Supporting Information

Riccaboni *et al.* 10.1073/pnas.0810478105

SI Text

The purpose of this supporting information is to describe the model presented by Fu and colleagues (1) and summarize the empirical evidence that supports it.

The Model. We model business firms as classes consisting of a random number of units of variable size. The number of units is defined as in the Simon model (2). The size of the units evolves according to the Gibrat growth process (3).

Firms grow by capturing new business opportunities and the probability that a new opportunity is assigned to a given firm is proportional to the number of opportunities it has already got (2, 4, 5). At each time t a new opportunity is assigned.

With probability b, the new opportunity is taken up by a new firm, so that the average number of firms at time t is $N(t) = N(0) + bt$.

With probability $1 - b$, the new opportunity is captured by an active firm α with probability $P_\alpha = (1 - \mathrm{b})K_\alpha(\mathrm{t})/\mathrm{t}$, where $K_\alpha(\mathrm{t})$ is the number of units of firm α at time t.

In the absence of the entry of new firms ($b = 0$) the probability distribution of the number of the units in the firms at large t, i.e., the distribution $P(K)$, is exponential:

$$P(K) \approx \frac{1}{K(t)}\exp(-K/K(t)), \quad [1]$$

where $K(t) = [n(0) + t]/N(0)$ is the average number of units in the classes, which linearly grows with time.

If $b > 0$, $P(K)$ becomes a Yule distribution that behaves as a power law for small K:

$$P(K) \sim K^{-\varphi}, \quad [2]$$

where $\varphi = 2 + b/(1 - b) \geq 2$, followed by the exponential decay of Eq. **1** for large K with $K(t) = [n(0) + t]^{1-\mathrm{b}}\mathrm{n}(0)^{\mathrm{b}}/\mathrm{N}(0)$ (2, 6).

In the Simon model opportunities are assumed to be of unit size so that $S_\alpha(\mathrm{t}) = K_\alpha(\mathrm{t})$. On the contrary, we assume that each opportunity has a randomly determined but finite size. To capture new opportunities firms launch new products, open up new establishments, divisions, or units. Each opportunity is assigned to exactly one firm and the size of the firm is measured by the sum of the sizes of the opportunities it has taken up (4). In this article, we consider products as the relevant constituent parts of the companies and measure their size in terms of sales, even if alternative decompositions of the firm into subunits (e.g., plants, divisions) and measures of size (e.g., the number of employees, total assets) can be applied.

At time t, the size of each product $\xi_\mathrm{i}(\mathrm{t}) > 0$ is decreased or increased by a random factor $\eta_\mathrm{i}(\mathrm{t}) > 0$ so that

$$\xi_i(t + 1) = \xi_i(t)\eta_i(t), \quad [3]$$

where $\eta_i(t)$, the growth rate of product i, is an independent random variable taken from a distribution $P_\eta(\eta_i)$, which has a finite mean and standard deviation.

Thus, at time t a firm α has $K_\alpha(\mathrm{t})$ products of size $\xi_\mathrm{i}(\mathrm{t})$, $i = 1, 2, \ldots, K_\alpha(\mathrm{t})$ so that its total size is defined as the sum of the sales of its products $S_\alpha(\mathrm{t}) \equiv \sum_{\mathrm{i}=1}^{\mathrm{K}_\alpha}\xi_\mathrm{i}(t)$ and its growth rate is measured as $g = \log(S_\alpha(t + 1)/S_\alpha(t))$.

The probability distribution of firm growth rates $P(g)$ is given by

$$P(g) \equiv \sum_{K=1}^{\infty} P(K)P(g|K), \quad [4]$$

where $P(g|K)$ is the distribution of the growth rates for a firm consisting of K products. By using central limit theorem, one can show that for large K and small g, $P(g|K)$ converge to a Gaussian distribution

$$P(g|K) \approx \frac{\sqrt{K}}{\sqrt{2\pi V}}\exp\left(-\frac{(g - m)^2 K}{2V}\right), \quad [5]$$

where V and m are functions of the distributions P_ξ and P_η. For the most natural assumption of the Gibrat process for the sizes of the products these distributions are lognormal:

$$P_\xi(\xi_i) = \frac{1}{\sqrt{2\pi V_\xi}}\frac{1}{\xi_i}\exp(-(\ln \xi_i - m_\xi)^2/2V_\xi), \quad [6]$$

$$P_\eta(\eta_i) = \frac{1}{\sqrt{2\pi V_\eta}}\frac{1}{\eta_i}\exp(-(\ln \eta_i - m_\eta)^2/2V_\eta). \quad [7]$$

In this case,

$$m = m_\eta + V_\eta/2 \quad [8]$$

and

$$V \equiv K\sigma^2 = \exp(V_\xi)(\exp(V_\eta) - 1), \quad [9]$$

but for large V_ξ the convergence to a Gaussian is an extremely slow process. Assuming that the convergence is achieved, one can analytically show (1) that $P(g)$ has similar behavior to the Laplace distribution for small g, i.e., $P(g) \approx \exp(-\sigma\sqrt{2}|g|/\sqrt{V})/\sqrt{2V}$, whereas for large g $P(g)$ has power-law wings $P(g) \sim g^{-3}$, which are eventually truncated for $g \to \infty$ by the distribution P_η of the growth rate of a single product.

Using the fact that the nth moment of the lognormal distribution

$$P_x(x) = \frac{1}{\sqrt{2\pi V_x}}\frac{1}{x}\exp(-(\ln x_i - m_x)^2/2V_x), \quad [10]$$

is equal to

$$\mu_{n,x} \equiv \langle x^n \rangle = \exp(nm_x + n^2V_x/2) \quad [11]$$

we can make an expansion of a logarithmic growth rate in inverse powers of K:

$$\begin{aligned} g &= \ln \frac{\sum_{i=1}^{K}\xi_i\eta_i}{\sum_{i=1}^{K}\xi_i} \\ &= \ln \mu_{1,\eta} + \ln\left(1 + \frac{A}{K(1 + B/K)}\right) \\ &= m_\eta + \frac{V_\eta}{2} + \frac{A(1 - B/K + B^2/K^2\ldots)}{K} \\ &\quad - \frac{A^2(1 - B/K + B^2/K^2\ldots)^2}{2K^2} + \ldots \end{aligned}$$

$$= m_\eta + \frac{V_\eta}{2} + \frac{A}{K} - \frac{AB + A^2/2}{K^2} + O(K^{-3})$$

$$A = \frac{\sum_{i=1}^{K} \xi_i(\eta_i - \mu_{1,\eta})}{\mu_{1,\eta}\mu_{1,\xi}} \qquad [12]$$

where

$$B = \frac{\sum_{i=1}^{K} \xi_i - \mu_{1,\xi}}{\mu_{1,\xi}}. \qquad [13]$$

Using the assumptions that ξ_i and η_i are independent: $\langle \xi_i \eta_i \rangle = \langle \xi_i \rangle \langle \eta_i \rangle$, $\langle \eta_i \eta_j \rangle = \langle \eta_i \rangle \langle \eta_j \rangle$, and $\langle \xi_i \xi_j \rangle = \langle \xi_i \rangle \langle \xi_j \rangle$ for $i \neq j$, we find $\langle A \rangle = 0$, $\langle AB \rangle = 0$, $\langle A^2 \rangle = CK$, where $C = a(b-1)$ with $a = \exp(V_\xi)$ and $b = \exp(V_\eta)$. Thus

$$\mu = <g> = \sum_{n=0}^{\infty} \frac{m_n}{K^n}$$

$$\sigma^2 = <g^2> - \mu^2 = \sum_{n=1}^{\infty} \frac{V_n}{K^n}, \qquad [14]$$

where $m_0 = m_\eta + V_\eta/2$, $m_1 = -C/2$, $V_1 = C$, $V_2 = C[a(5b+1)/2 - 1 - a^2 b(b+1)]$. The higher terms involve terms like $\langle A^n \rangle / K^n$, which will become sums of various products $\langle \xi_i^k (\eta_i - \mu_{1,\eta})^k \rangle$, where $2 \leq k \leq n$. The contribution from $k = n$ has exactly K terms of $\mu_{n,\xi}\mu_{1,\xi}^{-n}\sum_{j=0}^{n}\mu_{j,\eta}\mu_{1,\eta}^{-j}(-1)^{n-j}\binom{j}{n}$ with $\mu_{j,x}\mu_{1,x}^{-j} = \exp(V_x j(j-1)/2)$. Thus, there are contributions to m_n and V_n that grow as $(ab)^{n(n+1)/2}$ with $ab > 1$, which is faster than the nth power of any $\lambda > 0$. The radius of convergence of the expansions (14) is equal to zero, and these expansions have only a formal asymptotic meaning for $K \to \infty$. However, these expansions demonstrate that μ and σ do not depend on m_η and m_ξ except for the leading term in μ: $m_0 = m_\eta + V_\eta/2$.

Not being able to derive close-form expressions for σ we perform extensive computer simulations, in which ξ and η are independent random variables taken from lognormal distributions P_ξ and P_η with various V_ξ and V_η (supporting information (SI) Figs. S1 and S2).

Empirical Evidence. The model relies on the assumptions of independence of the growth of products from each other and from the number of products K. However, these assumptions could be violated and at least three alternative explanations must be analyzed:

1. Size dependence. The probability that an active firm captures a new market opportunity is more or less than proportional to its current size. In particular, there could be a positive relationship between the number of products of firm α (K_α) and the size ($\xi_i(\alpha)$) and growth ($\eta_i(\alpha)$) of its component parts due to monopolistic effects and economies of scale and scope. If large and small companies do not get access to the same distribution of market opportunities, large firms can be riskier than small firms simply because they tend to capture bigger opportunities.

2. Units interdependence. The growth processes of the constituent parts of a firm are not independent. One could expect product growth rates to be positively correlated at the level of firm portfolios, due to product similarities and common management, and negatively correlated at the level of relevant markets, due to substitution effects and competition. Based on these arguments, one would predict large companies to be less risky than small companies because their product portfolios tend to be more diversified.

3. Time dependence. The growth of firms' constituent units does not follow a pure Gibrat process because of serial autocorrelation and life cycles. Young products and firms are supposed to be more volatile than predicted by the Gibrat Law because of learning effects. If large firms are older and have more mature products, they should be less risky than small firms. On the contrary, aging and obsolescence would imply that incumbent firms are more unstable than newcomers.

The first two hypotheses are not falsified by our data (Fig. S3).

The number of products of a firm and their average size defined as $\langle \xi(K) \rangle = \langle 1/K \sum_{i=1}^{K} \xi_i \rangle$, where $\langle\ \rangle$ indicates averaging over all companies with K products, has an approximate power law dependence $\langle \xi(K) \rangle \sim K^\gamma$, where $\gamma = 0.38$.

The mean correlation coefficient of product growth rates at the firm level $\langle \rho(K) \rangle$ shows an approximate power-law dependence $\langle \rho(K) \rangle \sim K^\zeta$, where $\zeta = -0.36$.

Because larger firms are composed by bigger products and are more diversified than small firms, the two effects compensate each other. Thus, if products are randomly reassigned to companies, the size-variance relationship will not change.

As for the time dependence hypothesis, despite some departures from a Gibrat process at the product level (Fig. S4) because of life cycles and seasonal effects, they are too weak to account for the size-variance relationship. Moreover, asynchronous product life cycles are washed out on aggregation.

1. Fu D, *et al.* (2005) The growth of business firms: theoretical framework and empirical evidence. *Proc Natl Acad Sci USA* 102:18801–18806.
2. Ijiri Y, Simon HA (1977) *Skew Distributions and the Sizes of Business Firms* (North-Holland Publishing Co., Amsterdam).
3. Gibrat R (1931) *Les inégalités économiques* (Librairie du Recueil Sirey, Paris).
4. Sutton J (1998) *Technology and Market Structure: Theory and History* (MIT Press, Cambridge, MA).
5. Amaral LAN, Buldyrev SV, Havlin S, Salinger MA, Stanley HE (1998) Power law scaling for a system of interacting units with complex internal structure. *Phys Rev Lett* 80:1385–1388.
6. Yamasaki K, *et al.* (2006) Preferential attachment and growth dynamics in complex systems. *Phys Rev E* 74:035103.

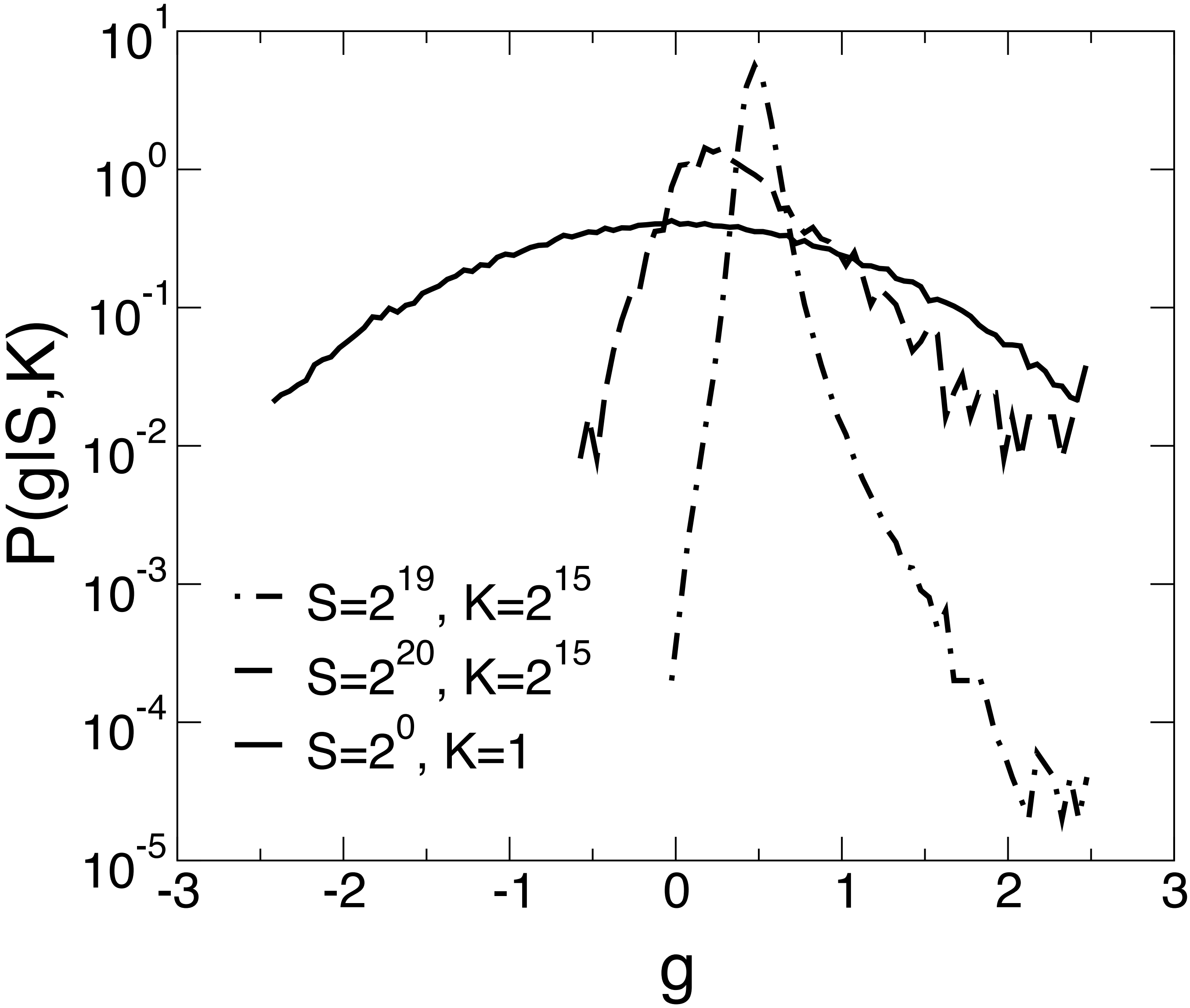

Fig. S1. Simulation results for the conditional growth rate distribution $P(g|S,K)$ for the case of lognormal P_ξ and P_η, with $V_\xi = 6$, $V_\eta = 1$, and $m_\xi = m_\eta = 0$. For $K = 1$ the distribution is perfectly Gaussian with $V_\eta = 1$ and $m_\eta = 0$. However, for large K the distribution develops a tent-shape form with the central part close to a Gaussian with mean $m = 1/2$ as predicted by Eq. **8**. The vast majority of firms (99.7%) have sizes in the vicinity of $K\mu_\xi$ which for $K = 2^{15}$ and $\mu_\xi = \exp(m_\xi + V_\xi/2) = 20.1$ belongs to the bin $[2^{19}, 2^{20}]$ and only 0.25% of firms belong to the next bin $[2^{20}, 2^{21}]$. These firms are due to a rare occurrence of extremely large products. The real number of products in these firms is $K_e = 2.4$, whereas the normally sized firms have $K_e = 31$. The fluctuations of these extremely large products dominate the fluctuations of the firm size and, hence, $P(g|S,K)$ for such abnormally large firms is broader than for normally sized firms. Accordingly, $\sigma = 0.09$ and $\sigma = 0.41$, respectively, for the normally sized and abnormally large firms.

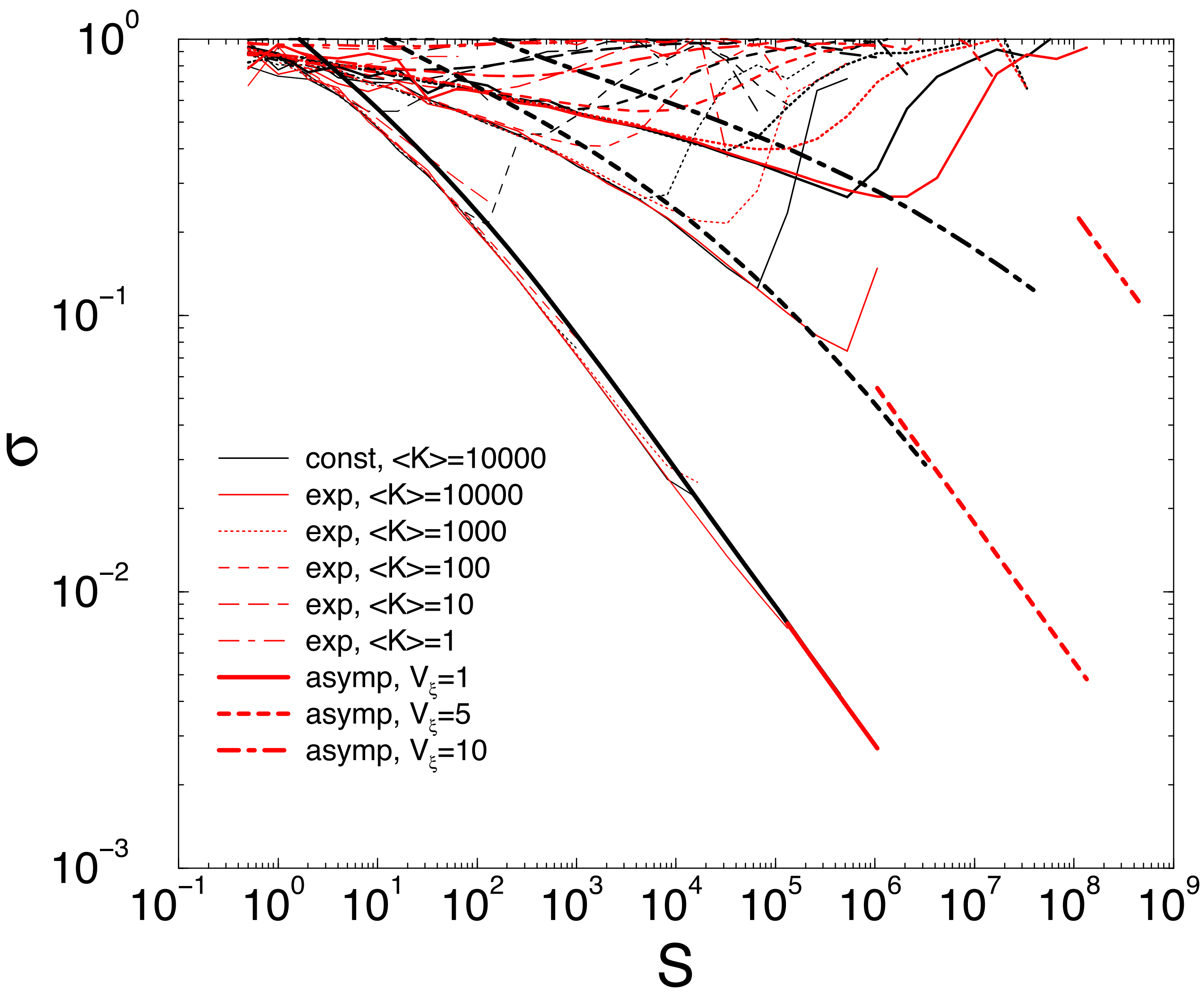

Fig. S2. The behavior of $\sigma(S)$ for the exponential distribution $P(K) = \exp(-K/\langle K\rangle)/\langle K\rangle$ and lognormal P_ξ and P_η. We show the results for $K_0 = 1$, 10, 100, 1,000, and 10,000 and $V_\xi = 1$, 5, and 10. The graphs $\sigma(K_S)$ and the asymptote given by

$$\sigma(S) = \sqrt{V/K_S} = \frac{\exp(3V_\xi/4 + m_\xi/2)\exp(V_\eta)-1}{\sqrt{S}}$$

are also given to illustrate our theoretical considerations. One can see that for $V_\xi = 1$, $\sigma(S)$ almost perfectly follows $\sigma(K_S)$ even for $\langle K\rangle = 10$. However, for $V_\xi = 5$, the deviations become large and $\sigma(S)$ converges to $\sigma(K_S)$ only for $\langle K\rangle > 100$. For $V_\xi = 10$ the convergence is never achieved.

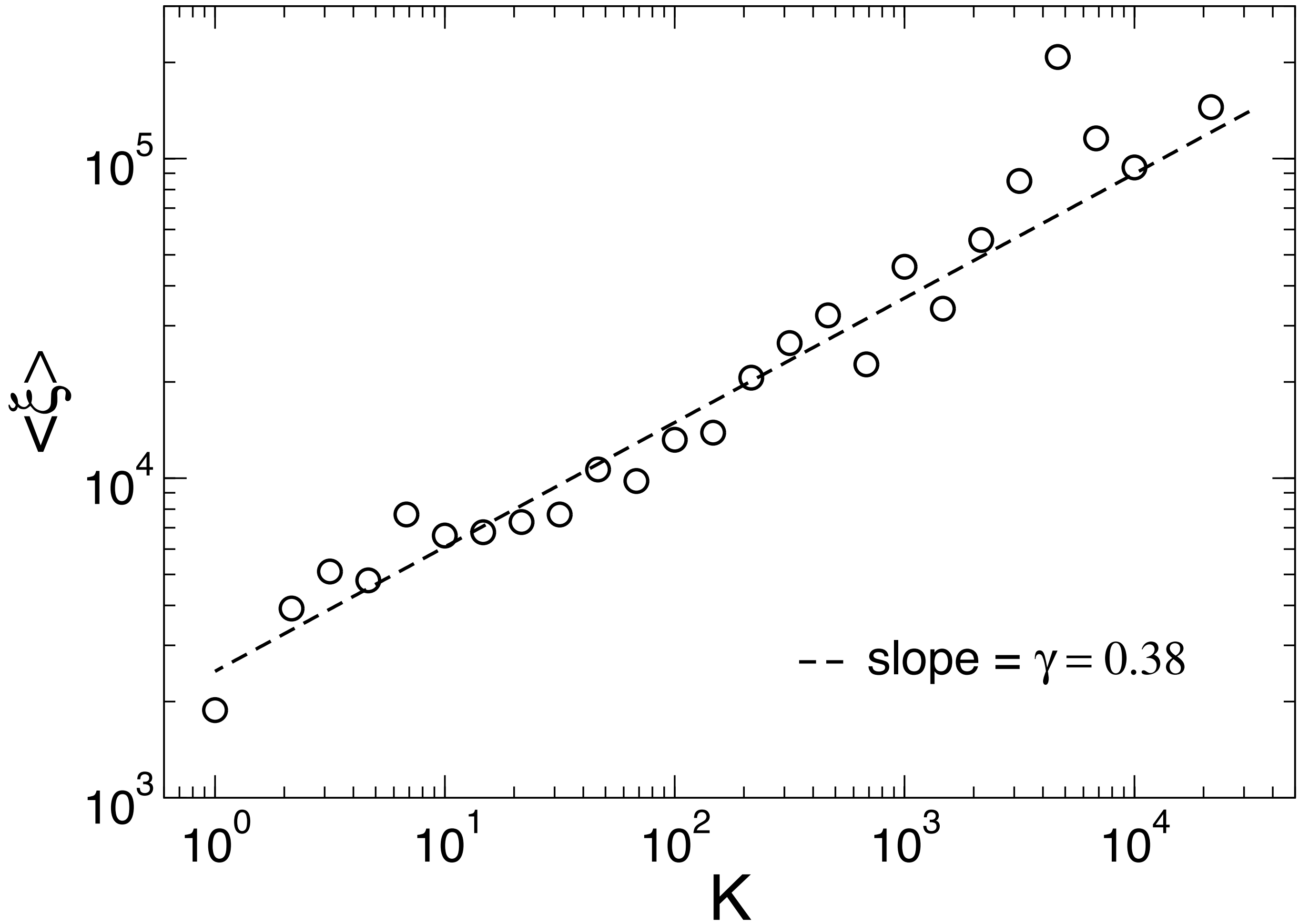

Fig. S3. The relationship between the average product size and the number of products of the firm. The log-log plot of $\langle \xi(K) \rangle$ vs. K shows power-law dependence $\langle \xi(K) \rangle \sim K^{0.38}$.

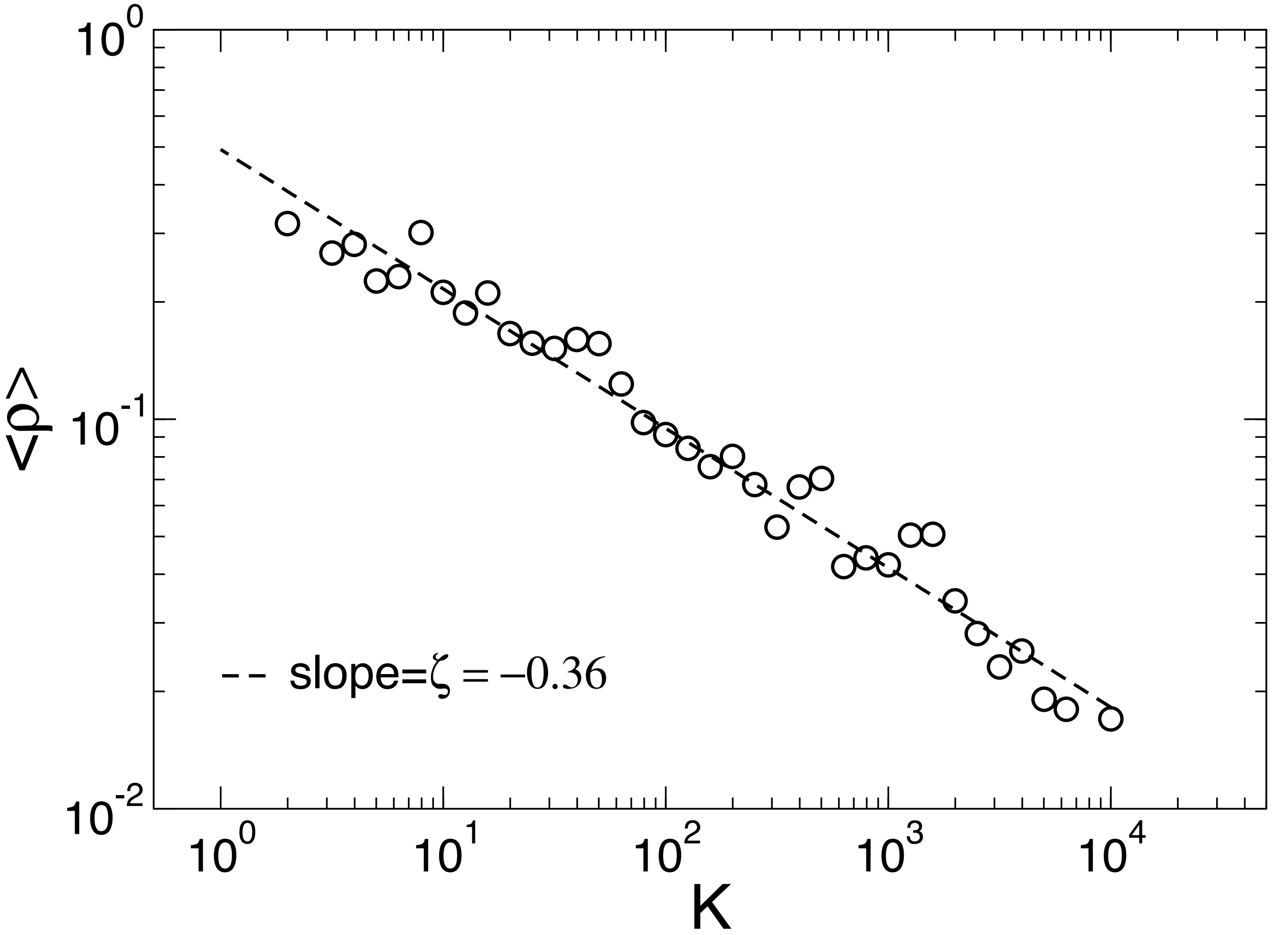

Fig. S4. The average growth rate and the autocorrelation coefficient of firms from entry. The departures of product growth from a Gibrat process are washed out on aggregation. The growth rates do not depend on age and do not show a significant autocorrelation.