A simple model of discontinuous firm's growth

D'Elia, Enrico

Ministry of Economy and Finance – Department of Treasury - Rome (Italy), ISTAT - Rome (Italy)

December 2011

A simple model of discontinuous firm's growth

Enrico D'Elia

Ministry of Economy and Finance – Department of Treasury

and

National Institute of Statistics

Address for correspondence: Enrico D'Elia

via Casalmonferrato 2/C

00181 Roma (Italia)

email: delia.enrico@gmail.com

Short abstract

A very essential model explaining the discontinuous growth of firms can be based on the following assumptions:

- in the short run, the firm's profit reaches its peak only for a given production level;
- in the long run, the firm adjusts its size *as if* the current equipment had to be exploited until profit exceeds the profit expected from the new desired plant.

The former two hypotheses imply that:

- maximum investment and disinvestment are bounded;
- the profitability by size exhibits a number of peaks;
- firms tend to invest when profit approaches a local minimum;
- the distribution of firms by growth rate is multimodal.

Keywords: Capacity utilization, Discontinuity, Firm's size, Growth, Lumpy investment.

JEL classification: D21, D92, L11

Abstract

Typically, firms change their size through a row of discrete leaps over time. Sunk costs, regulatory, financial and organizational constraints, talent distribution and other factors may explain this fact. However, firms tend to grow or fall discontinuously even if those inertial factors were removed. For instance, a very essential model of discontinuous growth can be based on a couple of assumptions concerning only technology and entrepreneurs' strategy, that is:

(a) in the short run, the firm's equipment and organization provide the maximum profit only for a given production level, and diverging form it is costly; and

(b) in the long run, the firm adjusts its size *as if* the current equipment had to be exploited until overall profit exceeds the profit expected from the new desired plant at the current production level.

Combining the latter two hypotheses entails a number of testable consequences, usually regarded as nuisance facts within the traditional theoretical framework. First of all, an upper bound constraints both investment and disinvestment. Secondly, the profitability is not a continuous function of the firms' size, but exhibits a number of peaks, each corresponding to a locally optimal size. Thirdly, firms tend to invest when profit approaches a local minimum, corresponding to the lowest profit claimed by the entrepreneur. Therefore, firm's level data would prove only weak statistical relationships among profitability, output and investment. Finally, the distribution of firms by growth rate is multimodal since, within each sector, every firm typically adjusts its size through the same sequence of leaps.

There are a number of analogies between the firm's growth process predicted by the model and some physical phenomena explained by the quantum theory.

1 Introduction (*)

Firms are not obliged to grow, but in case, grow through a row of discrete leaps over time, rather than changing their size continuously, adding small marginal units to their current equipments and staff.[1] These discontinuities have been explained assuming that a number of constraints and incentives (mainly technical, organizational, and regulatory factors) affect the entrepreneur's strategy. Also the discrete leaps in the firm's size have been justified sometimes by the hysteresis of the firm's organization, which entails large sunk costs for every change, so that size adjustments are profitable only if their expected advantage exceeds the related fixed cost of changing.[2] Further, size changes are often related to the economic background[3] of the firm and to different phases of the firm's evolution, which are strictly path dependent and generally require radical structural transformations, which entail size changes only as a side consequence.[4] Finally, the uncertainty about future prices and demand may

---

(*) The views expressed in this paper are those of the author and do not necessarily reflect views at ISTAT and MEF. The author gratefully acknowledges the valuable suggestions and criticisms by some early readers of the paper. Of course, errors or omissions are the responsibility of the author.

[1] See Caballero et al. (1997) for some evidence on the irregular dynamics of employment changes; Doms and Dunne (1998), Cooper, Haltiwanger and Power (1999) and Nilsen and Schiantarelli (2003) about lumpy investment; Santarelli and Vivarelli (2007) about the erraticness of small businesses growth.

[2] See Hammermesh and Pfann (1996) and Caballero and Engel (1999) about the discontinuity of factors demand and investment respectively induced by nonconvex adjustment costs.

[3] In particular, it matters if a firm is integrated in a district or not. See Pyke, Becattini and Sengenberger (1990) for a survey.

[4] See Traù (2000, 2003) on the role of firm specific factors and the interplay with macroeconomic changes.

induce firms to delay investment, concentrating them in few special periods, waiting for better information.[5]

Actual success or failure of each firm is undoubtedly related mainly to individual stories, however, this paper aims at showing that firms tend to grow or fall discontinuously even though they did not differ for opportunities, endowments, managers' talent, etc., and without making recourse to threshold models[6] and general equilibrium models.[7] Instead, here merely a couple of simple hypotheses are made:

(a) in the short run, the firm's equipment and organization provide the maximum profit only for a given production level, and diverging form it is feasible but costly; and
(b) in the long run, the firm adjusts its size *as if* the current equipment had to be exploited until overall profit exceeds profit expected from the new desired plant at the current production level.

The assumption (a) serves mainly to exclude some trivial source of discontinuity in the firm's behaviour, such as the impossibility to change the current output level without changing firm's size. Adjustment costs may be related to exogenous factors, such as technical, organizational and regulatory constraints, but even the simple cost of either overheating or under using the current equipments is sufficient for the short run profit function exhibits a single maximum. On its turn, the long run adjustment rule (b) sets some limits to the ability of firm to keep both its size unchanged and its profits above some minimum threshold in the long run. In particular, the assumption (b) implies that changing firm's size (in either direction) becomes almost unavoidable when output exceeds or falls below some given

[5] See the classical model by Dixit and Pindyck (1994) and the paper by Bulan, Mayer and Somerville (2006) about the optimal timing for investment projects.

[6] See Caballero e Engel (1999).

[7] See Khan e Thomas (2003).

thresholds permanently. Notably, this fact does not introduce a discontinuity in the firm's evolution in itself, but simply establishes a relationship between the profitable frontiers available to the firm in each moment. The leaps in the size adjustments raise only assuming, further, that the entrepreneur is unwilling to make less than that specific amount of profits related to the current output level.

The assumptions (a) and (b) and the combination thereof provide a number of consequences suitable to be tested empirically. First of all, the conjecture (a) implies that a one-to-one relationship between output and profit holds even in the short run only when the equipment is exploited in the most profitable conditions. Otherwise, at least a couple of different output levels, for each given firm's size, may provide the entrepreneur with the same overall profits. This fact weakens, or even breaks, the statistical relationship between production and performance at firm's level.

If the profit function and the minimum required profit are almost the same within each industry, the conjecture (b) suggests that investment and downsizing occur mainly when the output reaches some given thresholds, which are typical for each sector and market segment. Also the size changes desired when the output crosses those thresholds are likely similar among firms belonging to the same group. Thus the frequency distribution of firms by growth rate should show several concentration points, rather than the continuous smooth distribution predicted by the Gibrat's law and its variants,[8] and the power law frequency distributions

[8] See the seminal paper by Gibrat (1931), the comprehensive survey by Sutton (1997) and the critical views by Geroski (1995). Multimodal frequency distributions of firms by size were found, among the others, by Bottazzi and Secchi (2003) in the US industry; by Cabral and Mata (2003) in the Portuguese manufacturing industry; by Lotti, Santarelli and Vivarelli (2003) and Bottazzi, *et al.* (2007) for the Italian firms.

recently popularised by the econophysic literature.[9]

Furthermore, if the firm moves from a short run profit function conjectured by (a) to the other, the assumption (b) implies that the firm's profitability exhibits a number of peaks, each corresponding to different optimal production levels associated to the related size. Thus the profitability should not be a continuous function of the firms' size, as expected by taking into account only economies and diseconomies of scale. In fact, many studies acknowledge how it is difficult to explain the prevalence and better performance of medium size firms without assuming some exogenous factor. This fact determines, in turn, a fat tailed distribution of firm's growth rate.[10] In addition, if firms invest only when profit falls below a given threshold, a negative statistical relationship between profitability and investment should be expected as well. It is worth noticing that this conjecture does not reflects necessarily on the dynamics of macroeconomic aggregates.[11]

There are a number of impressive analogies between the discontinuous firm's growth process, assumed here, and some predictions of the quantum theory in physics, developed about a century ago.[12] Particularly, the concentration of firms' growth rates around some typical values, which are specific for each industry, resembles the fact that the frequency distribution of the light emitted by a pure gas overheated (i.e.: its spectrum) is discontinuous,

[9] See Sinha *et al.* (2010) for a recent survey of this branch of economics.

[10] See Clifford and Cavanagh (1985) and Simon (1996) among the others.

[11] Indeed, the topic of the macroeconomic consequences of lumpy investment is still under debate. By analyzing a general equilibrium model, Khan and Thomas (2008), among the others, showed that investment discontinuity has no special consequences on the dynamics of macroeconomic aggregates, but only on the heterogeneity and volatility of firm's behavior. Within a very similar theoretical framework, Bachmann and Bayer (2011) come to opposite conclusions, stressing that lumpy investment generate procyclical fluctuations of the corresponding aggregate indicator.

[12] See Nye (2002).

contrarily to other materials, and is characterised by the concentration of radiations around some specific frequencies, forming a number of "rows" in the spectrum. Furthermore, the uneven relationship between size, profitability and investment resembles the discontinuous behaviour of photoelectric materials, which is related exactly to their particular atomic structure. The atomic model explaining the rows in the spectrum of pure gases, due to Niels Bohr, and the explanation of the photoelectric effect by Albert Einstein gave rise to the quantum theory, in which the radiating energy is viewed as composed of discrete "granules" strictly resembling the atomic structure of the matter. It is possible that abandoning the assumption that firms are able or willing to grow continuously over time, by infinitesimal size adjustments, may prove a fruitful research programme in economics as well. Of course, this approach cannot take the place of analysing firm specific factors of development and decline.

The paper is organised as follows. The entailments of a general convex profit function in the short run are discussed in the next section. In particular, the range of profitable production is derived, and the complex relationship between output level and firm's performance is considered. The third section introduces a model that mimics the investment process based on the assumption that the entrepreneur wishes to change the size of his/her firm only when the current profit flow falls shorter than a given threshold. The corresponding growth pace of firms is also derived under special ideal conditions. Section 4 is devoted to the discussion of some entailments of the model for the firm's evolution over time and the structure of industries. In particular, the consequences for the shape of the distribution of firms by size, growth rate, profitability and degree of capacity utilization are analysed. The main conclusion is that the model predicts that those distributions are generally multimodal, contrarily to the deductions of most traditional models. Some conclusive remarks close the paper including, in particular, a number of conjectures suitable to be tested empirically and some possible extensions of the model.

2 The firm's profits in the short run

Let consider a representative firm that operates in a market where prices and demand are determined exogenously, and faces a short run profit function $\pi(q, E)$, where q is the output level and E is a vector of fixed production factors, with the following properties:

(a) is continuous and twice differentiable respect to the output level q and the inputs E;

(b) is a strictly concave function of q and achieves its unique maximum at $q_{max,E}$ for any given E;

(c) is not scale decreasing, i.e.: $\pi(q_{max,E}, E) \geq \pi(q_{max,F}, F)$ if $q_{max,E} \geq q_{max,F}$.

Every assumption on $\pi(.)$ is quite standard [13] and mainly contributes to make the problem analytically tractable. In particular, the assumption (a) simply ensures that the available technology allows the firm to change its output level continuously, within some admissible range, in order to exclude any possible discontinuity in the firm's growth process explained by a break in the range of production possibilities.

The features of profit function pointed out in (b) straightforwardly derive from the assumption that, in the short run, profits are constrained by the endowment of the fixed productive factors, and the marginal productivity of other inputs is not increasing. Thus, given the prices of output and intermediate inputs, it reads $\frac{\partial^2 \pi(.)}{\partial q^2} < 0$.

Finally, the assumption (c) argues that, exploiting plants at the most efficient conditions, profit does not decrease if the output increases. First of all, it serves to exclude several ordinary sources of discontinuities in the firm's behaviour, related to the indivisibility of equipment and large scale diseconomy. Furthermore (c) ensures that the size of firm can be measured univocally by the corresponding output level produced at the most efficient

[13] See Lau (1976) for a survey on profit functions.

combination of productive factors, so that a generic profit function fulfilling (c) can be written as $\pi(q, q_{max})$, without any loss of generality.

It is worth noticing that the function $\pi(.)$, even under the conditions (a) – (c), is quite general, since is capable to take into account many relevant factors, such as: capital amortization, financial charges and the "normal" capital remuneration; price elasticity to q and q_{max}; tax and incentives related to overall profit, q and q_{max}; possible scale economies and diseconomies, provided that the condition (c) holds; the discounted value of expected future gains and losses, possibly related to the dynamics of demand and prices.

The assumptions (a) – (c) suffice to draw a number conjectures on the firm's behaviour suitable of empirical tests. First of all, for a given endowment of fixed capital, the firm may make the same overall profits for different output levels, with the obvious exception of profits earned at q_{max}. Furthermore, the strict concavity of the profit function, postulated in (b), ensures that the entrepreneur whose main goal is attaining a given profit level, is indifferent between two (and only two) distinct production levels, with the obvious exception of the single value q_{max}. Formally, assumption (a) and (b) entail that a couple of output levels $q_a \neq q_b \neq q_{max}$ always exists such that $\pi(q_a, q_{max}) = \pi(q_b, q_{max})$. This fact introduces a source of unpredictability in the entrepreneur's behaviour and makes firm's production open to vary even substantially and suddenly without affecting profitability and investment, particularly depending on the expected dynamics of demand. Also it implies that the statistical relationships between output, on the one side, and profits and investment at firm's level, on the other, is expectedly weak and surely non linear. For the same reason, no one-to-one relationship can be generally established between profit and the capacity utilization, defined as the ratio of ongoing output q to the maximum output level corresponding to some minimum acceptable profit granted by the actual equipment. In other words, the same overall profit may be associated to (even very) different rates of capacity utilization and, what is more, the latter rate is not related univocally to profit at firm's level. However, this indeterminacy is compatible with a strong statistical association among profit and capacity utilization at sectoral or macroeconomic level, if the firms which produce more are also those whose profits increase as well.

The properties (a) and (b) imply that the following second order approximations of any profit function holds, in the neighbourhoods of q_{max}

$$\pi(q, q_{max}) \approx \pi(q_{max}, q_{max}) + \frac{1}{2}\pi'' \; (q - q_{max})^2 \qquad [1]$$

where $\pi'' = \left.\dfrac{\partial^2 \pi(q, q_{max})}{\partial q^2}\right|_{q=q_{max}} \leq 0$ owing to the assumption (b). According to [1], the firm makes profit higher than π^* only within the range of output

$$q_{max} \pm \sqrt{\frac{2(\pi^* - \pi(q_{max}, q_{max}))}{\pi''}} \qquad [2]$$

It follows that the range of profitable productions is as wider as π^* and $|\pi''|$ are smaller, that is if the minimum profit desired by the entrepreneur is lower and organization and technology are more flexible. In addition, the relationships between output, capacity utilization and profit are expectedly as weaker and unpredictable at firm's level, as the range [2] is larger. The range of profitable production may be quite large, so that the firm has a little incentive to change its current size even if it faces large demand fluctuations, in particular if the cost of the size adjustment does not compensate the risk that the demand bounces back to its previous level.

3 The long run adjustment of productive capacity

In the long run, the firm is assumedly able to choose its optimal size, taking into account also factors such as demand expectations, possible institutional constraints, the talents and attitude of managers, strategic considerations and the specific story of each firm.[14] Also financial constraints may influence the firm's decisions, but here the capital market is assumedly perfect, and the entrepreneurs and other stakeholders are indifferent to the source of investment funding,[15] to the aim of focusing on the desired size of firm when other restrictions and incentives are negligible. For the same reason, uncertainty about future demand and prices plays no role in what follows.

The condition [2] enables the entrepreneur to change output substantially without cutting profit below a given threshold, only by adjusting firm's size, namely investing or downsizing equipments. In particular, if the entrepreneur expects that the demand will be D during the next years, he may decide either to invest in new equipments that are more profitable for the output level D, or to be contented with the profit $\pi(D, q_t)$ granted by the actual plants. Let assume, further, that the entrepreneur makes his investment decisions as if was unable or unwilling to earn less than a minimum acceptable profit, say the "reservation" profit π^*, even during the size adjustment process. It implies that he allegedly wishes to change the firm's size only when profits fall below π^*. A rationale for this rule is provided by the managerialist theory of firm's growth,[16] according to which the manager (not necessarily being the entrepreneur) tends to expand the businesses under the constraint of ensuring a minimum remuneration to the shareholders.

[14] See Coad (2009) for a recent survey of the related literature.

[15] As postulated by Modigliani and Miller (1958).

[16] See Jensen and Meckling (1976).

Assuming the decision rule above, profit would not change continuously over time, but moves along the current profit function until the actual equipment grant at least π^* and then follow the profit function associated to the new equipment. Consequently, the optimal productive capacity of the firm tends to jump suddenly, say from q_t to q_{t+1}, while the actual output may change gradually. However, $\dot{q}_{t+1} = \frac{q_{t+1} - q_t}{q_t}$ can be regarded only as the relative pace between two successive steps of the firm's growth process, rather than a growth rate over time, since the subscripts t and $t+1$ here do not refer to the ordinary time scale, but simply to a couple of subsequent phases during the existence of a firm, likely of very different extent. It is worth noticing that the assumed decision process does not pretend to describe accurately the actual behaviour of the firm, but simply the final observable outcome of investment (or disinvestment) process.[17] In fact, the aim of this model is only to provide an explanation for lumpy size changes within an extremely simplified theoretical framework, without considering other factors likely influencing actual investment decisions on purpose.

The latter conjecture on the entrepreneur's strategy has some challenging and testable consequences about the expected size of investment or disinvestment. In fact, when the entrepreneur decides to invest, cannot expand the firm's size indefinitely, but has to choose equipments that grant at least the same profits of the installed plants for the ongoing output level, say q^*. Thus, the largest admissible size change is such that both the current and the new desired profit functions cross just when the output is q^*, that is

$$\pi(q^*, q_t) = \pi(q^*, q_{t+1}) - k = \pi^* \qquad [3]$$

where k is the sunk cost undergone by the firm every time its size changes.

Thus, in principle, q_{t+1} can be determined as a function of q_t, k and q^*, or the

[17] This approach aspires to follow the methodology of "positive economics" introduced by Friedman (1953).

corresponding threshold π*. In principle, the limiting size q_{t+1} implied by [3] would be the most convenient one as well, since any smaller size adjustment charges the firm with the same additional irrecoverable cost k anyway.

The approximation [1] of the profit functions in the neighbourhoods of q_t and q_{t+1} respectively gives

$$\pi^* \approx \pi(q_t, q_t) + \frac{1}{2}\pi_t^{''} (q^* - q_t)^2 \qquad [4]$$

and

$$\pi^* \approx \pi(q_{t+1}, q_{t+1}) + \frac{1}{2}\pi_{t+1}^{''} (q^* - q_{t+1})^2 \qquad [5]$$

Since during a growth process $\pi(q_{t+1}, q_{t+1}) \geq \pi(q_t, q_t)$ owing to the assumption (c), [3], [4] and [5] imply that

$$(q^* - q_{t+1})^2 - k \geq \frac{\pi_t^{''}}{\pi_{t+1}^{''}} (q^* - q_t)^2 \qquad [6]$$

which sets a floor to the largest profitable investment. The inequality [6] simply states that the entrepreneur is always able to achieve the same overall profit either by continuing to exploit the current plants devised to produce more efficiently q_t or by adjusting the firm's size to q_{t+1} and incurring in the cost k.

The condition [6] implies that

$$q_{t+1} - q_t \leq \sqrt{\frac{\pi_t^{''}(q_t - q^*)^2 + 2k}{\pi_{t+1}^{''}}} \leq \sqrt{\frac{\pi_t^{''}}{\pi_{t+1}^{''}}} (q^* - q_t) \qquad [7]$$

were the rightmost limit is identical to the first one if sunk costs are negligible. The equation [7] also gives

$$\dot{q}_{t+1} \leq \left(1+\sqrt{\frac{\pi_t^{''}}{\pi_{t+1}^{''}}}\right)\frac{q^*-q_t}{q_t} \qquad [8]$$

The Graph 1 provides an example of the upper limit to investment entailed by the decision process assumed here. Given the reservation profit π^* and the current endowment of fixed capital, the firm operates along the profit curve $\pi(q, q_t)$, symbolized by a bold line, until the output reaches q^* and expanding optimal firm's capacity to q_{t+1} becomes profitable. However, the concavity of the new profit curve $\pi(q, q_{t+1})$, represented by a thin line, ensures that the difference between q_t and q_{t+1} cannot be too large, as stated by [6].

[insert Graph 1 about here]

During a downsizing process, the assumption (c), and the approximations [4] and [5] give

$$(q^* - q_{t+1})^2 - k \geq \frac{\pi_t^{''}}{\pi_{t+1}^{''}} (q^* - q_t)^2 \qquad [9]$$

so that the limiting size cut is

$$\dot{q}_{t+1} \geq \left(1+\sqrt{\frac{\pi_t^{''}}{\pi_{t+1}^{''}}}\right)\frac{q^*-q_t}{q_t} \qquad [10]$$

Therefore, regardless to the direction of the planned size change, it reads

$$\left|\dot{q}_{t+1}\right| \leq \left(1+\sqrt{\frac{\pi_t^{''}}{\pi_{t+1}^{''}}}\right)\left|\frac{q^*-q_t}{q_t}\right| \qquad [11]$$

namely the largest size adjustment is proportional to the absolute value of the gap between q^* and the current productive capacity.

Since the right hand side of [11] is undetermined for $q_t = 0$, the model is virtually incapable to explain how firms born. Furthermore, the inequality [11] sets only a ceiling and a floor to the largest size changes consistent with profit never falling below π^*.[18] Those limits are identical to the actual size changes if the entrepreneur is forward looking and rationally prefers to save on sunk costs incurred for each investment or disinvestment decision. In fact, if market demand is expected to grow or fall steadily over time, there is only a minor incentive to adjust gradually the productive capacity in order to maximize profit only for the time being, incurring in the sunk cost k twice or more to attain finally the same productive capacity. Formally, lower sunk costs in [6] and [9] can take into account the possible advantages to invest or disinvest less than the limit [11], waiting for better information and technologies.

The condition [11] has a number of challenging implications. First of all, $\dot{q}_{t+1} = 0$ is always an admissible solution of [11], that is the entrepreneur is not obliged to change the firm's size, since he is allowed to content with the current profit limiting the production level to q^* deliberately even in face of larger market demand. Secondly, the largest relative size adjustment is proportional to the percentage difference between the threshold q^* and the current optimal output level. It confirms the intuition that the limiting investment and disinvestment is strictly related to the initial firm's size, that is a small business hardly may embark in an ambitious growth project while a large enterprise is much more flexible (and unpredictable) in adjusting its size. Furthermore, if the entrepreneur does not tolerate even a

[18] It is straightforward to demonstrate that the same limits hold if the entrepreneur incurs in costs proportional to the capacity change instead of the fixed sunk cost k, for each investment or disinvestment. For instance, during a growth process, the cost $k'(q_{t+1} - q_t)$ of changing size implies

$$q_{t+1} - q_t \le \sqrt{\frac{\pi_t^{"}(q_t - q^*)^2 - 2k'(q_t - q^*)}{\pi_{t+1}^{"}} - \left(\frac{k'}{\pi_{t+1}^{"}}\right)^2} + \frac{k'}{\pi_{t+1}^{"}} \le \sqrt{\frac{\pi_t^{"}}{\pi_{t+1}^{"}}}\ (q^* - q_t)$$

temporary profit reduction compared to the largest one attained at q_t, the condition [11] argues that no size adjustments is possible. In particular, in case of declining demand, such an ever-profit-maximizing entrepreneur could only continue exploiting existing plants at the most profitable rate or stop the production process. Although this is mainly an extreme consequence of having approximated the profit function in the neighbourhoods of q_t, it provides a rationale for the case of "hit and run" firms[19] entering the market only when profit is very high and exiting as soon as profit starts decreasing.

It is worth noticing that [11] holds also if the threshold q_t^* is associated to a reservation profit π_t^* changing over time. In particular, reformulating the right hand side of [8] as a function of the optimal degree of capacity utilization of plants compared to the largest profitable output, it reads

$$\dot{q}_{t+1} \leq \left(1+\sqrt{\frac{\pi_t''}{\pi_{t+1}''}}\right)\frac{1-r_t^*}{r_t^*} \qquad [12]$$

where $r_t^* = \frac{q_t}{q_t^*}$, that is less than 1 by definition. Even assuming that r_t^* is unreasonably low, [12] predicts that each firm may face substantial limits to its growth if the market demand is expected to increase very fast.

The absolute value of $\dot{q}_{t+1}$ rises with the root square of $\frac{\pi_t''}{\pi_{t+1}''}$, that is the ratio between the marginal cost of adjusting output in the short run before and after the desired size adjustment. Thus a firm that expects to gain flexibility after a size adjustment tends to make smaller adjustments over time. Combining the properties of [2] and [11], it follows that the firms eligible for larger relative size adjustments are those capable to change their output

[19] Popularized by Baumol, Panzar and Willig (1982).

much more exploiting their current plants, since the wider is the range $|q_t^* - q_t|$, the larger is $|\dot{q}_{t+1}|$. Since small and medium businesses are expectedly more flexible in the short run, it follows that investment is probably smaller in the industries where the average size of firms is smaller.

In any case, the maximum growth or downsizing pace of a firm could be much smaller than the expected demand change. In fact, adjusting the firm's size takes time, but the limiting step $\dot{q}_t$ given by [11] depends on the minimum profit tolerated by the entrepreneur, and nothing guarantees that π_t^* is small enough to determine the required size adjustment in due time. As a consequence, the model predicts that most firms have only a limited capacity to track the long run macroeconomic trends, even when trends are fully expected, especially when the demand falls or grows very fast. In particular, the constraints on $\dot{q}_t$ are more severe if the entrepreneur in not prepared (or in the condition) to incur in a profit reduction or loss during the transition process from the current size to the desired one. Therefore, the model argues that sometimes a firm could fall in a "size trap", since it faces a demand so dynamic that is unable to adjust its size without incurring in unendurable loss for some periods of time. This fact likely prevents many existing businesses from developing enough even within fast growing markets, giving space for other firms entering the market.

Looking only at the existing firms, the relation [11] and the possible "size trap" suggest that during a buoyant growth phase or a deep recession only modest changes in the size could be expected, while the correlation between demand and investment strengthens if the dynamics of demand is moderate enough. On the other hand, largest macroeconomic fluctuations should influence the number of new-born and dead firms much more than investment of existing firms. Consequently, according to the model, the statistical relationship between the dynamics of demand and the demography of firms may change abruptly with the growth rate of market demand. Under a given threshold, investment of existing firms are sufficient for tracking the market evolution, with minor effects on the number of new and dead firms; beyond this threshold, the demography of firms is dominant. Furthermore, the correlation between the dynamics of demand and the average size of firms is expectedly negative when the changes are very fast, and positive otherwise, since the new entries are expectedly smaller than the average.

4 The performance of firm over time and the distribution of firms by size

The very simple model sketched in the previous section introduces some "quantum" and "indeterminacy" elements in the growth of firms over time and predicts a number of facts suitable to be tested empirically. First of all, the inequality [11] is unable to predict the actual dynamics the firm's size, even facing a fully expected demand dynamics, unless saving on sunk costs is crucial for the entrepreneur. In fact, the decisions about the desired size adjustment may come early before profit has fallen below the threshold π_t^*. It is worth noticing that the indeterminacy entailed by the model is fully consistent with the assumption behind the Gibrat's law and its variants, envisaging that the representative firm grows at a rate that is independent from firm's size.

4.1 Investment, profit and capacity utilization

The model relates the limiting dynamics of $\dot{q}_t$ to the reservation profit π_t^* claimed by the entrepreneur and the parameters of the profit function [1]. If those parameters were observable and stable over time, and the size changes only when profit is exactly π_t^*, the evolution of the firm's size would be fully predictable on the basis of the expected dynamics of the market demand D_s, where the subscript s refers to the ordinary time scale, unlike the subscripts of $\dot{q}_t$, q_t^* and π_t^*.

Nevertheless, the firm's size develops regularly over time only under very special circumstances. For instance, if the minimum profit is always a percentage of the maximum one, and the ratio $\sqrt{\frac{\pi_t''}{\pi_{t+1}''}}$ does not vary with the size and over time, [8] implies that the firm grows through a sequence of identical relative steps $\dot{q}$. If, in addition, D_s grows at the constant rate $\dot{g}$, the firm's size changes when

$$D_0(1+\dot{g})^s = q_0(1+\dot{q})^t \qquad [13]$$

where t is an integer counting the number of size adjustments made by the firms until the time s. Thus investment or disinvestment would occur whenever

$$s \approx t\,\frac{\dot{q}}{\dot{g}} + \frac{1}{\dot{g}}\ln\frac{q_0}{D_0} \qquad [14]$$

where the second term in the right hand side of [14] vanishes if the initial size of firm matches the initial market demand exactly. Since t is an integer number, [14] foresees that the size changes happen only at some special points in time, that is when $s - \frac{1}{\dot{g}}\ln\frac{q_0}{D_0}$ is an integer multiple of the ratio $\frac{\dot{q}}{\dot{g}}$. In the remaining periods investment and disinvestment are likely null. Conformably, between two size changes, the degree of capacity utilization and profit vary from a local minimum, when $q_s = q_t^*$, to a peak, when $q_s = q_{t+1}$, to another local minimum when $q_s = q_{t+1}^*$.

If the firm faces a business cycle, the dynamics of the size adjustments is much more complicated and depends mainly on the amplitude of cycles. If D_s is predicted to vary only between q_t^* and q_t the entrepreneur could even leave the firm's size unchanged, accepting the corresponding profit fluctuations and saving on sunk costs of investing or disinvesting. Otherwise, a sequence of investment and downsizing should be envisaged to keep profit above the critical threshold. Continuing to assume that $\dot{q}$ is constant, only for sake of simplicity, the size adjustments would repeat over time with a pseudo-periodic cycle only in the unlikely event that the amplitude of demand fluctuations is an exact multiple of $q_t - q_{t-1}$. If this is not the case, an irregular sequence of size adjustments should be expected. In particular, when the market demand is accelerating, the constant adjustment $\dot{q}$ takes shorter time to be completed, thus the degree of capacity utilization and profit tend to exhibit cycles of decreasing length. The opposite happens during a decelerating phase. The timing and direction of size adjustments are particularly unpredictable close to the turning points of the business cycle, since then the sign of the adjustment crucially depends on whether the latest size change has set the current productive capacity already below or beyond the peak or low of demand. Also the timing of the size change is determined by how fast the cyclical phase changes.

Since the local cycles of capacity utilization and profit are likely shorter than the phases of the business cycle, the model predicts that the statistical correlations between D_s, r_s and π_s are time varying, even if D_s exhibits a purely deterministic dynamics. In particular profit may decrease when D_s and r_s increase and the other way round. In addition, nearby the turning points, the firm could even invest just before a downturn or downsize about an upturn. What is more, the relationship between profit and business cycle may be asymmetric, namely the same phase may generate a different profit dynamics, depending on the firm's size at the beginning of the phase. These facts contribute to make the short run predictions on profit and investment quite difficult at firm's level.

For the same reasons, the model predicts that investment can be weakly related to output, profit and capacity utilization at firm's level. More precisely, size changes occur when profit reaches local minima, regardless if the trend of profit is ascending or falling, and when the capacity utilization jumps abruptly to a maximum, related to the previous equipment, to a minimum, given the new size. As a consequence, the statistical relationships among investment, D_s, r_s, and π_s are likely to be all but linear at firm's level, even though they could keep on quite strong among the corresponding macroeconomic and sectorial aggregates.

4.2 The frequency distribution of firms

The discontinuous nature of firm's growth has some consequences also on the frequency distribution of size changes in each industry. Let suppose, only for sake of simplicity, that within each industry

(i) the profit function is the same and is time invariant;

(ii) the entrepreneurs content themselves with the same minimum profit for any given output level;

(iii) every firm faces the same dynamics of market demand.

Under the previous assumptions, according to [11], if the entrepreneur wishes to avoid unnecessary sunk costs, every firm tends to grow through the same sequence of size adjustments. In this framework, the only difference among the firms is the initial size and the initial position along the profit function. It follows that, in each point in time, firms may keep on their initial size or jump to the next or the previous step of the same sequence of jumps. In other words, the number of firms which invest or downsize may change over time, but not the admissible sequence of $\dot{q}_t$ which is the same within each industry and depends on the assumed decision rules. In addition, the sequence $\dot{q}_t$ should be stable over time, regardless to the trend of demand and the phase of the business cycle, since [11] depends only on π_t^* and not on D_s.

As a consequence, the model predicts that, in each industry, the distribution of firms by growth rate shows a number of peaks. Nevertheless the distribution of firms by size could be continuous as well, as assumed by the Gibrat's law and its variants, since the current firms' size is mainly conditioned by the initial distribution of size and capacity utilization of firms, which is likely continuous in turn. The distribution of firms by size would be discontinuous under very special conditions. For instance, this would be the case if each firm is assumed to have born with a different efficient size in some point of time in the past.

According to the model, the frequency associated to each peak of the growth rate distribution does not depend on the magnitude of $\dot{q}_t$, but mainly on the initial distribution of firms, which determines in turn the position of each firm along the sequence of $\dot{q}_t$. This conjecture contrasts with the Gibrat's hypothesis that $\dot{q}_t$ consists approximately of the sum of many independent relative size changes, so that its frequency distribution decays with $|\dot{q}_t|$. At variance, the model discussed here argues that large $\dot{q}_t$ could turn out to be even more frequent than smaller ones if the industry is mainly composed of very flexible firms planning large size adjustments, as argued in the section 3. This fact could explain why many empirical studies have found that the frequency distribution of size changes has tails fatter than expected according to the log-normal distribution.

Also, this model considers the discontinuity in the distribution of size changes as a structural characteristic of the growth process of firms, while it has been generally regarded as a nuisance factor in the empirical literature. As noticed in the introductory section, this discontinuity strictly resembles the "rows" in the spectrum of radiations emitted by a pure gas overheated. Exactly like the spectral rows, the model conjectures that the sequence of admissible $\dot{q}_t$ is typical for each industry and market segment, although the sequence of peaks in the distribution of size changes is hardly described by simple rules. In principle, the model argues that each firm could be classified by the specific sequence of size changes that characterizes its life, regardless to the branch of activity, current size and market position.

However, the peaks of the frequency distribution of $\dot{q}_t$ could be quite close to each other and, in the real world, firms likely differ also for their profit functions and entrepreneurs' preferences about π_t^* within the same industry and market segment. Thus the empirical distribution of size changes could be apparently continuous. Also analysing grouped data, instead of firm's level data, may provide evidence more favourable to the continuity of size distribution.

5 Conclusive remarks

Even disregarding many factors making the firm's growth discontinuous, the leaps in the size adjustment remain an intrinsic feature of the firm's evolution over time. This fact suggests that every model that aims at describing the firm's growth should be able to reproduce such discontinuities. Here this result has been achieved by assuming no more than few primitive hypotheses on the profit function and the reservation profit of entrepreneurs, intentionally disregarding other important factors of discontinuity considered in the literature.

In spite of its very simple structure, the model seems to give account of many stylised facts often pointed out by the literature as nuisance factors within the framework of rather general models. Therefore, the discontinuity arguably persists even removing the constraints and the inertial factors often referenced as its main determinants. Many other conjectures of the model are challenging and are suitable to be tested empirically. First of all, the abrupt changes over time from one locally optimal size to the other, that is the key feature of the model, may explain the loose statistical relationships among firm's output, profit, investment and degree of capacity utilization. The crucial role played by the minimum profit claimed by the entrepreneur makes clear why often firm's level investment may be large in market where current profit is small, and small elsewhere, contrarily to the intuition. In addition, the model provides an explanation for the paradox of very flexible and profitable firms, such as most SMEs, that nevertheless show a low propensity to grow but, when the entrepreneur decides to invest, expand their size by a large amount. On the other hand, the model envisages the risk that some firm falls in a "size trap" because it is very flexible or the entrepreneur contents himself with low profit. Furthermore, the minimum profit threshold puts a ceiling to the investment of a single firm even in face of a fast growing demand, leaving room for newborn firms enter the market. Lumpy investment contributes explaining why the distribution of firms by size is often multimodal. What is more, the model reconciles the results more favourable to the Gibrat's law achieved by analysing grouped data to the evidence coming from individual data and nonparametric estimates.

The model intentionally concentrate mainly on the role of the reservation profit during the size adjustment process. The profit function and the way the entrepreneurs make their decisions assumed here are very simplified on purpose, just in order to make the topic analytically tractable. In addition, possible credit constraints are not considered, recurring to the purely theoretical hypothesis that the financial market is perfect and the entrepreneur is indifferent to the source of funds. Also the time needed to adjust the firm to the desired size has not been considered explicitly as well, although a more realistic representation of the investment process should take into account the overall profit loss during the transition process, rather than the pure instantaneous profit flow. Indeed these topics deserve much more analytical work.

References

Bachmann R. and C. Bayer (2011), "Investment Dispersion and the Business Cycle", Working Paper n. 16861, National Bureau of Economic Research.

Baumol W., J.C. Panzar and R.D. Willig (1982), *Contestable Markets and Theory of Industry Structure*, Harcourt Brace Jovanovich, New York.

Pyke F., G. Becattini and W. Sengenberger (1990), *Industrial districts and inter-firm cooperation*, (eds.), International Institute for Labour Studies, Geneva.

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

Bottazzi G., E. Cefis, G. Dosi, and A. Secchi (2007), "Invariances and diversities in the patterns of industrial evolution: Some evidence from Italian manufacturing industries", *Small Business Economics*, vol. 29, n. 1, pp. 137–159.

Bulan L., C. J. Mayer and C. T. Somerville (2006), "Irreversible Investment, Real Options, and Competition: Evidence from Real Estate Development", *NBER Working Papers*, no. 12486, National Bureau of Economic Research.

Caballero R. J. and E. M. R. A. Engel (1999), "Explaining Investment Dynamics in U.S. Manufacturing: A Generalized (S,s) Approach", *Econometrica*, vol. 67, no. 4, pp. 783-826.

Caballero R. J., E. M. R. A. Engel, and J. Haltiwanger (1997), "Aggregate employment dynamics: Building from microeconomic evidence", *American Economic Review*, vol. 89, pp. 921–946.

Cabral L.M. and J. Mata (2003), "On the Evolution of the Firm Size Distribution: Facts and Theory", *American Economic Review*, Vol. 93, No. 5, pp. 1075-1091.

Caves R. E. (1998), "Industrial Organization and New Findings on the Turnover and Mobility of Firms", *Journal of Economic Literature*, vol. 36, n. 4, pp. 1947-1982.

Clifford D. K. and R. E. Cavanagh (1985), *The Winning Performance: How America's High-Growth Midsize Companies Succeed*, Bantam Books, New York.

Coad A. (2009), *The Growth of Firms: A Survey of Theories and Empirical Evidence*, Edward Elgar, Northampton.

Cooper R., J. Haltiwanger and L. Power (1999), "Machine Replacement and the Business Cycle: Lumps and Bumps", *American Economic Review*, vol. 89, no. 4, pp. 921-46.

Décamps J. P., T. Mariotti and S. Villeneuve (2006), "Irreversible Investment in Alternative Projects", *Economic Theory*, vol. 28, pp. 425-448.

Dixit A.K. and R.S. Pindyck (1994), *Investment under Uncertainty*, Princeton University Press, Princeton.

Doms M. E. and T. Dunne (1998), "Capital Adjustment Patterns in Manufacturing Plants", *Review of Economic Dynamics*, vol. 1, no. 2, pp. 409-29.

Friedman M. (1953), *Essays in Positive Economics*, The University of Chicago Press, Chicago.

Garnsey E., E. Stam, and P. Heffernan (2006), "New firm growth: Exploring processes and paths", *Industry and Innovation*, vol. 13, n. 1, pp. 1–20.

Geroski P. (1995), "What Do We Know About Entry?", *International Journal of Industrial Organisation*, Vol. 13, No. 4, pp. 421-440.

Gibrat R. (1931), *Les inégalités économiques*, Paris, Librairie du Receuil Sirey.

Hammermesh D. and G. Pfann (1996), "Adjustment costs in factor demand", *Journal of Economic Literature*, vol. 34, pp 67–140.

Jensen, M. C. and W. H. Meckling (1976), "Theory of the firm: Managerial behavior, agency costs and ownership structure", *Journal of Financial Economics*, vol. 3, pp. 305–360.

Khan A. and J. K. Thomas (2003), “Nonconvex Factor Adjustments in Equilibrium Business Cycle Models: Do Nonlinearities Matter?”, *Journal of Monetary Economics*, vol. 50, pp. 331–360.

Khan A. and J. K. Thomas (2008), “Idiosyncratic Shocks and the Role of Nonconvexities in Plant and Aggregate Investment Dynamics”, *Econometrica*, vol. 76, n. 2, pp. 395–436.

Lau L. J. (1976), “A Characterization of the Normalized Restricted Profit Function”, *Journal of Economic Theory*, 12, pp. 131-63.

Lotti F., E. Santarelli, and M. Vivarelli (2003), “Does Gibrat’s Law Hold in the Case of Young, Small Firms?”, *Journal of Evolutionary Economics*, Vol. 13, No. 3, pp. 213-235.

Modigliani, F. and M. Miller (1958). “The Cost of Capital, Corporation Finance and the Theory of Investment”, *American Economic Review*, vol. 48, No. 3, pp. 261–297.

Nilsen Ø. A. and F. Schiantarelli (2003), “Zeros and Lumps in Investment: Empirical Evidence on Irreversibilities and Nonconvexities”, *The Review of Economics and Statistics*, vol. 85, no. 4, pp. 1021-37.

Nye M. J. (ed.) (2002), *The Modern Physical and Mathematical Sciences, in The Cambridge History of Science*, vol. 5, Cambridge University Press, Cambridge.

Santarelli E. and M. Vivarelli (2007), “Entrepreneurship and the process of firms’ entry, survival and growth,” *Industrial and Corporate Change*, vol. 16, no. 3, pp. 455–488.

Simon H. (1996), *Hidden Champions-Lessons from 500 of the World's Best Unknown Companies*, Harvard Business School Press, Boston.

Sinha S., A. Chatterjee, A. Chakraborti and B.K. Chakrabarti (2010), *Econophysics: an introduction*, Wiley-VCH, Berlin.

Sutton J. (1997), “Gibrat’s Legacy”, *Journal of Economic Literature*, vol. 35, no. 1, pp. 40-59.

Traù F. (2000), "The Rise and Fall of the Size of Firms", *ESRC Working Paper*, no. 156.

Traù F. (2003), *Structural Macroeconomic Change and the Size Pattern of Manufacturing Firms*, Palgrave MacMillan, New York.

Graph 1 – Profit functions for a couple of alternative plants