

CHAPTER 2

A REVIEW OF THE LITERATURE OF FIRM GROWTH

A Review of the Literature of Firm Growth

TABLE OF CONTENTS

2.1. Introduction

2.2. Firm growth theory and firm size

2.2.1. Definition of the firm

2.2.2. Measure of firm size

2.2.3. Factors of firm growth

2.2.4. Different approaches to firm growth

2.3. Gibrat's Law

2.3.1. The evolution of Gibrat's Law

2.3.2. Empirical evidence

2.3.3. Spanish literature

2.4. Summary and concluding remarks

A Review of the Literature of Firm Growth

CHAPTER 2

A REVIEW OF THE LITERATURE OF FIRM GROWTH

“The growth and survival prospects of new firms will depend on their ability to learn about their environment, and to link changes in their strategy choices to the changing configuration of that environment”.

Geroski (1995)

2.1. INTRODUCTION

Firm Demography concerns the different stages in a firm’s life cycle. Firms appear in the market, survive, grow and eventually die, transferring their knowledge and information to surviving firms. In this sense, firm size reflects how the firm evolves and adapts to its environment. Changes in size are therefore extremely important events in Firm Demography (Wissen, 2002).

Firm growth has been one of most widely studied topics in economic literature. Several arguments highlight the crucial importance of this field. First, firm growth is related very closely to firm survival. Specifically, firm growth is positively correlated with the likelihood of

A Review of the Literature of Firm Growth

survival. Hence firms that experience continuous growth will have a higher probability of surviving in the market¹.

Second, firm growth has consequences for employment². A positive rate of growth implies a net creation of new jobs, while a negative rate implies the net destruction of jobs. Job creation and job destruction are closely related to the ability of incumbents and new entrants to grow. And, obviously, the evolution of employment therefore has obvious impacts on government budgets.

The third factor behind the importance of firm growth is its effect on economic growth. Backward and forward linkages will be higher or lower depending on the evolution of active firms. If we look at the general effect on an economy, an increase in firm growth may increase its demand towards other sectors, thus producing an increase in the economic activity of a region. This dynamism in the economy can lead to major growth. On the other hand, a decrease in the number of employees in a firm may indicate or cause a crisis³.

Fourth, firm growth is a way to introduce innovation and is a leitmotiv of technological change (Pagano and Schivardi, 2003). For example, if a firm wants to grow and survive in a competitive industry, it needs to

¹ It is a well known fact that firms are born undersized (Geroski, 1995). Firms, which adapt to the market process, will grow in size to take advantage of scale economies.

² When we speak about firm growth we refer to worker flows or the number of jobs created or destroyed over a period of time. However, we do not consider the number of people who changed of jobs over a period of time.

³ The causal relationship depends on the firm's expectations. Firms may foresee a crisis but they may also cause it if their expectations are incorrect about the economic evolution. For example, Penrose (1959: 40) pointed out the role of expectations in the productive opportunity of the firm.

incorporate new technologies in order to be more efficient. In this sense, growth is a challenge a firm must meet by introducing innovation⁴.

Fifth, the evolution of the size of incumbents and new entrants determines market concentration. If small firms grow at a high rate, market competitiveness will increase. Conversely, increases in the size of large firms will affect market concentration. The regulation of market concentration to avoid the creation of monopolies and oligopolies (Shepherd, 1979) has been one of the main interests of governments. The analysis of firm growth may therefore help to clarify the concentration of firms in a market.

Moreover, a study of firm growth can shed light on the importance of the selection process after a firm has entered the market (Audretsch and Mata, 1995). Once a firm enters a market a selection process takes place (Jovanovic, 1982) whereby less efficient firms decrease in size and disappear and more efficient ones survive and grow. The analysis of firm growth will therefore show how firms behave once they enter the market, their market opportunities, turbulence and level of efficiency.

Another important characteristic of this topic is that firm growth has practical consequences for policy-makers' decisions (Wagner, 1992). Firm growth can increase employment and economic activity and policy-makers can control these macroeconomic variables using firm growth policies. However, as the growth is heterogeneous between firms, it is crucial to know the internal and external characteristics of firms that affect their performance in the market. An ample knowledge of these

⁴ Audretsch and Lehman (2005) found that there is a positive impact on firm growth when a firm invests in R&D. Also, Thornhill (2005) confirmed that innovations are positively correlated with firm performance, as measured by revenue growth.

A Review of the Literature of Firm Growth

features will enhance the effectiveness of public policies as well as their impact.

Because of these important reasons, much of the literature has focused on the firm growth process. However, there has been no convergence of the theories. As Correa et al. (2003) pointed out, these varied approaches may be due to the complexity involved in defining the firm. Contributions from classic economic theory, the behaviourist, the stochastic growth theory and the learning models have helped to perceive the causes and effects of firm growth.

Our interest is to highlight contributions to the literature of stochastic firm growth. Since Gibrat's study (1931), several articles have sought to explain the relationship between firm growth and firm size. This approach characterises firm growth as a constant probability for a firm to grow independently of its initial size. As Simon and Bonini (1958) pointed out, the main consequence is that firm distribution has a skewed tail. Hence, the vast majority of firms in the market will be small and medium sized while a few firms will have the majority of the employees in the industry.

Gibrat's Law, or the Law of Proportionate Effect, is an alternative theory to classical economic theory which postulates that there is an optimal firm size. Classic economists found it difficult to explain the presence of firms with heterogeneous sizes. In this sense, Gibrat's Law explains the empirical evidence better. However, the classical and the stochastic theories offer different explanations for a firm's size and its performance in the market.

In the last few decades, the post-entry performance of firms has focused researchers' attention. Post-entry performance includes analysis of a

firm's growth and the likelihood of its survival. This chapter has two aims: first, to present several economic theories that have been used to analyse the firm growth process, and second, to analyse Gibrat's Law more carefully from the theoretical and empirical perspectives. We also consider the learning models (Jovanovic, 1982; Ericson and Pakes, 1995; Pakes and Ericson, 1998), which are closely related to Gibrat's Law. As in the literature there are many empirical contributions, we will analyse the most outstanding ones. For Spain, however, studies are scarce but these few contributions will also be reviewed.

In advance we can say that in the literature there is heterogeneity in the analysis of firm growth—in the measure used to analyse growth and in the results. Gibrat's Law is generally rejected in favour of the growth of small firms. Moreover, service industries have been largely ignored. Also, there are few analyses of the locational effects on firm growth.

This chapter is structured as follows. In section 2.2 I briefly summarise the approaches to the analysis of firm growth in the economic literature and describe the main differences between them. In section 2.3, I describe Gibrat's Law (1931) and describe, from a theoretical and empirical perspective, the evolution of the literature. Finally, in section 2.4 I summarise the main conclusions of this chapter.

2.2. FIRM GROWTH THEORY AND FIRM SIZE

Firm growth is one of the most analysed fields in economics. Its impact on employment, industry concentration, firm survival and economic activity are reasons enough for it to be considered an issue of crucial

A Review of the Literature of Firm Growth

interest⁵. However, there is no single theory to analyse the impact, causes or evolution of firm growth. As Correa (1999) pointed out, this may be because the definition of the firm is multiple and complex.

This complexity has led to the emergence of scholars with different perspectives and, more importantly, with different predictions of the evolution of growth. This is clearly seen from the variables used in the literature to measure firm growth and its determinants. Some theories focus on average size, some focus on internal characteristics and others focus on random variables.

In any case, we will see that firm size has been the link between the various theoretical approaches. The questions we focus on this section are as follows: which measures have been applied to the firm growth process? How has the economic literature approached the firm growth pattern? What is Gibrat's Law and what are its main empirical consequences?

2.2.1. Definition of the firm

As we intend to analyse the behaviour of firms, I shall describe what the literature understands by a firm, a definition that has evolved over the years. From the black box where a set of inputs enters production and transforms them into a set of outputs, the definition of the firm has widened its perspective and adopted a more ecological perspective in which firms interact with the other agents in society and have their own internal function. Here, we present the most relevant contributions and determine what our perspective will be.

⁵ Undoubtedly firm growth is an objective a firm needs to survive and be competitive and is the result of individual and collective effort. However, authors such as Suárez (1999) pointed out that in a more globalised economy, it is more important for firms to concentrate on the production of added value products than on oversizing.

First, a firm can be considered as the internal process superseded by the external price mechanism. In this sense, the firm is defined by the boundary from where the output leaves the production system and enters the market; at this point the firm does not have control of the output. Coase's (1937) seminal contribution considers that firms are created because of friction in the price mechanism. Firms are limited by a marginal rule and internalize activities up to the point where internal management costs equal the costs of transacting in the markets.

Second is the perspective that firms are a group of capabilities. Here we must mention Penrose (1959) and Richardson (1972). Penrose (1959) differentiated between resources and the services they render. Resources can provide a variety of productive services. In turn, the provision of these services can modify the attributes of the resources and enable the provision of new services. In this sense, the firm is considered as a collection of productive resources the disposal of which between different uses and over time is determined by administrative decisions (Penrose, 1959). The fact that there is heterogeneity rather than homogeneity of both human and material productive services implies that firms are unique. Finally, the limits are defined by the nature of the firm's managerial and administrative responsibilities.

Richardson (1972) replaced the Penrosian notion of productive services with capabilities and activities and widens the definition to the coordination of capabilities in industrial systems. He considered the firm as a network: the boundaries of the firm depend on the type of activities it carries out and how these activities fit with others. This means that the corporate ownership of a firm may control several autonomous firms

A Review of the Literature of Firm Growth

that depend to some extent on the main corporation. The main examples are the franchises that depend on the main corporation⁶.

Third, Hart (1995) defined a firm as the ownership of or the property rights to a firm. Therefore, the limit of the firm is when one person has all the risk of the economic activity. With this approach, the firm is conceived as a set of assets under common ownership and control. One problem with this definition is that, as employees are not a possession of the firm, they would not be considered as part of the firm.

Highly complementary assets should be owned in common and the owner of these assets should be the best person to provide investment incentives for the best use of these complementary assets (Hart, 1990). This view provides an answer to where the limits of the firm should lie since they coincide with decisions about physical asset ownership.

Finally, the firm can be defined in terms of its sphere of influence. Williamson (1985) extended the boundaries of the firm to other agents that are in direct contact with the firm, such as distributors, alliance partners and suppliers. From this perspective, the emergence of the firm is a response to problems causing delay (*hold-up* problems), given the intrinsic opportunistic nature of human actors and the specialized nature of assets required for efficient production.

As we can see, the firm is difficult to define because its influence is multiple. For the purpose of this thesis, we define the firm as the

⁶ For Richardson (1972), the degree of interdependence defines whether activities are complementary or closely complementary. Complementarity implies that activities must be matched in either level or specification and require some form of sequential coordination. Close complementarities arise when activities are to some extent specialized and require more careful coordination in the way they are combined.

ownership of assets i.e. the third definition above. However, we will consider employees as belonging to the firm. This means that we will not consider employees to be autonomous individuals with their own incentives. This definition, closely related to Hart's definition (1995), is often used in research (Kumar et al., 1999).

2.2.2. Measure of Firm Size

One of the main challenges in every discipline is to homogenise the criteria for classifying its units of observation. The analysis of firm growth is no different because there are different ways of measuring the growth of a firm. This diversification is sometimes due to the purposes of each author but, more usually, it is due to lag of data (Correa, 1999).

The main problem is therefore the difficulty in making comparisons with other studies. In fact, the empirical literature uses a wide range of measures whose use depends on the purpose and subject of the data.

Ardishvili et al. (1998) and Delmar (1997) found similar growth indicators used in the empirical literature. Some of these indicators) are⁷:

- The financial or stock market value
- The number of employees
- The sales and revenue⁸
- The productive capacity
- The value of production
- The added value of production

⁷ See Annex III for a review of the variables used in the stochastic empirical literature. Also see Audretsch et al.'s (2004) review of the literature.

⁸ Both variables, sales and revenues, are directly related since revenues are equivalent to sales less the costs of production.

A Review of the Literature of Firm Growth

Although all of these parameters are highly correlated (in other words, when a firm increases the added value of its production it also increases its stock market value), not all of them react so quickly to external or internal changes. For example, it is obvious that sales are more volatile than productive capacity because firms can generally modify their sales more often than they can modify their assets.

For example, Kirchhoff and Norton (1992) compared three measures (employment, assets and sales) and showed that they are interchangeable because they produce the same results when tested over a seven-year period.

However, each variable can paint a different picture of the firm. These may be interesting depending on the purpose of the research⁹. We can therefore select the most suitable variable for our interests. So, as our objective is to measure the firm's economic activity, we will see whether the above variables explain this internal process:

- The revenues of the firm do not provide any information about its internal process but show the prices and the quantities sold in the market.
- Sales are easily available and relatively insensitive to capital intensity. However, they are an unsatisfactory indicator because they can be influenced by a firm's arbitrary decisions (marketing strategies, financial decisions, etc.). Moreover, they can also be influenced by the decision to vertically integrate certain

⁹ Delmar et al. (2003) analysed different measures and conclude that firm growth can be expressed by different measures depending on the aim of the investigation.

production processes and are sensitive to inflation and currency exchange rates (Delmar et al., 2003).

- Added value may be a better variable since it explains the capacity of the process to increase the value of the output. It is therefore quite a good indicator of internal activity. Unfortunately, however, added value is sometimes not publicly available for individual firms.
- Assets can also define the size and growth of a firm. However, as we stated earlier, they are more rigid to changes in the internal process of the firm and may not be a good explanatory variable¹⁰.
- As Kimberley (1976) stated, the number of employees is the most widely used measure of size. The number of employees reflects how the internal process is organised and adapts to changes in activity. Moreover, employment is not sensitive to inflation or currency exchange rates. Scholars agree that this variable is a direct indicator of organizational complexity and is suitable for analysing the managerial implications of growth (Penrose, 1959).

The best variables for measuring firm size are therefore added value and the number of employees. As we have mentioned, the problem with added value is that there is usually a lack of information. The only problem with using the number of employees is that it does not consider the growth in labour productivity. Depending on the problem at hand, other variables may also be possible. For example, economic activities directly related to tourism may have alternative measures, such as the

¹⁰ It is true that assets can be a correct measure when manufacturing industries are taken into account. However, it can lose explanatory capacity when used with service industries.

A Review of the Literature of Firm Growth

capacity of supply (the number of beds, etc.). However, authors such as Delmar et al. (2003) pointed out that different growth “*measures and calculations affect model building and theory development differently*”.

Obviously, using a measure such as the number of employees has several disadvantages. Delmar et al. (2003) mentioned that the number of employees does not reflect “*labour productivity increases, machine-for-man substitution, degree of integration, and other make-or-buy decisions*”.

Firm growth as measured by a difference in the number of employees has led to analyses of aspects such as labour policies, labour market evolution and job creation. In this sense, Hart (2000) argued that the limitation of using only the number of employees is not important since all measures of size are highly correlated.

2.2.3. Factors of firm growth

In this section we analyse some of the factors in the literature on firm growth. We present several important theories of the determinants of firm growth. Although it is true that we could be more accurate in our presentation of these theories, the aim of this section is simply to sketch several of the factors behind firm post-entry performance.

The first studies on firm growth concentrated mainly on the impact of size and age. However, the characteristics that can influence post-entry firm behaviour are wider and authors such as Storey (1994) determined several factors affecting firm growth.

Following Storey's (1994a)¹¹ classification, a distinction is often made between three groups of growth determinants: (i) those related to the *entrepreneur* (also defined as founder-specific); (ii) those related to the *firm* (also defined as owner/manager specific); and (iii) those related to *strategy*.

Factors influencing growth in small firms

<i>The entrepreneur's resources</i>	<i>The firm</i>	<i>Strategy</i>
1 Motivation	1 Age	1 Workforce training
2 Unemployment	2 Sector	2 Management training
3 Education	3 Legal form	3 External equity
4 Management experience	4 Location	4 Technological sophistication
5 Number of founders	5 Size	4 Market positioning
6 Prior self-employment	6 Ownership	6 Market adjustments
7 Family history		7 Planning
8 Social marginality		8 New products
9 Functional skills		9 Management recruitment
10 Training		10 State support
11 Age		11 Customer concentration
12 Prior business failure		12 Competition
13 Prior sector experience		13 Information and advice
14 Prior firm size experience		14 Exporting
15 Gender		

Source: Storey (1994a)

With regard to (i) above, Storey defined the entrepreneur's inherited and learnt abilities e.g. motivation, experience and age¹². With regard to (ii), features of the firm are e.g. age (the experience of the firm in the

¹¹ Following Storey (1994a)'s classification, firms can be divided into three groups: "failures", "trundlers" and "flyers". "Failures" are those firms that disappear after entering the market. "Trundlers" are firms that survive to the observed period but do not significantly change size. "Flyers" are firms that really contribute to job creation and increase in size.

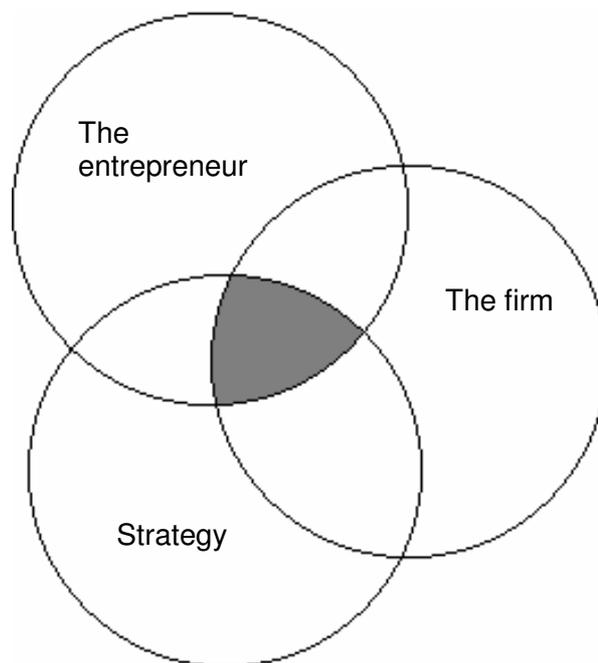
¹² Casson (1998) emphasised the key role of entrepreneurial ability as an explanatory variable of firm growth.

A Review of the Literature of Firm Growth

market), sector and size. With regard to (iii), firms may adopt strategies involving technology or exportation¹³. However, a lack of data in their analysis meant that they did not incorporate all of these dimensions¹⁴.

However, all of these elements should be combined in an appropriate way so that the firm grows rapidly. Graph 2.1 represents each of these factors (the entrepreneur, the firm and strategy) by a circle. The intersection of all of the circles (the shaded area) is where the fast-growing firms are located (Storey, 1994a). For example, a firm with a good strategy for tackling market competition may obtain low results if the manager does not have enough skills to cope with the new market situation or the ability to motivate his or her workers.

Graph 2.1 Interaction of factors on firm growth.



Source: Storey (1994a)

¹³ Barringer et al. (2005) recently analysed all these factors but grouped them into four explanatory vectors: founder characteristics, firm attributes, business practices and human resource management practices.

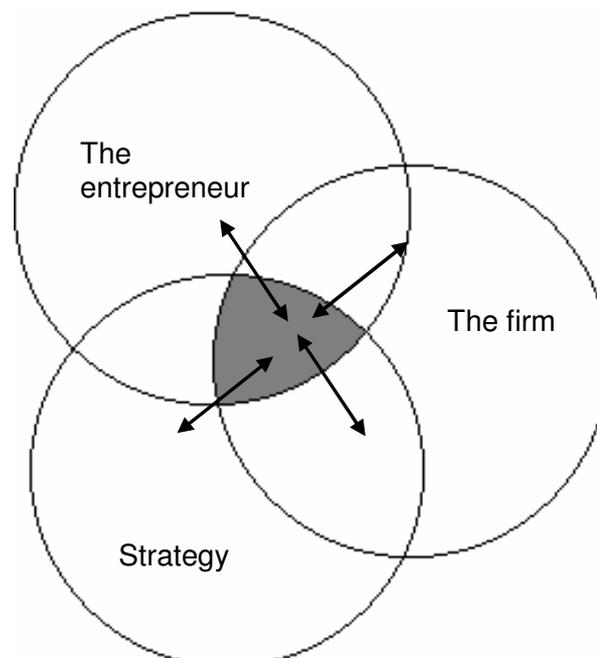
¹⁴ For example, Peña (2004) used the entrepreneur's characteristics to analyse growth in an incubation centre; other authors, such as Fariñas and Moreno (2000) and Correa et al. (2003), introduced variables related to the firm.

The interactions between these factors lead to a high heterogeneity in firm grow rates over time. Some researchers are therefore attempting to identify the relationships between all of these characteristics and firm growth.

A fourth factor also stands out: randomness. Some unexpected factors may affect the interaction of these factors. Geroski's (1999) theoretical study presents randomness as an event with an unknown form or a known fact whose date of occurrence is undetermined.

Graph 2.2 shows how this stochastic error term, represented by arrows, can affect this central core. The shaded area can increase or decrease depending on the effect of the error term.

Graph 2.2 Interaction of factors on firm growth and the error term.



Source: author's illustration based on Storey (1994a)

A Review of the Literature of Firm Growth

Therefore, some factors can be controlled by the firm, but other, unexpected, events that are internal or external to the firm can also affect firm growth. Not only are there factors that increase the growth of firms. Storey (1994a) pointed out the existence also of growth barriers. These barriers can be due to human failures but they may also be beyond the control of managers and owners.

Geroski (1995) mentioned that one of the most interesting subjects in firm dynamics is the ability to learn and respond to their changeable environment. *“The implication is that the growth and survival prospects of new firms will depend on their ability to learn about their environment, and to link changes in their strategy choices to the changing configuration of that environment”*. This ability to learn and adapt is crucial to firm growth and is highly correlated with the firm’s age or experience. As Geroski (1995) said, this is not the only factor but it is one of the most important.

Barriers to growth

Availability and cost of finance for expansion

Availability and cost of overdraft facilities

Overall growth of market demand

Increasing competition

Marketing and sales skills

Management skills

Skilled labour

Acquisition of new technology

Difficulties in implementing new technology

Availability of appropriate premises or site

Access to overseas markets

Source: Storey (1994a)

According to Scherer (1970), there are more factors that influence the size and growth of firms. These are:

Economies and diseconomies of scale

Classical economists explained that firm size and any changes in firm size depended on economies of scale. These economies of scale are due to diminishing costs when the firm increases. Therefore, the higher the economies of scale, the larger the optimum firm size¹⁵. Scherer (1970) claimed that diminishing unitary cost is not infinite. This is because:

- it implies the existence of a minimum optimal scale, but in the real economy there is a wide range of firm sizes in the same industry,
- when economies of scale exist, firm growth may reduce the unitary cost until other diseconomies appear: diseconomies of management due to a lack of harmony between different branches or the higher wages of executives, for example, should not exceed the benefits of economies of scale,
- transport costs are involved. This is because the larger the size of the firm, the larger its production and the higher its sales must be. To increase sales, the firm must sell at longer distances. Consequently, both unit shipping costs rise and prices rise. The limit is the price of the product in the market.

Mergers and acquisitions

Firms can grow internally or externally. They can grow organically, through acquisition, or by a combination of the two. Penrose (1959)

¹⁵ The concept of economies of scales was introduced by Adam Smith in the *Wealth of the Nations*, which reported that the specialisation of workers increased their productivity.

A Review of the Literature of Firm Growth

suggested smaller, younger firms in emerging industries mainly grow organically, whereas older, larger firms in mature industries mainly grow through acquisition.

When a firm acquires another firm, the concentration in the market and the market power of the firm increase. From Scherer's (1970) perspective, this situation responds to the "empire-building" desire of firms.

Ijiri and Simon (1977: 193) drew attention to two types of growth:

"The overall growth of firms consists of internal growth (due to mergers and acquisitions) and external growth (due to growth from sources outside the population). That overall growth satisfies Gibrat's law does not necessarily mean that internal growth and external growth each satisfy Gibrat's law individually, since deviations from the law may cancel with each other to produce an overall Gibrat's Effect."

The impact of government policies

Sometimes consciously and sometimes unconsciously, government policies can increase or decrease market concentration. Tax policies, for example, may make it difficult for Small and Medium-Sized Enterprises (SMEs) to attract capital if there is corporate income tax exemption. Also, when a government gives subsidies to firms with certain characteristics, this decision can influence the market structure through the disappearance of efficient firms that did not receive them.

Another example is the creation of incubating ventures whereby entrepreneurs receive public and private support to create their firm. This financial and administrative support modifies the distribution of

firms in the market. This modification can be positive, if the firms receiving the support are efficient, or negative, if the firms are inefficient. Firms located in the incubating centers may therefore increase their potential efficiency in the market artificially and take the place of efficient firms that do not receive external help. Consequently, public policy can cause inefficient firms to supplant efficient firms in the market. If these firms do not receive help, they may leave the market. The end result would be an increase in market concentration¹⁶.

Stochastic determinants of market structure

Other studies in the economic literature have focused on random determinants of firm growth such as managerial talent. In section 2.3 we will see that the heterogeneous pattern of firms in an industry is down to pure historical chance.

Authors such as Hoogstra and Dijk (2004) suggested that other *external factors* are related to a firm's location or environment. So far, most studies have mainly focused on firm- and entrepreneur-associated factors that influence firm growth. Almus and Nerlinger (1999) considered as external factors, as well as local factors, the average rates of wages and salaries. As these are cost factors, they can prevent the hiring of new employees and thus have a negative influence on growth¹⁷. Therefore, scholars have introduced internal and external firm characteristics to analyse firm growth.

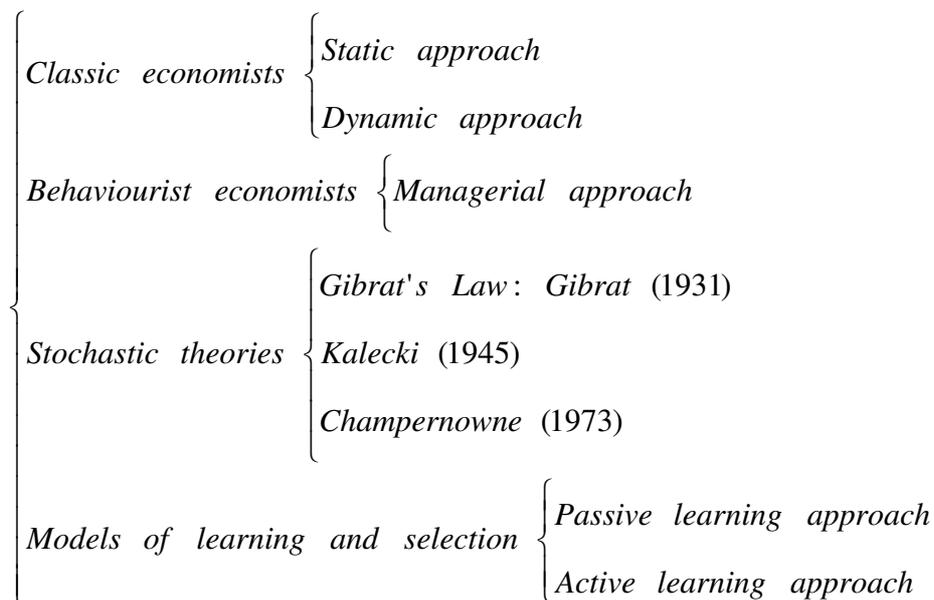
¹⁶ Hyytinen and Toivanen (2005) and Lerner (1999) showed that there is a significant positive relationship between public R&D subsidies and firm growth.

¹⁷ Almus and Nerlinger (1999) studied the new technology-based firms (NTBFs). As these firms are characterised as being capital intensive, we can assume that labour costs have only a minor impact on growth.

2.2.4. Different approaches to firm growth

This wide range of determinants is represented by several theories. These different theories are the result of the relevance of the topic and the difficulty in analysing it (Correa, 1999).

The main schools of thought can be divided into four groups¹⁸: (i) classical economists; (ii) behaviourist economists, who emphasise the role of managers on increases in firm size; (iii) stochastic theory, which assumes that firm growth follows a stochastic process; and (iv), models of learning and selection, which are linked to the stochastic firm growth theory.



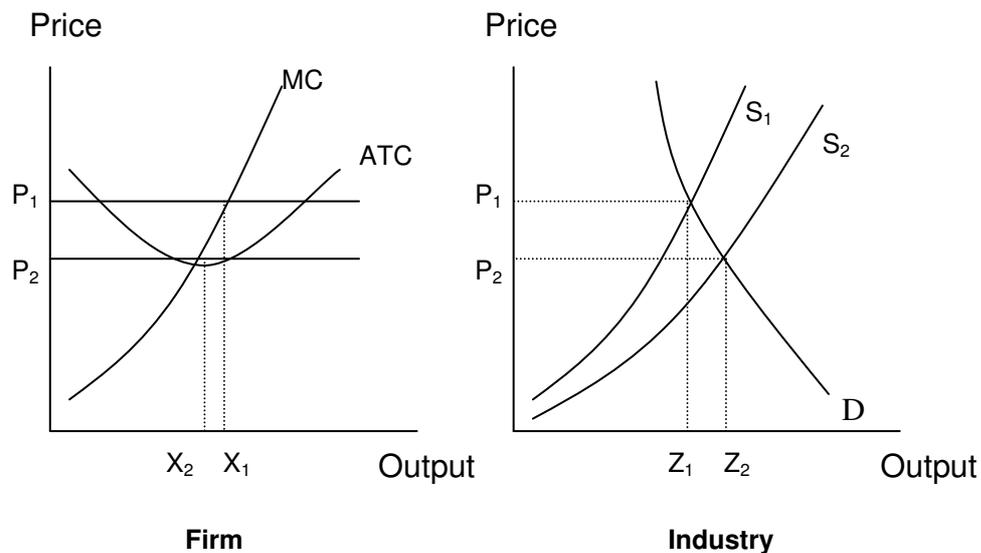
¹⁸ For a survey of the literature, see Hart (2000), Mazzucato (2000) and Geroski (1999). Hart (2000) analysed the neo-classical theory of the firm, imperfect competition, technical economies of scale, pecuniary economies of scale, external economies of scale and dynamic economies of scale. Mazzucato (2000) divided theories of firm growth and market structure into three: the Static Approach, the Dynamic Approach and the Stochastic Approach. Geroski (1999) focused on four types of growth theories: models of optimum firm size, stage theories of growth, models with Penrose effects and models of organizational capabilities.

A. Classical economists

Classical economic theory has studied firm growth indirectly because its aim is to find the optimum size (Viner, 1932). Firm growth is therefore the change between one equilibrium situation and another.

This approach shows a negative relationship between firm size and growth. The reason for this negative relationship is that firms search for the optimum, most efficient size. The benefits of efficiency are related to economies of scale: the larger the firm, the higher its profits.

Graph 2.3 Firm size equilibrium from a classical point of view.



Source: Scherer (1970: 12)

Note: MC is the Marginal Cost, S is the Supply, D is the demand for output and ATC is the Average Total Costs.

This kind of model suggests that competition will drive firms to the bottom of their U-shaped average cost curves. However, other

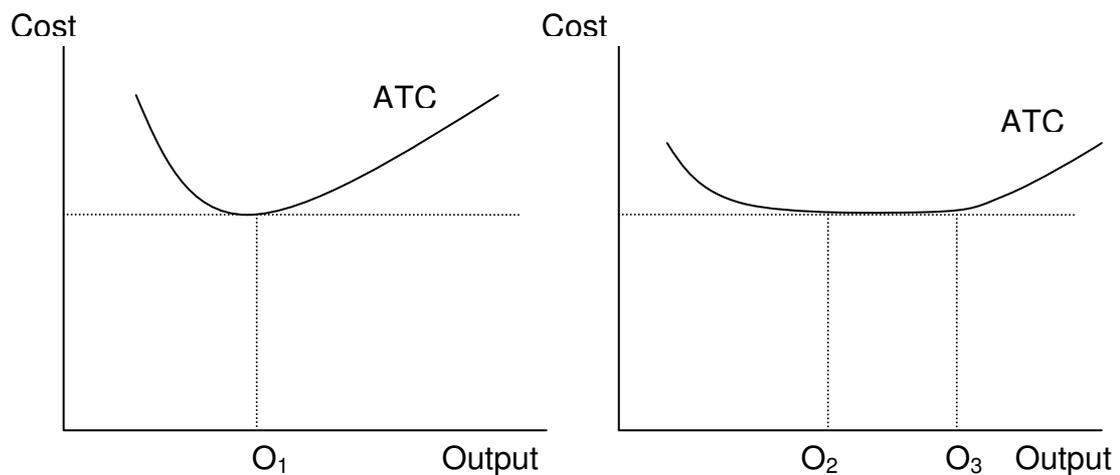
A Review of the Literature of Firm Growth

determinants, such as sunk costs, the intensity of the competition and organizational factors also play a key role in defining optimum size.

Graph 2.3 shows how the equilibrium size is reached. This is mainly due to market forces. For example, when an active firm in the market knows that the market price is p_I , it produces X_I units of output. The firm has positive profits per unit of output because the price is higher than the average total costs (ATC). Consequently, it cannot represent a long-run equilibrium position because these positive profits will attract new firms to produce this output and will add their marginal cost function to the industry's supply curve and shift the supply curve to the right. This increase in output will reduce the price in the market and new firms will continue to enter the market until the marginal cost equals the average total costs (ATC).

However, to adjust to reality, the cost-minimising part of the curve should cover a wide range of output levels and diseconomies appear when the firm reaches a high level of production (Graph 2.4).

Graph 2.4 Relationship between output and levels of cost in an industry.

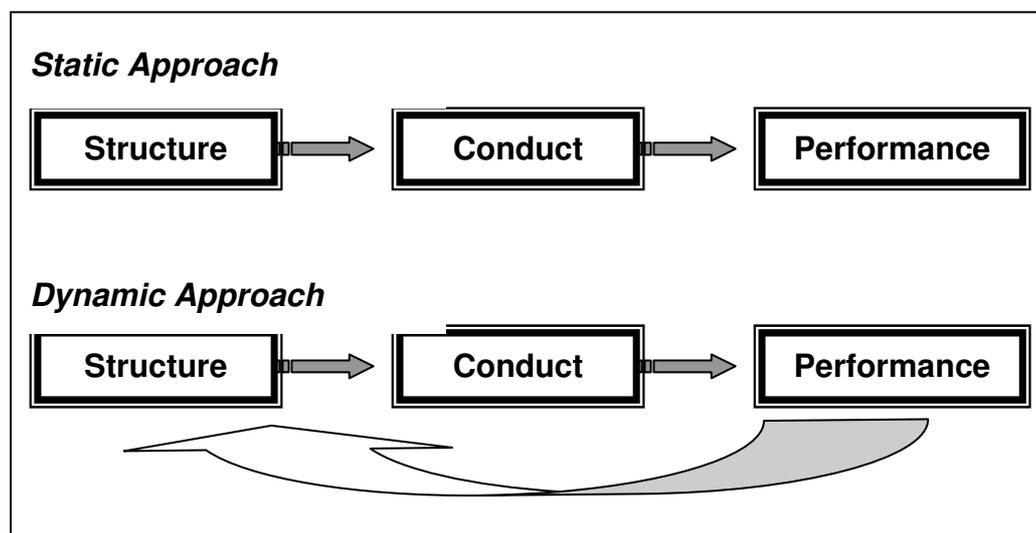


Source: Stigler (1958)

The new ATC curve reconciles the observation in the empirical evidence that heterogeneous firms coexist in the same industry. According to Stigler (1958), this would be because there is a wide range of outputs, between points O_2 and O_3 , for which the unit cost is more or less constant. Consequently, the hypothesis that diseconomies appear immediately after the point at which economies of scale disappear would be incorrect. There is a range of levels of output in which the firm can eliminate the different sources.

The relationship between firm size and the post-entry performance of firms has been analysed by classical economists. Graph 2.5 shows the differences between the Static Approach and the Dynamic Approach (Mazzucato, 2000).

Graph 2.5 Static and Dynamic approaches to the classical model of firm growth.



Source: author.

The Static Approach is based on microeconomic theory and refers to the Structure-Conduct-Performance framework, in which there is a linear

A Review of the Literature of Firm Growth

connection between the parameters. Given the structure of the market, it is possible to quantify a firm's production and hence the formation of prices.

The Dynamic Approach focuses on the feedback process between performance and structure. The central question is not the state of the production structure but the ongoing changes within it. It is no longer a matter of producing a specific quantity in order to maximise profits. With increasing returns to scale in a dynamic sense, we may see firms act in a way that would be considered irrational in the basic microeconomic context. However, this Dynamic Approach ignores the issue of long-run growth dynamics.

Geroski (1999) presented a model for reconciling this classical model with the stochastic firm growth model. He suggests that S^* drifts unpredictably over time and so the random variable $\mu_{i,t}$ will collect small, independent, firm-specific shocks that will alter the optimum size of reference. The proposed solution is therefore to let S^* vary unpredictably. To formulate the equation from a classical perspective, Geroski (1999) suggested a model based on the partial adjustment equation:

$$\Delta \log S_{i,t} = \lambda \{ \log S^* - \log S_{i,t-1} \} + \mu_{i,t}$$

This kind of model shows changes in firm size as a transitional process of convergence to S^* , the optimum size. The most interesting feature of the above equation is the parameter λ since it determines the velocity of convergence. When λ is equal to 0, convergence never occurs. When λ is equal to 1 there is a perfect coincidence between size and the optimum size.

The main implication of the classical model is that firm growth is always limited by this optimum size. However, evidence from the 1970s showed that there was a process of concentration inside industries. Consequently, one of the main criticisms of this model is that it cannot explain the presence of large firms whose size is greater than the optimum size or how the process of firm growth evolves over time.

B. Behaviour economists

The above criticisms in classical economic theory are addressed by the Behaviour approach. This theory considers that firms can be oversized because of the division between the objectives of control and ownership structures. When the owner does not control the firm, the managers maximize their own satisfaction instead of the firm's value. Behaviourist economists (Baumol, 1959, 1962; Penrose, 1959; Chandler, 1962; Marris, 1964) explain that managers can enhance their own satisfaction through an increase in the size of the firm.

This maximization of individual utility is a continuous process the limit to which is the managers' ability to coordinate and inspire confidence and security in others. Penrose (1959) argued that firms did not have a long-term optimum size but a constraint on current period growth rates. This means that rapid growth can imply a reduction in organisational efficiency (Penrose, 1959; Richardson, 1964; Williamson, 1967).

Another hypothesis is the *resources push* theory of growth, the basic premise of which is that competitive advantage lies in the possession of resources and routines, organizational capabilities or core competencies. This theory suggests that there are different types of firm behaviour and different levels of performance, so profitability (based on organizational

A Review of the Literature of Firm Growth

capabilities) is also different. However, this explanation does not explain why firm growth seems to be quite transitory, as if it were a random process.

The purpose of the Behaviour Approach is therefore to explain why some firms are more competitive. To do so, they focus on a firm's specificities and the internal elements that account for its performance. Heterogeneity in a firm's characteristics leads to a heterogeneity in performance.

This managerial approach highlights the importance of knowledge assets and processes of co-ordination within a firm (Penrose, 1959) and assumes that firm growth is due to an internal and endogenous creation and accumulation process of specific resources (Penrose, 1959; Chandler, 1962). However, this perspective has drawbacks such as the generality of managerial abilities and a lack of empirical studies that investigate the relationship between knowledge structures and firm growth.

C. Stochastic economists

The stochastic firm growth theory has developed simultaneously to these theories. Stochastic growth models have two main objectives: to detect the existence and persistence of the stochastic factors affecting firm behaviour and to detect the presence of inequality and concentration among firms. Three of the main stochastic growth models are those of Gibrat (1931), Kalecki (1945) and Champernowne (1973). Broadly speaking, the models of Gibrat (1931) and Kalecki (1945) follow a lognormal distribution of firm sizes, while that of Champernowne (1937) follows a Pareto distribution.

Gibrat (1931)¹⁹ suggested that there is no relationship between the size of a firm and its growth. This is known as Gibrat's Law or the Law of Proportionate Effect. In fact, firm growth is the result of a multiplicative process that affects the initial size. The factors that can affect firm growth relate not only to the firm, but also to its environment. The main consequences of Gibrat's Law²⁰ are as follows (Sutton, 1997).

- There is no optimum size to which firms will converge.
- The likelihood of growth is independent of initial size, so expected growth and its variability are the same for all firms²¹.
- Past growth does not affect current growth since there is no serial correlation (both between firms and over time).
- Firm size dispersion increases over time, so market concentration is higher if the number of firms remains constant.
- The variance of firm growth rates is equal for all sizes. This means that the variance of firm growth rates for small firms is equal to the variance of firm growth rates for large firms.

In other words, Gibrat's Law postulates that the “*probability that the next opportunity is taken up by any particular active firm is proportional to the current size of the firm*” (Sutton, 1997).

Kalecki (1945) formulated a stochastic growth model that assumes that the logarithmic variance of size is constant over time and, therefore, that the logarithm of size and the logarithm of the random variables are

¹⁹ In section 2.3 we will see a deeper approximation of Gibrat's Law.

²⁰ Gibrat tested this growth process not only with respect to firm size but also with respect to income distributions. We should also point out that the so-called Gibrat's law of firm growth is named differently in other research fields in reference to other “organisms”.

²¹ This is why Gibrat's Law has been called the Law of Proportionate Effects.

A Review of the Literature of Firm Growth

negatively correlated. Specifically, to express the current firm size he defines the following equation:

$$Y_n = Y_0 + \sum_{i=1}^n y_i = Y_0 \prod_{i=1}^n (1 - \alpha_i) + \sum_{j=1}^n \prod_{i=1}^j z_j (1 - \alpha_i)$$

where Y_n is the firm size at period “n”. Consequently the size will be the sum of small independent random increments $z_k \prod (1 - \alpha_k)$. When the central limit theorem is applied, the distribution to Y_n implies that as $n \rightarrow \infty$ the distribution is normal.

Kalecki’s (1945) model therefore has two main consequences.

- The stochastic process of firm growth is as in Gibrat’s Law.
- There is no increase in the concentration of the market since large firms find impediments to grow proportionate to their size.

Champernowne (1937) presented a model of income distribution that can be applied to the distribution of firm assets (Simon, 1955). In this case, the model assumes that firm size follows a Markov process that depends on the previous state and a random element. The possibility of changing state (firm size, in this case) is called the transition probability and all the possibilities are compiled in a transition matrix. The probabilities of changing firm size will depend on the distance: the greater the difference between the current size and the desired size, the lower the probability that this change occurs. Consequently, firm size in period $t+1$ will depend on the following formula:

$$X_x(t+1) = \sum_{u=-\infty}^1 p_u X_{(s-u)s}(t)$$

where u is the distance between the current size (r) and the future size (s). When there is a sufficiently long period, the distribution reaches an equilibrium called the stationary distribution. Champernowne (1937)

determined that the stationary distribution would approach a Pareto distribution:

$$\log F(Y_s) = \gamma - \alpha \log Y_s$$

where Y_s is the lower limit of size and $F(Y_s)$ is the number of firms whose size exceeds s . The main conclusions of Champernowne's model are as follows.

- As in Gibrat's Law, firm growth is independent of firm size.
- As in Kalecki, the "growth" process remains non-dissipative but in a much more restrictive sense. Champernowne imposes a stability condition that causes the expected value of variations to be negative for all firms. Consequently, concentration decreases.

Other authors, such as Ijiri and Simon (1977) and Scherer (1980), state that stochastic factors have a high impact on the distribution of firms in the market. The result is a highly skewed firm distribution in which a large number of small firms live with a low number of large firms. This approach has been criticised in the literature, however, because it assumes that the firm growth process is a random walk in which factors such as luck have a high weight.

Scherer (1970: 128) pointed out some firm determinants in which there is a component of stochasticity. Some of these are "*the hiring of key executives, research and new product development decisions, legal disputes involving critical patents, the choice of an advertising campaign theme, or a thousand and one other decisions among attractive but uncertain alternative courses of action. Given the operation of chance in these elemental decisions, high or low sales growth follows in a more*

A Review of the Literature of Firm Growth

traditionally deterministic manner". Therefore, from a managerial point of view, firm growth is full of characteristics that can cause the randomness of firm growth.

D. The models of learning and selection

More recently, the learning theory has appeared in the economic literature. Geroski (1995) emphasized that firm growth and survival depend on a firm's capacity to learn. Empirical evidence shows that the survival and post-entry performance of new firms depends on their capacity to adapt to the environment and apply the correct strategies. The learning and selection approach emphasizes the ability of firms to learn, their capacity for innovation and sectorial features.

There are several outstanding models of the learning and selection processes. These include those of Jovanovic (1982), Ericson and Pakes (1995) and Pakes and Ericson (1998). The main characteristics of these models are the fact that they take into account the dynamics of firms and their level of efficiency, which determine their chances to survive.

Jovanovic (1982) provided a model in which firms do not know their level of efficiency until they enter the market. This learning process is called a Bayesian or passive learning process. Once in the market, the most efficient firms grow faster until they reach a minimum efficient size. Inefficient firms disappear with the course of time. This is the *Theory of "noisy" selection*. These types of model introduce variables such as age to measure this ability of a firm to learn its economic efficiency.

Specifically, firms are created with a number of workers (l) and are affected by a productive shock. The distribution of the probability of profits is unknown at the initial moment and does not vary with time.

Entrepreneurs use past information to know their expected profits in the future. Their profit function is:

$$\Pi_t = \Pi(\eta_t; a_t, \xi) \equiv \max(l) \{ a_{1t} \eta_t F(l) - a_{2t} l \}$$

where Π_t are profits determined by a productive shock η , which affects the efficiency of all firms and is independent, l is the number of inputs and a is a vector of prices (1 implies the price of output and 2 implies the price of the input).

Each firm maximises the expected value of profits conditioned to the current information. The expected value of a firm, which lives for “ t ” periods and has been affected by η productive shocks, will be the expected value of the firm in the next period (closed down or still active):

$$V_t(n^t) = E[\Pi(\eta_{t+1}) | n^t] + \beta E[\max\{\phi, V_{t+1}(\eta^{t+1})\} | n^t]$$

The main result is that firms whose size is inferior to the minimum efficient size do not accept Gibrat’s Law. If these small firms survive, they will increase their size. However, for firms above the minimum efficient scale (MES), Gibrat’s Law is accepted. Jovanovic (1982) therefore models the heterogeneous behaviour of firm growths depending on firm size and their level of efficiency.

Ericson and Pakes (1995) and Pakes and Ericson (1998), on the other hand, presented an active learning process in which firms not only know their efficiency level when they participate in the market but can also can modify it through investment. During each period of time, firms

A Review of the Literature of Firm Growth

decide whether to continue in the market or to leave it, depending on both their own and their competitors' investment²².

In this model, active firms maximise the following profit function:

$$V(\omega, s) = \max_{x \geq 0} \left[\sup_{x \geq 0} \left\{ R(\omega, s; x) + \beta \sum_{\eta'} \sum_{s'} \sum_{\omega'} V(\omega', s') p(\omega' | \omega, x, \eta') q_{\omega}(\hat{s}' | s, \eta') p_{\eta'} \right\} \phi \right]$$

where $V(\cdot)$ represents the expected value of future profits given the probability of the state ω and the industrial structure s . This value depends on the decision of the firm to still be active or to close (the opportunity cost is ϕ in the case of closure).

The expected profits take into account the current revenue $R(\cdot)$ but also the future revenue. Future revenue is the discounted value of the expected profits $V(\omega', s')$ by the transition function $p(\cdot)$, which depends on the future probability (ω') conditioned to the current probability (ω), past investments (x) and the future productive shocks (η').

Moreover, the transition function $p(\cdot)$ is multiplied by the opinion of the firm with regard to the transition probabilities of the other firms ($q(\cdot)$), which depends on the number of future active firms \hat{s}' . The number of future active firms depends on the current number of firms (s) and the future productive shocks (η').

Moreover, two factors affect the expected future profits—the probability of future productive shocks ($p_{\eta'}$) and the discounting value β . New entrants maximise their expected profits:

²² This model is similar to the ecological perspective of Nelson and Winter (1982), who introduce investment in R&D.

$$V^e(s, m) \equiv \beta \sum_{\eta'} \sum_{s'} \sum_{\omega'_m} \sum_{\omega^0} V(\omega^0, s' + e_{\omega^0} + \hat{\omega}_m) \cdot \pi^e(\omega^0 - \eta')$$

$$\times \prod_{j=1}^{m-1} \pi^e(\omega_j^0 - \eta') \cdot q^0(\hat{s} | s, \eta') \cdot p_{\eta'} \phi$$

The entry decision is sequential since firms enter the market until the expected value of profits falls to a point at which the sunk costs are not recovered (x_m^e). This means that the entry of one more firm ($m+1$) represents a negative profit. Analytically:

$$V^e(s, m+1) - x_{m+1}^e \leq 0 \leq V^e(s, m) - x_m^e$$

Ericson and Pakes (1995) and Pakes and Ericson (1998) found an equilibrium with rational expectations and a finite number of heterogeneous firms in an environment with idiosyncratic *shocks*. Simultaneously, the equilibrium is an ergodic stochastic process i.e. a dynamic process is generated by optimal strategies of investment. The consequences of ergodicity are: a) the industrial structure evolves over time; b) there appears to be a regularity in the evolution of the industry; and c) the influence from the initial situation disappears with industrial development (i.e. the future is independent of the past; there is a Markov process).

Lucas (1978) presented a theoretical model where Gibrat's Law is the driving force. The production function depends on "production technology" and "managerial or entrepreneurial technology". Gibrat's Law is accomplished under the condition of heterogeneity of managers' skills levels and the presence of costs of rearranging assets between managers. The former presents constant returns to scale and the latter presents decreasing returns to scale. The results of this author show how every firm size depends on the level of managerial talent.

A Review of the Literature of Firm Growth

On the other hand, Cabral (1995) showed how small firms grow faster because of sunk costs. In this case, the initial investment is a small portion of the optimum production in the long run. In a recent paper from Aquilina et al. (2006), and based on Lucas's (1978) model, it is proved that an inverse relationship between the elasticity of substitution and average firm size exists. Their model therefore explains the positive relationship between the importance of SMEs in a country and the openness of the economy. In order to obtain this result, they introduce a normalized CES production technology to treat the elasticity of substitution as an explicit parameter of the model.

To sum up, our analysis of the different approaches to firm growth shows that there is a common aim: to measure firm growth. Moreover, these approaches have one common feature: the nature of the competencies and the process by which they are accumulated is difficult to reconcile with the erratic growth performance displayed by most firms (Geroski, 1999).

As a starting point, we will adopt the stochastic model. However, we will introduce a number of variables to provide a model of learning and selection. This evolution will be explained more thoroughly in the third point from this chapter. However, *"it would be unwise to reject more conventional explanations of market structure out of hand. Economies of scale, government policies, and the like are surely influential and not merely in a random way"* (Scherer, 1970: 130). For this reason next chapters will introduce other internal and external characteristics related to territory, economic growth, and the sector where the firm operates.

2.3. GIBRAT'S LAW

Gibrat (1931) develops a theoretical model to measure the relationship between firm growth and its initial size. Gibrat's Law shows how firm growth depends on random shocks that are independent of each other and on initial firm size. Gibrat's model is written in the following form:

$$\Delta \log S_{i,t} \equiv \log S_{i,t} - \log S_{i,t-1} = \mu_{it} \quad \text{where } \mu_{it} \approx N(0, \delta^2) \quad (2.1)$$

where $S_{i,t}$ is the number of employees working in a firm "i" in a period "t" and μ_{it} is a normally distributed random variable with a mean of zero and a variance of δ^2 .

Equation (2.1) implies that an unexpected shock can occur because: a) we do not know what will happen, or b) we know what will happen but we do not know when it will happen (Geroski, 1999). This means that the process is hard to predict.

This equation also means that the unexpected shocks have permanent effects on the size of the firm. Another way to illustrate this is to decompose each size until the period of creation ("t" = 0):

$$\log S_{i,t} = (1 + \mu_{i,t}) \log S_{i,t-1} = \log S_{i,0} (1 + \mu_{i,1})(1 + \mu_{i,2}) \dots (1 + \mu_{i,t}) \quad (2.2)$$

Rearranging this equation, we get:

$$\log S_{i,t} = \log S_{i,0} + \sum_{s=1}^t \mu_{i,s} \quad \text{if} \quad \log(1 + \mu_{i,s}) \cong \mu_{i,s} \quad (2.3)$$

where the logarithm of the employees working in a firm in a period "t" ($S_{i,t}$) depends on two factors: (i) the initial firm size ($S_{i,0}$) measured in

A Review of the Literature of Firm Growth

terms of the number of employees, and (ii) a set of random terms ($\mu_{i,t}$) which are the same for all the active firms in the market and independent of firm size²³.

Consequently, Gibrat's Law is characterised by a first-order Markov process, which does not imply a serial correlation between the different temporal rates of firm growth (Singh and Whittington, 1975).

Any firm "*i*" is therefore the sum of all the shocks (both expected and unexpected) the firm has received since its creation. Because of the unpredictable nature of these shocks, it is difficult to predict the future firm size.

We should mention certain aspects of random growth rates. First, the growth rates are assumed not to correlate with each other either across firms or over time. Second, the growth rates are assumed to be independent of firm size, which is why we also refer to this proposition as the Law of Proportionate Effect. Third, the random growth rates are normally distributed because random shocks are small effects from many forces.

Gibrat's Law predicts that all firms have the same likelihood of growth, regardless of their initial size. If we extrapolate this result to the future, we see that the market will tend to concentrate because the largest firms will increase their weight in the market. This means that firm size will inevitably become log-normal (right skewed) because of the central limit theorem. Due to random event, firms will eventually diverge in size, and the market concentration will increase even though the firms' growth prospects are still the same.

²³ This equality is accepted as long as the error terms are small enough (Sutton, 1997).

To test this result, the economic literature has estimated Gibrat's Law using various equations. There are three equations in particular that confirm Gibrat's Law.

Firstly, the logarithm of the number of employees belonging to firm "i" during the period "t" ($S_{i,t}$) depends on the logarithm of the number of employees from the previous period ($S_{i,t-1}$):

$$\log S_{i,t} = \alpha + \beta \log S_{i,t-1} + \mu_{i,t} \quad (2.4)$$

Gibrat's Law is accepted as long as coefficient β is equal to 1, so firm growth is independent of initial size. If β is less than 1, the smaller the firm, the higher the growth. If β is more than 1, the larger the firm, the faster the growth.

Secondly, firm growth has also been estimated as a function of initial size. Rather than obtaining the size for the following period, we obtain the growth of the firm during the periods "t-1" and "t" ($\Delta \log(S_{i,t})$):

$$\Delta \log S_{i,t} = \alpha + \beta \log S_{i,t-1} + \mu_{it} \quad (2.5)$$

If Gibrat's Law is satisfied, β will be equal to 0. A positive value implies that larger firms will grow more than smaller firms, so there will be a divergence in firm size. A negative value implies that smaller firms will have a higher growth rate than larger firms, so there will be convergence in the industry.

A Review of the Literature of Firm Growth

Thirdly, there is another dynamic model of firm growth that is linked to the implication of the absence of any dynamics associated with lagged dependent variables:

$$\Delta \log S_{i,t} = \alpha + \beta \Delta \log S_{i,t-1} + \mu_{i,t}$$

where the logarithm of the growth in the period “t” belonging to firm “i” depends on the firm growth rate from the previous period. In this case, Gibrat’s Law is accepted if β is equal to 0. Both equations have an error term ($\mu_{i,t}$) that depends not only on the period of time but that is also individual to each firm.

During the 1950s and 1960s, the literature developed stochastic models to explain firm growth and industrial dynamics. Ijiri and Simon (1974, 1977) introduced firm entries and exits to explain how market distribution evolves over time. These authors define equations that tackle firm distribution depending on the number of employees. Their results show the presence of concavity and a concentration of firms in the market due to autocorrelation of firm growth, new entries and firm acquisitions. In general, their lognormal distributions are right skewed, which means that they are asymmetric with much of the probability mass to the right of the modal value. The upper tail of the firm size distribution is described by the Pareto distribution, which is also known as a power law or scaling distribution²⁴.

²⁴ Ijiri and Simon (1977) estimated the growth equation for N firms and T time periods as:

$$\frac{S_{it}}{S_{it-1}} = \alpha S_{it-1}^{\beta-1} e_{it}$$

where S_{it} is the revenue of firm “i” in period “t”, α is a constant that drives the growth equation, β is the speed of growth, and e_{it} is the error term.

The initial empirical evidence confirmed Gibrat's Law, so the independence of firm growth and initial firm size was confirmed. However, Mansfield (1962) found that Gibrat's Law was only confirmed by firms that had survived the observed period and surpassed the minimum efficient scale (MES). Based on Mansfield's results, the studies described below attempted to determine the correct relationship between firm growth and firm size.

Gibrat's Law will have crucial implications for the evaluation of the firm growth process and economic growth. As well as the capacity of firms to create employment in the short and long runs, policy-makers should take into account the characteristics that enhance the capacity of firm to create new jobs. Furthermore, firm growth is rather important for firm survival.

In the rest of this section we will analyse the expansive body of empirical results on Gibrat's Law and classify them according to whether the Law is accepted or rejected. Furthermore, we classify the literature depending on some characteristics such as the database, the sectorial analysis, the econometric methodology and others. Finally, we will present the Spanish evidence.

2.3.1. The evolution of Gibrat's Law

As such studies are easy to compare, the literature contains many contributions. In general, the results do not confirm the independence between growth and size. Following McCloughan (1995), and based on equation (2.4), we summarize the main results of Gibrat's Law in points 1–5 below.

A Review of the Literature of Firm Growth

1) Gibrat's Law or Law of the Proportionate Effect

The first results of the relationship between firm growth and firm size accepted Gibrat's Law (for example, Mansfield, 1962; Ijiri and Simon, 1974, 1977; Wagner, 1992, 1994). As a result, the vast majority of active firms remain small and there is a scarcity of large firms in the market.

Mansfield (1962) showed that if we include all firms (surviving and non-surviving) in the sample, Gibrat's Law is rejected. If we group firms according to whether they survived or failed during the observed period²⁵, the results do not provide any conclusion about Gibrat's Law. Finally, Gibrat's Law holds when the number of employees in a firm is higher than the minimum efficient scale. Therefore, firms behave differently depending on whether they surpass the minimum efficient size.

Although Mansfield's study confirms Gibrat's Law, the author observes that the variability of firm growth depends on firm size. Specifically, the larger the size, the less variability. Some authors have argued that this is because large firms diversify their portfolio, so they can offset the results of one activity against another.

Ijiri and Simon (1974, 1977) suppose Gibrat's Law to obtain the concavity of the function of firm distribution. That implies the existence of a large number of small firms and a small number of large firms. Mansfield was therefore the first author to raise doubts about the existence of Gibrat's Law.

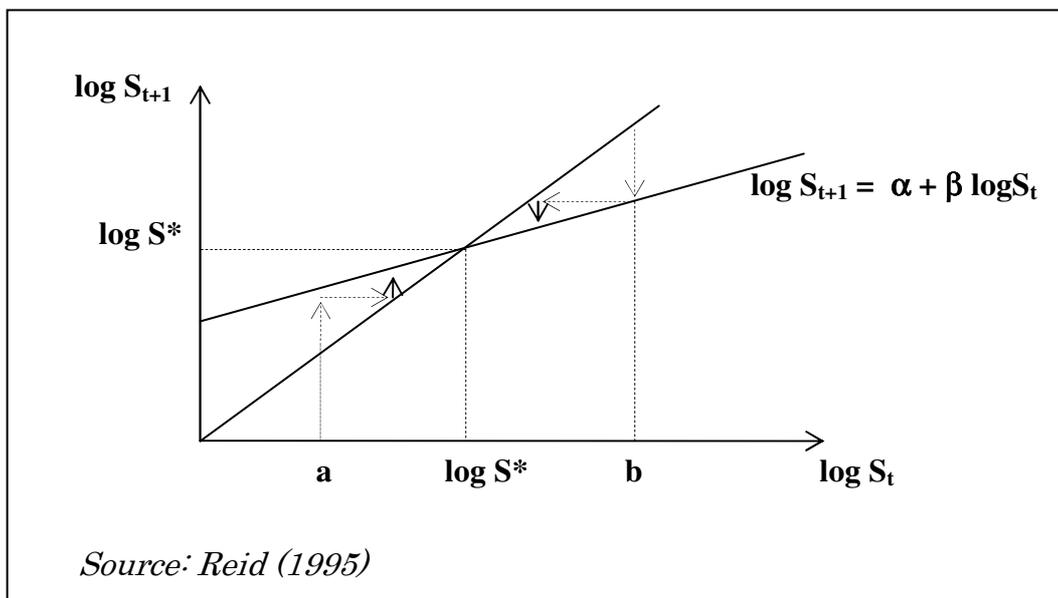
²⁵ The purpose is to avoid the market selection, because there is international empirical evidence which shows that small firms have a higher probability of failure but those that survive have higher growth rates. To prevent the two processes from cancelling each other out, Mansfield (1962) selected surviving firms.

2) Favourable perspectives for small firms

We encountered this situation when the coefficient β ranges from 0 to 1. In this case, small firms grow faster than larger firms up to a limit. This limit is the optimum size of the market. The asymptotic distribution by sizes is still lognormal, as in the previous situation, but the difference lies in the variability in the steady state $\sigma^2 = S^2 / (1 - \beta^2)$. Variability is therefore a multiple constant of firm size.

Evans (1987a, 1987b), Hall (1987), Dunne et al. (1989), Variyam and Kraybill (1992), Mata (1994), Mata and Portugal (2004), Dunne and Hughes (1994), Audretsch and Mahmood (1994a), Harhoff et al. (1998), Hart and Oulton (1999), Fariñas and Moreno (2000) and Lotti et al. (2001) later observed an inverse relationship between size and firm growth.

Graph 2.6 Perspective favourable to small firms



In graph 2.6, Reid (1995) showed that firm size converges to an equilibrium ($\log S^*$), which is the intersection between the 45° line and

A Review of the Literature of Firm Growth

the function that determines the relationship between the initial size (S_t) and the size in the following period (S_{t+1}). The horizontal axis shows the logarithm of firm size in the period “ t ”, and the vertical axis shows the logarithm of firm size in the period “ $t+1$ ”.

In this situation, a firm with a size smaller than the equilibrium (*case a*) will have positive growths but every time its capacity to grow will be smaller. The firm will stop growing when it reaches the optimum size. At this point, the size of the firm will remain constant. When it is higher than the optimum size (*case b*), the firm will have negative growths and reduce its size until it reaches equilibrium. The process of adjustment in sizes between one period and another is indicated by arrows. Once the firm is in equilibrium, the average firm size is obtained from the following equation:

$$\begin{aligned} \log S_{t+1} &= \alpha + \beta \log S_t \\ \text{if } \log S_{t+1} &= \log S_t, \text{ ie. } \log S^* \\ \log S^* &= \alpha + \beta \log S^* \\ \log S^* &= \frac{\alpha}{(1-\beta)} \\ S^* &= \exp\left(\frac{\alpha}{1-\beta}\right) \end{aligned}$$

Other contributions include that of Dunne et al. (1989), who also assumed Jovanovic (1982)’s theoretical framework with heterogeneous firms. Firm growth (g) is as follows (where S is the number of employees):

$$g'_t = \frac{(S_{t+1} - S_t)}{S_t}$$

This means that firm growth is the variation in the number of employees in the period “ $t+I$ ” with respect to the period “ t ”, divided by the number of employees in the initial period “ t ”.

Dunne et al. (1989) also defined a set of functions related to the probability function of growth:

- a) Probability function of density: $j(g'|x)$ belongs to firms with a set of characteristics x . This function represents the rates of the potential firm growth.
- b) The difference between potential growth and achieved growth is due to the existence of a critic value of size growth: g^* . If the firm has growth rates below this critic value, the firm will disappear and its growth will be recorded as -1. If the firm has growth rates above this critic value, its growth will be the same as the expected value.

$$\begin{cases} g_t = g'_t & \text{if } g'_t \geq g^* \\ g_t = -1 & \text{if } g'_t < g^* \end{cases}$$

- c) Since there are differences between potential growth and real growth, the real density function will include firms that disappeared before “ $t+I$ ” and will be $f(g|x) \in [-1, \infty)$. This function represents the interaction between the potential growth distribution and the exit condition from the market.

A Review of the Literature of Firm Growth

- d) The density function belonging to the surviving firms depending on its growth is represented by $h(g|x)$, which is equal to $f(g|x)$ but with a probability of 0 for firms with growth equal to -1 .

Dunne et al. (1989) represented the average growth as:

$$\mu_h(x) \equiv \frac{\left(\int_{-1}^{\infty} gh(g|x)dg \right)}{P_s(x)} \quad \text{where} \quad P_s(x) \equiv \int_{-1}^{\infty} h(g|x)dg$$

and where P_s is the probability that a firm does not disappear.

$$\mu_f(x) \equiv \int_{-1}^{\infty} gf(g|x)dg = \mu_h(x)P_s(x) - (1 - P_s(x))$$

This means that the average size growth (μ_f) coincides with the average growth probability of the surviving firms (μ_h) minus the average size growth probability of the non-surviving firms.

Dunne et al. (1989) concluded that the deviations of μ_h depending on “ x ” are a good sign of how μ_f changes with “ x ”, as long as failure rates do not change with “ x ”.

The previous definitions of density functions are similar to those of Jovanovic (1982), where firms maximize their profits and have to determine their production (q). This production depends on random shocks to the efficiency levels. Efficiency affects costs, so the cost function is: $\gamma(q_t)\theta(c + \varepsilon_t)$, where $\gamma(q_t)$ is a strictly convex cost function and where

$\theta(c + \varepsilon_t)$ is a multiplier that reflects different levels of efficiency. The properties of this multiplier are:

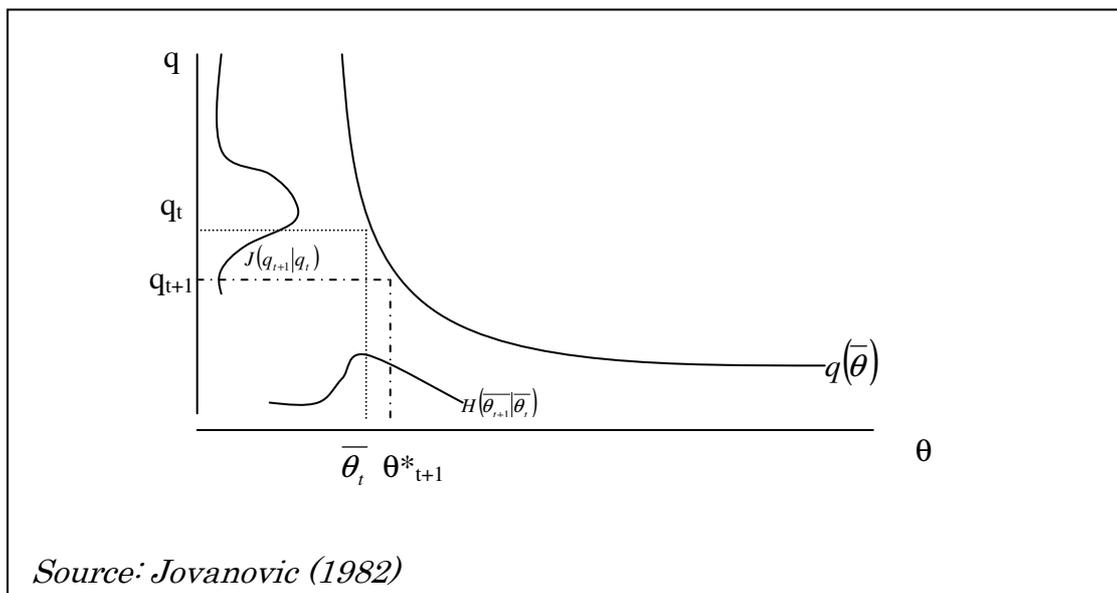
$$\lim_{z \rightarrow 0} \theta(z) = \vartheta_1 > 0 \quad \text{and} \quad \lim_{z \rightarrow \infty} \theta(z) = \vartheta_2 < \infty.$$

As the efficiency levels of the following period are unknown, there must be some hypotheses from a probability function $H(\bar{\theta}_{t+1} | \bar{\theta}_t)$ of the expected values in “ $t+1$ ” from variable θ where $E(\bar{\theta}_{t+1} | \bar{\theta}_t) = \bar{\theta}_t$.

Maximizing the production function gives a decreasing function from the maximizing output $q_t(\bar{\theta}_t)$. This means that firm production will decrease if the expected efficiency level is not the same as the real level.

Given the previous functions (the probability function of efficiency levels, $H(\bar{\theta}_{t+1} | \bar{\theta}_t)$, and the maximizing production function $q_t(\bar{\theta}_t)$), we can obtain the density function $J(q_{t+1} | q_t)$ of the production.

Graph 2.7. Maximization output function and distribution of efficiency



A Review of the Literature of Firm Growth

Graph 2.7 represents the function that maximizes the output ($q(\bar{\theta})$), where the independent variable is the efficiency level. If we draw the density function of the different efficiency levels ($H(\bar{\theta}_{t+1}|\bar{\theta}_t)$), and represent the efficiency level that maximizes the output ($\bar{\theta}_t$), we can obtain the value of the production function (q_t), which coincides with the maximum output from the density function of production ($J(q_{t+1}|q_t)$).

What would happen if our prediction does not correspond to reality i.e. if the efficiency level in the period “ $t+1$ ” is below the expected one (θ^*_{t+1} is higher than $\bar{\theta}_t$ and we therefore have higher costs)? The production would be smaller than the production for the previous period.

Therefore, Dunne et al. (1989)’s model explains why the efficiency and production levels are different. The reason is that expected efficiency levels may not correspond with the real level of efficiency. This mismatch produces that the expected and the final production of a firm will be different.

3) Perspectives favourable to large firms

One branch of the literature has found evidence of perspectives that are favourable to large firms. This implies that the value of β will be larger than 1. In this situation we have a lognormal function of distribution but there is a difference in the expression of the central limit theorem:

$$\beta^t y_t(0) \rightarrow \infty \text{ when } t \rightarrow \infty$$

The function therefore diverges and the inequalities remain. In this case, industrial concentration into a small number of firms is inevitable.

There is a scarcity of empirical evidence for this situation, though Singh and Whittington (1975) found that large firms grow faster. This means that as firms grow at a higher rate than unity ($\beta > 1$), market concentration of all the firms in the industry increases. These results may contain errors, however, since they have been estimated with ordinary least squares or even with a measure error of the initial firm size. Keating (1974) also found that Australian financial firms grow faster than smaller ones. Obviously, however, the scale of financial service industries is large (Audretsch et al., 2004).

4) Serial correlation of growth (Chesher, 1979)

The above results involved the relationship between current growth and past size. Another branch of the literature has studied the relationship between past firm growth and future firm growth i.e. whether firms that grew in the past will have a greater probability of growing in the future.

A certain relationship between these different growth rates has been found. Singh and Whittington (1975) began the analysis of growth persistence by introducing past growth to the initial Gibrat's Law. Their results show that firm size was not as important as the past growth. The analytical description of growth persistence was developed by Chesher's seminal work in 1979. The following equations determine the relationship:

$$\begin{aligned}y_{i,t} &= \beta y_{i,t-1} + u_{i,t} \\u_{i,t} &= \gamma u_{i,t-1} + \eta_{i,t}\end{aligned}\tag{2.6}$$

A Review of the Literature of Firm Growth

where $y_{i,t}$ is the size of firm “ i ”, β is the impact of firm size “ i ” in the period “ $t-1$ ”, η_{it} is a white noise, and γ measures the transmission of luck or success from “ $t-1$ ” to the following period and has a value of between 0 and 1. When γ is close to 1, the reiteration of past situations is greater. If we substitute reiteratively the last two equations until we reach the initial period, we obtain the following equation:

$$y_{i,t} = \beta^t y_{i,0} + \gamma \sum_{\tau=1}^t \beta^{t-\tau} u_{i,\tau-1} + \sum_{\tau=1}^t \beta^{t-\tau} \eta_{i,\tau} \quad (2.7)$$

This new equation is still lognormal because u and η are normal. So, for any β , the higher γ is the faster size inequalities will increase.

Chesher’s main conclusion is that serial correlation between error terms in the equation produces a dependence between past size ($y_{i,t-1}$) and the error terms ($u_{i,t}$). Since Chesher’s study, growth persistence has been further studied.

Growth persistence in one period with respect to another may produce a serial correlation. Moreover, with the ordinary least squares this serial correlation produces inconsistent estimators. The previous results, for example, would be biased downwards so small firms would grow at a higher rate (Dunne and Hughes, 1994).

In general, the proposed model determines the relationship between current firm growth and its past growth. If we express equation 2.6 in logarithms, we obtain the following expression:

$$z_{i,t} = \beta z_{i,t-1} + u_{i,t}$$

$$u_{i,t} = \gamma u_{i,t-1} + \eta_{i,t}$$

Where $z_{i,t}$ is the logarithm of firm growth belonging to firm “ i ” between “ t ” and “ $t-1$ ”. If we join the two equations, we get:

$$z_{i,t} = \beta z_{i,t-1} + \gamma u_{i,t-1} + \eta_{i,t} \quad (2.8)$$

If

$$z_{i,t} = \beta z_{i,t-1} + u_{i,t} ,$$

by isolating the error term $u_{i,t}$:

$$u_{i,t} = z_{i,t} - \beta z_{i,t-1} \quad (2.9)$$

and substituting 2.9 in 2.8, we get:

$$z_{i,t} = \beta z_{i,t-1} + \gamma \left(z_{i,t-1} - \beta z_{i,t-2} \right) + \eta_{i,t}$$

Finally;

$$z_{i,t} = g_1 z_{i,t-1} + g_2 z_{i,t-2} + \eta_{i,t} \quad (2.10)$$

where,

$$g_1 = \beta + \gamma$$

$$g_2 = -\beta * \gamma$$

We found a relationship between Chesher’s equation and Gibrat’s Law in the coefficients g_1 and g_2 of equation 2.10. Specifically, Gibrat’s Law postulates that the value of β is 1 and that the value of ρ is 0.

A Review of the Literature of Firm Growth

Empirical evidence in the literature has confirmed the presence of growth persistence. Therefore, firms that grew in the past will also grow in the future. Singh and Whittington (1975), Kumar (1985), Contini and Revelli (1989) and Wagner (1992, 1994) obtained a significant and positive sign for change persistence. Dunne and Hughes (1994), Almus and Nerlinger (2000) and Vennet (2001), on the other hand, found a non-significant relationship between growths in different periods, which indicates that firm growths are not serial correlated.

In the empirical research, authors analysing growth persistence have introduced different explanatory variables. Wagner (1994), for example, showed that macroeconomic conditions during entry do not influence future evolution, while industrial characteristics are not excessively important. Their results do not show growth persistence and they remembered the *Brown-Hamilton-Medoff Warning*: “*Do not judge firms by their size alone!*”. Vennet (2001) introduced variables such as macroeconomic growth, bank operational efficiency, the quality of credit and capitalization. He observed that between 1985 and 1989, there was a convergence of bank size, but between 1990 and 1994 there was a proportional growth by all banks.

The major contribution from growth persistence models is the introduction of the dynamic firm evolution. In other words, if we want to know how firms will grow in the future, we must not only analyse firms currently but also analyse how they grew in the past.

5) Variability of firm growth

In the 1980s new lines of research related to Gibrat's Law were opened. Sutton (1997) reported two main subjects. Firstly, economists took into consideration econometric problems such as sample censorship, the

functional function and the heterogeneity of firm behaviour (Hall, 1987; Evans, 1987a, 1987b; Dunne et al., 1989). Secondly, the presence of variability of firm growth.

The heterogeneity of the variability of firm growth produces heteroscedasticity in ordinary least square regression (inefficient estimators). The differences in firm growth due to firm size is caused by the fact that large firms are more diversified in products and markets, so reductions in some products and markets are compensated by others (Bottazi et al., 2002).

However, Dunne and Hughes (1994) pointed out that heteroscedasticity is not due to different firm sizes but to firm age. Age is the variable that introduces heteroscedasticity because the majority of young firms are small. These authors introduced the age of the firm into Gibrat's equation and incorporate the estimated error terms adjusted by heteroscedasticity:

$$\log S_{i,t} = \alpha + \beta \log S_{i,t-1} + \gamma \log A_{i,t} + e_{i,t}$$

Simultaneously, there is an evolution of theoretical models, which try to emphasize the empirically proven stylised facts. More importantly, they introduce stochastic growth into the traditional maximisation models (Jovanovic, 1982; Jovanovic and MacDonald, 1994).

Another characteristic we should point out is the introduction of firm characteristics that influence the profit function. The main consequence of this is that firm growth is not stochastic between the smaller firms (Sutton, 1997).

2.3.2. Empirical evidence of Gibrat's Law

Many studies have analysed the relationship between firm growth and firm size. In this section, therefore, we will present some of the evidence of firm post-entry performance. We will analyse the empirical evidence from several perspectives: the type of estimation, the analysed sectors, the geographical scope and the econometric methodology²⁶.

- **Type of estimation**

As many studies have involved Gibrat's Law, some kind of classification is needed. Economists have used two approaches to assess the contribution of firm size to firm growth. In accordance with Piergiovanni et al. (2002) and Audretsch et al.'s (2004) classification, we will divide the empirical evidence into static and dynamic analyses.

Static analyses relate firm growth to previous firm size. The relationship between these variables is observed by two methods. The first method involves dividing firms into categories that depend on their initial size and then examining whether firm growth rates are equally distributed between categories. This kind of empirical evidence is used by Mansfield (1962), Hymer and Pashigian (1962), Singh and Whittington (1975), Acs and Audretsch (1990) and Audretsch et al. (2004). To construct these categories, firms are ordered by their initial size and divided into quartiles. The firm growth rates belonging to each quartile are then calculated. Finally, if the growth rates are not significantly different between the groups, Gibrat's Law is supported. Econometric methods may also be suitable for analysing the relationship between firm growth

²⁶ See Annex III to a more detailed classification of the literature.

and initial firm size and, in fact, this is the most common way to apply the equation (2.5).

The second method is based on the Chesher's (1979) contribution, where the growth persistence indicates that Gibrat's Law is accepted. If growth turns out to be an autocorrelated process, on the other hand, Gibrat's Law is not accepted. Singh and Whittington (1975), Kumar (1985), Almus and Nerlinger (2000) and Lotti et al. (2001) used this method.

Both categories can be subdivided into three groups according to the sample they analysed. Following Mansfield's (1962) classification, these groups are:

Type 1: Surviving and non-surviving firms regardless of their initial size.

Type 2: Surviving firms only. This avoids the bias caused by the non-surviving firms.

Type 3: Surviving firms that surpass the minimum efficient size. There are behavioural differences between firms that surpass the minimum efficient size and firms that do not because the latter do not have scale economies that encourage them to increase.

To sum up, there are 6 classifications depending on sample selection and the method used to analyse firm growth behaviour:

	STATIC	DYNAMIC
<i>Surviving and non-surviving firms</i>	Static - 1	Dynamic - 1
<i>Surviving firms</i>	Static - 2	Dynamic - 2
<i>Surviving firms that surpass the minimum efficient size</i>	Static - 3	Dynamic - 3

A Review of the Literature of Firm Growth

Below we analyse the empirical evidence that agrees with Mansfield's (1962) classification.

For the static version that includes surviving and non-surviving firms (Static-1), we find Mansfield (1962), Acs and Audretsch (1990) and Stanley et al. (1995). The acceptance of Gibrat's Law is the main result with this type of estimation. However, these results are biased by sample censorship since they "*suppose that small firms with low growth rates are more likely to exit, then the proportional rate of growth, condition on survival, will be smaller for large firms*" (Sutton, 1997).

A second static version analyses surviving firms regardless of their size (Static-2). Mansfield (1962) interpreted Gibrat's Law differently depending on how the firms disappear from the sample.

Mansfield (1962), Variyam and Kraybill (1992), Evans (1987a, 1987b), Dunne and Hughes (1994) and Botazzi et al. (2001) included only surviving firms. Their results reject Gibrat's Law because of the introduction of small firms. These are smaller than the minimum efficient scale and tend to increase at a higher rate than large firms.

A third line of research, developed by Mansfield (1962), includes surviving firms whose size larger than the minimum efficient scale (Static-3). This version tends to accept Gibrat's Law (Mansfield, 1962; Wagner, 1994), whereas other studies tend to reject it ²⁷.

The fourth line of research is the dynamic version in which the whole pool of firms (surviving and non-surviving) is estimated regardless of their size (Dynamic-1). This version introduces past firm growth to

²⁷ Singh and Whittington (1975) showed that larger firms grow faster, while Hall (1987) and Faggio and Konings (1999) presented results where smaller firms grow faster

analyse its impact on current firm growth and contrast Gibrat's Law. Some of the main estimations are by Mansfield (1962), Contini and Revelli (1989), Wagner (1992), Fariñas and Moreno (2000), Vennet (2001), and Lotti et al. (2003). In general, their main results reject Gibrat's Law in favour of faster growth by smaller firms.

The fifth line of research involves the Dynamic-2 group of firms, which contains surviving firms regardless of size. Contributions include those from Hart and Prais (1956), Singh and Whittington (1975), Chesher (1979), Kumar (1985), Wagner (1992), Audretsch (1995a), Geroski et al. (2000) and Piergiovanni et al. (2002). Their main findings are the existence of a positive relationship between past and current firm growth. Hence Gibrat's Law is rejected. Note, however, that the results of Lazarova et al. (2003) accept Gibrat's Law for firms that survived thirty years.

Finally, Dunne et al. (1988, 1989), Mata (1994), Santarelli (1997), Almus and Nerlinger (2000) applied a dynamic estimation to surviving firms that are larger than the minimum efficient scale (Dynamic-3). Their main results reject Gibrat's Law. One exception is the contribution from Santarelli, who accepted it in 14 out of 20 Italian regions.

- **The measure of firm growth**

Heshmati (2001) analysed Gibrat's Law with sales, assets and employees. His results depended on the variable that was taken into account. This author found a negative relationship between firm growth and size (the number of employees) and a positive relationship between firm growth and sales, but found no relationship between firm growth

A Review of the Literature of Firm Growth

and assets. The positive relation between firm growth and sales is due to scale effects on sales.

However, the results of testing Gibrat's Law using different measures of size and growth rate depend on the time period (Kirchhoff and Norton, 1994). One of the most often used variables is therefore the number of employees, since this is not as volatile as sales or as rigid as productive capacity.

Obviously, the concept and measurement of firm growth can be looked at from several perspectives. We believe, however, that the number of employees is a much more stable variable for measuring firm growth since it avoids some of the above disadvantages. However, the performance of each variable will depend on the industry analysed (see section 2.2.2).

- **Sector**

Since the first studies of firm growth, researchers have focused mainly on the manufacturing sector (Simon and Bonini, 1958; Evans, 1987a, 1987b; Dunne et al., 1989; Mata, 1994; Fariñas and Moreno, 2000) and on one or more sectors differentiated by the degree of labour or technology intensity, etc. Examples of such studies in the literature are those by Mansfield (1962), who analysed the steel, petroleum and tyre industries, FirtzRoy and Kraft (1991), who analysed the metallurgic industry, Das (1995), who analysed the computer hardware industry, Botazzi et al. (2001), Botazzi et al. (2002) and Fabritiis et al. (2003), who analysed the pharmaceutical industry, Lotti et al. (2003), who analysed the telecommunication, radio and TV equipment industries, and Scherer et al. (2000), who analysed high technological firms in the United States.

The results of these early studies, which introduced variables such as age and size, were that small and new firms had an above-average growth potential (Evans, 1987a, 1987b; Hall, 1987; Dunne et al., 1989). This is because firms have to reach a size that enables them to exist in the market i.e. the minimum efficient scale (MES) of production. This MES varies from sector to sector, thus reflecting the heterogeneous behaviour of the production function.

Small firms operating in industries with a high MES should have a higher propensity to grow, since crossing the MES threshold ensures that the firm is large enough to survive. Audretsch (1995b) found a positive relationship between the MES and growth for various industries. We should therefore expect active firms in the service industries to behave differently from those in the manufacturing industries.

Service industries, however, have been largely ignored (Delmar, 1997). Some authors have placed firms in service industries in the same pool as firms in manufacturing industries. Clearly, as the service sector has different characteristics, joining the service sector with the manufacturing sector may lead to biased results since the pattern of firm growth is heterogeneous and it is difficult to observe service firms, which are usually smaller. In fact, Sutton (1997) analysed the evolution of firm growth applying states that within markets and submarkets there are exponential bounds to firm size, with parametric heterogeneity across markets and industries. Therefore, the analysis of different market would be justified in order to detect the heterogeneous behaviour of each industry.

Singh and Whittington (1975) studied the distribution of manufacturing and service industries. Chesher's (1979) sample for dynamic estimation

A Review of the Literature of Firm Growth

includes firms in both the service and manufacturing industries. Variyam and Kraybill (1992) also estimated the service and manufacturing industries together. Kumar (1985) and FirtzRoy and Kraft (1992) introduced a limited number of firms from the service industry in a pool of manufacturing firms. The manufacturing industry is analysed together with the construction and service industries by Harhoff et al. (1998).

Tschoegl (1996) and Vennet (2001) analysed Gibrat's Law with respect to the banking sector, while Santarelli (1997) analysed the hospitality sector. More recently, Piergiovanni et al. (2002) analysed the service industries. Finally, Audretsch et al. (2004) analysed the hospitality sector.

More recently, Oliveira and Fortunato (2004a, 2004b, 2005) and Bottazzi et al. (2006) analyse the possible heterogeneity between manufactures and service industries. Both authors determine that there are not differences between both industries at the aggregate level.

- **Countries and data bases**

There is a clear relationship between the frequency of studies in a country and the availability of large data bases.

Some studies, such as those by Fabritiis et al. (2003) using the *Pharmaceutical Industry Database* and Faggio and Konings (2003) using the *AMADEUS database*, analysed worldwide statistics.

However, most studies have concentrated on the United States because of the availability of large data bases (the *Small Business Data Base*, the

Census of Manufactures, the United States Establishment Longitudinal Microdata, Compustat, the OneSource database and the LEEM registry).

The United Kingdom also has a great deal of literature testing Gibrat's Law both with general data bases (the *Company Accounts Databank* and the *EXSTAT database*) and specialised data bases (the *Synthesis Life Database*).

Other European countries, such as Germany (*CREDITREFORM from the ZEW Foundation Panel*), Italy (*National Institute for Social Security and Mediocredito survey*), Portugal (*Ministry of Employment and Central Balance Sheet Office*) and Spain (*Encuesta sobre Estrategias Empresariales*) have also analysed the firm growth pattern. However, the Spanish studies are contemporary since until recently there was no tradition in Spain of compiling statistical information about firms. Spain therefore no data base with which to study firm growth in the long term.

There are also empirical contributions from non-western countries such as Japan (*Japan's Expanding US Manufacturing Presence by the Japan Economic Institute and Nikkei Kaisha Jouhou*) and India (*Dataquest*).

- **The time period**

Depending on their statistical tradition, some countries have more longitudinal data bases than others. Researchers usually focus on the previous decade and a time period of between 5 and 10 years. Some authors e.g. Amaral et al. (1997) and Tschoegl (1996) used data bases for the previous two decades. The longest data base is the one used by Hart and Prais (1956), which analysed the period between 1885 and 1950.

A Review of the Literature of Firm Growth

- **Econometric methodology**

The econometric methodology for testing Gibrat's Law has evolved over the years. The first method used (e.g. in Mansfield, 1962 and Hymer and Pashigian, 1962) was ordinary least squares.

In the 1980s researchers were interested in new issues such as sample censorship, the appropriate functional relationship and heteroscedasticity (Sutton, 1997). Specifically, Hall (1987), Evans (1987a, 1987b) and Dunne et al. (1989) analysed all these issues and found a double effect: large firms increase less but are more likely to survive than small firms.

Lazarova et al. (2003) used methods such as the Augmented Dickey-Fuller test to analyse the relationship between the logarithm of current firm growth and the logarithm of past firm growth²⁸. Scherer et al. (2000) and Goddard et al. (2002) analysed the relationship between initial firm investment and final firm investment using a Monte Carlo experiment. Fariñas and Moreno (2000) and Fotopoulos and Louri (2004) estimated kernel non-parametric density functions and quantile regressions to test Gibrat's Law.

Finally, Heshmati (2001) compared results obtained with ordinary least squares, generalised least squares and adjustment models, as well as within and between estimations. Del Monte and Papagni (2003) and Vennet (2001) applied a data panel to estimate Gibrat's Law.

²⁸These authors used an Augmented Dickey-Fuller test with a lagged dependent variable, with and without trend, to control the interdependence between firms.

Das (1995) estimated an OLS equation and panel data with fixed and random effects and compared the results in order to find the most suitable estimation.

Some authors have used the unit root test to estimate the relationship between firm growth and initial firm size. Oliveira and Fortunato (2003a) used a unit root test for a panel data of Portuguese manufacturing firms and showed that Gibrat's Law was not satisfied because of the existence of a unit root.

Other authors have recently emphasized the problem of endogeneity due to time-varying factors that are not included in the estimation but are correlated with the explanatory variables and with growth. To solve this problem, the Generalised Method of Moment (GMM) proposed by Arellano and Bond has been used. Nkurunziza (2005)²⁹ found convergence in size with all econometric methods except GMM equations. Oliveira and Fortunato (2004a) estimated Gibrat's Law for manufacturing and service industries using a GMM estimator and rejected Gibrat's Law in all cases because they found that smaller firms grew faster than larger ones.

More recently, Calvo (2006), Nkurunziza (2005), Niefert (2005) and Santarelli and Vivarelli (2002) used Heckman's equation to control for attrition bias. This method is useful for observing whether firms survive and for analysing the post-entry performance of the surviving firms in order to determine whether firms that survive have a greater propensity to grow than those that did not. Heckman's (1979) equations help to control the selection bias caused by the interdependence between past and present performance.

²⁹ Nkurunziza (2005) studied the African manufacturing industries.

A Review of the Literature of Firm Growth

Botazzi and Secchi (2005) applied the unit root test for the analysis of the pharmaceutical companies. Recently, Goddard et al. (2006) applied the Augmented dickey Fuller autoregressions to analyse the mean reversion of profit rates. Their results show that there is evidence of the existence of mean reversion among profit rates when controlling for individual heterogeneity. While there is contradictory results among the different tests applied to the mean size of the firm (measured in net assets).

Many studies have therefore used panel data econometric methods to estimate the relationship between firm size and firm growth. Recent contributions, however, have included methods to take into account the selection process of firms with more capacity to grow than firms which disappear from the market.

2.3.3. Spanish literature

Most studies, because of the available data bases, have analysed countries such as the United States, Germany and the United Kingdom. Some Spanish contributions have been made by Suárez (1977), Pisón (1983), González (1988), Fariñas and Moreno (2000), Correa et al. (2003) and Calvo (2006).

The Spanish empirical and theoretical contributions are scarce, however. This is not because the topic is unimportant but because of the lack of data. One of the challenges is therefore to improve the data bases. Extending the summary of the literature conducted by Correa (1999) and Correa et al. (2003), here we review some of the most outstanding studies.

Maravall's (1976) contribution was the first Spanish empirical estimation of Gibrat's Law. Previous studies had compared the distribution of firms in the market. The main aim of this study was to show the characteristics of the market structure. The relationship between the growth and size of manufacturing firms between 1964 and 1973 was analysed. The data-base comprised 254 Spanish firms from "Las grandes empresas industriales en España" published by the Spanish Department of Industry. All the firms survived during the period. Added value and sales were used as indicators of size. In this study Gibrat's Law was satisfied for medium-sized firms (sales of between 3 and 12 million euros), while the smallest and largest firms grew more rapidly.

However, analysis of variance showed a negative relationship between size and variance. This means that firms with a large number of workers, probably because of the diversification process, obtained a smaller variance. As Gibrat's Law assumes that the variance of firm growth is independent of firm size, Gibrat's Law was rejected. Moreover, study of the matrix of probabilities showed that the mobility of firms is scarce.

In another study, Suárez (1977) analysed Gibrat's Law for 46 firms between 1962 and 1972. The data came from "Agendas Financieras" provided by the Bank of Bilbao. The sectors included in the study were electrical, water and gas manufactures, real estate, chemical and textile manufactures, and miner manufactures.

The variable used to measure firm growth was total net assets. Suárez's (1977) results satisfied Gibrat's Law. However, as he pointed out, the small number of observations may have produced biased results.

A Review of the Literature of Firm Growth

In a third study, Pisón (1983) focused on large manufacturing firms in Galicia between 1975 and 1980. The sample comprised only 18 firms, which showed how difficult it is to obtain data and how few large firms there are in this Spanish region. As the small sample may have led to biased results, these should be viewed with caution. Gibrat's Law was not rejected because there was no relationship between profitability and size.

Recently, González (1988b) analysed Gibrat's Law for the Catalan textile industry between 1973 and 1983. This sample comprised 283 small firms and firm growth was measured in terms of the number of employees³⁰. The main conclusion was that Gibrat's Law was satisfied. This result contradicted that of a previous study by the same author (González, 1988a)³¹. However, the previous study was based on a sample of large banks. Clearly, as the characteristics of the textile and banking sectors are very different, the results are also expected to be different.

In 1983 Lafuente and Salas tested Gibrat's Law for firms present in the Bank of Bilbao between 1972 and 1978. Their results depended on the time period of time analysed. When the whole period was considered, Gibrat's Law was satisfied. For the period between 1974 and 1976, however, the relationship between profitability and size was significantly negative.

Fariñas and Rodríguez (1986) tested Gibrat's Law for the largest firms in the European Union. The sample comprised the largest 100 surviving firms among *Europe's 10,000 Largest Companies*. The results showed a

³⁰ Correa (1999) pointed out that measuring firm growth in terms of the number of employees introduced a certain bias because technological differences during the analysed period were omitted.

³¹ In this study Gibrat's Law was rejected because there was a negative relationship between size and growth.

significant negative relationship between firm growth and size for Spanish firms. For the firms from other European countries, however, Gibrat's Law was accepted. The rates of growth and profitability of Spanish firms were lower than those of the other European firms, though the causes were not related to differences in size.

The PhD thesis of Correa's (1999) analysed the relationship between the growth and size of 1,278 non-financial firms from Santa Cruz de Tenerife. The data were provided by the "Central de Balances" of the University of La Laguna, which collects the data of all firms that present their Annual Accounts to the Mercantile Register. The period analysed was between 1990 and 1996.

Her results showed that small firms grew faster than large ones, so Gibrat's Law was rejected. She suggested that these results may have been due to the greater ability to adapt of small firms and their lower initial investment.

The above study also analysed the learning model of introducing the age of the firm. The main results were an inverse relationship between firm growth and age for microfirms. Another conclusion was the importance of insularity: the insularity of the Canary Islands means that firms have a greater capacity for future growth.

In an article published in the *Review of Industrial Organization* in 2000, Fariñas and Moreno applied a non-parametric approach to analyse firm growth toward a mean size and showed that regression towards the mean does not justify the existence of a negative relationship between size and firm growth rates. The variable used to measure firm growth was the number of employees.

A Review of the Literature of Firm Growth

Their unbalanced panel comprised 1,971 manufacturing firms from the *Encuesta sobre Estrategias Empresariales* (ESEE). The main contributions were:

- a) They solved the problem of selection or censorship using Dunne et al.'s (1989) method.
- b) They applied non parametric methodology to analyse firm growth rates and survival likelihood. First they applied a standard regression and then applied a model based on estimates of kernel.
- c) They introduced heterogeneity in the firm, autocorrelation in the variance of growth and failure equations. They also introduced a linear model of failure probability.

Although their results show an inverse relationship between firm growth and size, this may also indicate a regression to the mean. Regression to the mean may be due to transitory or temporary fluctuations.

To avoid the fallacy to-the-mean³², they controlled firms' boundary-crossing from one size group to another by introducing dummies. Their results showed that the growth rates of surviving firms diminished sharply in the first three size groups. Consequently Gibrat's Law was rejected.

More recently, Correa et al. (2003)³³ examined the factors affecting the growth of small and medium-sized firms. In their study the main variables (size, age and economic activity) were used to determine the

³² The fallacy to-the-mean rises whenever measurement error introduces transitory fluctuations in observed size.

³³ This article is the publication of Correa's (1999) PhD Thesis.

existence of Gibrat's Law and the Learning Theory. The results showed that firms located in the Canary Islands have a higher propensity to grow than in the rest of Spain, probably because of the economic features of that region. They found a negative relationship between firm growth and size, while age produces ambiguous effects. Moreover belonging to a tertiary sector was not a significant explanatory variable for firm growth.

Peña (2004) analysed the survival and growth of incubating ventures in the Basque Country. The sample contained 114 start-ups that eventually became firms. The author described a firm growth model based on the entrepreneur's human capital, the firm's resources and strategies, and incubation variables. Firm growth is measured in terms of sales and employees³⁴. The relationship between firm growth and initial size is "puzzling": the larger the initial investment, the more negative the sales growth. Policy implications are that the enhancement of human capital and firm resources appears to be a more important goal than other policy strategies.

More recently, Calvo (2006) analysed the existence of Gibrat's Law for 1272 manufacturing firms from *Encuesta sobre Estrategias Empresariales* between 1990 and 2000. His results show that Gibrat's Law is not accomplished in favour to small firms and that the innovating behaviour has a positive impact in the survival likelihood and the firm growth. Since his database includes surviving and non-surviving firms, he controls for the selection bias estimating a survival likelihood function. Afterwards, he analysed Gibrat's Law with the Heckman's equation and a maximum likelihood estimator.

³⁴ Peña (2004) excluded profit growth because the explanatory power is low. This justifies his decision because profits are not an accurate measure of venture growth in the first 3-4 years.

A Review of the Literature of Firm Growth

The scarcity of observations has been the main characteristic of the studies in the last few decades. The only exceptions are the contributions by Calvo (2006) and Fariñas and Moreno (2000), which include Spanish manufacturing firms.

2.4. SUMMARY AND CONCLUDING REMARKS

Firm growth has been extensively studied. However, the variety of theoretical approaches in the literature reveals the complexity involved in studying this process. The difficulties in determining the factors as well as the consequences in the distribution of firms in the market are the main characteristics of the literature review.

This chapter is a review of the literature on Gibrat's Law. First we analysed the main approaches to tackling the issue of firm growth and then focused on Gibrat's Law. The evolution of the theoretical literature and the empirical evidence are the main aspects of this chapter. We have made a special reference to what little Spanish evidence there is.

Clearly, results since Mansfield (1962) have not been conclusive. Authors such as Das (1995) have pointed out that this heterogeneity may be due to the different industries analysed. Each industry has different technologies, and perhaps different growth processes, which might explain the mixed nature of the results (Das, 1995).

Gibrat's Law relies on an explanation of the firm growth process that depends on a stochastic process. Later models have attempted to be more accurate by introducing variables such as age in order to explain the trends in the growth of firms during the different phases of their life

cycle. Obviously, models try to explain firm behaviour in a stylized way and may sometimes be too unrealistic or too rigid. Empirical evidence has solved this lack by introducing different variables into the analysis of firm growth.

The most important conclusions of this chapter are as follows.

- Firm growth is a complex process that is affected by internal and external characteristics.
- Several theoretical approaches have analysed this process. However, the economic predictions depend on the economy concerned.
- Gibrat's (1931) Law, or the Law of Proportionate Effects, seems to be a good approach to the unequal distribution of firms in the market.
- The initial empirical studies seemed to accept Gibrat's Law (Ijiri and Simon, 1977). Lately studies (Mansfield, 1962) have been contradictory, however.
- These contradictory results were solved with Mansfield's (1962) article. The probability of growth is conditioned by firm survival and firm size. Surviving firms whose size is lower than the minimum efficient size will not satisfy Gibrat's Law, whereas surviving firms whose size is greater than the minimum efficient size will.
- The results and rate of growth depend on the country concerned, available databases, time period, etc.

This review of the literature has focused on firm growth. However, entry, exit, likelihood of survival and growth are different phases of the same process: the life cycle of a firm. Like live organisms, firms are born with certain characteristics and according to the conditions of a market. All these factors affect their growth and their survival or exit from the market. Future research should therefore consider the theoretical

A Review of the Literature of Firm Growth

perspective and the empirical evidence of all the phases in the firm life cycle that affect firm growth.

In the last fifty years we have seen the world's advanced industrial economies shift from primarily industrial to primarily service economies. After several technological revolutions, the service sectors have gained weight in national production. However, the literature on firm growth has not evolved at the same speed—sometimes for lack of information or interest—and the number of studies of service industries has not matched their economic development. In fact, it is important to note the lack of empirical studies related to the service industries (Audretsch et al., 2004).

There are two main reasons for studying the differences. First, each industry has different characteristics that lead to firms of different sizes. Second, a firm's initial characteristics can influence its long-term behaviour. Size and age are the variables that are mainly studied in the literature.

The Spanish evidence is scarce. There is therefore a field of research between firm growth and other significant variables which have been point out in different studies such as R&D, financial variables, etc.. Past contributions are characterised by the scarcity of data on Spanish firms and the majority of samples in these studies have included only firms in the manufacturing industries.