Tjalling C. Koopmans Research Institute

Discussion Paper Series nr: 10-23

Firm Size and Growth Rate Variance: the Effects of Data Truncation

Marco Capasso
Elena Cefis

Universiteit Utrecht

Utrecht School of Economics

Tjalling C. Koopmans Research Institute
Utrecht School of Economics
Utrecht University

Janskerkhof 12
3512 BL Utrecht
The Netherlands
telephone +31 30 253 9800
fax +31 30 253 7373
website www.koopmansinstitute.uu.nl

The Tjalling C. Koopmans Institute is the research institute and research school of Utrecht School of Economics.
It was founded in 2003, and named after Professor Tjalling C. Koopmans, Dutch-born Nobel Prize laureate in economics of 1975.

In the discussion papers series the Koopmans Institute publishes results of ongoing research for early dissemination of research results, and to enhance discussion with colleagues.

Please send any comments and suggestions on the Koopmans institute, or this series to J.M.vanDort@uu.nl

ontwerp voorblad: WRIK Utrecht

How to reach the authors

Please direct all correspondence to the last author.
Marco Capasso
Vakgroep Algemene Economie
School of Business and Economics
Maastricht University
Maastricht Economic Research Institute of Innovation and Technology
United Nations University (UNU-MERIT)
P.O. Box 616
6200 MD Maastricht - The Netherlands
E-mail:mailto:marco.capasso@gmail.com
Elena Cefis
Utrecht University
Utrecht School of Economics
Janskerkhof 12
3512 BL Utrecht
The Netherlands.
Dep. of Economics
University of Bergamo
Via dei Caniana 2
24127 Bergamo,
Italy
E-mail: elena.cefis@unibg.it

Utrecht School of Economics
Tjalling C. Koopmans Research Institute
Discussion Paper Series 10-23

Firm Size and Growth Rate Variance: the Effects of Data Truncation

Marco Capasso[a]
Elena Cefis[b]

[a]School of Business and Economics and UNU-MERIT,
Maastricht University.

[b]Department of Economics,
University of Bergamo
Utrecht School of Economics
Utrecht University

November 2010

Abstract

This paper discusses the effects of the existence of natural and/or exogenously imposed thresholds in firm size distributions, on estimations of the relation between firm size and variance in firm growth rates. We explain why the results in the literature on this relationship are not consistent. We argue that a natural threshold (0 number of employees or 0 total sales) and/or the existence of truncating thresholds in the dataset, can lead to upwardly biased estimations of the relation. We show the potential impact of the bias on simulated data, suggest a methodology to improve these estimations, and present an empirical analysis based on a comprehensive dataset of Dutch manufacturing and service firms. The only stable relation between firm size and growth rate variance is negative regardless of how we define the measure of firm growth.

Keywords: firm's growth, growth rates variance; truncation; thresholds

JEL classification: L25, C21.

Acknowledgements

The authors thank Andrew Bernard, Koen Frenken and all the participants at the DIME Workshop on "The dynamics of firm evolution: productivity, profitability and growth", Pisa, 3-4 October 2008. The empirical work was carried out at the Centre for Research of Economic Microdata at Statistics Netherlands (CBS). The views expressed in this paper are those of the authors and do not necessarily reflect the policies of Statistics Netherlands. The authors thank Gerhard Meinen and on-site staff of CBS for their collaboration. This work was supported by Utrecht University [High Potential Grant (HIPO) to E. Cefis and K. Frenken]; and the University of Bergamo [grant ex 60%, n. 60CEFI10, Dept. of Economics, to E.Cefis].

1. Introduction

Several studies on industrial dynamics and firm growth in particular, highlight that the variance in growth rates decreases with increasing firm size. This result was observed especially in studies that find a “reversion to the mean” pattern of firm growth: small firms grow faster than large firms and consequently the variance in growth rates decreases as firm size increases. A negative relation between firm size and growth rate variance would be in accordance with this growth pattern, and would contradict the predictions of the Law of Proportionate Effects that the size of a firm and its growth rate are independent. Meyer and Kuh (1957) were the first to observe this negative relation, which was confirmed by other researchers using different datasets over different time periods, for example Hymer and Pashigian (1962), Stanley et al. (1996), Amaral et al. (1997), Amaral et al. (1998), Bottazzi and Secchi (2003., 2006), Matia et al. (2004). According to Axtell (2001, p. 1820), the fact that "the standard deviation in growth rates falls with initial firm size according to a power law" places "important limits on models of firm dynamics". Studies where size is approximated by total sales or numbers of employees focus only on the growth of medium and large firms. Works that includes small firms in the analyses, in particular Bottazzi, Cefis and Dosi (2002), Perline, Axtell and Teitelbaum (2006) and Bottazzi et al. (2007), find either no relation or find a positive relation, between firm size and growth rate variance. Coad (2008, p. 2) argues that the reason why Bottazzi et al. (2007) fail to find any significant relationship between firm size and firm growth rate “could well be due to the fact that the firms analyzed in Bottazzi et al. (2007) are smaller than those firms in the empirical analyses discussed above”. Perline, Axtell and Teitelbaum (2006, p. 8) observe that, beyond the possible economic explanations, a simple statistical phenomenon emerges when medium or small firms are included in the sample: “The high concentration of small establishments […] highlights the issue of establishments (or corporations in other studies) that exit from a longitudinal database because they drop to size 0”.

The researcher’s need for a cross-section (or a panel) of firm growth rates measured between two time periods (e.g. two adjacent years) generally implies that firms must persist over two time periods i.e. all the firms that exited the market in the time between the two observations are generally excluded from the database. The effect of this is negligible in a sample of large and/or publicly traded firms, but can become important when micro and small firms are being considered. The problem of dropouts does not depend only on real exits, but also on the existence of an exogenous threshold in the data collection process, which means that data are collected only for firms whose size exceeds a given threshold. A common example is industry panel datasets that supply data only on firms with at least 10 or 20 employees. Annual production statistics such as those produced by national statistical offices (e.g. CBS in the Netherlands, ISTAT in Italy) provide data on firms with a minimum of 20 employees; the Community Innovation Surveys (CIS) collect data on representative samples of firms with at least 10 or 20 employees (the different thresholds depend on the country). In these data, an exit from the database may not correspond to a real exit from the market, but only to a decrease in the number of a firm's employees below the threshold. An example here is the database used by Bottazzi et al. (2007), who analyse Italian firms with at least 20 employees. When number of employees is used as a proxy for size, they find a clear positive relation between size and growth rate variance. This might be explained by the fact that for each two-year time span considered to compute firm growth rates, the same firm must have at least 20 employees in both years in order to be included in the analysis. On the basis of this example we would suggest that the existence of a "database threshold" of 20 employees, as in Coad (2007), Bottazzi et al. (2007) and Bottazzi et al. (2009), could have the same effect as the "natural threshold" of 1 employee described in Perline, Axtell and Teitelbaum (2006). In other words, it is immaterial whether dropouts are caused by real market exits or database exits: in both cases there is an upward bias in the estimation of the relation between firm size and growth rate variance.

In our study, we highlight the problems that arise when estimating the relation between size and growth rate variance on panel databases of firms. We argue that the presence of a left-truncation of the firm size empirical distribution, a truncation that may be due to a natural or endogenous threshold (e.g. zero when size is proxied by the number of employees) or to an exogenous threshold when data are collected only for firms above a certain threshold (as in the cases of databases that do not include micro and small firms), causes an upward bias in the estimation. We propose an improved methodology to avoid this problem, based on the method suggested by Perline, Axtell and Teitelbaum (2006). We apply this methodology to an empirical analysis of manufacturing and services firms in the Netherlands to show how the impact of slight differences in procedures can have a dramatic impact on results.

The paper is structured as follows. Section 2 describes the model to be estimated. Section 3 defines the three cases that may emerge if firms exit the panel dataset during the two consecutive periods on which firm growth is calculated. Section 4 uses a simulation to provide a numerical example. Section 5 describes an alternative methodology to avoid estimation bias and in Section 6 this methodology is applied to a Dutch panel dataset of manufacturing and service firms. Section 7 concludes.

2. The model

In this study, the main variables of interest are the size of the firm and the firm's rate of growth. For each firm *j* and each year *t*, we approximate the size $S_j(t)$ by 1 plus the number of employees (we add 1 to allow for logarithmic transformations, since self-employment is treated as a firm with 0 employees). Following previous studies, for example Bottazzi et al. (2007), Bottazzi et al. (forthcoming), Coad (2007), and Coad and Hölzl (2009), the rate of growth of firm size is calculated as the difference in the log size across two consecutive years, namely

$$\widehat{g}_j(t) = \log(S_j(t)) - \log(S_j(t-1)) \qquad (1)$$

The literature commonly tests the relation between firm size and variance in growth rates to divide a sample of firms into several equi-populated size classes. The relation between firm size and variance in growth rates then is measured by estimating the model

$$y_i = \alpha + \beta x_i + \varepsilon_i \qquad (2)$$

where x_i and y_i respectively are average size and the log standard deviation of growth rates computed within the size class *i*, and ε_i is an error term. Size and variance in growth rates are correlated in case the slope β is different from zero.

3. Source of the bias

To illustrate the problem that can arise in the presence of endogenous or exogenous thresholds we assume a very simple scenario in which firms can have only two sizes. Suppose our sample contains two equipopulated groups of n_{A} firms. All the $n_{\mathrm{A1}} = n_{\mathrm{A}}$ firms belonging to the first group, called A1, have a size equal to x_{A1}, and all the $n_{\mathrm{A2}} = n_{\mathrm{A}}$ firms belonging to the second group, called A2, have a size equal to x_{A2}, with $x_{\mathrm{A1}} < x_{\mathrm{A2}}$. Suppose that the stochastic process generating firm growth, as defined in (1), is the same for all firms, and the real distribution of the relative frequencies of growth rates between periods 1 and 2, for each size class, is standard normal. Then, for each group the logarithm of the standard deviation of the growth rates is equal to 0, that is, there is no relation between firm size and the variance in growth rates.

CASE A: the researcher can observe all firms in both of years 1 and 2, i.e. no firms exit the observed set.

For our two-point case, the estimation of (2) provides the slope

$$\beta_{\text{A}} = \frac{y_{\text{A2}} - y_{\text{A1}}}{x_{\text{A2}} - x_{\text{A1}}} = 0 \qquad (3)$$

where β_{A} properly indicates that there is no relation between firm size and variance in growth rates.

CASE B: the researcher can observe only the firms whose size in both periods is larger than a given threshold, and where size classes are not redefined if some firms have exited the observed set in year 2.

Suppose now that we cannot observe firms whose size is smaller than a given threshold $\tau < x_{\text{A1}}$. In this case, we are not able to observe (in period 2) all the firms from the first group with a growth rate smaller than $\kappa_1 = \log(\tau) - \log(x_{\text{A1}})$ nor all firms in the second group whose growth rate is smaller than $\kappa_2 = \log(\tau) - \log(x_{\text{A2}})$.

As a consequence, the observed growth rate distributions are standard normal truncated, respectively at $\kappa_1 < 0$ for the first group (which we call B1 after the truncation) and at $\kappa_2 < 0$ for the second group (called B2 after the truncation), where $\kappa_1 > \kappa_2$. The group B1 is composed of $n_{\text{B1}} < n_{\text{A1}}$ firms, and the group B2 of $n_{\text{B2}} < n_{\text{A2}}$ firms. Notice that $n_{\text{B1}} < n_{\text{B2}}$ as $\kappa_1 > \kappa_2$.

Barr and Sherrill (1999) prove that the variance of a truncated standard normal distribution Z is:

$$V(Z) = c(\kappa)\left[\sqrt{\frac{\pi}{2}}[1 + C_3(\kappa^2)] - c(\kappa)e^{-\kappa^2}\right] \qquad (4)$$

where κ is the negative truncation point in the distribution, $[1 + C_3(\kappa^2)]$ is the integral of a chi-square density function with three degrees of freedom over the two branches $(\kappa, 0)$ and $(0, +\infty)$, and $c(\kappa) = 1/\left[\sqrt{2\pi}(1 - \Phi(\kappa))\right]$, $\Phi(\cdot)$ being the standard normal cumulative distribution function. Barr and Sherrill (1999) find that V(Z) is decreasing with κ, as depicted in Figure 1. In our case, substituting respectively κ_1 and κ_2 in (4) for each size class we obtain the respective variances $V(Z_1)$ and $V(Z_2)$, where $V(Z_1) < V(Z_2)$ as $\kappa_1 > \kappa_2$. The log standard deviations used to estimate model (2) now become $y_{B1} = \log\sqrt{(V(Z_1))}$ and $y_{B2} = \log\sqrt{(V(Z_2))}$, where $y_{B1} < y_{B2}$, and the corresponding average sizes will be $x_{B1} = x_{A1}$ and $x_{B2} = x_{A2}$. The observed slope is now positive and equal to

$$\beta_B = \frac{y_{B2} - y_{B1}}{x_{B2} - x_{B1}} = \frac{y_{B2} - y_{B1}}{x_{A2} - x_{A1}} > 0\,, \qquad (5)$$

which does not properly measure the real phenomenon, suggesting instead the existence of a positive relation between firm size and variance in growth rates.

CASE C: the researcher can observe only firms whose size is in both periods above a given threshold, and where size classes are redefined after some firms have exited the observed set in year 2, in order to have equipopulated size classes after firm exits.

Suppose that, after the truncation and thus after some firms are excluded from the original population, we redefine the size classes in order to keep them equipopulated. We assume that there will be a larger number of exits from the group of firms with an initial size x_{A1} than from the group with an initial size x_{A2}. Given that the redefinition of the size classes must still be based on initial firm size, some firms, randomly chosen from those with an initial size equal to x_{A2}, must move from the second to the first size class (with the firms with an initial size x_{A1}), leaving the remaining group of the firms with an initial size equal to x_{A2} in the second class.
We now have two new groups. Group C2 is composed of the firms from group B2 that survived the truncation and were not reclassified into the first group; the growth rate distribution of C2 is still normal truncated in κ_2 (then $y_{C2} = y_{B2}$) and the average size value x_{C2} associated with the group is exactly equal to $x_{B2} = x_{A2}$. Group C1, then, is composed of the "surviving" firms from

group B1 plus some firms from group B2. Its average size value x_{C2} is a weighted average of x_{A1} and x_{A2}, thus $0 < x_{C2} - x_{C1} = x_{B2} - x_{C1} < x_{B2} - x_{B1} = x_{A2} - x_{A1}$; since its growth rate distribution results from the sum of the two truncated normal distributions Z_1 and Z_2, the corresponding log standard deviation y_{C1} should be lower than zero (i.e. lower than for a normal distribution), and should be between y_{B1} and y_{B2}. Thus (see Figure 2), the observed slope β_C will be positive (i.e. upward biased), but whether it is lower or higher than β_B will depend on the value of the derivative of (4) with respect to κ, given the position of the threshold in our dataset.
We provide a numerical example for the previous three cases by running a simulation.

4. The simulation

We assume that the two original groups, A1 and A2 respectively, include firms whose size is equal to $x_{A1} = 5$ and $x_{A2} = 10$. We cannot observe firms whose size becomes smaller than 1. Between periods 1 and 2, the distribution of the growth rates (meant as log size differences) is normal with standard deviation equal to $\bar{\sigma}$. Table (1) presents the results for different values of $\bar{\sigma}$.[1] Each result is the arithmetic mean of 1,000 replications of the simulation for each given value of $\bar{\sigma}$ between 2.0 and 3.5, in 0.1 steps. Figure 2 shows that the slope β_A, obtained if we do not consider the truncation, is close to zero, which is to be expected since growth rate distributions do not depend on size. The slope β_B, obtained when we cannot observe those firms whose size has become lower than 1, is always positive. The slope β_C, obtained when we cannot observe those firms whose size has become lower than 1, and when after the truncation we have made the two size classes equipopulated, is still positive although slightly lower than β_B.

We have shown that the existence of a truncation (the same for both periods) creates a positive bias in the estimation of coefficient β in equation (2). This is because we cannot observe small firms that experience very low (high in absolute value) growth rates, thus creating selection bias

[1] Simulations were performed using the Matlab software package. The program code is available on request.

in our sample. Indeed, if the smallest size admitted into the database is 1, a firm of size $S_1(t-1)$ at time *t* can have only a growth rate that is larger than $-\log S_1(t-1)$, a firm of size $S_2(t-1)$ can have only a growth rate that is larger than $-\log S_2(t-1)$, and so on. This is akin to implicitly imposing that the stochastic process of growth is different for each firm, and the associated probability density distribution is always left truncated with a left truncation that depends on the initial size of the firm. If in the subsampling procedure, we approximate these probability distributions by means of the frequency distributions of the growth rates within each subsample, and then estimate the subsample variances of growth rates, we are implicitly maintaining the arbitrary assumption of a left-truncation in the growth rate distribution, where the left-truncation moves according to firm size.

5. Alternative methodology

In empirical terms, the truncation can derive from reality (firms cannot have less than zero employees) or from exogenously imposed thresholds in the construction of the database (e.g. if the database collects data only on firms with at least 10 or 20 employees, such as the CIS for all European countries). To limit the biases deriving from truncation, we suggest the exclusion from the dataset of firms that are below a given size threshold *M* (slightly higher than the threshold already imposed on the data) at time *t-1*, and that the remaining firms are used to build the balanced sample of growth rates between time *t-1* and time *t*. Notice that our artificial threshold is applied only at time *t-1*: no constraint on firm size should be imposed at time t. In the next section we describe the calibration of *M* in order to reduce the number of firms excluded from the dataset. The resulting distribution of growth rates is not bounded from above because, in theory, a firm could grow indefinitely and still belong to the sample, and is not strictly bounded from below because the firm would have to experience a very low negative growth rate (high in absolute value) in order to approach the natural (or endogenous) threshold of zero employees or any exogenous threshold imposed by the database construction. This methodology extends the proposal in Perline, Axtell and Teitelbaum (2006) to simply exclude from the regression the first size class (i.e. associated with the smallest average size).

We also suggest an alternative way to measure growth rates to lower the bias associated with the left-truncation of the growth rate distribution. Using logarithmic differences, as in some of the studies already referred to, is useful in terms of such properties as the additivity of growth rates over time; however, it causes the lower boundary of the growth rate distribution to become closer to zero as the average size of the firms under analysis become smaller, which exacerbates the problem of endogenous and exogenous thresholds. As we will see in the empirical analysis, alternative definitions of growth rates, such as the weighted difference of the plain size

$$\tilde{g}_i(t) = \frac{S_i(t) - S_i(t-1)}{\frac{1}{2}\left(S_i(t) + S_i(t-1)\right)} \ , \qquad (7)$$

which confines the growth rate distribution between -2 and 2, allows us to adopt a lower value of the artificial threshold *M*, and to remove the bias without loss of too many data. While the definition in eq.1 is adopted by most of the literature on firm growth rates, an application in the same context of the definition in eq.7 can be found in Davis et al. (2006).

In order to test our methodology and to get an approximate value for *M*, we conduct the empirical analysis.

6. The empirical analysis

The data in this paper were collected by the Centraal Bureau voor de Statistiek (CBS) and stem from the Business Register of enterprises. The Business Register database includes all firms registered for fiscal purposes in the Netherlands in the year considered. The population includes firms with zero employees, referred to as self-employment. We consider growth rates between 2002 and 2003 for all Dutch firms in manufacturing and services (approximately 60,000 manufacturing and 1 million services firms), considered separately, that is, as belonging to two different macro-sectors..

The first part of the analysis is performed using firm growth rates calculated as log size differences (eq. 1). In a first step, we exclude from the sample all firms with less than *M* employees at time *t-1* (i.e. in year 2002) and retain all the remaining firms that still exist at time *t* (i.e. in 2003). *M* is an integer value varying between 0 and 20 and we consider all the cases between 0 and 20: when *M* is equal to 0 no firms are excluded by the artificial threshold, we retain all firms that exist in 2002 and 2003.

In a second step we estimate the relation between firm size and growth rate variance (eq.2) by minimizing the Least Absolute Deviation of the error terms ε_i, assuming that the residuals are Laplace distributed as in Bottazzi, Cefis, Dosi and Secchi (2007).[2]

Table 2 shows that the estimated coefficient β seems to decrease with *M*, with a clear tendency to move from positive to negative values as the threshold increases. Our methodology consists of excluding from the database firms whose number of employees is, at time *t-1*, below a threshold *M*, letting *M* increase until the estimated coefficient is stable (i.e. keeping the smallest *M* for which the estimated β finds a plateau). It should be possible to achieve stability for a reasonably small *M*, in our case M=9 for the manufacturing and service sectors. If estimation of the coefficient β does not stabilize at small values of *M*, model (3) is probably misspecified (i.e. there are nonlinearities in the relation being studied).

The second part of the analysis repeats the above procedure but using firm's growth rates as weighted size differences, that is, calculated according (eq.7). Table 3 shows that the main results obtained using weighted size differences are much less dependent on the artificial threshold imposed at time *t-1*: we obtain negative values for services at most values of *M*, and negative values for manufacturing at all values of *M*. In other words, it would seem that using the definition of growth rates expressed in (eq.7) allows us to choose a very low threshold M which inevitably excludes a much smaller number of firms from the database and produces an unbiased estimation of the coefficient of the relation between firm size and variance in growth rates.

---

[2] The variables used in the regressions were built using the R software package. The LAD regressions were run using the *qreg* function of the Stata software package.

Figures 3 and 4, which refer to growth rates computed respectively as log size differences (eq.1) and as weighted size differences (eq.7), graphically summarize the relation between the estimation of the coefficient β and the artificial threshold *M*. It is immediately clear that in both cases the estimation of this coefficient becomes insensitive to the artificial threshold *M* after this threshold becomes sufficiently high, but that while in the first case, stability is achieved after excluding all firms with less than nine employees, in the second case stability is achieved after excluding firms with less than two employees for manufacturing, and less than 3 employees for services. Therefore, our methodology would suggest a threshold $M=9$ in the first case, and thresholds of $M=2$ for manufacturing and $M=3$ for services, in the second case.

Our methodology allows us to observe that the relationship between firm size and variance in firm's growth rates is stable only for negative values of the coefficient. A positive relation is due only to the effects of truncation derived from the existence of endogenous and/or exogenous thresholds that bias the estimation upwards. This result is robust regardless of how firm growth rates are defined, although the use of weighted size differences, as proxy for growth rates, allow us to choose a very low artificial threshold, and thus to exclude fewer firms than when log size differences are used.

In our view, the contrasting results (i.e. the null or positive relationship between firm size and variance in growth rates) in the literature are caused by truncation problems exacerbated by the use of the log definition of firm growth.

7. Conclusions

We have discussed some of the biases that can arise when estimating the relation between firm size and variance in growth rates. We focused in particular on the problem of dropouts when endogenous and/or exogenous thresholds truncate the firm size distribution, that is, when micro firms are included in the analysis or when the dataset considers only firms whose size is above a certain threshold. This problem was highlighted by Perline, Axtell and Teitelbaum (2006), but is ignored in most of the literature. However, it seems to be a determinant in explaining the contrasting evidence from previous studies, that is, that the variance in firms' growth rates decreases or increases with firm size. After pointing to the biases that can derive from a truncated firm's size distribution, from a theoretical perspective, we suggested a simple methodology to estimate the relation between firm size and variance in firm growth rates that avoids these biases. We tested our methodology using two different definitions of firm growth rates, on the population of manufacturing and service firms in the Netherlands.

The results show that the methodology we propose allows us to observe how the estimated coefficients of the size-growth rate variance relation change depending on the firm size threshold. We suggest using an artificial threshold M that can increase until the estimated coefficient is stable and retaining the lowest M for which the estimated β finds a plateau.

Our results show also that using firm growth rates defined as weighted size differences reduces the bias that arises from truncation and that using our methodology excludes a much smaller number of firms, to obtain an unbiased coefficient of the firm size-growth rate relation.

Our empirical analysis shows there is a stable, negative relationship between firm size and variance in growth rates: the dynamics of smaller firms is characterized by a stronger turbulence. Most previous studies show a positive relationship, but this is due to the presence of natural/endogenous and exogenous thresholds in the firm size distributions.

References

Amaral, L. A. N., Buldyrev, S. V., Havlin, S., Salinger, M. A., & Stanley, H. E. (1998). Power law scaling for a system of interacting units with complex internal structure. *Physical Review Letters*, 80(7):1385–1388.

Amaral, L. A. N., Buldyrev, S. V., Havlin, S., Salinger, M. A., Stanley, H. E., & Stanley, M. H. R. (1997). Scaling behavior in economics: the problem of quantifying company growth. *Physica A*, 244, 1–24.

Axtell, R.L. (2001) Zipf Distribution of U.S. Firm Sizes. *Science*, 293, 1818-1820

Barr, D.R. & Sherrill, E.T. (1999). Mean and variance of truncated normal distributions. *The American Statistician*, 53(4), 357–361.

Bottazzi, G., Cefis, E., & Dosi, G. (2002). Corporate growth and industrial structure: Some evidence from the Italian manufacturing industry. *Industrial and Corporate Change*, 11, 705–723.

Bottazzi, G., Cefis, E., Dosi, G., & Secchi, A. (2007). Invariances and diversities in the evolution of manufacturing industries. *Small Business Economics*, 29, 137–159.

Bottazzi, G., Coad, A., Jacoby, N., & Secchi, A. (2009). Corporate growth and industrial dynamics: Evidence from French manufacturing. *Applied Economics*, forthcoming, DOI: 10.1080/00036840802400454.

Bottazzi, G., & Secchi, A. (2003). Common properties and sectoral specificities in the dynamics of U. S. manufacturing companies. *Review of Industrial Organization*, 23, 217–232.

Bottazzi, G., & Secchi, A. (2006). Explaining the distribution of firms growth rates. *Rand Journal of Economics*, 37(2), 235–256.

Coad, A. (2007), A closer look at serial growth rate correlation. *Review of Industrial Organization*, **31** (1), 69-82.

Coad, A. (2008), Firm growth and scaling of growth rate variance in multiplant firms. *Economics Bulletin*, **12** (9), 1-15.

Coad, A., Hölzl, W. (2009), On the autocorrelation of growth rates. *Journal of Industry, Competition and Trade,* **9** (2), 139-166.

Coad, A., Rao, R. (2010). Firm growth and R&D expenditure. *Economics of Innovation and New Technology*, **19** (2), 127-145.? Not in text

Davis, S.J., Haltiwanger, J., Jarmin, R. & Miranda, J. (2006). Volatility and dispersion in business growth rates: publicly traded versus privately held firms. National Bureau of Economic Research, Inc., NBER Working Papers 12354.

Hymer, S., & Pashigian, P. (1962). Firm Size and the Rate of Growth. *Journal of Political Economy*, 70(4), 556-569.

Matia, K., Fu, D., Buldyrev, S. V., Pammolli, F., Riccaboni, M., & Stanley, H. E. (2004). Statistical properties of business firms structure and growth. *Europhysics letters*, 67(3), 498-503.

Meyer, J.R. and Kuh, E. (1957). *The Investment Decision*. Harvard University Press, Cambridge, Massachusetts.

Palestrini, A. (2007)**.** Analysis of industrial dynamics: a note on the relationship between firms' size and growth rate, *Economics Letters*, 94(3), 367-371.? Not in text

Perline, R., Axtell, R, & Teitelbaum, D. (2006), Volatility and asymmetry of small firm growth rates over increasing time frames. U.S. Small Business Administration, The Office of Advocacy Small Business Working Papers 06/rarpdt.

Stanley, M.H.R., Amaral, L.A.N., Buldyrev, S.V., Havlin, S., Leschorn, H., Maass, P., Salinger, M.A,. & Stanley, H.E. (1996). Scaling behavior in the growth of companies. *Nature*, 379, 804-806.

Sutton, J. (2002). The variance of firm growth rates: the 'scaling' puzzle. *Physica A*, 312, 577–590.

Tornqvist, L., Vartia, P., & Vartia, Y. (1985). How should relative change be measured? *The American Statistician,* 39(1), 43-46.

Table 1: Simulated (β_{A}) and estimated (β_{B}, β_{C}) values for the relation between size and log standard deviation of growth rates ($\log\sigma(g)$)

$\bar{\sigma}$	β_{A}	β_{B}	β_{C}
2.00	0.000094	0.016486	0.016267
2.10	0.000037	0.015990	0.015759
2.20	-0.000045	0.015434	0.015195
2.30	-0.000022	0.014975	0.014818
2.40	-0.000085	0.014506	0.014308
2.50	0.000065	0.014249	0.014079
2.60	0.000024	0.013808	0.013629
2.70	-0.000013	0.013453	0.013294
2.80	-0.000036	0.013063	0.012902
2.90	-0.000047	0.012664	0.012555
3.00	-0.000050	0.012265	0.012128
3.10	-0.000052	0.012010	0.011859
3.20	-0.000103	0.011484	0.011309
3.30	-0.000028	0.011421	0.011292
3.40	0.000066	0.011180	0.011001
3.50	0.000111	0.010869	0.010751

Table 2: Empirical relation between size and growth rate variance: growth rates as log size differences.

	Manufacturing				Services			
Threshold M	Coefficient	Standard error	t-ratio	p-value	Coefficient	Standard error	t-ratio	p-value
0	0.230674	0.0725512	3.18	0.004	0.2062013	0.0684692	3.01	0.006
1	0.2639847	0.0556176	4.75	0	0.4570136	0.023385	19.54	0
2	0.171481	0.0429158	4	0.001	0.3309451	0.0695685	4.76	0
3	0.0394362	0.0657523	0.6	0.555	0.2090258	0.0694974	3.01	0.006
4	-0.0201472	0.0545533	-0.37	0.715	0.0998723	0.0581483	1.72	0.099
5	-0.0199713	0.039089	-0.51	0.614	0.0807111	0.0452961	1.78	0.088
6	-0.0979089	0.0718298	-1.36	0.186	0.0690582	0.037816	1.83	0.081
7	-0.1411341	0.0602617	-2.34	0.028	0.0010362	0.0635136	0.02	0.987
8	-0.156124	0.0729555	-2.14	0.043	-0.024666	0.0402654	-0.61	0.546
9	-0.2203155	0.0491147	-4.49	0	-0.1058003	0.0253304	-4.18	0
10	-0.2025012	0.0343726	-5.89	0	-0.1140678	0.0387135	-2.95	0.007
11	-0.2400626	0.0493007	-4.87	0	-0.1062662	0.0154259	-6.89	0
12	-0.263497	0.0494163	-5.33	0	-0.1112175	0.0410291	-2.71	0.012
13	-0.2076971	0.0474368	-4.38	0	-0.1283899	0.0208001	-6.17	0
14	-0.2164968	0.0640041	-3.38	0.003	-0.1125624	0.0458423	-2.46	0.022
15	-0.2920069	0.0523248	-5.58	0	-0.1472411	0.0275546	-5.34	0
16	-0.2044645	0.0453665	-4.51	0	-0.0638025	0.0419402	-1.52	0.142
17	-0.2138333	0.0500104	-4.28	0	-0.0471097	0.0377778	-1.25	0.225
18	-0.2699747	0.0786009	-3.43	0.002	-0.0198923	0.0393109	-0.51	0.618
19	-0.1955001	0.062721	-3.12	0.005	-0.0832989	0.0258271	-3.23	0.004
20	-0.2154891	0.0848024	-2.54	0.018	-0.060316	0.0324907	-1.86	0.076

Table 3: Empirical relation between size and growth rate variance: growth rates as weighted size differences.

	Manufacturing				Services			
Threshold	**Coefficient**	**Standard error**	**t-ratio**	**p-value**	**Coefficient**	**Standard error**	**t-ratio**	**p-value**
0	-0.0003496	0.0007301	-0.48	0.637	0.0037164	0.0006175	6.02	0
1	-2.79E-08	0.0007055	0	1	0.0028246	0.000593	4.76	0
2	-0.0008141	0.0002126	-3.83	0.001	0.0007106	0.0003575	1.99	0.059
3	-0.0008908	0.0001913	-4.66	0	-0.0001123	0.0001708	-0.66	0.517
4	-0.0006529	0.0001598	-4.08	0	-0.0002659	0.0001339	-1.99	0.059
5	-0.0006385	0.0000992	-6.43	0	-0.0003142	0.0000508	-6.18	0
6	-0.0007816	0.000163	-4.79	0	-0.0002884	0.0000569	-5.07	0
7	-0.0007569	0.0002326	-3.25	0.003	-0.0003079	0.0001076	-2.86	0.009
8	-0.0006967	0.0002711	-2.57	0.017	-0.0003331	0.0000959	-3.47	0.002
9	-0.0006581	0.0002302	-2.86	0.009	-0.0002723	0.0000981	-2.77	0.011
10	-0.0005728	0.0001412	-4.06	0	-0.0002288	0.0000655	-3.49	0.002
11	-0.0006626	0.0002395	-2.77	0.011	-0.0002284	0.0000898	-2.54	0.018
12	-0.0006369	0.0002146	-2.97	0.007	-0.0001954	0.0000592	-3.3	0.003
13	-0.0005606	0.0001754	-3.2	0.004	-0.0002095	0.0000867	-2.42	0.024
14	-0.0005767	0.0002266	-2.55	0.018	-0.0001454	0.0000537	-2.7	0.013
15	-0.0005211	0.0001483	-3.51	0.002	-0.000152	0.0000508	-2.99	0.007
16	-0.0004817	0.0001501	-3.21	0.004	-0.0001051	0.0000264	-3.99	0.001
17	-0.0011216	0.0003538	-3.17	0.004	-0.0001023	0.0000261	-3.92	0.001
18	-0.00038	0.0001994	-1.91	0.069	-0.0000873	0.0000263	-3.32	0.003
19	-0.0011174	0.0002984	-3.75	0.001	-0.0000877	0.0000188	-4.66	0
20	-0.0003689	0.000111	-3.32	0.003	-0.0000597	0.0000301	-1.98	0.059

Figure 1 Relation between truncation point and observed variance. (our elaboration of the original Figure in Barr and Sherrill, 1999)

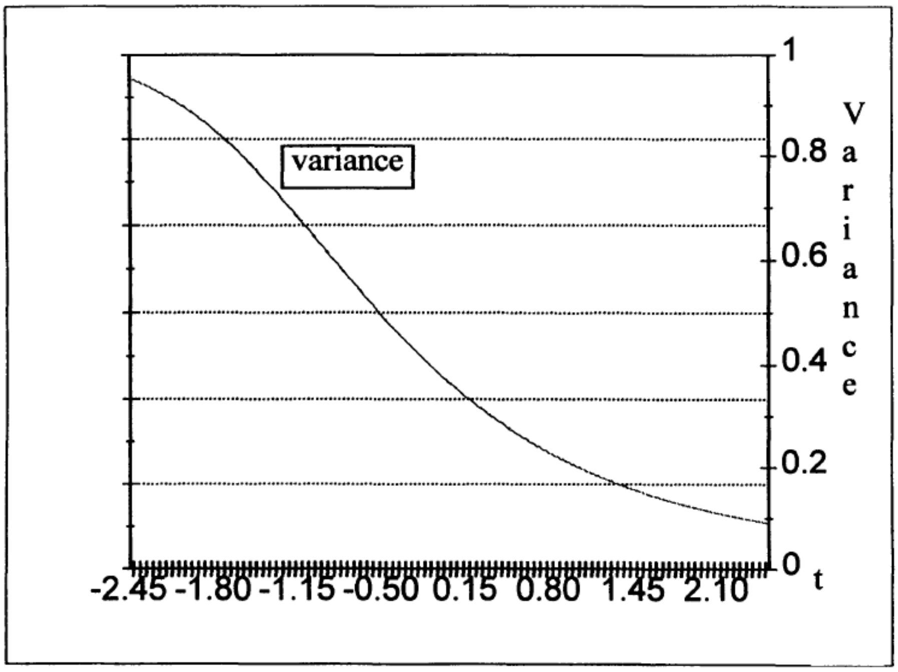

Figure 2: Difference between real (β_A) and observed (β_B, β_C) slope

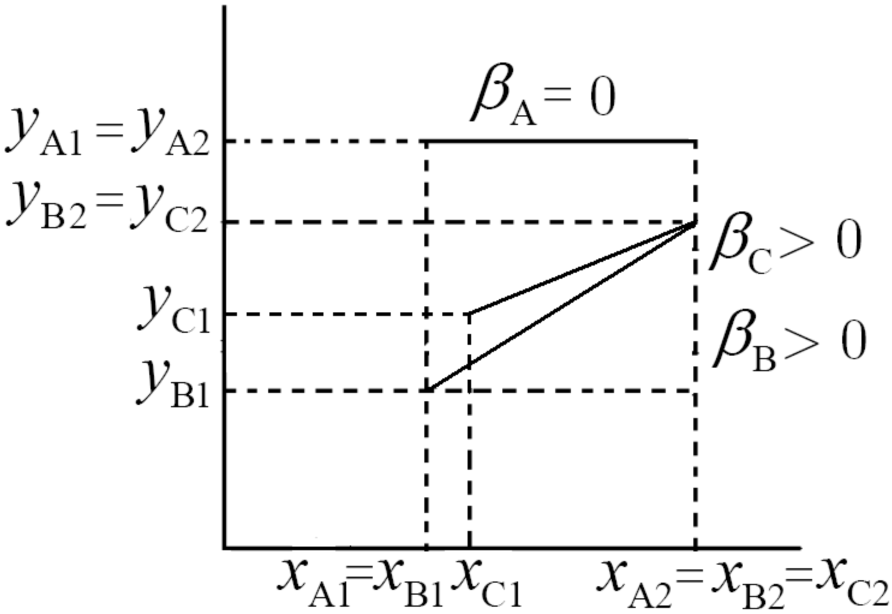

Figure 3: Growth rates as log size differences. Relation between estimated β (vertical axis) and artificial threshold M (horizontal axis)

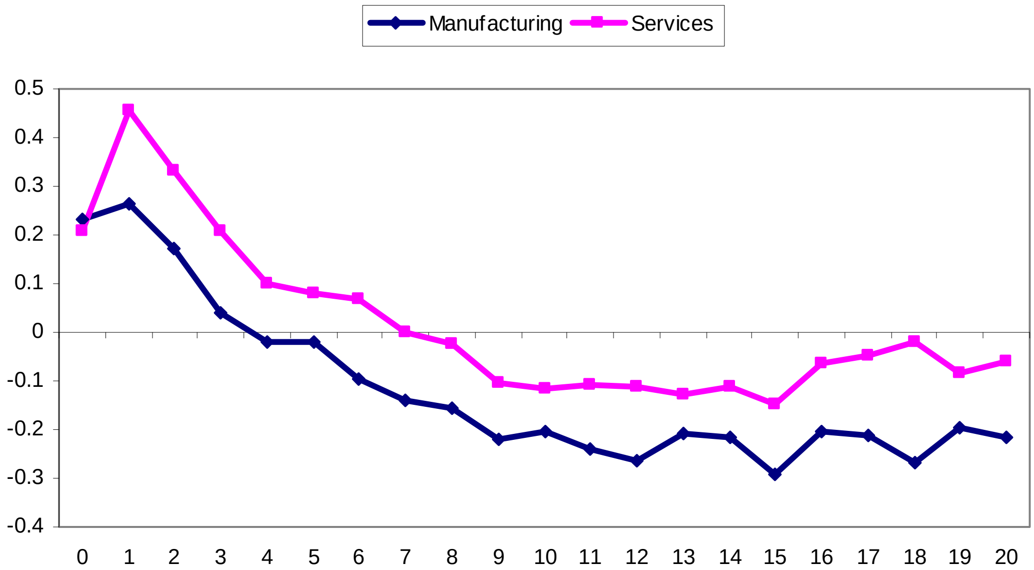

Figure 4: Growth rates as weighted size differences. Relation between estimated β (vertical axis) and artificial threshold M (horizontal axis)