence $K(t_0, t) \sim 1/R^{1-b}$.

(2) The conditional distribution of the logarithmic growth rates $P(g|K)$ for the classes consisting of a fixed number K of units converges to a Gaussian distribution (35) for $K \to \infty$. The width of this distribution, $\sigma(K)$, decreases as $1/K^\beta$, with $\beta = 1/2$. Note that due to slow convergence of the sum of lognormal variables to the Gaussian in case of a wide lognormal distribution of unit sizes computed from the empirical data [26]($V_\xi = 5.13$), we have $\beta = 0.2$ for relatively small classes. This result is consistent with the observation that large firms with many production units fluctuate less than small firms [1,5,9,29]. Interestingly, in case of large V_ξ, $P(g|K)$ converges to the Gaussian in the central interval which grows with K, but outside this interval it develops tent-shape wings, which are becoming increasingly wider, as $K \to \infty$. However, they remain limited by the distribution of the logarithmic growth rates of the units, $P_\eta(\ln \eta)$.

(3) For $g \gg V_\eta$, the distribution $P(g)$ coincides with the distribution of the logarithms of the growth rates of the units:

$$P(g) \approx P_\eta(\ln \eta). \qquad (44)$$

In the case of power law distribution $P(K) \sim K^{-\varphi}$ which dramatically increases for $K \to 1$, the distribution $P(g)$ is dominated by the growth rates of classes consisting of a single unit $K = 1$, thus the distribution $P(g)$ practically coincides with $P_\eta(\ln \eta)$ for all g. Indeed, empirical observations of reference [26] confirm this result.

(4) If the distribution $P(K) \sim K^{-\varphi}$, $\varphi > 2$ for $K \to \infty$, as happens in the presence of the influx of new units $b \neq 0$, $P(g) = C_1 - C_2|g|^{2\varphi-3}$, for $g \to 0$ which in the limiting case $b \to 0$, $\varphi \to 2$ gives the cusp $P(g) \sim C_1 - C_2|g|$ (C_1 and C_2 are positive constants), similar to the behavior of the Laplace distribution $P_L(g) \sim \exp(-|g|C_2)$ for $g \to 0$.

(5) If the distribution $P(K)$ weakly depends on K for $K \to 1$, the distribution of $P(g)$ can be approximated by a power law of g: $P(g) \sim g^{-3}$ in a wide range $\sqrt{V/K(t)} \ll g \ll \sqrt{V_\eta}$, where $K(t)$ is the average number of units in a class. This case is realized for $b = 0$, $t \to \infty$ when the distribution of $P(K)$ is dominated by the exponential distribution and $K(t) \to \infty$

as defined by equation (1). In this particular case, $P(g)$ for $g \ll \sqrt{V_\eta}$ can be approximated by equation (38).

(6) In the case in which the distribution $P(K)$ is not dominated by one-unit classes but for $K \to \infty$ behaves as a power law, which is the result of the mean field solution for our model when $t \to \infty$, the resulting distribution $P(g)$ has three regimes, $P(g) \sim C_1 - C_2|g|^{2\varphi-3}$ for small g, $P(g) \sim g^{-3}$ for intermediate g, and $P(g) \sim P(\ln \eta)$ for $g \to \infty$. The approximate solution of $P(g)$ in this case is given by (40). For $b \neq 0$, equation (40) can not be expressed in elementary functions. In the $b \to 0$ case, equation (40) yields the main result, equation (42), which combines the Laplace cusp for $g \to 0$ and the power law decay for $g \to \infty$. Note that due to replacement of summation by integration in equation (28), the approximation equation (42) holds only for $g < \sqrt{V_\eta}$.

In conclusion we want to emphasize that although the derivations of the distributions (38), (40), and (42) are not rigorous they satisfactory reproduce the shape of empirical data, especially the $1/g^3$ behavior of the wings of the distribution of the growth rates and the sharp cusp near the center.

References

1. L.A.N. Amaral, S.V. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M.A. Salinger, H.E. Stanley, M.H.R. Stanley, J. Phys. I France **7**, 621 (1997)
2. S.V. Buldyrev, L.A.N. Amaral, S. Havlin, H. Leschhorn, P. Maass, M.A. Salinger, H.E. Stanley, M.H.R. Stanley, J. Phys. I France **7**, 635 (1997)
3. J. Sutton, Physica A **312**, 577 (2002)
4. G.D. Fabritiis, F. Pammolli, M. Riccaboni, Physica A **324**, 38 (2003)
5. L.A.N. Amaral, S.V. Buldyrev, S. Havlin, M.A. Salinger, H.E. Stanley, Phys. Rev. Lett. **80**, 1385 (1998)
6. H. Takayasu, K. Okuyama, Fractals **6**, 67 (1998)
7. D. Canning, L.A.N. Amaral, Y. Lee, M. Meyer, H.E. Stanley, Econ. Lett. **60**, 335 (1998)
8. S.V. Buldyrev, N.V. Dokholyan, S. Erramilli, M. Hong, J.Y. Kim, G. Malescio, H.E. Stanley, Physica A **330**, 653 (2003)
9. J. Sutton, J. Econ. Lit. **35**, 40 (1997)
10. Y. Ijiri, H.A. Simon, Proc. Natl. Acad. Sci. **72**, 1654 (1975)
11. M.R. Kalecki, Econometrica **13**, 161 (1945)
12. D.E. Mansfield, Am. Econ. Rev. **52**, 1024 (1962)
13. B.H. Hall, J. Ind. Econ. **35**, 583 (1987)
14. S. Kotz, T.J. Kozubowski, K. Podgórski, *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance* (Birkhauser, Boston, 2001)
15. M.H.R. Stanley, L.A.N. Amaral, S.V. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M.A. Salinger, H.E. Stanley, Nature **379**, 804 (1996)
16. W.J. Reed, Econ. Lett. **74**, 15 (2001)
17. W.J. Reed, B.D. Hughes, Phys. Rev. E **66**, 067103.(2002)
18. K. Yamasaki, K. Matia, S.V. Buldyrev, D. Fu, F. Pammolli, M. Riccaboni, H.E. Stanley, Phys. Rev. E **74**, 035103 (2006)
19. N.L. Johnson, S. Kotz, *Urn Models and Their Applications* (Wiley, New York, 1977)
20. S. Kotz, H. Mahmoud, P. Robert, Statist. Probab. Lett. **49**, 163 (2000)
21. W.J. Reed, B.D. Hughes, Math. Biosci. **189**, 97 (2004)
22. H.E. Stanley, *Introduction to Phase Transitions and Critical Phenomena* (Oxford University Press, Oxford, 1971)
23. D.R. Cox, H.D. Miller, *The Theory of Stochastic Processes* (Chapman and Hall, London, 1968)
24. R. Gibrat, Bulletin de Statistique Général, France **19**, 469 (1930)
25. R. Gibrat, *Les Inégalités Économiques* (Librairie du Recueil Sirey, Paris, 1931)
26. D. Fu, F. Pammolli, S.V. Buldyrev, M. Riccaboni, K. Matia, K. Yamasaki, H.E. Stanley, Proc. Natl. Acad. Sci. **102**, 18801 (2005)
27. K. Matia, L.A.N. Amaral, M. Luwel, H.F. Moed, H.E. Stanley, J. Am. Soc. Inf. Sci. Technol. **56**, 893 (2005)
28. G. Zipf, *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Cambridge, MA, 1949)
29. S. Hymer, P. Pashigian, J. Pol. Econ. **70**, 556 (1962)