

EUROPHYSICS LETTERS

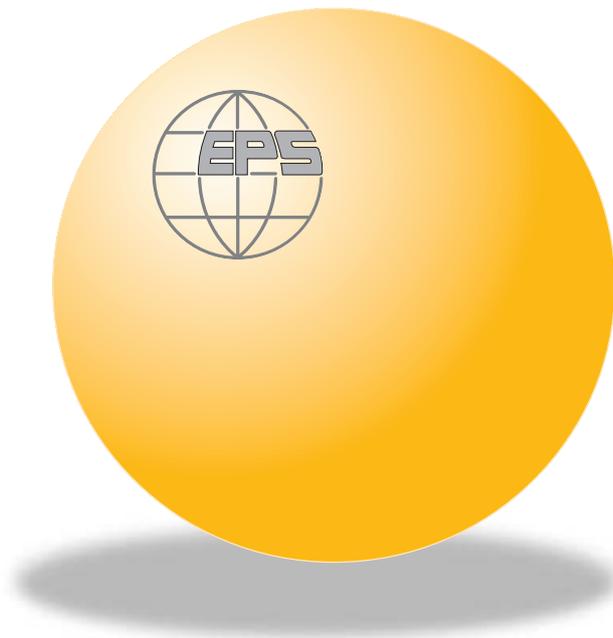
OFFPRINT

Vol. 67 • Number 3 • pp. 498–503

Statistical properties of business firms structure and growth

* * *

K. MATIA, DONGFENG FU, S. V. BULDYREV, F. PAMMOLLI,
M. RICCABONI and H. E. STANLEY



Published under the scientific responsibility of the
EUROPEAN PHYSICAL SOCIETY
Incorporating
JOURNAL DE PHYSIQUE LETTRES • LETTERE AL NUOVO CIMENTO

Statistical properties of business firms structure and growth

K. MATIA¹, DONGFENG FU¹, S. V. BULDYREV¹, F. PAMMOLLI²,
M. RICCABONI² and H. E. STANLEY¹

¹ *Center for Polymer Studies and Department of Physics, Boston University
Boston, MA 02215 USA*

² *Facoltà di Economia, Università di Firenze and CERM
Via Banchi di Sotto 55, 53100 Siena Italy*

(received 10 March 2004; accepted in final form 17 May 2004)

PACS. 89.90.+n – Other topics in areas of applied and interdisciplinary physics.

PACS. 05.45.Tp – Time series analysis.

PACS. 05.40.Fb – Random walks and Levy flights.

Abstract. – We analyze a database comprising quarterly sales of 55624 pharmaceutical products commercialized by 3939 pharmaceutical firms in the period 1992–2001. We study the probability density function (PDF) of growth in firms and product sales and find that the width of the PDF of growth decays with the sales as a power law with exponent $\beta = 0.20 \pm 0.01$. We also find that the average sales of products scales with the firm sales as a power law with exponent $\alpha = 0.57 \pm 0.02$. And that the average number products of a firm scales with the firm sales as a power law with exponent $\gamma = 0.42 \pm 0.02$. We compare these findings with the predictions of models proposed till date on growth of business firms.

In economics there are unsolved problems that involve interactions among a large number of subunits [1–3]. One of these problems is the structure of a business firm and its growth [2,4]. As in many physical models, decomposition of a firm into its constituent parts is an appropriate starting place for constructing a model. Indeed, the total sales of a firm is comprised of a large number of product sales. Previously accurate data on the “microscopic” product sales have been unavailable, and hence it has been impossible to test the predictions of various models. Here we analyze a new database, the Pharmaceutical Industry Database (PHID) which records quarterly sales figures of 55624 pharmaceutical products commercialized by 3939 firms in the European Union and North America from September 1991 to June 2001. We shall see that these data support the predictions of a simple model, and at the same time the data do not support the microscopic assumptions of that model. In this sense, the model has the same status as many statistical physics models, in that the predictions can be in accord with data even though the details of microscopic interactions are not. The assumptions of this simple model given in ref. [5] are as follows: i) Firms tends to organize themselves into multiple divisions once they attain a particular size. ii) The minimum size of firms in a particular economy comes from a broad distribution. iii) Growth rates of divisions are independent of each other and there is no temporal correlation in their growth. With these assumptions the model builds a diversified multi-divisional structure. Starting from a single product evolving

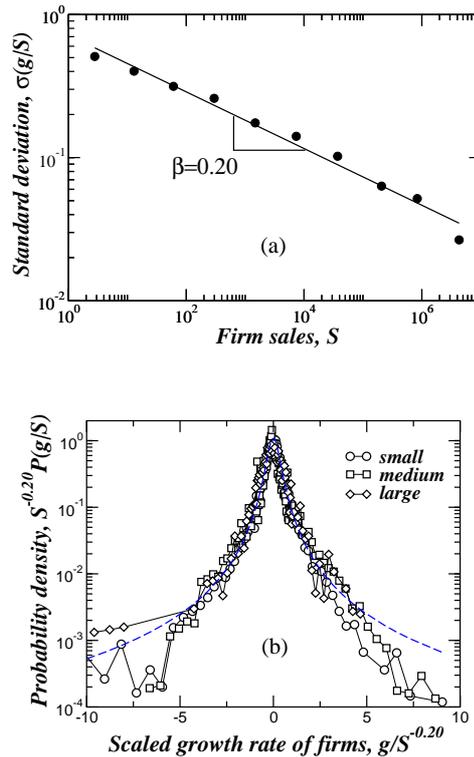


Fig. 1 – (a) Firms are divided into 10 groups according to firm sale S . We find the standard deviation $\sigma(g|S)$ of the growth rates scales as a power law, $\sigma(g|S) \sim S^{-\beta}$, with $\beta = 0.20 \pm 0.01$. Symbols are data points and the solid line is a regression fit. (b) The PDF of the growth rates for small ($S < 10^2$), medium ($10^2 < S < 10^4$), and large ($10^4 < S$) values of S is scaled by their standard deviation. Note the collapse of the histograms of the three groups which confirms the scaling exponent β . The dashed line is f_0 as predicted by the model approximating the results of ref. [5] given by $f_0(x) \approx a_0 \exp[-a_1 [1 + 0.75 \ln(1 + a_2 x^2)]] [1 + 0.75 \ln(1 + a_2 x^2)]^{1/2}$, where a_0 , a_1 , and a_2 are parameters of the model.

to a multi-product firm, this model reproduces a number of empirical observations and makes some predictions which we discuss in detail below along with results and predictions of other models which attempt to address the problem of business firm growth.

Consider a firm i of sales S_i with N_i products whose sales are $\xi_{i,j}$, where $j = 1, 2, \dots, N_i$. Thus the firm size in terms of the sales of its products is given as $S_i = \sum_{j=1}^{N_i} \xi_{i,j}$. The growth rate is

$$g_i(t) \equiv \log \left(\frac{S_i(t + \Delta t)}{S_i(t)} \right) = \log S_i(t + \Delta t) - \log S_i(t), \tag{1}$$

where $S_i(t)$ and $S_i(t + \Delta t)$ are the sales, in units of 10^3 British pounds, of firm i being considered in the year t and $t + \Delta t$, respectively. Pharmaceutical data has seasonal effect, and hence the analysis of quarterly data will have effects due to seasonality. To remove any seasonality, that might be present, we analyze the annual data instead of the quarterly data.

Recent studies have demonstrated power law scaling in economic systems [6]. In particular

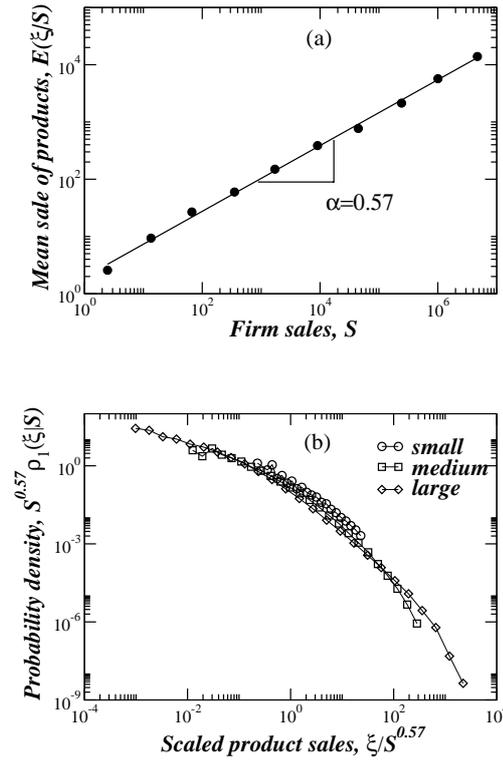


Fig. 2 – (a) Mean, $E(\xi|S)$, of the product sale conditioned for a firm of sale S . We observe that the mean scales as $E(\xi|S) \sim S^\alpha$ with $\alpha = 0.57 \pm 0.02$. Symbols are data points and the solid line is a regression fit. (b) PDF of the product sales for small ($S < 10^2$), medium ($10^2 < S < 10^4$), and large ($10^4 < S$) values of S scaled by $S^{0.57}$. Note the collapse of the PDFs of the three groups which confirms the scaling exponent α .

the standard deviation σ of the growth rates of diverse systems including firm sales [6] or gross domestic product (GDP) of countries [7] scales as a power law of S .

The models of refs. [1–3,5,8] all predict that the standard deviation of growth rates amongst all firms with the same sales scales as a power law $\sigma(g|S) \sim S^{-\beta}$. Further, the model of ref. [5] predicts that the probability density function PDF $p(g|S)$ of growth rates for a size of firm S scales as a function of S as

$$p(g|S) \sim \frac{1}{S^{-\beta}} f_0\left(\frac{g}{S^{-\beta}}\right), \quad (2)$$

where f_0 is a symmetric function of a specific “tent-shaped” form resulting from a convolution of log-normal distributions and a Gaussian distribution, with parameters dependent on the parameters of the model. Figure 1a plots the scaling of the standard deviation $\sigma(g|S)$. We observe $\sigma(g|S) \sim S^{-\beta}$ with $\beta = 0.19 \pm 0.01$. Figure 1b plots the scaled PDF as given by eq. (2) for three sales groups; small ($S < 10^2$), medium ($10^2 < S < 10^4$) and large ($10^4 < S$). The figure also plots f_0 as predicted by ref. [5]. The model of ref. [5] further predicts that the PDF of the product size ξ for a fixed firm size S , $\rho_1(\xi|S)$ should scale as

$$\rho_1(\xi|S) \sim \frac{1}{S^\alpha} f_1\left(\frac{\xi}{S^\alpha}\right), \quad (3)$$

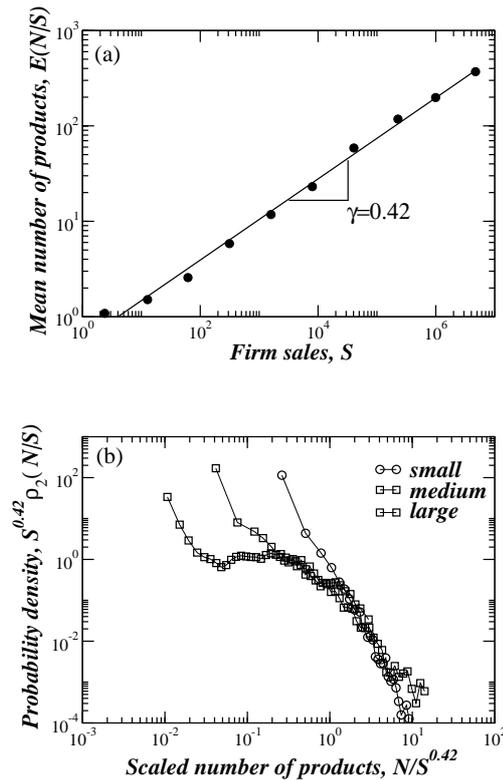


Fig. 3 – (a) Mean, $E(N|S)$, of the number of products N for a firm of sale S . Symbols are data points and the solid line is a regression fit. We observe that the mean scales as $E(N|S) \sim S^\gamma$ with $\gamma = 0.42 \pm 0.01$. (b) PDF of the number of product for small ($S < 10^2$), medium ($10^2 < S < 10^4$), and large ($10^4 < S$) values of S scaled by $S^{0.42}$. Note the partial collapse of the PDFs of the three groups which confirms the scaling exponent γ . For small values of N , which also corresponds to small values of S , the statistics become poor. This statistical errors gets even more amplified when we divide small values of N by $S^{0.42}$. Thus we observe poor quality of data collapse for $N/S^{0.42} < 1$. The data collapse is better for $N/S^{0.42} > 1$, where we have good statistics.

where again f_1 depends on the parameters of the model. According to the model discussed in refs. [5, 8], f_1 is a log-normal PDF. We then evaluate the average product size $E(\xi|S)$ in a firm of size S , defined as $E(\xi|S) = \int d\xi \rho_1(\xi|S) \xi \sim S^\alpha$. Figure 2a plots $E(\xi|S)$; we observe $E(\xi|S) \sim S^\alpha$ with $\alpha = 0.57 \pm 0.02$. Figure 2b plots the scaled PDF as given by eq. (3) for three sales groups: small ($S < 10^2$), medium ($10^2 < S < 10^4$) and large ($10^4 < S$). We observe that for each of the groups the PDF $\rho_1(\xi|S)$ is consistent with a log-normal distribution by noting in a log-log plot the PDF $\rho_1(\xi|S)$ is parabolic, which is tested by performing a regression fit.

According to ref. [5], the PDF $\rho_2(N|S)$ of number of products N in a firm of size S should obey the scaling relation

$$\rho_2(N|S) \sim \frac{1}{S^\gamma} f_2\left(\frac{N}{S^\gamma}\right), \tag{4}$$

where the function f_2 is log-normal and depends on the parameters of the model. We evaluate the average number of products $E(N|S)$ for a firm of size S . Using eq. (4) we note that $E(N|S) = \int dN \rho_1(N|S) N \sim S^\gamma$. Figure 3a plots the expectation $E(N|S)$ and we observe

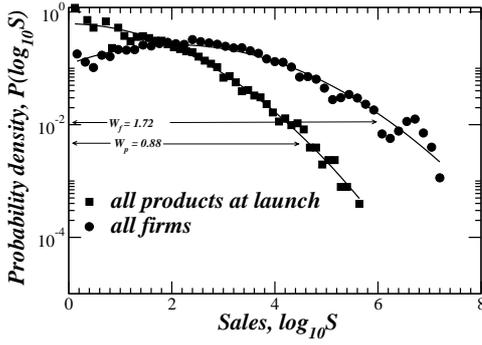


Fig. 4

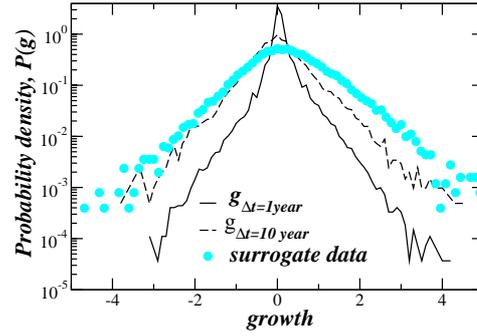


Fig. 5

Fig. 4 – PDF of products sales (squares) firm sales (circles) between 1990–2001. The variance of the PDFs of products sales at launch and firm sales are estimated to be $W_p = 0.88$ and $W_f = 1.72$, respectively. This gives $W_f - W_p = 0.84$ and $\beta = 0.24$ (cf. eq. (7)), which is approximately what is observed empirically as predicted by [5].

Fig. 5 – Probability density function (PDF), $P(g_{\Delta t})$, of products for $\Delta t = 1$ year (solid line) and $\Delta t = 10$ years (dashed line). Circles represent the surrogate data [9]. In the absence of correlation we expect the data for $\Delta t = 10$ to coincide with the PDF of the surrogate points.

that $E(N|S) \sim S^\gamma$ with $\gamma = 0.42 \pm 0.01$. Figure 3b plots the scaled PDF $\rho_2(N|S)$ as given by eq. (4) for three groups: small ($S < 10^2$), medium ($10^2 < S < 10^4$) and large ($10^4 < S$).

According to [5], the relations between the scaling exponents α , β , and γ are given by

$$\gamma = 1 - \alpha, \quad (5)$$

$$\beta = \frac{1 - \alpha}{2}, \quad (6)$$

which we find to be approximately valid for the PHID database.

According to the model discussed in ref. [5], the distribution of product sizes in each firm scales with the product size at launch, $S_{\text{at launch}}$, according to $\rho_0(\xi/S_{\text{at launch}})$, which is approximately log-normal. Model [5] postulates that the PDF of $S_{\text{at launch}}$ is log-normal, *i.e.*, $P(\log S_{\text{at launch}})$ is Gaussian with variance W_p and each firm is characterized by a fixed value of $S_{\text{at launch}}$. Furthermore, ref. [5] predicts that the distribution of firm sales is close to log-normal, *i.e.* the PDF $P(\log S)$ is Gaussian with variance W_f . With these hypotheses ref. [5] derives that

$$\beta = \frac{W_f - W_p}{2 W_f}. \quad (7)$$

Figure 4 plots PDF of annual products sales, products sales at launch $P(\log S_{\text{at launch}})$, and firm sales $P(\log S)$ between 1990–2001. The variance of the PDFs of products sales at launch and firm sales are estimated to be $W_p = 0.88$, $W_f = 1.72$, respectively. This gives $\beta = (W_f - W_p)/2 W_f = 0.24$ which is approximately what is observed empirically. We employ two methods to estimate W (the standard deviation): i) Estimate W from the definition, *i.e.* $W^2 = (1/(N-1)) \sum_{i=1}^N (x_i - \langle x \rangle)^2$, where $\{x_1, x_2, \dots, x_N\}$ is a set of data and $\langle x \rangle$ is the mean of the set $\{x_i\}$. ii) First estimate the PDF from the set $\{x_i\}$, then perform a regression fit with a log-normal function to the PDF. The standard deviation W will be one of the fitting

parameter. Hence estimate W from the estimated parameter value from the least-square log-normal fit to the PDF. We observe both this method gives similar values of W and the ratio β (cf. eq. (7)) remains unchanged as long as we consistently use one of the 2 methods described above. Our estimate of W presented here is using the former method.

Reference [5] postulates products growth rate to be Gaussian and temporally uncorrelated. To test this postulate, fig. 5 plots the PDF $P(g)$ of the growth g of the products $\Delta t = 1$ and $\Delta t = 10$ year [9]. We see that the empirical distribution is not growing via random multiplicative process as ref. [5] postulates, but has the same tent shape distribution as the distribution of firm sales growth rate, suggesting that the products themselves may not be elementary units but may be comprised of smaller interacting subunits. Figure 5 also plots PDF $P(g')$ surrogate data obtained by summation of the 10 annual growth rates from the empirical distribution. We observe that the $P(g_{\Delta t=10})$ for products differs from the surrogate data implying there are significant anti-correlations in the growth dynamics between successive years.

In summary we study the statistical properties of the internal structure of a firm and its growth. We identify three scaling exponents relating the i) sales of the products, ξ , ii) the number of products, N , and iii) the standard deviation of the growth rates, σ , of a firm with its sales S . Our analysis confirms the features predicted in ref. [5]. However, we find that the postulate of the model, namely, the growth rate of the products is uncorrelated and Gaussian, is not accurate. Thus the model of ref. [5] can be regarded as a first step towards the explanation of the dynamics of the firm growth.

* * *

We thank L. A. N. AMARAL and S. HAVLIN for helpful discussions and suggestions, and NSF and Merck Foundation (EPRIS Program) for financial support.

REFERENCES

- [1] IJIRI Y. and SIMON H. A., *Skew Distributions and the Sizes of Business Firms* (North-Holland, Amsterdam) 1997.
- [2] SUTTON J., *Physica A*, **312** (2002) 577.
- [3] WYART M. and BOUCHAUD J. P., cond-math/0210479v2 (19 Nov. 2002).
- [4] COASE R. H., *Economica*, **4** (1937) 386; COASE R. H., *The Nature of Firm: Origins, Evolution and Development* (Oxford University Press, New York) 1993, pp. 34-74; MANSFIELD E., *Research Policy*, **20** (1991) 1; PAKES A. and SOKOLOFF K. L., *Proc. Nat. Ac. Sci. USA*, **93** (1996) 12655; ERIKSON R. and PAKES A., *Rev. of Eco. Studies*, **62** (1995) 53.
- [5] AMARAL L. A. N., BULDYREV S. V., HAVLIN S., SALINGER M. A. and STANLEY H. E., *Phys. Rev. Lett.*, **80** (1998) 1385.
- [6] STANLEY M. H. R., AMARAL L. A. N., BULDYREV S. V., HAVLIN S., LESCHHORN H., MAASS P., SALINGER M. A. and STANLEY H. E., *Nature (London)*, **379** (1996) 804; AMARAL L. A. N., BULDYREV S. V., HAVLIN S., LESCHHORN H., MAASS P., SALINGER M. A., STANLEY H. E. and STANLEY M. H. R., *J. Phys. I*, **7** (1997) 621; BULDYREV S. V., AMARAL L. A. N., HAVLIN S., LESCHHORN H., MAASS P., SALINGER M. A., STANLEY H. E. and STANLEY M. H. R., *J. Phys. I*, **7** (1997) 635; LEE Y., AMARAL L. A. N., CANNING D., MEYER M. and STANLEY H. E., *Phys. Rev. Lett.*, **81** (1998) 3275; CANNING D., AMARAL L. A. N., LEE Y., MEYER M. and STANLEY H. E., *Econ. Lett.*, **60** (1998) 335; KEITT T. and STANLEY H. E., *Nature (London)*, **393** (1998) 257.
- [7] PLEROU V., AMARAL L. A. N., GOPIKRISHNAN P., MEYER M. and STANLEY H. E., *Nature (London)*, **400** (1999) 433.
- [8] DE FABRITIIS G., PAMMOLLI G. and RICCABONI M., *Physica A*, **324** (2003) 38.
- [9] $g_i(t) \equiv \log(S_i(t+1)/S_i(t))$. And $g_{\Delta t=10} = \sum_{i=1}^{10} g_i, \Delta t=1$.