Università degli Studi di Firenze

DOTTORATO DI RICERCA IN
"*Statistica Applicata*"

CICLO XXIII

COORDINATORE Prof. FABIO CORRADI

Size distributions and the analysis of the size-growth relationship

Settore Scientifico Disciplinare SECS/S01

Dottorando
Dott. (*Tortolini Valentina*)

Tutore
Prof. (*Rampichini Carla*)

Prof. (*Velucchi Margherita*)

Anni 2008/2012 (*di inizio e fine corso*)

Contents

Chapter 1

Introduction

The aim of this work is to analyze the size distributions of economic phenomena and to investigate the relationship between size and growth.

The study of the size distributions is not exclusive of economics. Beyond economics and finance (Gabaix (2009)), several type of size distributions are studied in computer science (Mitzemacher (2003)), physics, biology, social systems (Newman (2005)). The results of those different disciplines point out similar behaviors in terms of size distribution and in terms of dynamics of size distribution. Obviously, to study the size distribution of a phenomenon and the dynamic of the size of this phenomenon is equivalent to study the size and the growth process. Since the size distributions are formed as an outcome of underlying dynamics (for instance, the firm size distribution is the results of the firm underlying growth process involving entry and exit from the market of firms and products, innovation, merges, acquisitions, spin-out, etc.), several models have been proposed in literature to account for these dynamics and each of these models generate a certain equilibrium size distribution[1].

According to these results, in literature, two different empirical tests are commonly used (Hall (1987a)) to check the consistency of the theoretical stochastic models of growth with the empirical data. The first approach consists in assuming a stationary (over time) growth model, and then studying the size distribution obtained from the growth model. The second approach consists in investigating the determinants of growth rates by means of regression methods.

[1] An equilibrium size distribution is a probability distribution that remains constant over time. In many stochastic models used to describe the growth process, the stochastic matrix is assumed to remain constant over time. Under this assumption, and provided other properties, the distribution will tend to an equilibrium distribution dependent on the stochastic matrix but not on the initial distribution (Champernowne (1953)).

When the the first approach is used for testing growth models, usually, the lognormal distribution is compared with the Pareto distribution. As a matter of fact, these two distributions are obtained as size distributions of the two most popular growth processes: the Gibrat's rule of proportionate effects (Gibrat (1931)) and the Simon growth process (Simon and Bonini (1958b)).

The simplest growth model was proposed by Gibrat in 1931 (Gibrat (1931)). The Gibrat's law of proportionate effects states that the mean growth rates are independent on the size. This model generates a lognormal distribution as size distribution. More complex models (which account for boundary conditions, exit and entry of units) generate, as equilibrium size distribution, other distributions like the Pareto one (Simon and Bonini (1958b)). Gibrat showed that the size distribution is approximately lognormal for a broad range of data (Gibrat (1931), Sutton (1997)). Simon and co-authors, on the other hand, argued that the observed size distributions are well approximated by a Pareto distribution, at least in the upper tail (Simon and Bonini (1958a)). The exact shape of the size distribution is still debated and, notwithstanding the Pareto distribution and the lognormal distribution are typically retained useful benchmarks (Hall (1987b), Cabral and Mata (2003), Growiec et al. (2008)), further works tried to develop new models with a better fitting to the empirical data (L.C.Thurow (1970), Salem and Mount (1974)), Singh and Maddala (1976), McDonald (1984), Azzalini (1985), Azzalini and Capitanio (2003)). The drawback of this approach is that the relationship between growth rates and size is not explicitly investigated.

The second approach for testing a growth model consists in investigating the determinants of growth rates by means of regression methods. A great contribution to this second approach has been provided by industrial economists. An early prominent contribution on the investigation of Gibrat's Law was made by Mansfield (1962) who studied the growth rates of steel, petroleum, rubber, tire and automobile industries. According to Mansfield, the majority of empirical studies has rejected the Gibrat law claiming that small firms grow faster than larger firms. The negative relation between firm size and growth rates has been found using data for different countries[2], different level of industrial aggregation[3], and different industrial sectors[4]. Only

[2]Dunne and Hughes (1994), Kumar (1985) studied the quoted UK manufacturing firms, Bottazzi and Secchi (2003), Hall (1987b) the quoted US manufacturing firms, J. Goddard (2002) the quoted Japanese firms, Gabe and Kraybill (2002) establishments in Ohio.

[3]Dunne et al. (1989) analyze plant-level data as opposed to typical firm-level data.

[4]Typically the manufacturing sector but for example Barron et al. (1994) study New York Credit Unions, Weiss (1998) Austrian farms, Liu et al. (1999) Taiwanese electronic

few studies find some support in favour of the Gibrat law[5]. However, when a specific discrimination between small and large firms is done, results are somewhat different. While for small firms the negative relation holds nearly always[6], for large firms a flat relation is typically observed and whenever the Gibrat law is rejected even a positive dependence is found[7]. Some studies have tested the Gibrat law for firms above a certain size threshold. For example Droucopoulos (1983) focuses on a sample of the world's largest firms and finds support for it[8]. Mowery (1983) analyzes two samples of small and large firms and finds a negative relation for the former while the Gibrat's law holds for the latter. Cefis et al. (2006), for the worldwide pharmaceutical industry, and Hart and Oulton (1996), for a data set of independent U.K. companies, find a negative relation for pooled data but once the sample is restricted to only large firms the dependence vanishes.

In this work we used both the approaches commonly used in literature for testing growth models. First we analyzed some of the theoretical size distributions proposed in literature. Then we check the fitting of these models to the empirical data considering Italian wage data for private sector. We used these data since the renewed interest on income/wage distribution observed since 1980s (Forster (2000)). We decided to use wage data instead of income data for two main reasons. Firstly, the wage is a source of revenue more homogeneous than the income. As a matter of fact, income is given by the sum of wages, salaries, profits, interest payments, rents and other forms of earnings received in a given period of time. Secondly, at the best of our knowledge, in literature exist (for Italy or across Countries) many comparisons of parametric models of income distributions over time (see Bandourian et al. (2003), Azzalini et al. (2002), Dastrup et al. (2007)), but does not exists a comparison for parametric models of wage distribution in Italy over time. As pointed out before, the analysis of the size distribution is one of the approaches commonly used to investigate the size growth relation and to study the stochastic models of growth proposed in literature. Furthermore this analysis allows us to know the entire distribution of a certain phenomenon (in this case the Italian daily wage). Once the wage distribution is known,

---

plants.

[5]Bottazzi et al. (2005) for French manufacturing firms, Hardwick and Adams (2002) for UK life insurance companies.

[6]See Evans (1987a), Evans (1987b), Yasuda (2005), Calvo (2006), McPherson (1996), Wagner (1992), Almus and Nerlinger (2000). However Audretsch et al. (2004) find results in favour of the Gibrat's law for small-scale Dutch services.

[7]A positive relation was found by early studies of Hart (1962), Samuels (1965), Prais (1974), Singh and Whittington (1975) on data for UK manufacturing firms.

[8]See also Becchetti and Trovato (2002), Geroski and Gugler (2004), Lotti et al. (2003).

we can investigate some important features of wage, as the concentration and the inequality, that, at least in the last 25 years, are become relevant research topics (Forster (2000)). After performing the analysis of the size distributions we use an econometric approach to test for the dependence between growth and size. For this analysis a data set of firms belonging to the pharmaceutical sector is used. The expected growth is modeled as a function of lagged sizes and of some other covariates (that are considered relevant, in industrial literature, to explain the firms dynamics). This approach allows to purge the size/growth relation of the effect due to other relevant economic variables which we can observe in our dataset. Moreover, we make use of dynamic panel data estimators in order to identify the coefficient of lagged sales in the growth regression. In fact, this strategy allows us also to remove the confounding effects due to unobserved determinants of growth and to the correlation between lagged sales and the idiosyncratic error.

This work is structured as follows. In Chapter 2 we briefly describe the models proposed by Gibrat and Simon and the related equilibrium distributions, then we analyzed the regression methods used to investigate the determinants of growth.

In Chapter 3 we study the size distribution of Italian wages between 1985 and 2004. The empirical results are based on a database called Work Histories Italian Panel (WHIP), implemented by Laboratorio Revelli (Revelli (2010)). This data set was builded starting from the pieces of information of the National Social Security Institute's administrative archives. The WHIP database contains a great amount of individual information about the work histories of people who have worked in Italy (in private sector). After a brief description of the Italian labour market between 1985 and 2004 (Section 3.2), the data set used for the analysis is described in section 3.3 (alongside with some preliminary evidences). In Section 3.4, the empirical wage distribution (by year and by gender) is fitted with different models belonging to the generalized beta-family and with a skew-normal model and a skew-t model. The distributions are compared by means of different measures of goodness of fit (see Dastrup et al. (2007) and Bandourian et al. (2003)). In the last section of Chapter 3 we performed an analysis of the inequality in the Italian wage distribution between 1985 and 2004. The analysis of inequality is performed by means of four inequality indices belonging to the generalized entropy class of inequality measures. The inequality indices are calculated by gender and over time. With this approach it is possible to analyze the dynamic of the inequality within different groups (the whole sample, the male sub-sample and the female sub-sample). Usually, the inequality analysis proposed in literature for the Italian data does not distinguish between male and female (see Brandolini et al. (2002), Manacorda (2004), Jappelli and Pistaferri (2010) and Devicienti (2003)). The results we found for the male sub-sample agree with the findings proposed in literature while the results for the female sub-sample are different. For this reason we decided also to analyze the dynamic of inequality between the two samples. To perform this analysis we study the dynamic of the gender gap by means of regression methods.

In Chapter 4 we investigate the size-growth relation analyzing a set of micro data for firm sales. To perform this analysis we used a data set of 1,152 USA pharmaceutical firms for the period 1996-2007. In particular we applied a growth regression approach to deal with a number of econometric issue which have been arisen on literature. First, the panel dimension allows us to account for the effect of time-constant unobserved heterogeneity. Then, making use of an instrumental variables approach, we allow for the violation of the strict exogeneity assumption which comes with the inclusion of size on the right-hand side. This approach is carried out within a Generalized Method of Moments (GMM) framework which delivers efficiency gains in estimation. Estimates are also robust to heteroskedasticity and autocorrelation in the error term. Moreover, information contained in our data set allow us to control for the age and the innovation of firms, which are considered the most relevant determinants of growth (Jovanonic (1982), Cabral and Mata (2003), Cirillo (2010), Evans (1987b), Dunne and Hughes (1994), Geroski and Gugler (2004), Yasuda (2005), Coad (2007), Growiec et al. (2008)). Age is an important variable to study firms behavior over time since the study of the relationship between size and age generally shows that the size distribution of firms varies with firms age. A large number of studies find that age reduces the growth rate of firms (see Evans (1987a), Evans (1987b), Dunne and Hughes (1994), Geroski and Gugler (2004), Yasuda (2005)).

As regards innovation the relation with growth is not as clear as the relation between age and growth. While the role of innovation is considered central to the growth of firms (Carden (2005), Hay and Kamshad (1994), Geroski (2000), Geroski (2005)), empirical studies find difficulties in modeling such relation. Moreover Coad (2007) suggests, according to the literature, that another problem consists in the definition of innovation itself. Two popular innovation proxies used are the expenditure in R&D and the patents count but, both these measures have drawbacks though. Expenditure in R&D may not be well associated with the actual output of an innovative process. Patents count cannot discriminate between patents with substantial and marginal economic impact, while it is typically found that the former are a negligible amount. On the other hand, in many theoretical models, innovation is represented by the entry in the market of new business opportunities (Ijiri and Simon (1964), Kalecki (1945), Pammolli et al. (2007). In such a case the innovation can be proxies both with new product launches and with the opening of new product lines, divisions, subsidiaries, and plants (Bottazzi et al. (2001), Pammolli et al. (2007), Growiec et al. (2008)). Unfortunately it is not easy to test empirically the effect of the entry of the new business opportunities in the market . As a matter of fact it is not easy to have data provided at the product level. Our data are provided at the product level, so we can account for the entry in the market of new products and firms.

In our model specification we include a variable which attempts to overcome both limits outlined above. In particular this variable synthesizes both the net inflow of products and the change in Anatomical Therapeutic Chemical (ATC) classification[9]. In this way, on the one side we allow for products which are actually marketed, on the other side it is likely that a change in ATC captures at least in part the output of the innovation process, unless associated to a negative inflow, so that the problem of temporal lag should be mitigated.

For the best of our knowledge, only in other few cases (see Oliveira and Fortunato (2003), Ribeiro (2007)) the dynamic panel estimators were used to test the Gibrat's rule in the industrial context[10]. Furthermore we want to explore the effect of the innovation on the size-growth relationship in pharmaceutical industry. In last years many authors have dealt with innovation in pharmaceutical industry (see DiMasi et al. (1991), Masi et al. (2003), Ornaghi (2006), Comanor and Scherer (2011)) focusing their research on the relationship between innovation and merges. According to these results our contribution here is to explore whether the innovation benefits are different for smaller and larger firms. There are two different levels of innovation related to the firms size in pharmaceutical industry: the small firms are start-ups using the so called "genetic engineering" or biotechnologies while the big pharma companies developed licensing, sponsored R&D and partnerships with biotech in order to join biotechnological innovation (Bobulescu and Soulas (2006)). Most of the authors studied the relationship between innovation and size from the point of view of the existence of scale economies in pharmaceutical industry R&D (Jensen (1987), Graves and Langowitz (1993),Bobulescu and Soulas (2006), Cockburn and Hender-

[9]The classification system divides drugs into different groups according to the organ or system on which they act and/or their therapeutic and chemical characteristics. The anatomical first level of the code contains 14 main groups: Alimentary tract and metabolism, Blood and blood forming organs, Cardiovascular system, Dermatologicals, Genito-urinary system and sex hormones, Systemic hormonal preparations, excluding sex hormones and insulins, Antiinfectives for systemic use, Antineoplastic and immunomodulating agents, Musculo-skeletal system, Nervous system, Antiparasitic products, insecticides and repellents, Respiratory system, Sensory organs, Various.

[10]Soo (2011) used a dynamic panel estimation to test the relation between size and growth of state population in the United States.

son (2001), Miyashige et al. (2007)). The results diverge from one to another especially in relation to the measure of innovation used, however seems that scale economies exist in pharmaceutical innovation. Moreover, Masi and A. (1995) and R.Henderson and Cockburn (1996) showed that R&D costs per new drug approved in the U.S. decrease with firm size, while sales per new drug increase with firm size, but the relationship between innovation and growth by firms size is not investigated.

The Chapter 4 is organized as follows. Section 4.1 describes the data and provide some preliminary evidence. Section 4.2 discusses the methodology employed to estimate the relation between growth and size. Section 4.3 and 4.4 discuss results respectively when the whole sample of firms is used and when two sub-groups of small and large firms are selected. Section 4.5 concludes.

Chapter 2

Size and Growth: theoretical framework

Stochastic models of growth have been developed since 1931, when Gibrat (1931) proposed his law of proportionate effects. Nowadays the Gibrat's model, together whit the Simon's model (Simon (1955)), are considered benchmark cases for the growth models. These two growth models generate two different equilibrium size distributions. An equilibrium size distribution is a probability distribution that remains constant over time. In many stochastic models used to describe the growth process, the stochastic matrix is assumed to remain constant over time. Under this assumption, and provided other properties, the distribution will tend to an equilibrium distribution dependent on the stochastic matrix but not on the initial distribution (Champernowne (1953)). The equilibrium size distribution generated by the Gibrat's model is the lognormal distribution while the Simon growth process converges to a Pareto distribution.

To check the validity of the stochastic models of growth two different approaches has been proposed in literature. The first approach consist on studying the size distribution of a phenomenon and check if this distribution is coherent with the distribution predicted by the stochastic model. For instance, to test the Gibrat's model we can study the size distribution and check if it is lognormal.

The second approach consists on investigating the determinants of growth rates by means of regression methods. For instance to check the validity of the Gibrat's model we can test if the growth rate are independents on the size.

In this chapter we briefly describe the Gibrat's model and the Simon's model and the related equilibrium distributions. Finally we analyzed the regression methods used to investigate the determinants of growth rates.

2.1 Growth models and related equilibrium distributions

Gibrat (Gibrat (1931)) in 1931 presented the book *"Inegalites Des Economiques"* in which he postulated the theory of a *"law of proportionate effects"* that was the first formal model of the dynamics of the size of a firm (or the income of a individual). Gibrat asserted that the size of certain phenomena at time t was the result of a joint effect of a large number of mutually independent causes that have worked for a long period of time. At a certain time t the size of a certain unit, x_t, can be expressed in terms of size of the same unit at time $t-1$, namely x_{t-1}, and of proportionate rate growth over the period $t-1$ to t, namely ϵ_t, so that:

$$x_t - x_{t-1} = \epsilon_t x_{t-1} \tag{2.1}$$

whence

$$x_t = (1+\epsilon_t)x_{t-1} = x_0(1+\epsilon_1)(1+\epsilon_2)...(1+\epsilon_t) \tag{2.2}$$

if ϵ_t is small enough, and it can be reasonably choosing a short period of time t (Sutton (1997)), then is possible to approximate $log(1+\epsilon_t) \simeq \epsilon_t$. Taking logarithms we thus obtain

$$log(x_t) \simeq log(x_0) + \epsilon_1 + \epsilon_2 ... + \epsilon_t \tag{2.3}$$

As $t \to \infty$ the term $log(x_0)$ becomes insignificant, and we obtain

$$log(x_t) \simeq \sum_{k=1}^{t} \epsilon_k \tag{2.4}$$

This imply that the size of a unit at time t can be explained in terms of its idiosyncratic history of multiplicative shocks. Assuming that all the units are independent realizations of these shocks and assuming that the shocks ϵ_t are independent and normally distributed with means m and variance σ^2, than the distribution of $log(x_t)$ is approximated by a normal distribution and the limit distribution of x_t is lognormal.

This simple model is affected by several limitation. The growth rates distribution of many economic phenomena is not normal but it is a fat-tailed distribution (Buldyrev et al. (2007)). Furthermore, as pointed out by Kalecki (1945), it is not reasonable to suppose, as predicted by the Gibrat's model, that the variance of units tends to infinity.

Simon and co-authors (Simon and Bonini (1958b), Ijiri and Simon (1964)) extended Gibrat's model by accounting for the entry of new firms. The Simon growth model was developed to explain the growth process of firms but it is a general model to describe the dynamics of a system of elements with associated counters (Barabasi and Albert (2002)).

In Simon's framework, the market consists of a sequence of many independent "opportunities" which arise over time, each of size unity. At time t a new firm born with probability α or an existing firm, randomly selected with probability β proportional to its size, increases its size with probability $1-\alpha$. For this stochastic process, Simon found a stationary solution exhibiting a power-law distribution. Several models of proportional growth have been subsequently introduced in economics to explain the growth process, especially to explain the growth of of business firms (Gabaix (1999), Sutton (1997)).

2.1.1 Pareto Distribution

The statistical study of size distributions started after the publication in 1896 of the Pareto's *Cours d'economique politique* (Pareto (1896)). He showed that the relation between the logarithm of the number of taxpayers N_x, with incomes above a level x, and the log value of income x was close to a straight line of slope $-\gamma$ for some $\gamma > 0$. Formally:

$$log(N_x) = A + log(x^{-\gamma}) \tag{2.5}$$

where $A, > 0$. The Pareto distribution is power-law distribution (Mitzemacher (2003)).

A power law is a relation of the type $Y = kX^{\alpha}$, where X and Y are the variables of interest and α is called the *exponent* of the power law (Gabaix (2009)). A variable X has a power-law distribution if its probability of taking a value greater than x varies at the power of α. Formally, the complementary cumulative distribution function[1] (ccdf) is:

$$Pr(X \geq x) = Cx^{-\alpha} \tag{2.6}$$

where $\alpha > 0$, $C > 0$. Power law distributions are scale free distributions[2] characterized by heavy tails (heavier than other distribution such as the exponential distribution) that decay, asymptotically, according to the power of α.

An empirical test to check if a random variable follows a power law is to plot the ccdf in log-y scale. Asymptotically the behavior of the CCDF of a

[1]The complementary cumulative distribution function is given by $1 - Pr(X \leq x)$

[2]Mathematically a scale free distribution satisfies $p(bx) = g(p)p(x)$

power law whit exponent α will be a straight line whit slope $-\alpha$ (Mitzemacher (2003)) since:

$$log[Pr(X \geq x)] = log(Cx^{-\alpha}) = log(C) - \alpha log(x). \tag{2.7}$$

Power laws distributions are used to describe a large number of empirical regularities in economics and finance (Gabaix (2009)), computer science (Mitzemacher (2003)), physics, biology, social systems (Newman (2005)).

Pareto distribution (Pareto (1896)) and Zipf's (Zipf (1949)) law are two common power-laws distributions. The Zipf's law (after George Kingsley Zipf, a Harvard linguistics professor) usually refers to the size y of an occurrence of an event relative to it's rank r and states that the size of the r-th largest occurrence of an event is inversely proportional to its rank r

$$y \sim r^{-b} \tag{2.8}$$

If y is a certain income, equation 2.8 means that the r-th richest person has an income equal to y plus a certain constant. So, reminding that from equation 2.8 follows that $r \sim y^{-1/b}$, the probability that the variable Y is equal to a certain income y can be written as follow

$$P(Y = y) = \frac{dr}{dy} \sim y^{-(1+1/b)} \tag{2.9}$$

Expression (2.9) represents the PDF of a Pareto distribution. Since the Zipf's law and the Pareto law can be regarded as equivalent, from this point we will refer to the Pareto law.

In the classical version, the c.d.f of the Pareto distribution is defined as:

$$F(x) = 1 - \left(\frac{x}{x_0}\right)^{-\gamma}, \quad x \geq x_0 > 0, \tag{2.10}$$

where γ is the shape parameter and x_0 is the location and the scale parameter. The density of a classical Pareto distribution is given by

$$f(x) = \frac{\gamma x_0^{\gamma}}{x^{\gamma+1}}, \quad x \geq x_0 > 0. \tag{2.11}$$

Note that $\gamma = \alpha - 1$ where α is the power-law slope. The parameter γ gives the heaviness of the right tail of the distribution. An higher coefficient γ is associated with a less fat upper tail, i.e. faster convergence of density towards zero. In economics γ is called the Pareto index and it is usually used as a measure of the breadth of the income or wealth distribution. When the

Pareto distribution is used to describe the income or the wealth distribution, The Pareto index represents a measure of inequality: the larger is γ the smaller is the proportion of very high-income people. In economics is often used the inverted Pareto coefficient $\lambda = \gamma/(\gamma - 1)$ rather than the standard Pareto coefficient γ. The λ coefficient represent the ratio between the average income of individuals with income above a certain threshold and the threshold. The characteristic of the Pareto distribution (and of the Power laws) is that λ does not depend od the threshold (Atkinson et al. (2011)).

The raw moment μ'_k[3] is given by

$$\mu'_k = \frac{\gamma x_0^k}{\gamma - k}, \tag{2.12}$$

and exists only if $k < \gamma$. From eq. (2.12) follows the expressions for the mean and the variance of a Pareto distribution. The expected value is given by

$$E(X) = \frac{\gamma x_0}{\gamma - 1}, \quad \gamma \neq 1 \tag{2.13}$$

and exists only if $\gamma > 1$ [4]. The variance is given by:

$$var(X) = \frac{\gamma x_0^2}{(\gamma - 1)^2(\gamma - 2)}, \tag{2.14}$$

and exists only if $\gamma > 2$, while the mode is at x_0.

The Pareto distribution is linked with the Exponential distribution in fact it could be shown that if $X \sim Par(x_0, \gamma)$ then

$$Y = log(\frac{X}{x_0}) \sim Exp(\gamma), \tag{2.15}$$

and equivalently, if $Y \sim Exp(\gamma)$, then $x_0 e^Y \sim Par(x_0, \gamma)$, as showed in (fig.2.1).

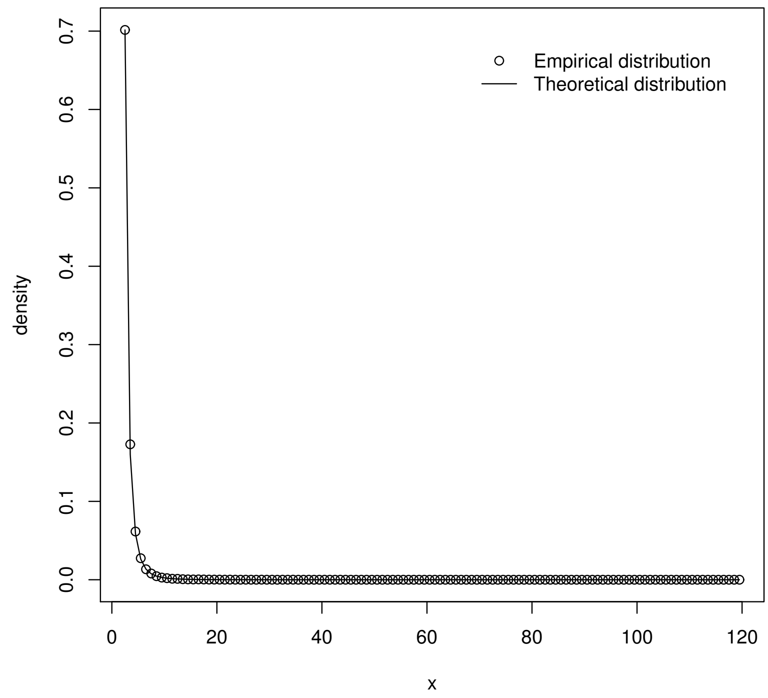
The Pareto distribution was proposed in three different variants. The first is the classical Pareto distribution defined in 2.10. The cdf of the second Pareto model is given by

$$F(x) = 1 - \left(1 + \frac{x - \mu}{x_0}\right)^{-\gamma}, \quad x \geq \gamma. \tag{2.16}$$

[3] The k-th raw moment of a distribution with continuous pdf $f(x)$ is defined as $\mu'_k = \int_{-\infty}^{+\infty} x^k f(x)dx$

[4] From extremely heavy-tailed distribution of this class, other measure of location must be used (C. and S. (2003))

Figure 2.1
PARETO DISTRIBUTION



Notes:

1. Points represent the density distribution of random variable X obtained as $X = x_0e^Y$ where x_0 is setting equal to 3 and Y is a random sample drawn from an exponential distribution with parameter $\gamma = 2$. The line represent the theoretical density distribution of a Pareto distribution with parameters $x_0 = 3$ and $\gamma = 2$.

By setting $\mu = 0$ in the 2.16 we obtain the cdf of a Pareto typeII distribution

$$F(x) = 1 - \left(1 + \frac{x}{x_0}\right)^{-\gamma}, \quad x \geq 0, \quad x_0, \gamma > 0. \tag{2.17}$$

The Pareto type II distribution is called also Lomax distribution since Lomax (Lomax (1954)) rediscovered it in a different contest. The Pareto type II distribution is also considered a beta-type distribution (C. and S. (2003)) since it is a special case of Singh-Maddala distribution. The relation between a classical Pareto (Pareto type I model) and a Pareto type II model is the following

$$X \sim ParII(x_0, \gamma) \iff X + x_0 \sim Par(x_0, \gamma). \tag{2.18}$$

The cdf of a Pareto type III distribution is given by

$$F(x) = 1 - \frac{Ce^{-\beta x}}{(x-\mu)^{\gamma}}, \quad x \geq \mu, \quad \mu \in \mathbb{R}, \quad \beta, \gamma > 0. \tag{2.19}$$

2.1.2 Lognormal Distribution

The lognormal distribution is often presented in terms of normal distribution since a random variable X has a lognormal distribution if $Y = log(X)$ has a normal distribution. The pdf of a normal distribution is given by

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}, \tag{2.20}$$

where μ is the mean, σ^2 is the variance and $-\infty < y < \infty$. Therefore the pdf of the lognormal distribution is given by

$$f(x) = \frac{1}{x\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(logx-\mu)^2}, \quad x > 0, \tag{2.21}$$

while the cdf is given by

$$F(x) = \Phi\left(\frac{logx - \mu}{\sigma}\right), \quad x > 0, \tag{2.22}$$

where Φ is the cdf of a standard normal distribution.

The lognormal distribution is unimodal, with the mode being at $x_m = e^{\mu-\sigma^2}$, and right skewed so that mode<median<mean[5].

It is convenient to obtain the moment generating function in terms of the moment generating function of a normal distribution

$$E(X^k) = E(e^{kY}) = e^{k\mu + \frac{1}{2}k^2\sigma 2}. \tag{2.23}$$

From eq.(2.23) follows that the mean of a lognormal distribution is

$$E(X) = e^{\frac{\mu+\sigma^2}{2}}, \tag{2.24}$$

and the variance is

$$Var(X) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1). \tag{2.25}$$

[5]This follows from the expressions of the three positions indices:

- Mode=$exp(\mu - \sigma^2)$,
- Median=$exp(\mu)$
- Mean=$exp(\mu + \sigma/2)$

From the moment generating function (eq.2.23) it is also clear that the log-normal distribution has moments of all orders.

As a consequence of the close relationship with the normal distribution, some basic properties of the lognormal distribution follow from properties of the normal distribution. For example, from the stability property under summation of the normal distribution[6] follows the multiplicative stability property for the lognormal distribution. Formally, if X_1 and X_2 are two independent variables with distribution $X_1 \sim LN(\mu_1, \sigma_1^2)$ and $X_2 \sim LN(\mu_2, \sigma_2^2)$, respectively, then

$$X_1 X_2 \sim LN(\mu 1 + \mu 1, \sigma_1^2 + \sigma_1^2). \tag{2.26}$$

Unfortunately, sums of lognormal random variables are not very tractable (C. and S. (2003)).

There are some similarities between the Pareto distribution and the log-normal distribution. First, both distributions can be obtained via exponentiation of another random variable, namely the Pareto distribution from an exponential and the lognormal from a normal distribution. Second, the behavior of the log-log plot of the ccdf (or of the pdf) of the two distributions, will be very similar (Mitzemacher (2003)). In the case of the Pareto the behavior is exactly linear while in the lognormal case, for large value of σ^2, the behavior will be almost linear for a large portion of the distribution (see 2.2). Using the pdf for simplicity, we have for the Pareto distribution

$$lnf(x) = (-\gamma - 1)lnx + \gamma lnx_0 + ln\gamma, \tag{2.27}$$

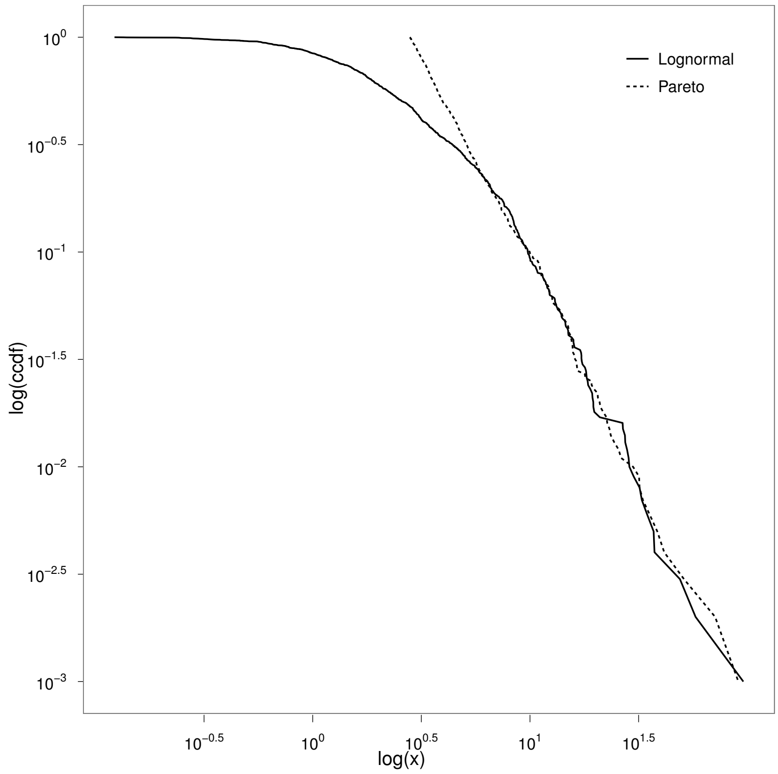
and for the lognormal

$$lnf(x) = -\frac{(lnx)^2}{2\sigma^2} + \left(\frac{\mu}{\sigma^2} - 1\right) lnx - ln\sqrt{2\pi}\sigma - \frac{\mu^2}{2\sigma^2}. \tag{2.28}$$

Since the Pareto and the Lognormal distributions often fitted only a part of the empirical size distribution, other authors tried to develop new models to fit the whole range of income. In recent years, empirical literature proposed distributions with an high number of parameters: three parameters Beta (L.C.Thurow (1970)), Gamma (Salem and Mount (1974)), Singh and Maddala (1976) and generalizations of these densities, such as first kind and second kind generalized gamma (GB1 and GB2) families densities (McDonald (1984)).

[6] The stability property under summation of the normal distribution means that the sums of independent normal variables are again normal.

Figure 2.2
CCDF for a Pareto distribution and a Lognormal distribution



Notes:

1. Both axes are in log-scale

2.2 Regression methods

As pointed out in the previous sections, an approach proposed in literature (Hall (1987b)) to check the validity of a stochastic growth models, consists on investigating the determinants of the growth by mean of regression methods. Since the 1950s the Gibrat law of proportionate effects has stimulated a multiplicity of this kind of empirical works[7]. Empirical investigation of Gibrat's law by means of regression methods rely on estimation of equation of type:

$$ln(S_{i,t}) - ln(S_{i,t-1}) = \beta ln(S_{i,t-1}) + X_{i,t}\delta + \mu_i + u_{i,t}, \tag{2.29}$$

[7]An exhaustive survey on firms growth is provided by Coad (2007).

Where $S_{i,t}$ is the size at time t of the $i - th$ unit. The composite error term $\nu_{i,t} = \mu_i + u_{i,t}$ consists of a time constant unobserved heterogeneity μ_i and of an idiosyncratic component $u_{i,t}$. X_{it} is a matrix of regressors which can correlate with μ_i and $u_{i,t}$

The coefficient β is the "Gibrat coefficient" in the sense that evidence of $\beta = 0$ supports the Gibrat's law, while evidence of either positive or negative β is at odds with it. Equation 4.1 can be rewritten as

$$s_{i,t} = \tilde{\beta} s_{i,t-1} + X_{i,t}\delta + \mu_i + u_{i,t}, \tag{2.30}$$

where $\tilde{\beta} = 1 + \beta$, and $s_{i,t} = ln(S_{i,t})$. Equation 4.2 makes it clear that estimating equation 4.1 is equivalent to estimating a dynamic equation of logarithmic sizes with a lagged-dependent variable on the right-hand side. We estimate this equation but interpretation of parameters can more easily recovered from equation 4.1[8]. Testing for $\tilde{\beta} = (>, <)1$ is equivalent to testing for $\beta = (>, <)0$.

In this equation it is clear that endogeneity may come from two sources. First, lagged sizes are necessarily positively correlated with the fixed effect μ_i so that, regardless of regressors in $X_{i,t}$ being strictly exogenous, pooled OLS applied to equation 4.2 would inflate estimates of $\tilde{\beta}$. For example, imagine a large negative sizes shock for a given unit in a certain year t which cannot be ascribed to any variables in the model. *Ceteris paribus*, the time constant fixed effect over the whole period for that unit would appear lower so that in $t + 1$ both lagged sizes and the fixed effect will be lower. In the end, the correlation between $s_{i,t}$ and $s_{i,t-1}$ would appear stronger though it would be simply driven by the fixed effect (Roodman (2009a)). Moreover, also the regressors in $X_{i,t}$ may be correlated with the unobserved heterogeneity. For example, some units may have an intrinsic capability to grow faster than others and they may be as well more strongly associated to characteristics included in the model, such as a higher propensity to innovation, which in turn may affect growth. In such case the estimated innovation effect on the growth would simply pick up, at least in part, the effect of unobserved capability.

Standard panel data methods overcome this problem wiping out the fixed effects by ad hoc transformation of data. For example, the *Within* and first differences methods transform data so that the estimating equation would appear like this

$$s^*_{i,t} = \tilde{\beta} s^*_{i,t-1} + X^*_{i,t}\delta + u^*_{i,t}, \tag{2.31}$$

[8]Estimates of parameters δ are to be interpreted as regressors effects on the sales growth since are estimated in equation 4.2 for given $s_{i,t-1}$.

where the apex * denotes transformed data. The typical panel approach consists in applying pooled OLS to this equation. Unfortunately, even though one assumes that the regressors in $X_{i,t}$ are strictly exogenous[9], $s^*_{i,t-1}$ is necessarily correlated with $u^*_{i,t}$ so consistency of panel estimates is lost. In particular, sizes and the error term in levels can appear, respectively, in the $s^*_{i,t-1}$ and in the $u^*_{i,t}$ expressions at some periods, (depending on the kind of transformation) of which at least one is the same in both expressions. For the fixed effects case, the correlation between $s^*_{i,t-1}$ and $u^*_{i,t}$ turns out to be negative and its strength is inversely related to T (see Nickell (1981)). As a consequence, fixed effects estimates would underestimate $\tilde{\beta}$ and the magnitude of the bias would be smaller the larger T. In the fixed effect case $s^*_{i,t-1} = s_{i,t-1} - \frac{1}{T_i-1}(s_{i,1} + \ldots + s_{i,t} + \ldots + s_{i,T_i})$ and $u^*_{i,t} = u_{i,t} - \frac{1}{T_i-1}(u_{i,2} + \ldots + u_{i,t-1} + \ldots + u_{i,T_i})$. The component $-\frac{s_{i,t}}{T_i-1}$ in $s^*_{i,t-1}$ is correlated with $u_{i,t}$ in $u^*_{i,t}$, and the component $-\frac{u_{i,t-1}}{T_i-1}$ in $u^*_{i,t}$ is correlated with $s_{i,t-1}$ in $s^*_{i,t-1}$ (see Bond (2002) and Roodman (2009a)). If T_i were large the component above would be negligible and the correlation would disappear.

The typical solution in panel data when strict exogeneity is violated is to apply instrumental variables methods. For example, a tempting approach may be to apply two stage least squares to the transformed equation where lagged sales are instrumented with variables that are both correlated with $s^*_{i,t-1}$ and orthogonal to $u^*_{i,t}$. Natural candidates as instruments of $s^*_{i,t-1}$ may be deeper lags of $s_{i,t-1}$. However, in the fixed-effects transform, which is based on a time demeaning, lags of $s_{i,t-1}$ are still not orthogonal to the transformed error. This is the reason why other kinds of transformations are applied when dealing with dynamic panels, such as the first differences and the "orthogonal deviations". In dynamic panel estimation, efficiency gains can be achieved within the Generalized Method of Moments (GMM) framework (see Hansen (1982)). Building on the work of Anderson and Hsiao (1982) and Holtz-Eakin et al. (1988), Arellano and Bond (1991) derived one-step and two-step GMM estimators using moments conditions in which lagged levels of the dependent and predetermined variables are instruments for the first-differenced equation (other than strictly exogenous variables). In a first step, where some reasonable but arbitrary covariance matrix for the errors is chosen, an initial GMM regression is performed to get a preliminary consistent estimate of parameters. Then residual from the first step are used to estimate the sand-

[9]In panel data jargon, variables are strictly exogenous when they are uncorrelated with current, past and future realizations of the disturbance term. When this condition does not hold, the variable is said endogenous, but when the variable remains uncorrelated with future values it is said predetermined.

wich covariance matrix of the errors which is used to perform second step. This two-step estimator is asymptotically efficient and robust to whatever patterns of heteroskedasticity and cross-correlation the sandwich covariance estimator models. However, though two-step is asymptotically efficient, robust two-step estimator of standard errors tends to be severely downward biased. Windmeijer (2005) derived a bias-corrected robust estimator for the two step covariance matrix so that the two-step efficient GMM is in general preferred to the one-step. Windmeijer finds that the two-step efficient GMM performs somewhat better than one-step in estimating coefficients, with lower bias and standard errors. Also corrected two-step standard errors are quite precise.

The Arellano-Bond estimator has been further developed by Arellano and Bover (1995) and Blundell and Bond (1998) who pointed out that in some applications lagged levels may be poor instruments for transformed variables. In particular, they proposed a system GMM estimator in which the transformed equation is augmented by adding the original equation in levels to the system. In this equation, variables in *levels* are instrumented with suitable lags of their own *first differences* (no matter the kind of transformation), and the transformed equation is still instrumented with all available lags of endogenous and predetermined variables. This method is known as the system GMM, in opposition to the difference GMM of Arellano-Bond[10].

[10]The Arellano-Bond estimator is generally called difference GMM since in the original formulation data are transformed by first differences, but when orthogonal deviations are carried out the term is not appropriate though sometimes still used.

Chapter 3

The Size Distribution of Italian Wage

Studying size distribution is not exclusive of economics. Beyond economics and finance (Gabaix (2009)), several type of size distribution are studied in computer science (Mitzemacher (2003)), physics, biology, social systems (Newman (2005)). And, the results of those different disciplines point out similar behaviors in terms of size and growth distribution. The aim of this work is to analyzed some of the theoretical size distributions proposed in literature and to check the fitting of these models to empirical data. To perform this analysis Italian wage data for private sector are used. We used these data since the renewed interest on income/wage distribution observed since 1980s (Forster (2000)). We decided to use wage data instead of income data for two main reasons. Firstly, the wage is a source of revenue more homogeneous than the income. As a matter of fact, income is given by the sum of wages, salaries, profits, interest payments, rents and other forms of earnings received in a given period of time. Secondly, at the best of our knowledge, in literature exist (for Italy or across Countries) many comparisons of parametric models of income distributions over time (see Bandourian et al. (2003), Azzalini et al. (2002), Dastrup et al. (2007)), but does not exists a comparison of parametric models of wage distribution in Italy over time.

Income distribution analysis has always been a relevant research topic in Italy. From the beginning of Twenty century eminent Italian economists and statisticians were interested in studying income distributions, concentration and inequality (Benini (1897), Gini (1909b), Gini (1909a), Gini (1932) Amoroso (1925), Cantelli (1929), D′Addario (1932)). In other western countries, on the contrary, the focus on income distribution has been developed only in the last 25 years. Quoting the web site of the Distributional Analysis Research Programme (DARP) at London School of Economics and Political

Science (LSE): "*the study of income distribution is enjoying an extraordinary renaissance: interest in the history of eighties, the recent development of theoretical models of economic growth that persistent wealth inequality, and the contemporary policy focus upon the concept of social exclusion are evidence of new found concern with distributional issues*".

The renewed interest on income distribution is, almost in part, a consequence of the significant increase in income inequality observed in many countries since 1980s. Forster (2000) studying inequality in OECD countries summarized that *"there has been no generalised long-term trend in the distribution of disposable household incomes since the mid-1970s. However, during the more recent period (mid-1980s to mid-1990s), income inequality has increased in about half the countries, while non of the remaining countries recorded an unambiguous decrease in inequality."*

The first income distribution was proposed by Pareto (1896). In his studies Pareto argued that the income distribution is hyperbolic and skewed (see also Ammon (1898)). Furthermore, Pareto sustained that the distribution does not change significantly in space and time. The author also concluded that inequality shrink as the income rise. Gini (1909b) proposed a new inequality measure, scale free, called Gini index. The empirical results of Gini was at variance with the conclusions of Pareto. Gini showed that inequality in income distribution raises as income increases. Furthermore empirical studies showed that the Pareto distribution has a good fitting with empirical data only for high income levels.

Gibrat (1931) introduced the assumption that income distribution is log-normal. The author argued that income and wealth are governed by a multiplicative random process. The log-normal model has a god fitting for the bulk of empirical income distribution but not for the tails. Following the Gibrat example, Simon and Bonini (1958a) ,Ijiri and Simon (1964), Mandelbrot (1960), Parker (1999), Draculescu and Yakovenko (2001b) and Draculescu and Yakovenko (2001a) proposed densities that arise from stochastic processes of income growth.

In recent years, empirical literature proposed distributions with an high number of parameters: three parameters Beta (L.C.Thurow (1970)), Gamma (Salem and Mount (1974)), Singh and Maddala (1976) and generalizations of this densities, such as first kind and second kind generalized gamma (GB1 and GB2) families densities (McDonald (1984)).

Nowadays, the GB2 is largely considered a good description of income distributions, with a fine goodness-of-fit. A large number of empirical studies showed that GB2 outperforms other densities in income fitting, in particular for USA income data (Butler and McDonald (1989), McDonald and Xu (1995) and F.Bordley et al. (1996)), for German data (Brachmann et al.

(1996)) and for UK data (Jenkins (2007) and Jenkins (2009)).

GB2 is an extremely flexible model that can assume the shape of several well-known distributions of the social statistics literature and of the industrial statistical literature (Cirillo (2010)). Moreover, it has been shown that the GB2 distribution family also includes several flexible models recently introduced in the literature, such as the generalized k-distribution (Clementi et al. (2008)).

A theoretical justification for the use of a GB2 as personal income distribution was provided by Parker (1999). The author proposed a neoclassic optimizing model, based on micro-foundations, which predicts the earnings distribution to follow a GB2. In Parker model a representative firm has to choose the optimal number of workers to employ at each human capital level, considering that the firm must pay earnings to induce workers to invest in human capital. This decision problem leads to a GB2 density function as the optimal earnings distribution.

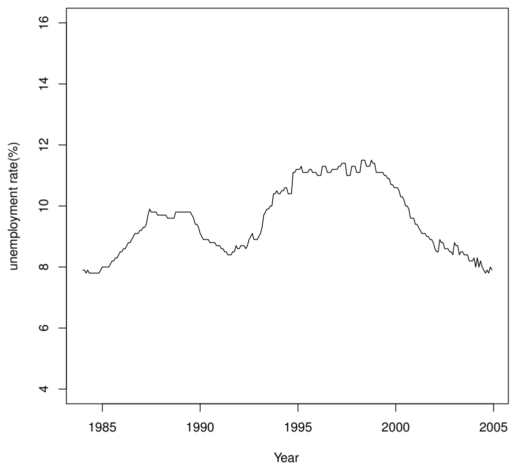
An alternative to the GB2 density was proposed by Azzalini et al. (2002) introducing a log-skew normal and a skew-t as income family distributions for US and European data.

In this work we analyze the distribution of wage in Italy for the private sector. The empirical results are based on a database called Work Histories Italian Panel (WHIP) implemented by Laboratorio Revelli (Revelli (2010)). After a brief description of the Italian labour market between 1985 and 2004 (Section 3.2), the data set used for the analysis is described in Section 3.3 (alongside with some preliminary evidences). In Section 3.4 the empirical wage distribution (by year and by gender) was fitted with different models belonging to the generalized beta-family and with a skew-normal model and a skew-t model.The distributions are compared by means of different measures of goodness of fit (Dastrup et al. (2007), Bandourian et al. (2003)). As pointed out in Chapter 2, the analysis of the size distribution is one of the approaches commonly used to investigate the size growth relation and to study the stochastic models of growth proposed in literature. Furthermore this approach allows us to know the entire distribution of a certain phenomenon (in this case the Italian daily wage). Once the wage distribution is known, we can investigate some important features of wage, as the concentration and the inequality, that, at least in the last 25 years, are become relevant research topics (Forster (2000)). In the last section we performed an analysis of the dynamic of the inequality in the Italian wage between 1985 and 2004. The analysis of inequality is performed by means of four inequality indices belonging to the generalized entropy class of inequality measures. The inequality indices are calculated by gender and over time. With this approach is possible to analyzed the dynamic of the inequality within different groups (the whole sample, the male sub-sample and the female sub-sample). Usually, the inequality analysis proposed in literature for the Italian data does not distinguish between male and female (see Brandolini et al. (2002), Manacorda (2004), Jappelli and Pistaferri (2010) and Devicienti (2003)). The results we found for the male sub-sample agree with the findings proposed in literature while the results for the female sub-sample are different. For this reason we decided also to analyzed the dynamic of inequality between the two samples. To perform this analysis we study the dynamic of the gender gap by means of regression methods.

3.1 The Italian labour market between 1985 and 2004

As argued by Fabiani et al. (2000), Italy, between 1980s and 2000s, experienced *"one of worst labour market performances among the countries of the European Union."*. In 1980s and 1990s levels of unemployment rate[1] were so high as those observed in the post-war period.

Figure 3.1
UNEMPLOYMENT RATE IN ITALY BETWEEN 1985-2004.

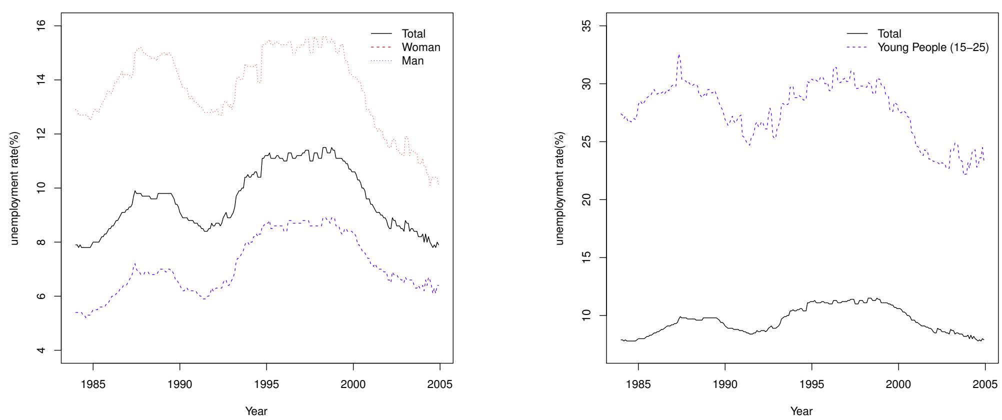


Notes:

1. Plot represent the harmonized unemployment rate monthly series based on the results of the EU Labour Force Survey.
2. Data are provided by Eurostat on the web site: $http : //epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database$.

The dynamic of aggregate unemployment rate in Italy between 1985 and 2004 is showed in Figure 3.1. A great increase in unemployment took place from the mid 1970s to the late of 1990s. This rise could be divided in two separate episodes (see Bertola and Garibaldi (2003)). An early stage, between 1975 and 1988, over which unemployment increased constantly up to

[1]Unemployment rate is the number of people unemployed as a percentage of the total population aged 15-64. Where unemployed persons are all persons 15 to 74 years of age who were not employed during the reference week, had actively sought work during the past four weeks and were ready to begin working immediately or within two weeks.



(a) Unemployment rate by gender. (b) Unemployment rate by age.

Figure 3.2

UNEMPLOYMENT RATE BETWEEN 1985 AND 2004. MONTHLY DATA.

Notes:

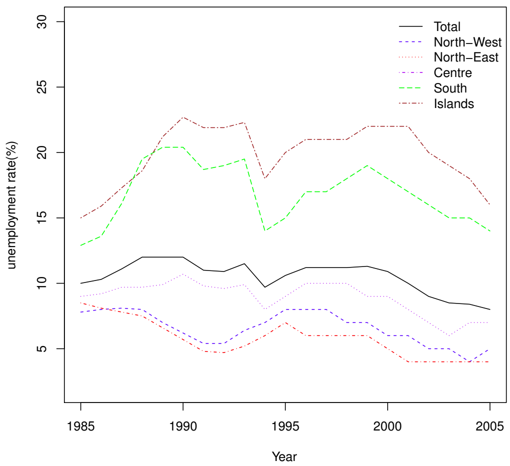
1. Figure a represent the harmonized unemployment rate monthly series by gender based on the results of the EU Labour Force Survey.
2. Figure b represent the harmonized unemployment rate monthly series by age based on the results of the EU Labour Force Survey.
3. Data are provided by Eurostat on the web site: *http : //epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search*$_d$*atabase.*

some 10% in 1989. And a second stage between 1993 and 1998 when the increase of unemployment rate was remarkably fast, particularly in 1993-95. Between this two episodes, in 1989-92, a fall in unemployment rate is observable. In most recent years, from 1998 to 2004, unemployment rate showed a decline. The dynamic of the aggregate Italian unemployment is the results of a complex structure of disaggregated unemployment rates. The main dimensions of heterogeneity are a gender effect, an age effect and a geographical effect. Figures 3.2(a), 3.2(b) and 3.3 shows the unemployment rate in Italy by gender, age and geographical area respectively.

Gender is a relevant dimensions. Female unemployment rate is never lower of 10% of the labour force while male unemployment rate is never higher than 9% of the labour force. Also the age dimension is relevant. The unemployment rate of young people is remarkable higher than the national average reaching a maximum of 32% of the labour force in 1987. In addition to the gender dimension and to the age dimension, Italian unemployment is affected by a geographical dimension. Unemployment rate in North areas is lower than in the other geographical areas. In the Center the unemployment is above the Italian average but larger than in the Northern areas, while in South and Islands the unemployment is much larger than the national average.

Figure 3.3

Unemployment rate in Italy between 1985-2004 by geographical area.



Notes:

1. Plot represents the annual unemployment rate based on data provided by ISTAT
2. Data are from: Istat *Rilevazione trimestrale sulle forze di lavoro* until 2003 and *Rilevazione sulle forze di lavoro* for 2004.
3. Between 1985 and 2004 definitions of employed and unemployed change. Between 1993 and 2003 data have been updated according to revised population series within inter-censual period 1991-2001.
4. Until 1992 people of 14 year old are included in labor force. From 1993 labor force included people of at least 15 year old.

Aggregate and disaggregated unemployment rate dynamics among OECD countries have been explained in terms of interaction between institutional features and macroeconomics shocks (Layard et al. (2001), Grubb and Wells (1993), Saint-Paul (1997), Nickell and Layard (1999), Nickell (1997), Blanchard and Wolfers (2000), Bertola and Garibaldi (2003), Belot and Ours (2000), Bassanini and Duval (2006a), Bassanini and Duval (2006b), Barbieri (2009)). The incapacity of European labor markets to cope with the unemployment was ascribed to their lack of flexibility and to the lack of flexibility of the the social security systems. The most common solution of EU countries to face rising unemployment has been to introduce new contracts that allowed for a progressive deregulation of the labour market. The introduction of new type of temporary and non standard contracts favored the entry in the labour market of some particular categories of workers characterized by a high rate of exclusion (especially young people and women). The new contracts in fact, thanks to lower levels of social protection, allowed to reduce the labour cost. In addition, the entry of new contracts contributed to the segmentation of the labour market. Therefore different kinds of workers coexist: regular workers with a higher social protection and a higher bargain power and temporary workers, potentially, with an higher risk of been firing and low wages.

In early 1980s Italian labour market was characterized by strong job protection and downward wage rigidity. The stringent regulation of employment relationship was regulated by the 1970 *Statuto dei lavoratori* and by subsequent reforms such as the reform of labor litigation of 1974 (see Bertola and Ichino (1995)). Still during 1970s regionally differentiated wage were replaced by industry-specific negotiated wage structured and in 1975 the wage indexation system was introduced (the so called *Scala Mobile*). Since early 1980s flexibility-oriented policies have been introduced. On the wage side, the indexation system was progressively reformed and then abolished in 1992. Further new legislations on temporary and non standard contracts were introduced.

In Italy the atypical contracts were introduced since 1955 (Law n.25), when the first regulation related to apprenticeship contracts was established. The fixed-term contracts were introduced in 1962 by Law n. 230. The regulation on temporary contracts limited fixed terms contracts to seasonal workers, unusual activities and top management. The Law n. 863/84 introduced the *on the job training contracts* (Contratto di formazione lavoro CFL) in order to ease entry into the labour market. Like the fixed-term contracts, the CFL contracts had a determined duration: one year in order to acquire low qualification and two years in order to acquire high qualification. The legislation of CFL contracts was modified in 1987, with the Law n. 56 that made the CFL contract applicable to all economic sectors, and in 1994 (Law 451/94) with the raising of the age limit of their applicability from 29 years old to 32 years old. In 1995 the coordinate and continuous collaboration (Co.Co.Co.[2]) contracts were introduced. In 1997, in order to bring flexibility and dynamism to the Italian labour market, the so called Treu Law (Law n. 196) was introduced. The Treu Law (named after the Labor Minister Tiziano Treu) introduced temporary contracts (without age limitation) and

[2]A Co.Co.Co is a self-employment with some specific relationship with the company featuring.

Temporary Work Agency (TWA[3]). Furthermore the reform modified: the statuatory discipline of fixed term contracts, the apprenticeship relationship and the CFL applicability.

In 2003 the Treu Law were replaced by the Law n.30/2003 or Biagi reform (named after Marco Biagi, advisor on labor market reform under the 2001-2006 Berlusconi government). The law introduced new contractual forms and innovated some existing ones, affecting mainly subordinated jobs. Major innovations concerned apprenticeship and training contracts. The CFL contracts were substituted by the insertion contracts[4]. Finally, the Biagi Law changed the Co.Co.Co contracts, substituting them with the Co.Co.Pro. relationships, in which workers are put into a specific project or plan. The dynamics of contracts in private sector between 1985 and 2004 is showed in Table 3.1. The number of workers with a temporary or atypical contract is increased constantly since the introduction of the new type of contracts until 2003. The increase of temporary contracts is remarkably evident especially among young workers (see Table 3.2)

Table 3.1

Evolution of share of contracts from 1985 to 2004 in private sector

	Year									
Contract	**1985**	**1990**	**1995**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**
Permenent	93.81	88.67	92.48	82.10	74.65	74.65	74.18	72.95	71.05	71.12
Fixed-term				5.78	6.50	7.10	7.08	7.91	8.57	9.81
Apprendiceshipe	5.30	5.25	3.97	4.87	5.04	5.42	5.39	5.06	4.99	5.48
CFL	0.89	6.08	3.55	3.70	2.95	2.12	1.94	1.53	1.23	0.57
TWA				0.07	0.61	1.09	1.46	1.41	1.66	1.87
Collaborator				1.77	8.67	8.12	8.63	10.01	11.34	9.98
Seasonal				1.72	1.58	1.49	1.32	1.12	1.16	1.17
Totale	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Notes:

1. Data are from: Work Histories Italian Panel (WHIP).

As described above, between 1985 and 2004, the Italian aggregate unemployment rate showed different dynamics. Two episodes of rise (the first episode started in 1975 and ended in 1988, while the second took place between 1993 and 1998) and two episodes of decline (the first took place between 1989 and 1992 and the second started after 1998) are observable. The two episodes of increasing unemployment correspond with two economic shocks: the oil price hikes in 1970s and the productivity slowdown of 1980s; and

[3]TWA employment represents a triangular contract, in which an agency hires a worker for the purpose of making him available to a client firm for a temporary assignment

[4]Insertion contracts can not by applied in public administration.

Table 3.2
Evolution of share of contracts from 1985 to 2004 sector for young people (15-25 years)

	Year									
Contract	**1985**	**1990**	**1995**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**
Permanent	77.64	63.53	70.44	48.19	42.91	44.33	40.81	39.97	37.22	34.46
Temporary contracts*	22.36	36.47	29.53	42.99	46.30	48.20	48.86	46.40	47.24	50.94
Collaborator				8.83	9.37	8.89	10.33	13.63	15.54	14.60
Totale	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Notes:

1. *Termporary contract comprehends all type of contracts except permanent and collaborator

the fiscal and monetary restrictions, due to reach the prerequisite for EMU (European Monetary Union) membership in 1992.

In the first episode unemployment rate rose lower than in the 1990s since the employers were protected from dismissing redundant employees (Bertola and Garibaldi (2003)). The growth of unemployment rate was particulary high for the young people (who were not able to enter in the labor market) and for the South and Island (i.e in geographical areas with a low productivity).

From 1998 unemployment rate shows a decline despite a sharp cyclical slowdown and despite rather restrictive fiscal policy. A possible explanation to this phenomenon can be found in the institutional changes (such as a more flexible labour market, the introduction of the wage moderation and the introduction of the hiring subsidies) occurred from the mid of 1990s (Bertola and Garibaldi (2003)).

3.2 The Work Histories Italian Panel (WHIP) data base

The Work Histories Italian Panel (WHIP) is a data base implemented by Laboratorio Revelli. WHIP is builded starting from the pieces of information of the National Social Security Institute's administrative archives. The reference population is made up by all the people, Italian and foreign, who have worked in Italy even only for a part of their working career.[5]

[5] The standard version of the data base is a Public version of WHIP. The sampling coefficient in the standard file is about 1:180, for a dynamic population of about 370,000 people.

Table 3.3
SAMPLE CONSISTENCY BY YEAR AND GENDER BETWEEN 1985 AND 2004

Year	*Freq.*	*Male*	*Female*	*Year*	*Freq.*	*Male*	*Female*
1985	56,397	69.76%	30.24%	1995	59,740	65.66%	34.34%
1986	56,741	69.46%	30.54%	1996	60,537	65.38%	34.62%
1987	57,954	68.89%	31.11%	1997	60,693	64.73%	35.27%
1988	59,440	68.48%	31.52%	1998	61,347	64.32%	35.68%
1989	60,628	67.91%	32.09%	1999	63,614	64.06%	35.94%
1990	62,392	67.80%	32.20%	2000	66,451	63.78%	36.22%
1991	63,011	67.56%	32.44%	2001	68,798	63.38%	36.62%
1992	62,742	67.36%	32.64%	2002	71,243	63.35%	36.65%
1993	60,090	66.93%	33.07%	2003	72,464	63.04%	36.96%
1994	59,332	66.48%	33.52%	2004	73,023	62.59%	37.41%

For each of these people the main episodes of their working careers are observed. The complete list of observations includes: private employee working contracts, dependent self employment contracts[6] (the so called *parasubordinati*), self-employment activities (i.e artisan, trader and some activities as freelancer), retirement spells and non-working spells. During non-working spells the individual received social benefits, such as unemployment subsidies or mobility benefits. The workers for whom activity is not observed in WHIP are those who have an autonomous security fund (i.e people who worked in the public sector or as freelancers). In our analysis we dropped out the atypical contracts and the self-employment. We do this for the following reasons. Firstly, most of the available variables for the other contracts are not available for atypical contracts and self-employment. Secondly, it is not possible to calculate a reliable daily wage for the atypical contracts. The WHIP section concerning employee contracts is a Linked Employer Employee Database. In addition to the data about the contract, thanks to a linkage with the Inps Firm Observatory, data concerning the firm in which the worker is employed are also available.

For each individual within the database, several variables are available, distinguishable between time-constant and time varying variables. The time-constants variables are: birth data; birth area, which indicates the geographical area[7] of Italy where the individual was born, and gender. The time-

[6]There are two main categories of workers: professionals and co-workers (the so called *collaboratori*). The archive for dependent self employment contracts covers the period 1996 to 2004 co-workers contract were introduced in Italian Labour market in 1995.

[7]Four geographical macro areas are available in the standard version: North West -which includes Valle d'Aosta, Piemonte, Lombardia and Liguria- North East - which includes Trentino Alto Adige, Veneto, Friuli Venezia Giulia and Emilia Romagna- Centre -which includes Toscana, Marche, Umbria and Lazio - South - which includes Campania, Molise, Abruzzo, Basilicata, Puglia and Calabria - and Islands - which includes Sardegna

varying variables are: work area, which indicates the geographical area of Italy where employment was performed [8]; number of paid working days equivalent to full time [9]; number of paid weeks equivalent to full time work; skill level, distinguishes between various employment positions (Apprentice, Blue Collar Worker, White Collar Worker, Cadre[10] -high skilled White Collar-, Manager); classification of economic activity into 18 sections according to the Ateco91 classification (ISTAT (1991)); total annual compensation in euro[11]; TFR fund which indicates the amount accrued by the employee in the end-of-service fund (Trattamento di Fine Rapporto, TFR)[12]; a code which indicates the type of contribution rebate eventually applied to the worker's contract; the starting date of the job spell, deduced from the contributions paid monthly by the employee (the variable is left censored at January 1, 1985); the date of ending of the job spell (the variable is right censored at 31 December 1999).

In addition, the data base contains the following dummy variables: maternity benefit, part time position, wage supplement for temporary layoffs (Cassa Integrazione Guadagni, CIG), illness benefits.

Table 3.3 shows the sample consistency by year and gender. In the observed sample males represent at least 62.5% of total observations. In the twenty years of observations females weight rises of 7%.

---

e Sicilia

[8]During the whole period of employment the employee can modify the geographical location in which he/she works; partition of italian regions in macro areas is the same used for birth area

[9]A day is considered paid when the employer paid compensation subject to tax; A week or month is considered paid if they contain at least one paid day. Conventionally Inps (Italian Social Security Organization) reports paid days based on a 6 day working week; for example a 40 hour week 5 working days corresponds to 6 days 'paid.' The conversion, justified by insurance specifications, implying that one month and one year completely 'paid' are 26 days and 312 days respectively.In the case of part-time work, paid days and weeks are converted into days and weeks equivalent to full time.

[10]The position of Cadre is distinguishable from that of White Collar Worker only since 1997 on. Before 1997 Cadres have the same code as White Collar Workers.

[11]Total annual compensation in euro (top coded at 1.100 euro, applied to the average weekly compensation). At the fiscal/accounting level it represents the base for calculating social security and insurance contributions paid by the firm, the social burden of the employee and the eventual tax relief applied to employment. Therefore, it represents the annual net compensation received by the employee, net of the social security and health benefit contributions paid by the firm but gross of the social security and health benefit contributions that have to be paid by the employee.

[12]Top coded at 60,000 euro.

3.2.1 Preliminary evidences on WHIP data

In this section a preliminary analysis of the data described above is performed. The dynamics of the age and of the daily wage of the sample between 1985 and 2004 are showed by means of some descriptive statistics and non parametric distributions. To make the value of daily wage comparable between different years we adjusted all values to 1985 euros by GDP price deflator. The sample was then divided in sub sample by age, geographical area, type of contract, skills and gender and the wage distribution is analyzed for each sub sample.

Table 3.4
AGE DISTRIBUTION: MEAN AND PERCENTILES. 1985-2004

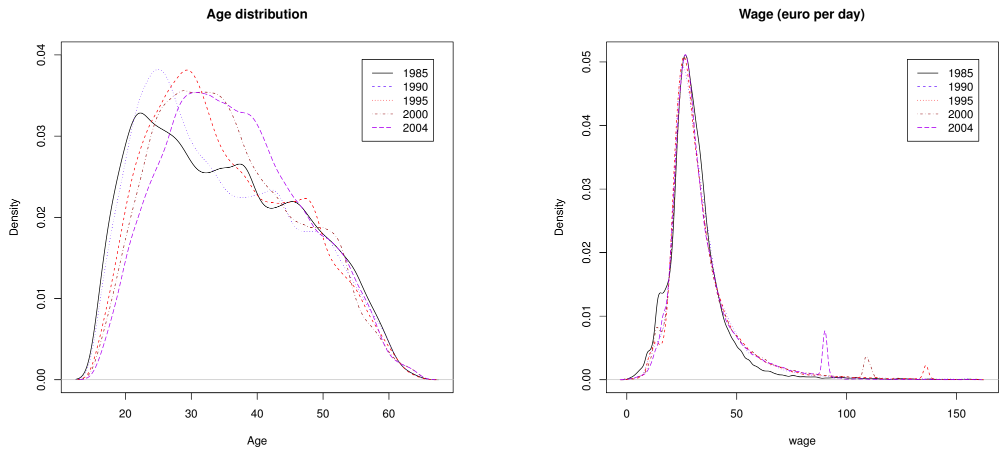
Stats	*1985*	*1990*	*1995*	*2000*	*2004*
			Age		
mean	35	35	35	36	37
p5	18	19	20	20	21
p10	20	21	22	23	23
p50	34	33	33	34	36
p75	44	44	44	43	44
p95	55	55	55	54	55
p99	60	60	60	60	59

Table 3.4 reports the average and percentiles of the age distribution (see Figure 3.4(a)) from 1985 to 2004. The trend underlines an ageing of the sample. Average age changes from 35 years old in 1985 to 37 years old in 2004. In Fig. 3.4(a), we can see the right shift of the median age and of the percentiles below the median, and the invariance of superior order percentiles (graphically a reduction of asymmetry) This effect could be the result of many phenomena:

- a delayed entry in labour market;
- an increase of atypical contracts as main type of contract for young workers (as suggested by the increase in the number of collaborator contracts for young people showed in Table 3.2);
- a delayed exit from the labour market due to the increase in age of retirement[13].

Dividing the sample in two groups by age (between 15 and 25 and more than 25 years old) the ageing of the sample become clear. In 1985 workers younger

[13]The increase in the age of retirement started in Italy in 1992 with the Law 503/1992, the so called Riforma Amato.



(a) Evolution of age distribution by year (b) Evolution of wage distribution by year

Figure 3.4
KERNEL DISTRIBUTIONS OF AGE AND WAGE BETWEEN 1985-2004

than 26 years old represent about the 27% of the whole sample while in 2004 the percentage of young worker shrinks to 14.86%. (see Table 3.5)

Table 3.5
SAMPLE DISTRIBUTION BY AGE BETWEEN 1985-2004

Age	*1985*	*1990*	*1995*	*2000*	*2004*
15-25	26.93%	27.26%	21.08%	18.23%	14.86%
over 25	73.07%	72.74%	78.92%	83.42%	85.14%

Figure 3.4(b) illustrates the daily wage distribution[14] between 1985 and 2004. The wage distribution results to be right asymmetric with the right tail heavier than the left one. This is a consequence that low wage workers are fewer than high wage workers. As time goes by, right tail enlarges itself and since 1995 it appears a wage clumping. On the contrary, left tail shows a wage clumping in the first year of observation, which shrinks without disappearing in the following years.

The average daily wage varies from 32.31 euros in 1985 to 34.58 in 2004. The median varies from 29.07 euros in 1985 to 29.76 in 2004. Average wage time series (see Figure 3.5) shows that daily wage reaches its maximum in 1992 and then it decreases following Italian economic cycles.

[14]Epanechnikov kernel

Figure 3.5
AVERAGE DAILY WAGE AND MEDIAN DAILY WAGE: 1985-2004.

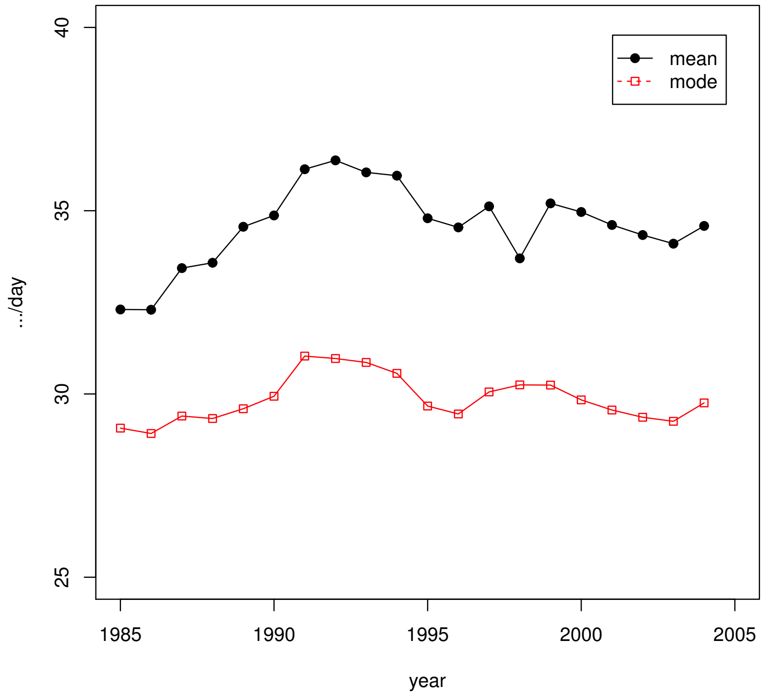


Table 3.6 shows workers distributions by geographical area in three years: 1985, 1995 and 2004. Around 60% of the sample works in North Italy, while the remaining 40% is distributed in the other areas (19.5% in the Centre, between 13.90% and 15.4% in the South, and 6.5% in the Islands). Most of the natives of North and Centre work, during their lives, in the same geographical areas. Around 35% of people coming form the South and the Islands work in different macro areas. The North West is the geographical area where average daily wage is higher, while the South is where it is lower. In the 20 observed years, the North West area and the North East one register higher increments in average wage (respectively 2.91 euros and 3.5 euros), than others geographical areas, (the average wage rise less than 1.50 euros). In the same period, standard deviation of average daily wage decreases. In particular, Islands standard deviations shrinks from 45.28 euros in 1985, to 21.18 in 2004 (see Table 3.7).

The sample was then divided into groups on the basis of the employment

Table 3.6
WORKERS BY WORK AREA AND BORN AREA

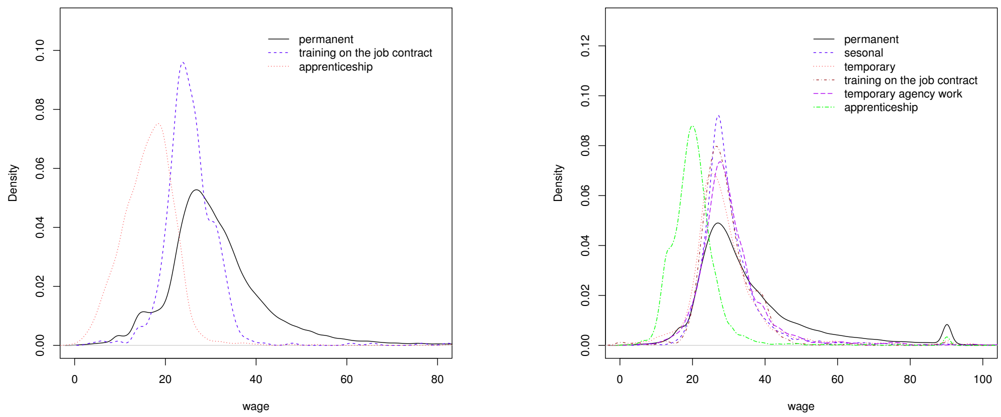
Born Area	Work Area				
	North-West	North-East	Central	South	Islands
		1985			
North-West	95.7%	2.3%	1.4%	0.3%	0.2%
North-East	11.7%	86.3%	1.5%	0.3%	0.2%
Central	5.0%	2.5%	91.0%	1.2%	0.3%
South	22.5%	4.2%	8.6%	64.1%	0.6%
Islands	23.1%	3.4%	6.2%	1.4%	65.9%
Abroad	32.3%	32.4%	19.8%	11.7%	3.9%
Total	36.0%	23.0%	19.7%	14.6%	6.7%
		1995			
North-West	94.6%	3.0%	1.5%	0.6%	0.4%
North-East	7.1%	91.2%	1.4%	0.3%	0.1%
Central	4.2%	2.9%	90.9%	1.6%	0.4%
South	20.4%	7.0%	8.7%	63.2%	0.6%
Islands	20.8%	5.1%	6.5%	1.1%	66.5%
Abroad	31.6%	35.4%	18.7%	10.6%	3.7%
Total	34.8%	25.5%	19.2%	13.9%	6.5%
		2004			
North-West	93.3%	3.5%	1.9%	0.8%	0.5%
North-East	5.8%	92.1%	1.5%	0.4%	0.2%
Centre	3.7%	3.0%	91.0%	2.0%	0.4%
South	15.9%	8.4%	9.2%	65.8%	0.7%
Islands	17.7%	7.2%	6.1%	1.6%	67.4%
Abroad	35.7%	33.5%	20.6%	8.0%	2.2%
Total	33.0%	24.9%	19.8%	15.4%	6.8%

contract. Until 2000, only three groups were observed: Permanent, Training on-the-job (TOJ) and Apprenticeshipe. Since 2000 other three types of contracts are provided: Seasonal, Temporary and Temporary Work Agency (TWA). Figures 3.6(a) and 3.6(b) shows the distributions of average daily wages by contract for the years 1985 and 2004. In 1985, we observe three distinct distributions for the three types of contracts. Apprenticeship contract workers are those who earn, on average, the lowest wage, while workers with permanent contracts receive, on average, the highest wage. TOJ contract workers are in an intermediate position.

In the following years, distributions change, even if the differences between the three distributions do not change. New contracts are placed between the distribution of Apprenticeshipe and Permanent job workers.

Table 3.7
WAGE DYNAMICS BY WORK AREA (EURO PER DAY) BETWEEN1985-2004.

work area	1985		1990		1995		2000		2004	
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
North-West	33.92	24.78	36.99	27.54	36.69	23.30	37.07	22.22	36.83	21.43
North-East	30.63	28.34	32.66	20.54	33.11	19.81	34.27	22.27	34.13	17.36
Central	32.99	38.38	35.62	25.85	35.33	23.40	35.15	21.66	34.42	20.41
South	30.56	19.50	33.40	17.85	33.33	21.98	32.18	15.75	31.67	18.78
Islands	31.20	45.28	32.62	18.06	32.78	17.19	32.30	20.52	32.45	21.18



(a) Year 1985. (Euro/day) (b) Year 2004. (Euro/day)

Figure 3.6
DAILY WAGE BY CONTRACTS. (EURO/DAY)

Table 3.8 illustrates the average and the most relevant percentiles of daily wage by gender. Female wage is consistently lower than the wage received by men, both in daily average and in the percentiles. This difference in wage by gender could be a consequence of a different distribution by gender among different work positions as showed in the Section 3.3.5. Density distributions by gender for the years 1985 and 2004 are showed in Figures 3.7(a) and 3.7(b), respectively.

In 1985, female distribution is more centered on the mode than the male one, with a lower modal value. In the same year, left tail shape of male distribution shows that men wages are more scattered than women ones. In addition, the heavier left tail of the female distribution with respect to the male distribution suggests that the percentage of low payed women is higher than the percentage of low payed men. In the following years, female

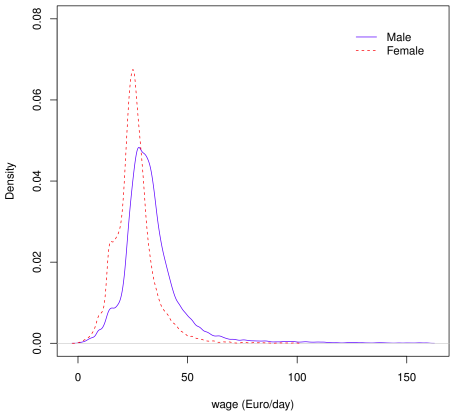
Table 3.8
WAGE DYNAMIC BY GENDER (EURO PER DAY)

Male					
Stats	**1985**	**1990**	**1995**	**2000**	**2004**
mean	34.71	37.55	37.28	37.10	36.51
p5	16.02	18.49	18.88	18.03	18.49
p10	20.99	22.18	21.87	21.66	21.60
p50	31.14	32.25	31.82	31.64	31.38
p75	37.85	41.13	41.42	41.36	40.97
p90	47.89	55.22	57.49	58.79	58.34
p95	57.01	67.94	72.29	75.81	75.54
p99	97.69	111.83	134.82	109.07	90.08
Female					
Stats	**1985**	**1990**	**1995**	**2000**	**2004**
mean	26.75	29.22	30.05	31.20	31.35
p5	13.22	15.59	16.04	15.09	16.12
p10	15.14	18.90	19.68	19.14	19.72
p50	25.19	26.11	26.57	27.35	27.45
p75	29.64	31.59	32.40	33.59	34.07
p90	35.52	40.55	42.05	44.29	45.12
p95	41.50	48.81	50.95	55.96	55.72
p99	55.47	68.76	71.15	75.64	84.53

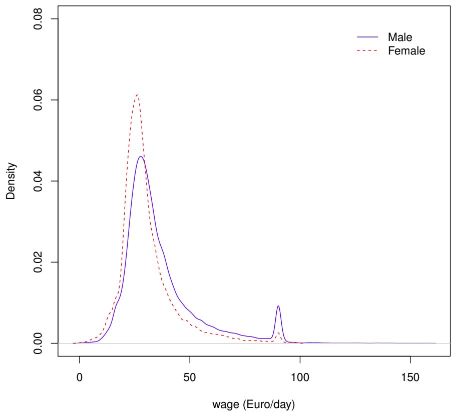
distribution shifts on the right, becoming more similar to male distribution. Nevertheless, a difference in concentration near the mode, a modal value for women lower than the men and different tails heaviness, remains evident.

In 1985, male earned 34.71 euros as average daily wage, while female earned on average only 26.75 euros. In 2004, the average daily wage for men was 36.51 euros, increasing of 1.8 euros since 1985. In the same time period (1985-2004), female average daily wage increases of 4.6 euros reaching, in 2004, the value of 31.35 euros.

Figures 3.8(a) and 3.8(b) summarize the two facts stands out from the analysis of wage by gender. Firstly, the average daily wage received by women is lower than the male daily wage for the whole observation period. Secondly, female average daily wage growths more than male daily wage between 1985 and 2004. It is possible to analyze the evolution in the wage gender gap plotting the ratio between the male wage and the female wage over 20 years. The plot (see fig. 3.9) shows that the gender gap, between 1985 and 2004, shrinks. This relation can be skewed since essential variables to explain the wage and its evolution over time are no taken into account. For this reason, in the section (3.3.5), the wage gender gap is analyzed after controlling for some individual covariates.



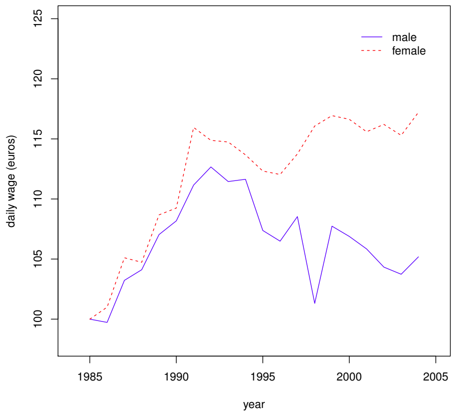
(a) Daily wage by gender in 1985. (Euro/day)



(b) Daily wage by gender in 2004. (Euro/day)

Figure 3.7

(a) Daily wage dynamics by sex. (Euro/day)



(b) Daily wage growth by gender. (Euro/day)

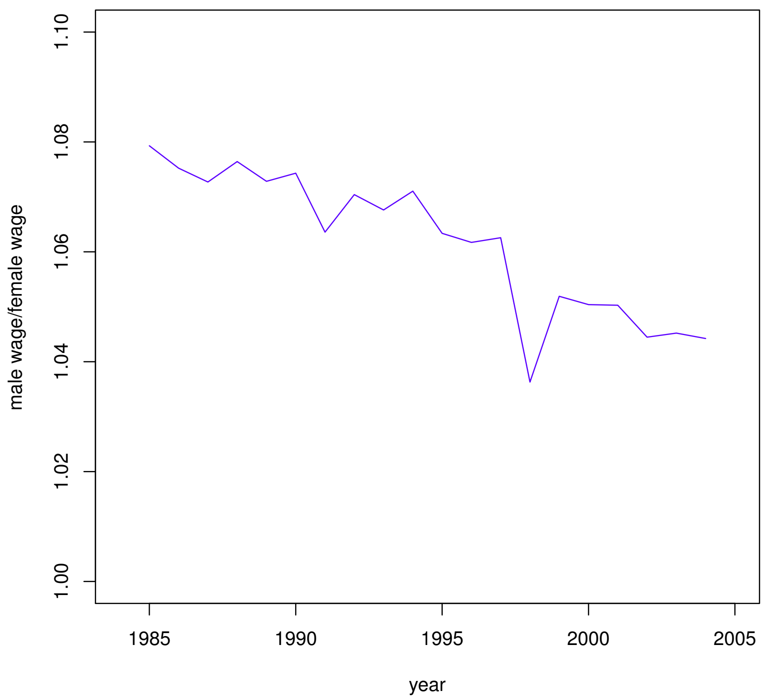
Figure 3.8

3.3 Statistical models for wage distribution

In this section some parametric models to describe the wage distribution are analyzed. Different models (with two three and four parameters) were fitted to the WHIP data. For the best fitting model inequality indices were

Figure 3.9

DIFFERENCE BETWEEN LOGARITHMIC MALE WAGE AND LOGARITHMIC FEMALE WAGE (1985-2004)



calculated. The analysis is performed separately for the whole sample, the male sub-sample and the female sub-sample. Furthermore, the wage distribution is analyzed in three different years (1985, 1995 and 2004), so that it can be possible to evaluate the wage distribution dynamics over time and by gender. Wage distribution has been fitted with statistical models belonging to the generalized beta family (C. and S. (2003)) and with distribution generated by perturbation of symmetry of a normal distribution (Azzalini and Capitanio (2003)). In the following section the models used in the analysis are briefly described, than the results of the fitting are showed. Finally, the inequality dynamic is analyzed by the means of four inequality indices.

3.3.1 The generalized beta family

The generalized Beta distribution of the second kind (GB2) is a four-parameters distribution belonging to the Beta-type family. The probability density function of a generalized beta (GB) distribution is given by:

$$GB(y;a,b,c,p,q) = \frac{ay^{ap-1}(1-(1-c)(y/b)^a)^{q-1}}{b^{ap}B(p,q)(1+c(y/b)^a)^{p+q})} \text{ for } 0 < y^a < \frac{b^a}{1-c} \quad (3.1)$$

and zero otherwise, where $B(.,.)$ is the Beta function and $0 \leq c \leq 1$, $0 \leq b, p, q$.

A Generalized Beta of Second kind (GB2) is obtained from a GB setting the c parameter equal to one:

$$GB2(y;a,b,p,q) = \frac{ay^{ap-1}}{b^{ap}B(p,q)[1+(\frac{y}{b})^a]^{p+q}}, x, a, b, p, q > 0 \quad (3.2)$$

where a, p and q are shape parameters, b is a scale parameter, $B(.,.)$ is the Beta function and $\Gamma(.)$ is the Gamma function. The thickness of the tails is given by a, the larger is the value of a the thinner the tails of density are, while p and q determinate the skewness of the distribution (see C. and S. (2003)).

The K^{th} moment of the distribution is:

$$E\left(y^k\right) = \frac{b^k\Gamma(p+\frac{k}{a})\Gamma(q-\frac{k}{a})}{\Gamma(p)\Gamma(q)} \quad (3.3)$$

and exists only if $-ap < k < aq$.

The GB2 includes the Singh-Maddala, the Beta of second Kind and the Dagum distribution as special cases, corresponding to q=1, a=1 and p=1, respectively. Setting $a = b = p = q = 1$ the GB2 can also considered a generalized log-logistic or Fisk distribution and for $a = p = 1$ a Lomax distribution is obtained (seeC. and S. (2003)). The Generalized Gamma (GG) distribution is obtained as a limiting case of the GB2:

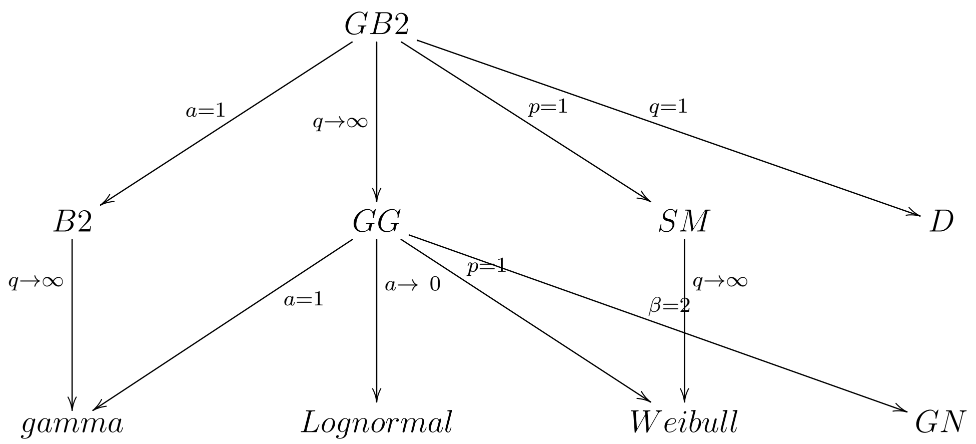
$$GG(y;a,\beta,p) = \lim_{q\to\infty} GB(y;a,b = q^{1/a}\beta, c = 1, p, q) \quad (3.4)$$

The Weibull and the Gamma distribution are obtained from a GG setting $p = 1$ and $a = 1$, respectively while the lognormal distribution is a limiting case of GG.

$$LN(y;\mu,\sigma) = \lim_{a\to 0} GG(y;a,\beta = (\sigma^2 a^2)^{1/a}, p = (a\mu+1)/\sigma^2 a^2) \quad (3.5)$$

The GG family includes the generalized normal (GN) family ($\beta = 2$), that is itself a flexbile family and includes half normal (HN) distribution ($a = 1/2$ and $p^2 = 2\sigma^2$). The relationships between these distributions are summarized in Fig. 3.10.

Figure 3.10
DISTRIBUTION TREE



3.3.2 Skew-normal distribution and skew t-distribution

Another distributions used as parametric models to describe the empirical income and wage belong to the class of the distributions generated by perturbation of symmetry of a normal distribution (Azzalini et al. (2002)) The class of skew-normal distributions was introduced in Azzalini (1985). The density of skew-normal (SN) is defined by:

$$\phi(z;\alpha) = \phi(z)\Phi(\alpha z), \text{ for } -\infty < y < \infty \tag{3.6}$$

where ϕ and Φ are the density function and the distribution function of a $N(0,1)$ variate, respectively and α is the shape parameter. From a SN a $N(0,1)$ is obtained setting $\alpha = 0$ and a half normal is obtained as a limiting case setting $\alpha \to \infty$.

It is possible to generalized the eq. 3.6 introducing a scale parameter ω and a location paramater ψ. If $Z \sim SN(\alpha)$ and $Y = \psi + \omega Z$ where $\psi \in R$, $\omega \in R^+$, then $Y \sim SN(\psi, \omega^2, \alpha)$ and its pdf is:

$$f(y) = 2\phi(y - \psi; \omega)\Phi[\alpha(y - 1\psi)\omega^{-1}]. \tag{3.7}$$

Azzalini (1985) proposed an alternative parametrization, the so called centered parametrization (CP parametrization), where the new three parameters (μ, σ, γ_3) have the usual meaning of mean, variance and skewness.

Centred parameters can be expressed in terms of direct parameters:

$$\begin{aligned}
\mu &= \psi + \omega\delta\sqrt{\frac{2}{\pi}}, \\
\sigma^2 &= \omega^2\left(1 + \frac{2\delta^2}{\pi}\right), \\
\gamma_3 &= \frac{4-\pi}{2}\frac{(\delta\sqrt{2/\pi})^3}{(1-2\delta^2/\pi)^{3/2}},
\end{aligned} \tag{3.8}$$

where $\delta = \frac{\alpha}{\sqrt{1+\alpha^2}}$.

The skew-t distribution is an asymmetric version of a Student's t distribution. In the Azzalini and Capitanio (2003) formulation a skew t-distribution is obtained as the ratio of a skew normal variable and a transformation of a χ^2 variable. A skew t-distribution can be defined as:

$$\tilde{Z} = V^{-\frac{1}{2}}Z, \tag{3.9}$$

where Z is a skew normal variate with $\psi = 0$ and $V \sim \chi^2_\nu/\nu$ is independent of Z. It is possible to generalize this expression by introducing other two parameters ψ and ω so that $\tilde{Y} = \psi + \omega\tilde{Z}$. A skew t-distribution can be obtained also mixing a scale mixture of SN variables whit scale mixing factor $V^{1/2}$. The pdf of a standard skew t-distribution is:

$$f(y) = 2t(z;\nu)T\left(\alpha\left(\frac{1+\nu}{z^2-\nu}\right)^{1/2}; \nu-1\right) \tag{3.10}$$

where

$$t(z;\nu) = \frac{\Gamma\left((\nu+1)/2\right)}{(\pi\nu)^{1/2}\Gamma(\nu/2)}\left(1+\frac{z^2}{\nu}\right)^{-(\nu+1)/2} \tag{3.11}$$

is the density function of a t variable with ν degrees of freedom, and $T(x;\nu{+}1)$ is the t distribution function with $\nu+1$ degrees of freedom.

A skew t-distribution can be derived also as a transformation of a beta density. The skew t family introduced by Jones and Faddy (2003) has density:

$$f(x;a,b) = \frac{1}{B(a,b)2^{a+b-1}\sqrt{a+b}}\left(1+\frac{x}{\sqrt{a+b+x^2}}\right)^{a+1/2}\left(1-\frac{x}{\sqrt{a+b+x^2}}\right)^{b+1/2}, \tag{3.12}$$

where $a, b > 0$ and $B(.,.)$ is the Beta function. As Jones and Faddy (2003) underlined the skew-t distribution that they derived behaves rather differently from the skew-t distribution derived by Azzalini. The authors stressed the different way in which skewness is controlled by the two models. Jones model considers two separate left-tail and right-tail power parameters. In this model the skewness arises through differences between tail-weight parameters. In the skew t-distribution proposed by Azzalini there is just one parameter controlling the tails weight. The skewness is introduced by different scales in each tails.

3.3.3 Parameters estimation

Parameters estimation of models described in the Sections 3.3.1 and 3.3.2 is based on the log-likelihood(see C. and S. (2003)). The parameters are estimated to maximize:

$$l(\theta) = \sum_{i=1}^{N} ln(f_d(y_i : \theta)), \tag{3.13}$$

where f is the pdf of the distribution and θ is the vector of the distributional parameters. When the equations for the estimation are not in a closed form unknown parameters are estimated with numerical optimization methods[15].

To compare nested models (such as GB2 and Sing-Maddala) the likelihood ratio test is used. To compare non-nested models (such as Singh-Maddala and Dagum) the chi-square goodness-of-fit measures is computed (see Dastrup et al. (2007)).

Parameters estimation for the GB2 and the skew t-distribution

GB2 and skew t-distribution are best performing models in terms of goodness of fit for our data set, as showed in Section 3.3.3. For this reason GB2 and skew t specification are described in more datails.

Consider a GB2 model, the log-likelihood for a complete sample is:

$$\begin{aligned} l =& n \log \Gamma(p+q) + n \log(a) + (ap-1) \sum_{i=1}^{n} \log(x_i) - nap \log(b) - n \log \Gamma(p) - \\ & - n \log \Gamma(q) - (p+q) \sum_{i=1}^{n} \log \left[1 + \left(\frac{x_i}{b} \right)^a \right] \end{aligned} \tag{3.14}$$

Deriving with respect to a, b, p, q the following partial derivatives are obtained

[15] R users may referred to VGAM package, GB2 package and SN package

(C. and S. (2003)):

$$\frac{n}{a} + p\sum_{i=1}^{n} \log\left(fracx_i b\right) = (p+q)\sum_{i=1}^{n} \log\left(\frac{x_i}{b}\right)\left[\left(\frac{b}{x_i}\right)^{a} + 1\right]^{-1}, \tag{3.15}$$

$$np = (p+q)\sum_{i=1}^{n}\left[\left(\frac{b}{x_i}\right)^{a} + 1\right]^{-1}, \tag{3.16}$$

$$n\psi(p+q) + a\sum_{i=1}^{n} \log\left(\frac{x_i}{b}\right) = n\psi(p) + \sum_{i=1}^{n} \log\left[\left(\frac{x_i}{b}\right)^{a} + 1\right], \tag{3.17}$$

$$n\psi(p+q) = n\psi(q) = \sum_{i=1}^{n} \log\left[\left(\frac{x_i}{b}\right)^{a} + 1\right] \tag{3.18}$$

where $\psi(.)$ is digamma function, the derivative of the function $\Gamma(.)$

$$\psi(x) = \frac{d}{dx} ln\Gamma(x) = \Gamma'(x)/\Gamma(x) \tag{3.19}$$

The system can be solved using a numerical optimization. Algorithms for numerical optimization are used to find the maximum (or minimum) of a function when the function has not explicit solutions. One of the most used iterative method to calculate an approximation of the maximum loglikelihood estimates is the Newton-Raphson method. This method of optimization consist on to find the root of the equation $f'(x) = 0$. The first-order approximation of this function around the n-th approximation x_n of the true solution x^* is (Dennis and Schnabel (1983)):

$$f'(x^*) \sim f'(x_n) + f''(x_n)(x^* - x_n). \tag{3.20}$$

where $f'(x_n)$ is the gradient of f at x_n and $f''(x_n)$ is the Hessian at x_n. Since $f'(x^*) = 0$, we can solve for $\Delta x_n = x^* - x_n$ by solving:

$$f''(x_n)\Delta x - n = -f'(x_n) \tag{3.21}$$

One of the main disadvantages of the Newton-Raphson method is that it requires second derivatives. When to calculate the Hessian is impractical or costly quasi-Newton algorithm can be used. In quasi$-$Newton methods, the idea is to use matrices which approximate the Hessian matrix, instead of exact computing of the Hessian matrix. To solve the system in eq.(3.15)-eq.(3.18) we use the Broyden$-$Fletcher$-$Goldfarb$-$Shanno (BFGS) method[16] (Avriel (2003)).

[16]R user may referred to GB2 package

The parameters standard errors are estimated with a sandwich estimator and with the bootstrap method (see Gentle (1994)). A sandwich variance estimator for the vector of parameters $\hat{\theta}$ is given by:

$$\hat{Var}(\hat{\theta}) = [l''(\hat{\theta})]^{-1}\hat{Var}(l'(\hat{\theta}))[l''(\hat{\theta})]^{-1} \tag{3.22}$$

where $l'(\hat{\theta})$ and $l''(\hat{\theta})$ are the sums of first and second derivatives of the log density (Freedman (2006), Pfeffermann and Sverchov (2003), Graf and Nedyalkova (2010)).

The log likelihood of skew t-distribution has been developed by Azzalini and Capitanio (2003) . Azzalini also provided a suite of R routines for evaluating the log-likelihood and its derivatives. [17].

Moments for the skew t-distribution can be calculated remembering that a variable $Y \sim st(\mu, \omega, \alpha, \nu)$ could be expressed as: $Y = \mu + V^{-\frac{1}{2}}X$ where $V^2 \sim \chi_\nu/\nu$, $X \sim SN(\mu + \omega Z)$ and V is independent on X. If $E(Y^{(m)})$ is a moment of order m, than:

$$E(Y^{(m)}) = \mu + E(V^{-\frac{(m)}{2}})E(X^{(m)}). \tag{3.23}$$

It is known that:

$$E(V^{-\frac{(m)}{2}}) = \frac{(\nu/2)^{(m/2)}\Gamma\left(\frac{\nu-m}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \tag{3.24}$$

while the moment generating function for a $SN(\mu, \omega^2\alpha)$ is given by:

$$mgf_y(t) = 2exp(\mu t + \frac{\omega^2 t^2}{2})\Phi(\delta\omega t) \tag{3.25}$$

where $\delta = \alpha/\sqrt{1+\alpha^2}$.

3.3.4 Inequality Indices

The literature on Lorenz curves and inequality measures is very wide. In this section we shall present only the basic results. The Lorenz curve was introduced in 1905 as a powerful method of illustrating the inequality of the wealth distribution. Suppose to have n people with income $y_1 ... y_n$ to plot the Lorenz curve we must calculate:

$$p_i = \frac{i}{n}, \tag{3.26}$$

[17]The SN package is available at http://azzalini.stat.unipd.it/SN.

$$q_i = \frac{\sum_{j=1}^{i} y_{(j)}}{\sum_{j=1}^{n} y_{(j)}}, \tag{3.27}$$

where p_i and q_i represent the cumulate share of the first i people and the cumulate share of the income of the first i people respectively. To drawn the Lorenz curve the points with coordinates (p_i, q_i) are interpolated linearly. The diagonal of the unit square corresponds to the Lorenz curve of a society in which everybody receives the same income and thus serves as a benchmark case against which actual income distributions may be compared with.

The most commonly used measure of inequality is the Gini coefficient (Gini (1909b)). The coefficient varies between 0, which reflects complete equality and 1, which indicates complete inequality (one person has all the income or consumption, all others have none). Graphically, the Gini coefficient can be easily represented as twice the area between the Lorenz curve and the line of equality (C. and S. (2003)):

$$G = 1 - 2\int_0^1 L(u)du \tag{3.28}$$

where $L(u)$ is the Lorenz curve. Among the several formula of the Gini coefficient one of the most used is (Xu (2004)):

$$G = \frac{2}{N^2\overline{y}} \sum_{i=1}^{N} i(y_{(i)} - \overline{y}) \tag{3.29}$$

where N is the population size, $\overline{y}$ is the average income and individuals are ordered in non decreasing order. In this formulation the Gini coefficient can be interpreted as the espected gap between the wages of two random selected individuals. As stressed by Jenkins (2007) the main disadvantage of Gini coefficient is that it is relatively sensitive to income difference only around the mode of the distribution. Other inequality indices, sensitive to income difference in different areas of the income distribution, are the indices belonging to the generalized entropy (GE) class of inequality measures. The GE class of inequality measures $I(\alpha)$ is defined, for $\alpha \neq 0, 1$, as (Cowell and Kuga (1981)):

$$I(\alpha) = \frac{\nu^a \mu^{-1} - 1}{\alpha(\alpha - 1)}, \tag{3.30}$$

where

$$\nu^\alpha = \int y^\alpha dF(y). \tag{3.31}$$

and $F(y)$ is the cdf for y. The index for $\alpha = 0$ and for $\alpha = 1$ are obtained as limiting case of $I(\alpha)$.

$$I(0) = \lim_{\alpha \to 0} I(\alpha) = log\mu + \nu_0 \tag{3.32}$$

where $\nu_0 = \int logydF(y)$ and μ is the expected value of y.

$$I(1) = \lim_{\alpha \to 1} I(\alpha) = \frac{\mu}{\nu_1} - log\mu \tag{3.33}$$

where $\nu_1 = \int ylogydF(y)$. $I(0)$ is the mean logarithmic deviation (MLD) and $I(1)$ is the Theil index. Measures from the GE class are sensitive to changes at the lower end of the distribution for α close to zero, equally sensitive to changes across the distribution for α equal to one (Theil index), and sensitive to changes at the higher end of the distribution for higher values. Following Jenkins (2007) and Jenkins (2009) GE class of inequality indices are provided for $\alpha = (-1, 0, 1, 2)$. Inequality in earnings distributions has been summarized in terms of the Gini coefficient and of four inequality indices derived from the General Entropy (GE) class of inequality measures $I(\alpha)$ (see Jenkins (2007)). Bootstrap standard errors for all indices used to summarized inequality are computed.

Inequality indices for the GB2 distribution and for the skew t-distribution

The expressions for the inequality indices for a GB2 model has been provided by Jenkins (2007), in particular: the bottom sensitive index $I(-1)$ is given by:

$$I(-1) = -\frac{1}{2} + \frac{\Gamma(p - \frac{1}{a})\Gamma(q + \frac{1}{a})\Gamma(q - \frac{1}{a})}{2\Gamma^2(p)\Gamma^2(q)}. \tag{3.34}$$

The top sensitive index $I(2)$ is given by:

$$I(2) = -\frac{1}{2} + \frac{\Gamma(p)\Gamma(q)\Gamma(p + \frac{2}{a})\Gamma(q - \frac{2}{a})}{2\Gamma^2(p + \frac{1}{a})\Gamma^2(p - \frac{1}{q})} \tag{3.35}$$

The mean logarithmic deviation index is given by:

$$I(0) = \Gamma(p + \frac{1}{a})\Gamma(q - \frac{1}{a}) - \Gamma(p) - \Gamma(q) - \frac{\psi(p)}{a} + \frac{\psi(q)}{a} \tag{3.36}$$

and the Theil index is:

$$I(1) = \frac{\psi(p + \frac{1}{a})}{a} - \frac{\psi(q - \frac{1}{a})}{a} - \Gamma(p + \frac{1}{a}) - \Gamma(q - \frac{1}{a}) + \Gamma(p) + \Gamma(q). \tag{3.37}$$

For the skew t-distribution is not possible to calculate $I(-1)$ from eq.3.30, since it is necessary to know explicitly ν^{-1}. It could be proof (see Cressie et al. (1981)) that for a positive random variable X,

$$E(X^{-1}) = \int_{-\infty}^{0} M_X(t)dt. \tag{3.38}$$

where $M_X(t)$ is the moment generating function for X. Substituting in eq. 3.38 the expression for the mgf for the $SN(\mu, \omega^2\alpha)$ given in eq. 3.25, the expression for the the moment of order $m = -1$ is obtained:

$$\int_{-\infty}^{0} 2e^{\left(\mu t + \frac{\omega^2 t^2}{2}\right)} \Phi(\delta\omega t)dt \tag{3.39}$$

The integral in eq.(3.39) is equal to:

$$\int_{-\infty}^{0} 2e^{\left(\mu t + \frac{\omega^2 t^2}{2}\right)} \left[\frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_{\delta\omega t}^{\infty} e^{-x^2} dx\right] dt =$$
$$\int_{-\infty}^{0} e^{\left(\mu t + \frac{\omega^2 t^2}{2}\right)} dt + \int_{-\infty}^{0} \left[\frac{2}{\sqrt{2\pi}} e^{\left(\mu t + \frac{\omega^2 t^2}{2}\right)} \int_{\delta\omega t}^{\infty} e^{-x^2} dx\right] dt,$$

but first integral, calculate in $t = -\infty$, is not finite

$$\int_{-\infty}^{0} e^{\left(\mu t + \frac{\omega^2 t^2}{2}\right)} dt = \left[\frac{e^{\mu t + \frac{\omega^2 t^2}{2}}}{\mu + \omega^2 t}\right]_{-\infty}^{0}.$$

For this reason is not possible to calculate $I(-1)$. To overcame this problem is possible to calculate inequality index $I(-1)$ from a random sample drawn from a skew t-distribution. To check if $I(-1)$ calculated from a random sample is a good approximation to $I(-1)$ calculated directly from the moment generating function the two index has been compared for the GB2 model. The approximation is reasonable for all the inequality indices (results are showed in Appendix). For this reason and for homogeneity purpose all the four indices are calculated from a random sample. In details, when the skew t-distribution was the best-fitting model, 1000 random samples with the same size of the empirical sample has been drawn. Then the inequality indices has been calculate, with their standard errors. In particular the formula for inequality indices for $\alpha \neq 0, 1$ is (C. and S. (2003)):

$$I(\alpha) = \frac{1}{N\alpha(\alpha - 1)} \sum_{i=1}^{N} \left[\left(\frac{y_i}{\bar{y}}\right)^{\alpha} 1\right], \tag{3.40}$$

while for $\alpha = 1$ the inequality index is given by:

$$I(1) = \frac{1}{N} \sum_{i=1}^{N} \left[\frac{y_i}{\bar{y}} ln \left(\frac{y_i}{\bar{y}} \right) \right], \tag{3.41}$$

and for $\alpha = 0$

$$I(0) = \frac{1}{N} \sum_{i=1}^{N} ln \left(\frac{y_i}{\bar{y}} \right). \tag{3.42}$$

3.3.5 Empirical results

The daily wage is fitted with all the distributions described in the previous sections. The analysis was performed for the whole sample, for the men sub-sample and for the women sub-sample. For each sample three years are analyzed: the first and the last available (1985 and 2004) and an intermediate one (1995). Thus the distributions was fit for 9 different data sets.

Distributions were compared on the basis of the likelihood ratio test (where applicable) and on the basis of other goodness of fit measures. Following Dastrup et al. (2007), table 3.9 shows the log likelihood value for the "best fitting" two - three - and four parameters models. The lognormal distribution results to be the best two parameters model in seven cases on nine, while among the three parameters distribution Sing-Maddala and Dagum are equally frequent the best fit. Among the four parameters model the GB2 is the best fitting model in 6 cases on 9, while in three cases the skew-t distribution is preferable. The four parameters GB2 provides a statistically significant improvement relative to its nested distributions in all the 8 cases where was possible to perform a log-likelihood test.

Table 3.9
Best fitting models

		Two parameters		Three-Parameters		Four Parameters	
		model	Log-L	model	Log-L	model	Log-L
Whole sample	1985	Gamma	-212620.8	Dagum	-210339.4	GB2	-209889.5
	1995	Lognormal	-230987.5	SM	-228495.3	GB2	-228035.0
	2004	Lognormal	-274593.6	Dagum	-271948.6	Skew t	-271145.8
Male	1985	Gamma	-148996.0	Dagum	-147218.1	GB2	-146805.7
	1995	Lognormal	-155494.5	SM	-154021.6	GB2	-153842.2
	2004	Lognormal	-172063.0	SM	-170945.0	Skew t	-1700787.0
Female	1985	Lognormal	-60352.77	Dagum	-58534.5	GB2	-58424.9
	1995	Lognormal	-70240.32	Skew-Normal	-62152.77	GB2	-59776.2
	2004	Lognormal	-100078.0	SM	-98885.72	Skew t	-98379.9

Tables from 3.10 to 3.13 show the estimated parameters, and their standard errors, for the best fitting models for the wage distribution by gender and year. For the years 1985 and 1995 the GB2 is the best fitting model for all the three samples (all workers, male and female), while in 2004 the skewed-t distribution is the best model to describe the data.

It is interesting to analyzed the evolution of the wage distribution for the three samples. To perform this analysis we compare the estimated parameters of the GB2 distribution by gender and year. The overall shape of the distribution is governed by the parameter α, while p and g govern the left tail and the right tail of the distribution, respectively. The lower are the values of p and q the heavier are the tails of the distribution. Comparing the distribution for all the workers in 1985 and ten years later, we notice that a and p increase while q decreases. Between 1985 and 1995 the wage distribution shift on the right and the heaviness of the right tail increases. The effect on the wage is an increase of the modal wage, furthermore the probability to select randomly a high wage was higher in 1995 than in 1985.

Table 3.10

Parameters estimation for a GB2 model. Whole sample.

Par	1985				1995			
	Coef	S.E	Z	S.E boot	Coef	S.E	Z	S.E boot
a	13.44	0.407	33.01	0.73	14.61	0.52	28.27	0.75
b	30.32	0.089	341.14	0.15	27.27	0.07	378.81	0.11
p	0.251	0.009	30.12	0.02	0.331	0.01	26.01	0.02
q	0.309	0.012	27.58	0.02	0.217	0.01	25.11	0.02

The parameters dynamic for Italian data set between 1985 and 1995 is different from the dynamic showed in other Countries. For instance, Brachmann et al. (1996), who analyzed the household income in Germany between 1984 and 1993, show a different dynamics where a decreases while p and q increases, so that the tails of the distribution become less heavy and the variance of the data shrinks.

The dynamic founded by Brachmann and his coauthor is similar to the dynamic of the GB2 parameters, between 1985 and 1995, for the male subsample (Table 3.11) where: a decreases, p increases while q remains stable. The effects of this dynamic on the size distribution of the wage are a decreasing of the modal value, a lighter left tail (i.e. the probability of select randomly a low wage is lower in 1995 than ten years before) and a substantial invariance of the right tail.

The female sub-sample shows, between 1985 and 1995 a different dynamic respect to the male sub-sample (see Table 3.12). Parameters a and p increase while q decreases, so that, between 1985 and 1995, the modal value of daily wage increases, the left tail become lighter while the right tail become heavier

Table 3.11
PARAMETERS ESTIMATION FOR A GB2 MODEL. MEN

Par	1985				1995			
	Coef	S.E	Z	S.E boot	Coef	S.E	Z	S.E boot
a	15.81	0.613	25.77	0.657	11.35	0.46	24.75	0.511
b	31.84	0.098	324.90	0.102	28.86	0.10	280.53	0.112
p	0.237	0.010	23.24	0.012	0.419	0.02	21.79	0.022
q	0.273	0.012	22.56	0.014	0.262	0.01	21.41	0.011

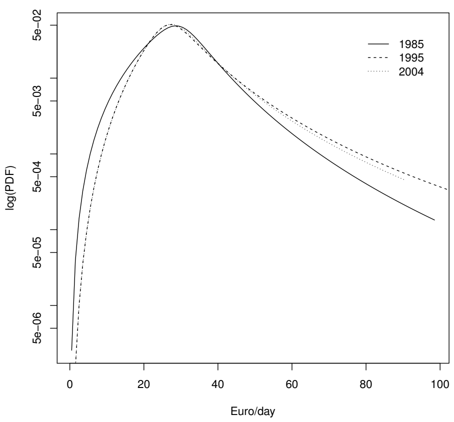
(i.e. the probability of select randomly a low wage is lower in 1995 than in 1985, while the probability of select randomly a high wage is higher in 1995 than in 1985).

Table 3.12
PARAMETERS ESTIMATION FOR A GB2 MODEL. WOMEN

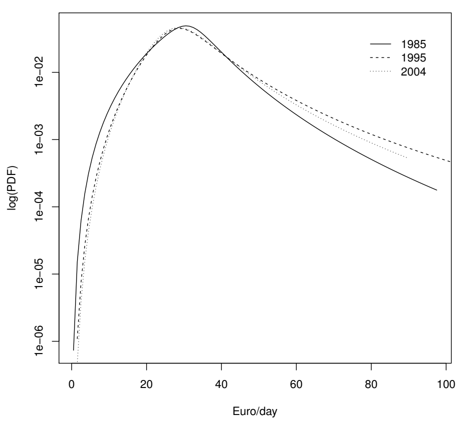
Par	1985				1995			
	Coef	S.E	Z	S.E boot	Coef	S.E	Z	S.E boot
a	21.73	2.710	8.02	1.620	28.524	0.41	70.07	0.53
b	27.14	0.128	212.53	0.130	25.741	0.09	289.55	0.11
p	0.154	0.020	7.74	0.014	0.190	0.01	19.18	0.015
q	0.257	0.036	7.14	0.028	0.162	0.01	11.40	0.011

Even if for 2004 the best theoretical model to describe the daily wage distribution is the skewed-t, we have estimated also the parameters of the GB2 model. In this way we can compare the 2004 estimates with the estimates of the previous years. The results are summarized in Figure 3.11. The comparison between the daily wage distribution by gender and years points out three facts:

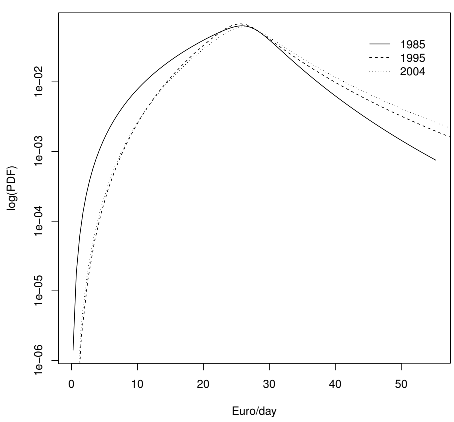
- for all the three samples (the whole sample, the men subsample and the female subsample), the main changes in the wage distribution occurred between 1985 and 1995 for all the three samples;
- The dynamic of the whole sample is mainly driven by the dynamic of the male sub sample, since men represent, at least, the 63% of the entire sample;
- The dynamic of the female sample, in terms of parameters q, is different respect with the dynamic of the male sample. The estimated parameter q for the female distribution decreases monotonically between 1985 and 2004, so that the heaviness of the right tail of the distribution is monotonically increasing between 1985 and 2004.



(a) Whole sample



(b) Male



(c) Female

Figure 3.11
Daily wage distribution by gender and year.

Inequality

Table 3.14 shows the Gini coefficent, the bottom sensitive index $I(-1)$, the mean logarithmic deviation $I(0)$, the Theil index $(I(-1))$ and the top sensitive index or the half the squared coefficient of variation $I(2)$. For the first two years indices are calculated for a GB2 model for the last year (2004) indices are calculated for a skew t-distribution.

The dynamic of the inequality indices points out that the wage distribution changes, between 1985 and 2004, especially in the right tail, namely for

Table 3.13
Parameters estimation for a Skew t-distribution. Whole sample.

Par	Whole Sample			
	Coef	S.E	Z	S.E boot
location	21.23	0.06	85.02	0.07
scale	11.73	0.10	37.91	0.11
shape	2.43	0.03	15.30	0.03
df	3.20	0.05	14.81	0.05

Par	Male sample			
	Coef	S.E	Z	S.E boot
location	21.56	0.08	76.19	0.07
scale	13.51	0.138	36.10	0.145
shape	2.81	0.037	14.61	0.041
df	3.77	0.09	12.64	0.09

Par	Female sample			
	Coef	S.E	Z	S.E boot
location	21.46	0.096	69.23	0.083
scale	8.59	0.118	24.94	0.128
shape	2.68	0.030	9.72	0.036
df	3.69	0.059	11.01	0.072

the higher wages. Inequality between 1985 and 2004 decreases among the low wage (I(-1) index is the only inequality index that decreases from 1985 to 2004) and increases for the high wage. The dynamic of the inequality indices can be divided in two different stages. An early stage between 1985 and 1995 characterized by a high increase in inequality especially on the top of the distribution. And a last stage, from 1995 to 2004, characterized by a general decreasing of the inequality indices. These results are coherent with the standard findings for Italian inequality. The wage inequality dynamic in Italy was analyzed by Brandolini et al. (2002) which used the survey of Bank of Italy (SHIW). The authors investigated the dynamics of net wages in Italy between 1977 and 1998 and pointed out that at the beginning of 1990s an increase in earnings inequality took place (especially in 1992-1993) due to the economic crisis (1992) and the abolition of the wage index mechanism, the so-called Scala Mobile[18] (see also Manacorda (2004)). From 1993 to 1998 Brandolini et al. (2002) argued that inequality remains unchanged. Jappelli and Pistaferri (2010) analyzed the SHIW data from 1980 and 2006 and found that "*inequality grows quite dramatically between 1989 and 1995 and is basically flat thereafter*". The same results was pointed out by Devicienti (2003) analyzing the WHIP data set from 1985-1996.

[18] With the Scala Mobile mechanism the wage was indexed to the cost of life

Table 3.14
INEQUALITY INDICES. WHOLE SAMPLE

		I(-1)	I(2)	I(0)	I(1)	Gini
1985	coeff	0.0825	0.0852	0.0723	0.0730	0.20
	s.e boot	(0.002)	(0.002)	(0.043)	(0.034)	(0.002)
1995	coeff	0.0929	0.1816	0.0943	0.1125	0.23
	s.e boot	(0.002)	(0.003)	(0.03)	(0.003)	(0.002)
2004	coeff	0.0741	0.1044	0.0801	0.1021	0.22
	s.e boot	(0.001)	(0.005)	(0.05)	(0.001)	(0.001)

The inequality analysis proposed in literature for the Italian data does not distinguish between male and female (see Brandolini et al. (2002), Manacorda (2004), Jappelli and Pistaferri (2010) and Devicienti (2003)). The results we found for the whole sample and for male sub-sample agree with the findings proposed in literature (the dynamic of the whole sample is mainly driven by the male sub-sample since men represent, at least, the 63% of the entire sample) while the results for the female sub-sample are different. Inequality for men wage is summarized in Table 3.15. Over the first 10 years of observation, inequality increases in particular for the high wages. Between 1995 and 2004 a general decline of inequality is observable.

Table 3.15
INEQUALITY INDICES. MEN

		I(-1)	I(2)	I(0)	I(1)	Gini
1985	coeff	0.0741	0.0809	0.0668	0.0685	0.194
	s.e	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
1995	coeff	0.0976	0.1946	0.0989	0.1185	0.241
	s.e	(0.001)	(0.006)	(0.001)	(0.002)	(0.002)
2004	coeff	0.0832	0.0923	0.07174	0.0759	0.216
	s.e	(0.004)	(0.003)	(0.001)	(0.001)	(0.002)

Inequality within woman shows a different pattern (Table 3.16). Between 1985 and 1995 all the inequality indices increase except I(-1) that decreases, as happened for the whole sample, but from 1995 and 2004 inequality continues to rise slightly.

Some theories, explaing the dynamics of inequality in OECD countries, pointed out the importance of the skill based technological change. Quoting The New Palgrave Dictionary of Economics (Durlauf and Blume (2008)): *"Skill-biased technical change is a shift in the production technology that favours skilled over unskilled labour by increasing its relative productivity and, therefore, its relative demand. Traditionally, technical change is viewed as factor-neutral. However, recent technological change has been skill-biased.*

Table 3.16
Inequality indices. Women

		I(-1)	I(2)	I(0)	I(1)	Gini
1985	coeff	0.0702	0.0572	0.0582	0.0548	0.1770
	s.e	(0.001)	(0.001)	(0.052)	(0.054)	(0.001)
1995	coeff	0.0651	0.0984	0.0650	0.0732	0.2410
	s.e	(0.002)	(0.002)	(0.041)	(0.032)	(0.002)
2004	coeff	0.093	0.1093	0.0689	0.0759	0.2160
	s.e	(0.015)	(0.024)	(0.003)	(0.002)	(0.002)

Theories and data suggest that new information technologies are complementary with skilled labour, at least in their adoption phase. Whether new capital complements skilled or unskilled labour may be determined endogenously by innovators economic incentives shaped by relative prices, the size of the market, and institutions. The factor bias attribute puts technological change at the center of the income-distribution debate." (the so called STBC), see also Acemoglu (2002)). Standards STBC theories are associated to an increase of wage inequality. Furthermore, recent empirical works (see Autor et al. (2006) Goos and Manning (2007) and Goos et al. (2009)) show that upper tail and the lower tail of the wage distribution are characterized by different patterns of the inequality trends. An explanation of these phenomena can be related with the changes in the structure of the job quality distribution. Autor et al. (2006) and Goos and Manning (2007) found that non-routine tasks are concentrated in high-paid and low-paid service jobs (the top and the bottom of the distribution) while routine tasks are concentrated in manufacturing and clerical works (the middle of the distribution). Autor et al. (2006), Goos and Manning (2007) and Goos et al. (2009) argued that from the early 1990s in USA and in Europe technologies are becoming more intense in the use of non-routine tasks concentrated in high-paid and low-paid service job.

The different trends of the inequality recorded within male sub-sample and and female sub-sample suggests us to investigate the dynamics of wage among the two groups defined by the gender. To analyzed the dynamics of the wage between 1985 and 2004 a regression is used.

The regression model is builded starting from the Mincerian wage regression (Mincer (1974)). Mincer developed a model of earnings to estimate returns to schooling, returns to schooling quality, and to measure the impact of work experience on male-female wage gaps (Heckman et al. (2003)) In the standard form of the Mincer earnings model, log earnings are regressed on a constant term, a linear term in years of schooling, and linear and quadratic terms in years of labor market experience. Formally a standard Mincerian wage regression can be expressed as:

$$log(W_t) = \beta_0 + \beta_1 s + \beta_2 z + \beta_3 z^2, \tag{3.43}$$

where W_t is the individual wage al time t, s represents the years of education and z keeps the years of post-schooling investment in human capital into account (Heckman et al. (2003)). Since, in our dataset, the informations about the years of education and the years of experience are not available, we have choose the skill level and the age as proxies for the education and labour market experience respectively.

The analysis is performed with a fixed effects model. To decide between a fixed effects model or a random effects model a Hausman test was performed. Since the null hypothesys is rejected (i.e. the error components are correlated with the regressors) we choose to perform this analysis with a fixed effects method. The estimated model is:

$$\begin{aligned} y_{it} = \beta_0 + \beta_1 age_{it} + \beta_2 age_{it}^2 + \beta_3 contract_{it} + \beta_4 illness_{it} + \beta_5 migration_{it} + \\ + \gamma skills_{it} + \delta work\_area_{it} + \alpha sector_{it} + \psi_t year_{it} * sex + \omega_t year_{it} + \epsilon_{it} \end{aligned} \tag{3.44}$$

where y_{it} is the logarithm of the daily wage; age_{it} is the age at time t of the $i-th$ individuals and age_{it}^2 is the square of age_{it}. The quadratic term is used to model a non linear effect of age. $Contract_{it}$ is a dummy variable taking value 1 if the $i-th$ individual at time t has a permanent contract and zero otherwise; $illness_{it}$ is a dummy variable taking value 1 whether in the reference year the $i-th$ worker received an illness benefit; $migration_{it}$ is a dummy variable that takes value 1 if the working geographical area differs from the geographical born area; $skills_{it}$ indicates the skill of individual $i-th$ at time t; $work\_area_{it}$ indicates the geographical area of Italy where employment was performed; $sector_{it}$ indicates the industrial sector and $year$ is the calendar year. Since the fixed effects model removes variables that are time constant is not possible to include the gender as one of the covariates in the model. To overcome this problem and to estimate how the gender gap has changed over time we add to the model the interaction between gender and time. Even thought is not possible to identify the effect of gender in any particular time period t, the coefficients $\psi_2 \dots \psi_{20}$ are identified, and therefore the differences in partial effects on time-constant variables relative to the base period can be estimated (see Wooldridge (2010)). The estimated coefficients of the model in eq.3.44 are showed in table 3.17.

The estimated coefficients are generally in line with the standard economic interpretation. The age has a quadratic effect with the linear term positive and the squared term negative. The parabolic function given by the

Table 3.17
Fixed effect estimates of the logarithmic daily wage

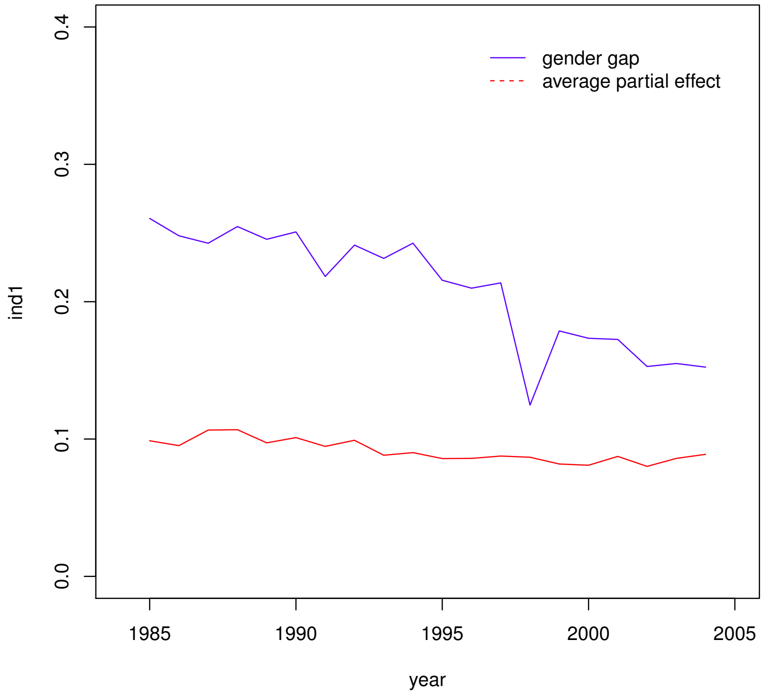
lwage	Coef.	Std.
Age	0.0416***	0.0011
Age2	-0.0003***	0.0000
Ptime	0.0713***	0.0013
Contract	0.0309***	0.0011
Illness Benefit	-0.0435***	0.0008
Migration	-0.0135***	0.0014
Skill: Blue Collar	0.2125***	0.0018
Skill: White Collar	0.2999***	0.0021
Skill: Manager	0.5515***	0.0033
Work area: North East	-0.0067	0.0027
Work area: Centre	-0.0211***	0.0029
Work area: South-Island	-0.033***	0.0028
Sector D	0.0446***	0.0029
Sector E	0.0928***	0.0081
Sector F	0.0949***	0.0032
Sector G	0.031***	0.0030
Sector H	0.0594***	0.0034
Sector J-I	-0.0389***	0.0029
Sector L-M-N-O-P-Q	-0.0589***	0.0031
Cons	2.1055	0.02978
Interaction sex-year dummies	✓	
Year dummies	✓	
sigma_u	0.36627822	
sigma_e	0.24892871	
rho	0.68405136	

Notes:

1. ** Significant at the 5% level; *** Significant at the 1% level.
2. Base categories for categorical variables are: Apprentice (Skill), North-West (Work area), Sector A-B-C-E (Sector).
3. Sectors A-B-C-E include: Agriculture, hunting and forestry; Fishing; Mining and quarrying. Sector D includes: Manufacturing. Sector E includes: Electricity, gas and water supply. Sector F includes:Construction. Sector G includes: Commerce. Sector H includes: Hotels and restaurants. Sectors I-J-K includes: Transport and communications; Financial intermediation;Business services. Sectors L-M-M-O-P-Q includes:Public administration and defence, compulsory social security; Education; Health and social work; Other community, social and personal service activities; Activities of households; Extra-territorial organizations and bodies.
4. Coefficients for Interaction between gender and year dummies and coefficients for the year dummies are all significant at the 5% level.

coefficients of *age* and age^2 reachs its maximum for an age of 69 years (i.e out of the working age range). The implication is that wage increases with age at decreasing rates until retirement. Coefficients for *Ptime* and *Contract* are positive and significant, while coefficients for *Illines* and *Migration* are negative. The employment positions *Blue Collar, White Collar* and *Manager* are associated with higher wages respect to the base category *Apprentice.* The working areas Center and South and Islands are associated with lower wages respect to the base category North-East, while the difference between North West and North East is not significative. From the estimated model the average partial effects[19] of the interaction between gender and time are calculated.

Figure 3.12
PREDICTIVE MARGINS



In Fig. 3.12 the gender gap calculated as the difference between the logarithmic male wage and the logarithmic female wage (blue line in the graph) is compared with the gender gap obtained as difference in the average partial effects (red line in the graph). The estimation results, after controlling for some individual specific time-varying variables, highlights to facts. Firstly, the average wage is still higher for males over 20 years. Wage gender gap, calculated as difference in average partial effects, is lower than gender gap calculated as the logarithmic difference between male and female wage. Secondly, the gender gap is constant between 1985 and 2004. Considering a men and a women of the same age and with the same skills, working in the same

[19]To calculate the average partial effects a marginal effect is computed for each case, and the effects are then averaged.

industrial sector and in the same geographical area, the men wage is always higher than the women wage for each time period considered in the analysis. Furthermore, the wage gap is not shrinking over the time. Therefore the shrink of the gender gap showed in Fig. 3.9 is not the effect of better wage conditions for women.

This phenomena could be a consequence of the female dynamic within the labour market. Tab 3.18 shows the distribution of workers by gender and skill. Men are concentrated in the blu-collar category both in 1985 and in 2004. In 1985, 54% of women belong to the blu-collar category while in 2004 this percentage shrinks to about 44%. In the same time the percentage of women working as white-collar rise from about 40% to in 1985 to 48% in 2004. Between 1985 and 2004, women have increased their skills more than men while the gender gap remained constant over time. As a consequence it could be possible to detected a higher growth in average female wage due to an higher percentage of women working in high-paid job positions.

Table 3.18

Distribution of workers by gender and skills

Men			
Skill	**1985**	**1995**	**2004**
Apprentice	4.83	3.91	5.66
Blu-Collar	69.00	66.42	66.67
White-Collar	24.97	28.24	23.96
Cadre and Manager	1.23	1.44	4.31
Women			
Skill	**1985**	**1995**	**2004**
Apprentice	6.39	4.10	6.69
Blu-Collar	53.88	42.56	43.78
White-Collar	39.67	47.16	48.00
Cadre and Manager	0.09	0.18	1.57

3.4 Conclusions

This chapter analyzed the wage distribution in Italy between 1985 and 2004. The preliminary analysis of the data underlines two main changes in the sample composition between 1985 and 2004. Firstly, in 1985 women represented the 30% of the total sample, after 20 years this percentage rises at 37%. Secondly, in the twenty years of observation, an ageing of the sample occurs. Changes in the age distribution from 1985 to 2004 can be the results of a lagged entry in labour market and/or of an increase of atypical contract as main contract for young people, and/or of a delayed exit from the labour market. The analysis of the wage by gender highlights that female average daily wage growths more than the male daily wage between 1985 and 2004. Nevertheless the average daily wage perceived by women is lower than the males ones for the whole observational period.

The empirical wage distributions (by year and by gender) are fitted with 10 different models with two, three or four parameters. The distributions are compared on the basis of the likelihood ratio test (for nested models) and on the basis of other goodness-of-fit measure (such as the chi-square). For 1985 and 1995 the best fitting model is the GB2 distribution, while in 2004 data are better represented by a skewed-t distribution. For these two distributions, four inequality indices are calculated.

The analysis of inequality for the whole sample and the men sub-sample highlights two main facts. On the one hand, the dynamic of daily wage can be divided in two stages: i) an early stage (between 1985 and early 1990s) characterized by a high increase in inequality indices at the top of the distribution and a decrease at the bottom; ii) a last stage characterized by a general decreasing of the inequality indices.

On the other hand, the main changes regard the top of the distribution. These results agree with the results obtained by other authors (see Brandolini et al. (2002), Manacorda (2004), Jappelli and Pistaferri (2010) and Devicienti (2003)). The same analysis on the female sub-sample reveals a different trend of the inequality. The women inequality measures seems to rise slightly also after the 1995. This fact could be a consequence of the skill based technological change (STBC). The different trends of the inequality recorded within male sub-sample and within female sub-sample suggests to investigate the dynamics of wage among the two groups defined by the gender. At a first look seems that the wage gender gap, between 1995 and 2004, was shrinking but a deeper analysis (performed by the means of a fixed effects regression model) shows that the shrinking in the difference between men wage and women wage could be due to an increasing of the percentage of women working in high-paid job and not to an "approaching" of the female wage to the male wage.

Chapter 4

The Size Growth Relation in Pharmaceutical Industry

The first formal model of the firm size dynamics dates back to the well known law of proportionate effects presented by Gibrat in 1931 (Gibrat (1931)). He argued that the firm size distribution follows a lognormal process, which implies the independence between size and growth.

As pointed out in Chapter 2, an approach proposed in literature to check the validity of a stochastic growth model, consists on investigating the determinants of the growth by means of regression methods. Since the 1950s the Gibrat law of proportionate effects has stimulated a multiplicity of empirical works[1]. An early prominent contribution on the investigation of Gibrat's Law was made by Mansfield (1962) who presented the law in three versions and tested its validity accordingly. First, he included all surviving and exiting firms in the sample, setting to -100% the growth rates of firms just dropped off, and observed a negative relationship in seven of the ten cases. Mansfield argued that one of the principal reasons of this failure was that the probability of surviving is lower for smaller firms. Second he considered only surviving firms and showed that the growth was higher for smaller firms. Finally he considered only firms exceeding a "minimum efficient size" [2] and found that in this case "results are quite consistent with Gibrat's Law" (Mansfield (1962)).

According to the second definition of Mansfield, the majority of empirical studies has rejected the Gibrat law claiming that small firms grow faster than larger firms. This negative relation has been found using data for different

[1] An exhaustive survey on firms growth is provided by Coad (2007).

[2] The minimum efficient size was first proposed by Simon and Bonini (1958a).

countries[3], different level of industrial aggregation[4], and different industrial sectors[5]. Only few studies find some support in favour of the Gibrat law[6]. However, when a specific discrimination between small and large firms is done, results are somewhat different. While for small firms the negative relation holds nearly always[7], for large firms a flat relation is typically observed and whenever the Gibrat law is rejected even a positive dependence is found[8].

According to Mansfield's third definition, some studies have tested the Gibrat law for firms above a certain size threshold. For example Droucopoulos (1983) focuses on a sample of the world's largest firms and finds support for it[9]. Mowery (1983) analyzes two samples of small and large firms and finds a negative relation for the former while the Gibrat's law holds for the latter. Cefis et al. (2006), for the worldwide pharmaceutical industry, and Hart and Oulton (1996), for a data set of independent U.K. companies, find a negative relation for pooled data but once the sample is restricted to only large firms the dependence vanishes.

The relationship between size and growth can be skewed if essential variables to explain the size and its evolution over time are no taking into account. Two of the most relevant determinants of growth are considered the age and the innovation of firms. Age is an important variable to study firms behavior over time. The study of the relationship between size and age generally shows that the size distribution of firms varies with firms age. In particular as firms age, their size distribution shifts towards the right-hand side; the mode, as well as the thickness of the right tail, increase (Cabral and Mata (2003), Cirillo (2010)). However, the rate of increase in size generally shrinks as firms get older. A large number of studies find that age reduces the growth rate of firms (see Jovanonic (1982), Evans (1987a), Evans (1987b), Dunne and Hughes (1994), Geroski and Gugler (2004), Yasuda (2005)).

---

[3]Dunne and Hughes (1994), Kumar (1985) studied the quoted UK manufacturing firms, Bottazzi and Secchi (2003), Hall (1987b) the quoted US manufacturing firms, J. Goddard (2002) the quoted Japanese firms, Gabe and Kraybill (2002) establishments in Ohio.

[4]Dunne et al. (1989) analyze plant-level data as opposed to typical firm-level data.

[5]Typically the manufacturing sector but for example Barron et al. (1994) study New York Credit Unions, Weiss (1998) Austrian farms, Liu et al. (1999) Taiwanese electronic plants.

[6]Bottazzi et al. (2005) for French manufacturing firms, Hardwick and Adams (2002) for UK life insurance companies.

[7]See Evans (1987a), Evans (1987b), Yasuda (2005), Calvo (2006), McPherson (1996), Wagner (1992), Almus and Nerlinger (2000). However Audretsch et al. (2004) find results in favour of the Gibrat's law for small-scale Dutch services.

[8]A positive relation was found by early studies of Hart (1962), Samuels (1965), Prais (1974), Singh and Whittington (1975) on data for UK manufacturing firms.

[9]See also Becchetti and Trovato (2002), Geroski and Gugler (2004), Lotti et al. (2003).

As regards innovation the relation with growth is not as clear as the relation between age and growth. While the role of innovation is considered central to the growth of firms (Carden (2005), Hay and Kamshad (1994), Geroski (2000), Geroski (2005)), empirical studies find difficulties in modeling such relation. As suggested by Coad (2007), one of the main problems is the consistent time delay between investments in innovation and the conversion of investments into economic performance. Mansfield et al. (1977) identified innovation as a three-stages process, on the ground of which, only those firms able to get through it are successful innovators. Mansfield (1962) found that successful innovators grow faster especially if their starting size is small.

Some later studies (Scherer (1965), Geroski and Machin (1992), Geroski and Toker (1996)) find a positive relationship between innovation and growth, while others have shown that the relation is not clear (Freel (2000)) or even absent (Bottazzi et al. (2001)).

Moreover Coad (2007) suggests, according to the literature, that another problem consists in the definition of innovation itself. Two popular innovation proxies used are the expenditure in R&D and the patents count but, both these measures have drawbacks though. Expenditure in R&D may not be well associated with the actual output of an innovative process. Patents count cannot discriminate between patents with substantial and marginal economic impact, while it is typically found that the former are a negligible amount. On the other hand, in many theoretical models, innovation is represented by the entry in the market of new business opportunities (Ijiri and Simon (1964), Kalecki (1945), Pammolli et al. (2007). In such a case the innovation can be proxies both with new product launches and with the opening of new product lines, divisions, subsidiaries, and plants (Bottazzi et al. (2001), Pammolli et al. (2007), Growiec et al. (2008)). Unfortunately is not easy to test empirically the effect of the entry of new business opportunities in the market since it is not so common have available this information in empirical data.

The purpose of this work is to investigate the size-growth relation and to test the validity of the Gibrat's rule. Furthermore we want to explore the effect of the innovation on the size growth relation. To perform this analysis we used a set of micro data for firms sales. The data set is an unbalanced panel of 1,152 USA pharmaceutical firms for the period 1996-2007, developed by the IMS.

The USA is nowadays the leading country for the pharmaceutical industry: in 2008, 5 of the top 10 pharmaceutical companies in terms of sales was based in USA. Even if the European population is larger than the American population, the European pharmaceutical sales are a 10% lower than the U.S. sales (Daemmrich (2011)). The America leadership in term of sales is also a consequence of a less rigid regulation especially for the price system respect to the Europe

Since the early 1980s the U.S. pharmaceutical sector has been characterized by a significative consolidation of large firms as well as the entry into the R&D process by small and early stage biopharmaceutical firms. From the middle 1980s the pharmaceutical industry has been experienced also significative legislative changes and market developments. The most important legislative novelty was the Waxman-Hatch Act in 1984 that allowed generic drug[10] to enter the market without the need to do clinical test of safety and efficacy. Generics can be marketed only by demonstrating bio-equivalence with existing and patented drugs (Masi et al. (2003)). In the same period, the market experienced the expiration of patents on major products, the rising cost of R&D and the rise in price competition for generics. The legislative changes and the market developments have caused a decline in sales and profit particularly evident from the middle of 1990s (Grabowski and Kyle (2007)). Nowadays, one of the main issue regarding the pharmaceutical industry is related to the R&D and the innovation. In pharmaceutical industry the cost of innovation is large, the average cost of bringing a new efficacious molecule is estimated between 800 million dollars[11], and risky, to produce a new molecule it can take 12-13 years and only one out of 10,000 molecules will be marketed. From the early 2000s a decline in the productivity of innovation is observable: the annual number of new active substances approved in major markets fell by 50 percent during the 1990s, while private-sector pharmaceutical R&D spending tripled (Cockburn (2004)). In last years many authors have dealt with innovation in pharmaceutical industry (see DiMasi et al. (1991), Masi et al. (2003), Ornaghi (2006), Comanor and Scherer (2011)) focusing their research on the relationship between innovation and merges. The literature claims that conflicting trends confound the pharmaceutical industry. One the one hand, high R&D real costs and the decline of the rate of innovation have been cited (DiMasi et al. (1991), Masi et al. (2003)) as one of the main explanation of the concentration trend showed by the pharmaceutical industry from the 1990s. On the other hand, the increased concentration brought on by recent mergers may have contributed to the declining rate of innovation(Comanor and Scherer (2011)).

According to this results our contribution here is to explore whether the innovation benefits are different for smaller and larger firms. There are two different levels of innovation related to the firms size: the small

[10]A generic drug is defined from the U.S. Food and Drug Administration as "a drug product that is comparable to brand/reference listed drug in dosage form, strength, route of administration, quality and performance characteristics, and intended use"

[11]In year 2000 dollars.

firms are start-ups using the so called "genetic engineering" or biotechnologies while the big pharma companies developed licensing, sponsored R&D and partnerships with biotech in order to join biotechnological innovation (Bobulescu and Soulas (2006)). Most of the authors studied the relationship between innovation and size from the point of view of the existence of scale economies in pharmaceutical industry R&D (Jensen (1987), Graves and Langowitz (1993),Bobulescu and Soulas (2006), Cockburn and Henderson (2001), Miyashige et al. (2007)). The results diverge from one to another especially in relation to the measure of innovation used, however seems that scale economies exist in pharmaceutical innovation. Moreover, Masi and A. (1995) and R.Henderson and Cockburn (1996) showed that R&D costs per new drug approved in the U.S. decrease with firm size, while sales per new drug increase with firm size, but the relationship between innovation and growth by firms size is not investigated.

The paper is organized as follows. Section 4.1 describes the data and provide some preliminary evidence. Section 4.2 discusses the methodology employed to estimate the relation between growth and size. In particular, we apply a growth regression approach to deal with a number of econometric issues which have been arisen in the literature.

First, the panel dimension allows us to account for the effect of time-constant unobserved heterogeneity. Then, making use of an instrumental variables approach, we allow for the violation of the strict exogeneity assumption which comes with the inclusion of size on the right-hand side. This approach is carried out within a Generalized Method of Moments (GMM) framework which delivers efficiency gains in estimation. Estimates are also robust to heteroskedasticity and autocorrelation in the error term. Moreover, information contained in our data set allow us to control for the age and the innovation of firms, which are considered the most relevant determinants of growth. In our model specification we include a variable which proxies the innovation and attempts to overcome both limits of the other proxies of the innovation outlined above. In particular this variable synthesizes both the net inflow of products and the change in Anatomical Therapeutic Chemical (ATC) classification[12]. In this way, on the one side we allow for products

[12]The classification system divides drugs into different groups according to the organ or system on which they act and/or their therapeutic and chemical characteristics. The anatomical first level of the code contains 14 main groups: Alimentary tract and metabolism, Blood and blood forming organs, Cardiovascular system, Dermatologicals, Genito-urinary system and sex hormones, Systemic hormonal preparations, excluding sex hormones and insulins, Anti-infectives for systemic use, Antineoplastic and immunomodulating agents, Musculo-skeletal system, Nervous system, Antiparasitic products, insecticides and repellents, Respiratory system, Sensory organs, Various.

which are actually marketed, on the other side it is likely that a change in ATC captures at least in part the output of the innovation process, unless associated to a negative inflow, so that the problem of temporal lag should be mitigated. As further regressors for our model specification we use a dummy which controls for exiting firms. This is a crude way to account for the sample selection problem, but at least can purge the size effect on growth of at least some part of the contribute to large decrements due to firms embarking on a final stage of their life cycle. In the related empirical literature sample selection has been dealt with by using either the Heckman correction (Harhoff et al. (1998)) or Tobit procedure (Hall (1987b)). Unfortunately it is a somewhat hard task to apply these techniques in conjunction with IV panel methods. Anyway, empirical findings point out that when sample selection is accounted for, the Gibrat coefficient changes only modestly and the negative effect is still supported (Hall (1987b), Evans (1987a)). For the best of our knowledge, only in other few cases (see Oliveira and Fortunato (2003), Ribeiro (2007)) the dynamic panel estimators were used to test the Gibrat's rule in the industrial context[13]. Furthermore we are able to insert in our model important variables usually not available in other models, as the entry and the exit of products in the market and the level of innovation of each new product.

Section 4.3 and 4.4 discuss results respectively when the whole sample of firms is used and when two sub-groups of small and large firms are selected. Section 4.5 concludes.

4.1 Data and Preliminary Evidence

We draw data from the PHID database, which provides information on sales of USA pharmaceutical firms and on other dimensions particularly appropriate for the nature of our study. The database provides sales data for different levels of aggregation. In particular we have information at firm level as well as product level. The different level of aggregation allows us to analyze, at the best of our knowledge for the first time, the role of inflows and outflows of products within the firm, which we will exploit to proxy firm innovation. Furthermore the data base provides for each product the date of the launch in the market (in month and year) and this information allows us to analyzed the role of age.

The data consist in quarterly sales for the US pharmaceutical market at the smallest level of aggregation from 1997 to 2008. Our window of data

[13]Soo (2011) used a dynamic panel estimation to test the relation between size and growth of state population in the United States.

starts in the 3rd quarter of 1997 and ends in the 2nd quarter of 2008 so we have 48 quarterly observations. Our unit of observation consists in packs for a same product. For each unit we have the id of the product and the id of the firm, so we recover products data by aggregation of sales at the unit level and firms data by aggregation at the product level. The data provide information on the launch date of the unit so we can define the age of both products and firms according to it. First, we define the age of the unit as the time elapsed in days from the launch date, then days are divided by 365. Then we define the age of each product as the age of the oldest unit, and the age of the firm as the age of the oldest product. Firms quarterly sales are collapsed by age to yield yearly sales.

Sales data for units which do not have all the four quarterly observations are dropped in order to prevent that in the yearly aggregation some units have yearly sales which derive from the sum of 3 or less quarterly sales. In particular, first, for units born before 1/7/1996, we delete for units born in the 4th, 1st, 2nd quarter respectively 1, 2 and 3 quarterly observations at the beginning of our window of data, and 3, 2 or 1 quarterly observations at the end of our window of data. Second, for units born after 1/7/1996, we delete for units born in the 2nd, 1st and 4th quarter respectively 1, 2 and 3 quarterly observations at the end of our window of data. Anyway, the problems of yearly sales as sum of 3 or less quarterly sales remains both for units which die during the sample window (with regards to the year of death), and for units born again after a death during the sample window (with regards to the year of birth). It is important to remark that, in the light of this data cleaning, we are sure that this underestimation of the yearly sales does not apply to the first year of life (but can apply to the last year of life and to others). Moreover this problem of within-sample death and rebirth is less important for firms[14]. At this stage we aggregate quarterly sales data to have yearly sales. The same issue of aggregation described above applies here. Then, firms with gaps in yearly observations are removed from the sample.

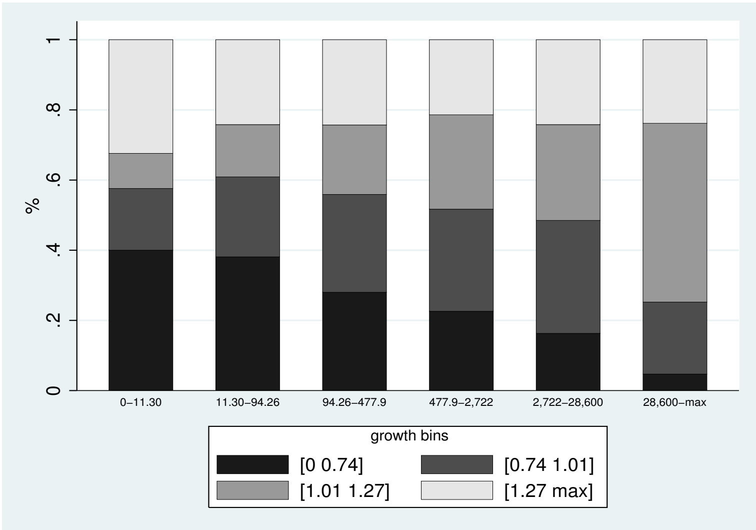
We report some descriptive statistics on the relation between the sales growth and both sales and age.

Fig 4.1 shows the pattern of growth distribution by sales classes. First of all, it is clear that the importance of extreme growth bins drops as the sales increase, which is in accordance with the typical finding that the variance of growth rates is higher for smaller sales. Moreover, though the portion of

[14]We remark that in order to have this hierarchical model we have treated a given product belonging to 2 or more different firms as 2 or 3 different products (yet this happens for only a small portion of cases).

Figure 4.1

GROWTH DISTRIBUTION BY SIZE (THOUSANDS OF POUNDS)



Notes:

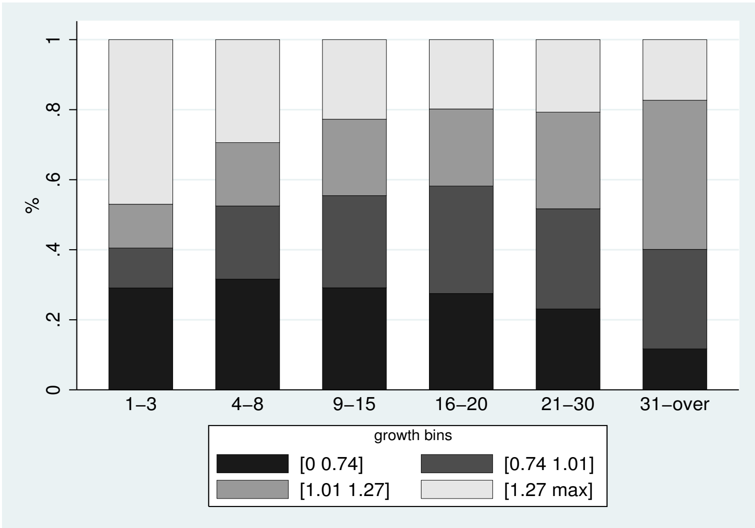
1. The growth is calculated as logarithmic difference between yearly sales. The total observations are divided in 4 growth bins with approximately equal frequencies. Then for each growth bin, observations are distributed over six bins of sales.

positive growths for the first class of sales is higher than the second class, it definitely increases as sales increase. Apparently, the sales growth relationship is negative only for very small sales value, but above a certain threshold it turns out to be positive. As regards the growth distribution by age classes (Fig. 4.2), the portion of positive growths exhibits an hump-shaped movement, in that it shrinks when firms are young, but then increases for firms above 20. The pattern of the portion of growth mid bins is instead more similar to the sales pattern since it is thinner for higher classes.

The data provide the product ATC code which we can use to identify firms patterns in innovation. We attach to each firm the ATC code of the most important products in term of sales share and we interpret variations in ATC in conjunction with changes in the number of products as proxies for firms innovation[15]. Our definition of firm's main ATC may be criticized as for firms with a highly fragmented products portfolio the largest sales share may be small and not representative of the firm. However in our data this problem arises only for a small number of cases. As Fig. 4.3 shows, for less than 20% of firms the share is below 50%, and for less than 10% the share is

[15]The idea to look simultaneously at changes in ATC and changes in number of products is due to the modest variation over time of the first.

Figure 4.2
GROWTH DISTRIBUTION BY AGE



Notes:

1. The growth is calculated as logarithmic difference between yearly sales. The total observations are divided in 4 growth bins with approximately equal frequencies. Then for each growth bin, observations are distributed over six bins of age.

below 30%.

Table 4.1
MEAN GROWTH BY INNOVATION PATTERNS

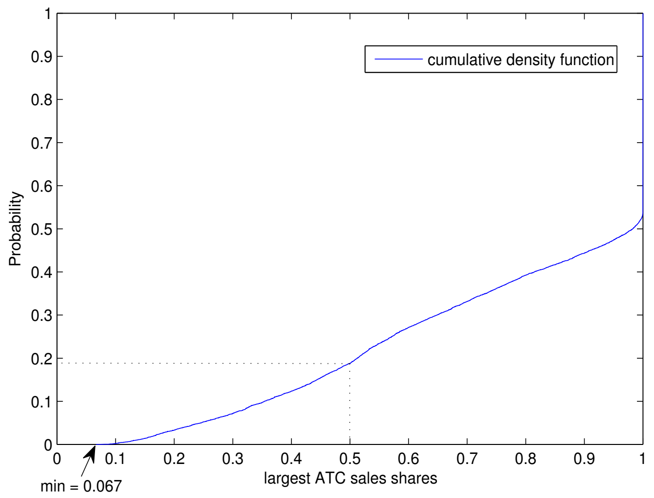
	Mean	*Std. Dev.*	*Min*	*Max*	*Obs.*
$k^{+,c}$	0.67	1.30	-3.41	11.20	335
$k^{=,c}$	-0.10	1.35	-9.06	7.78	286
$k^{-,c}$	-0.76	1.53	-7.86	4.98	335
$k^{+,nc}$	0.30	0.78	-2.97	8.20	1,031
$k^{=,nc}$	-0.15	1.05	-10.37	8.91	3,919
$k^{-,nc}$	-0.27	0.97	-8.81	5.12	1,033

Notes:

1. The $k^{\mathbf{p},\mathbf{q}}$ variables are dummies which indicate firms whose number of products can increase ($k^{+,\mathbf{q}}$), be constant ($k^{=,\mathbf{q}}$) or decrease ($k^{-,\mathbf{q}}$) between two consecutive years, and whose ATC code can simultaneously either change ($k^{\mathbf{p},c}$) or not ($k^{\mathbf{p},nc}$).
2. The growth is calculated as logarithmic difference between yearly sales.

An increase in the number of products may be an important signal of innovation of a firm, but when this does not come with a change in the

Figure 4.3
Size of the largest ATC sales share within firms



Notes:

1. The ATC sales share is the sales share within a firm of products with the same ATC code. For each firm we select the largest one.

ATC it may represent just marginal innovations. However, when a change in products comes with a change in ATC it is likely that some structural innovations are taking place within the firm. Thus we create a discrete variable which capture simultaneously the sign of variation in the number of products ($\Delta k \in \{+, =, -\}$) and whether the ATC change ($\{c, nc\}$). Specifically, the variable can take on six values which represent all possible combinations, and we define the following six dummies accordingly: $k^{+,c}$, $k^{=,c}$, $k^{-,c}$, $k^{+,nc}$, $k^{=,nc}$, $k^{-,nc}$.

In Table 4.1 we see that the mean growth in the U.S pharmacy industry is higher and positive for firms that increase the number of products between two consecutive years. Out of these, the growth is higher for firms which also change ATC code. Then we observe a slightly negative growth for firms which have a zero net flow. The lowest means are for firms which lose some products, where the worst performance is for firms which also change ATC. From Table 4.2 it is evident that firms about to exit from the market have a far lower growth. Interestingly, the mean growth for surviving firms is just below zero which suggests that overall the growth in our database is negative.

Table 4.2
THE EXIT EFFECT ON THE MEAN GROWTH

	Mean	*Std. Dev.*	*Min*	*Max*	*Obs.*
$exit = 0$	-0.03	1.00	-9.06	11.20	6,716
$exit = 1$	-1.38	2.14	-10.37	7.78	333

Notes:

1. $Exit = 1$ for firms in the last year before exit.
2. The growth is calculated as logarithmic difference between yearly sales.

4.2 Econometrics Methodology

In this section we employ an econometric approach to test for the dependence between growth and size. We model the expected growth as a function of lagged sales and of some other covariates which proxy the age and innovation pattern of U.S. pharmaceutical firms. This approach allows to purge the size/growth relation of the effect due to other relevant economic variables which we can observe in our dataset. Moreover, we make use of dynamic panel data estimators in order to identify the coefficient of lagged sales in the growth regression. In fact, this strategy allows us also to remove the confounding effects due to unobserved determinants of growth and to the correlation between lagged sales and the idiosyncratic error (Arellano and Bond (1991), Arellano and Bover (1995)).

The model we develop to test the Gibrat's rule (Jovanonic (1982)) is as follows:

$$ln(S_{i,t}) - ln(S_{i,t-1}) = \beta ln(S_{i,t-1}) + X_{i,t}\delta + \mu_i + u_{i,t}, \tag{4.1}$$

where S_i are cross sections of firms yearly sales which span, at maximum, the period 1996-2007. The composite error term $\nu_{i,t} = \mu_i + u_{i,t}$ consists of a time constant unobserved heterogeneity μ_i and of an idiosyncratic component $u_{i,t}$. X_{it} is a matrix of regressors which can correlate with μ_i and $u_{i,t}$ and can contain time dummies.

The coefficient β is the "Gibrat coefficient" in the sense that evidence of $\beta = 0$ supports the Gibrat's law, while evidence of either positive or negative β is at odds with it. If $\beta = 0$ the growth rate at time t (i.e. $ln(S_{i,t}) - ln(S_{i,t-1})$) will be independent on the size at time $t-1$ (i.e. $S_{i,t-1}$). An estimated value of $\beta < 0$ means that the mean growth rate at time t is dependent on the size at time $t-1$ and that this relation is negative (i.e. the small firms growth faster than the big firms). Conversely, an estimated value $\beta > 0$ means that a relationship between size and growth exists and that big firms are associated with higher growth rate than the small firms. Equation 4.1 can be rewritten as

$$s_{i,t} = \tilde{\beta} s_{i,t-1} + X_{i,t}\delta + \mu_i + u_{i,t}, \tag{4.2}$$

where $\tilde{\beta} = 1 + \beta$, and $s_{i,t} = ln(S_{i,t})$. We estimate this equation but interpretation of parameters can more easily recovered from equation 4.1. Estimates of parameters δ are to be interpreted as regressors effects on the sales growth since are estimated in equation 4.2 for given $s_{i,t-1}$. Testing for $\tilde{\beta} = (>,<)1$ is equivalent to testing for $\beta = (>,<)0$.

In details we estimate the following model:

$$\begin{aligned} s_{i,t} &= \tilde{\beta} s_{i,t-1} + \delta_1 lage_{i,t} + \delta_2 exit_{i,t} + \\ &\delta_3 k_{i,t}^{+,c} + \delta_4 k_{i,t}^{=,c} + \delta_5 k_{i,t}^{+,nc} + \delta_6 k_{i,t}^{=,nc} + \delta_7 k_{i,t}^{-,nc} + \\ &\gamma_t y_t + \mu_i + u_{i,t}, \end{aligned} \tag{4.3}$$

where $s_{i,t-1}$ is the logarithm of sales at time $t-1$, $lage_{i,t}$ is the logarithm of the age at time t, $exit_{i,t}$ is a dummy variable which takes value 1 the last year of existence of the i-*th* firm, $k^{\cdot,\cdot}$ are dummies that capture simultaneously the sign of variation in number of products and whether the ATC change, and y_t are yearly dummies. To estimate the model described in eq. 4.2 we used a Generalized Method of Moments (GMM) as described in Section 2.2.

In the next section we report results of both Arellano-Bond and system GMM estimators.

4.3 Results

In Table 4.3 we report estimates of equation 4.2 when we ignore endogeneity of lagged sales. Coefficients of both the pooled OLS and the fixed effects must be biased, anyway we report them for the purpose of indicating, respectively, a higher threshold and a lower threshold within which, or at least close to which, the unbiased coefficient of lagged sales should fall. As Bond (2002) points out this interval can be used for a first broad-brush inspection of the reliability of consistent estimators. The matrix of regressors $X_{i,t}$ contains logarithmic *age*, the dummy *exit*, which indicates the last year before exit, the $k^{(\cdot,\cdot)}$ dummies and *yearly dummies*. Point estimates of the pooled OLS and the fixed effects are respectively 0.997, with a 95% confidence interval $0.988 < \tilde{\beta} < 1.007$, and 0.721, with a 95% confidence interval $0.672 < \tilde{\beta} < 0.769$. It appears that while in the pooled OLS model the Gibrat law is not rejected, in the fixed-effects model it is.

Table 4.3

SIZE GROWTH REGRESSION. IGNORING ENDOGENEITY OF LAGGED SALES

	Pooled OLS					Fixed-Effects				
s	*Coeff.*	*Std. Err.*	*P-value*	*[95% Conf.*	*Interval]*	*Coeff.*	*Std. Err.*	*P-value*	*[95% Conf.*	*Interval]*
s_{-1}	0.997**	*0.0049*	0.001	0.988	1.007	0.721**	*0.0248*	0.001	0.672	0.769
$ln(age)_{-1}$	-0.028	*0.0161*	0.082	-0.059	0.004	-0.369**	*0.0635*	0.001	-0.494	-0.245
$exit$	-1.303**	*0.1228*	0.001	-1.544	-1.062	-1.382**	*0.1270*	0.001	-1.631	-1.133
$k^{+,c}$	1.376**	*0.1083*	0.001	1.164	1.589	1.091**	*0.1128*	0.001	0.870	1.312
$k^{=,c}$	0.644**	*0.1156*	0.001	0.418	0.871	0.607**	*0.1178*	0.001	0.376	0.838
$k^{+,nc}$	0.997**	*0.0842*	0.001	0.832	1.162	0.826**	*0.0927*	0.001	0.644	1.008
$k^{=,nc}$	0.609**	*0.0829*	0.001	0.446	0.771	0.608**	*0.0913*	0.001	0.429	0.787
$k^{-,nc}$	0.477**	*0.0857*	0.001	0.309	0.645	0.461**	*0.0899*	0.001	0.284	0.637
yearly dummies			✓					✓		
Observations			6,939					6,939		
Firms			994					994		
ρ								0.75		

Notes:

1. ** Significant at 1%, * significant at 5%.
2. Standard errors are robust to heterokedasticity and autocorrelation. For the fixed-effects estimates the ρ statistic is the fraction of errors variance due to the time-constant component. Yearly dummies are not significant in both regressions.

Table 4.4 reports results of the GMM Arellano-Bond and system estimators which do allow for endogeneity, instead. In both cases data are transformed using orthogonal deviations (Arellano and Bover (1995)). With this method, the transformed observation is obtained as difference between the current and the average of all future available observations, instead of differencing with respect to the the previous one. On balanced panels the two transforms yield numerically identical coefficients. When the panel is unbalanced and with gaps, orthogonal deviations is preferred to first-differences since it saves sample size given that, unlike the latter, it is computable for all observations except one regardless of the gaps (see Arellano and Bover (1995)). However, when there are no gaps in the panel, as it is our case, this gain is lost. In our data set the orthogonal deviations transform performs slightly better so we will focus on it[16]

Both estimators are built on the assumption that errors are only correlated within firms, not across them. Since errors are assumed $i.i.d$ across firms, Roodman (2009a) recommends to include time dummies in order to remove universal time-related shocks from the error. As covariance matrix for the idiosyncratic error in the first step of GMM we choose the identity matrix.

When choosing instruments we should keep in mind that they must be uncorrelated with the transformed error in the transformed equation. Thus lags of the endogenous and predetermined variables can be valid instruments as long as they are not correlated either with $u_{i,t}$ or with $u_{i,t-1}$. For lagged sales, the first lag of sales in levels is correlated with $u_{i,t-1}$ but from second lag to the last available they may not be correlated so they are all good candidates. We use in total $(T-2)(T-1)/2 = 55$ instruments for sales. Similarly, for predetermined variables, lags from the first up to the last available can be valid instruments. Since we treat the $k^{\cdot\cdot}$ dummies as predetermined, in the instruments count we use 55 instruments for each so in total they account for $55 * 5 = 275$ instruments[17]. In fact, it is likely that past unobserved sales shocks affect firms strategies with regards both to the launch or phasing off of products, and to the positioning of its products within the market. Even if we think that *exit* may be a predetermined variable too, we cannot include its lags as instruments for the transformed equation since the variable is always zero apart from the last observation available for firms which do not exit during the sample window. Anyway, for the system GMM we can include 10 first differences of exit in the levels equation, as well as 10 first differences of

[16]Estimates and diagnostics using first-differences are available upon request

[17]We stress that for each $k^{\cdot\cdot}$, unlike sales, we can use also the first lag as instrument, but the first observation is missing by construction, so we have $55 - 10 + 10 = 55$ instruments just like sales.

the $k^{\cdot,\cdot}$ dummies. Strictly exogenous variables such as $ln(age_{-1})$ and *yearly dummies* are used as instruments in the transformed equation and in the levels equation for system GMM.

In Table 4.4 we report the estimated coefficients for the Arellano-Bond GMM and for the system GMM. The coefficients of lagged sales are 0.725, with a 95% confidence interval $1.056 < \tilde{\beta} < 1.084$ for the Arellano-Bond GMM and 1.056, with a 95% confidence interval $0.618 < \tilde{\beta} < 0.833$, for the system GMM, so in both cases the Gibrat hypothesis is rejected, but surprisingly the effect on the growth is negative in the former while positive in the latter. However, while the point estimate delivered by the Arellano-Bond method is plausible, the coefficient of the system GMM is higher than the confidence interval of the pooled OLS coefficient, which is likely to represent an upper threshold for consistent estimates.

Table 4.4

SIZE GROWTH REGRESSION. GMM ESTIMATES (ORTHOGONAL DEVIATIONS TRANSFORM)

	Arellano-Bond GMM					System GMM				
s	*Coeff.*	*Std. Err.*	*P-value*	*[95% Conf.*	*Interval]*	*Coeff.*	*Std. Err.*	*P-value*	*[95% Conf.*	*Interval]*
s_{-1}	0.725**	*0.0548*	0.001	0.618	0.833	1.056**	*0.0141*	0.001	1.029	1.084
$ln(age)_{-1}$	-0.345**	*0.0593*	0.001	-0.461	-0.229	-0.122**	*0.0256*	0.001	-0.172	-0.072
$exit$	-1.426**	*0.1691*	0.001	-1.757	-1.094	-0.993**	*0.1308*	0.001	-1.249	-0.736
$k^{+,c}$	0.597**	*0.1490*	0.001	0.305	0.888	1.297**	*0.1441*	0.001	1.014	1.579
$k^{=,c}$	0.373*	*0.1482*	0.012	0.082	0.663	0.800**	*0.1460*	0.001	0.514	1.086
$k^{+,nc}$	0.390**	*0.1309*	0.003	0.133	0.647	0.939**	*0.1237*	0.001	0.697	1.182
$k^{=,nc}$	0.335**	*0.1307*	0.010	0.078	0.591	0.757**	*0.1258*	0.001	0.511	1.004
$k^{-,nc}$	0.298**	*0.1147*	0.009	0.074	0.523	0.588**	*0.1166*	0.001	0.360	0.817
yearly dummies			✓					✓		
Observations			5945					6939		
Firms			908					994		
Nr. of instruments			341					412		

Notes:

1. ** Significant at 1%, * significant at 5%.
2. Transform of data to expunge the fixed effect is performed by means of orthogonal deviations. Estimates of both models are two-step. In a first step an initial consistent GMM regression is performed assuming a scalar covariance matrix for the errors. Then residuals from the first step are used to estimate the sandwich covariance matrix of the errors which is used to perform second step. Standard errors are robust to whatever pattern of heteroskedasticity and autocorrelations, and the Windmeijer correction is applied. The $k^{\cdot,\cdot}$ dummies and *exit* are assumed predetermined while $ln(age)_{-1}$ and yearly dummies are strictly exogenous. Yearly dummies are generally not significant in both regressions.
3. We use as instruments the strictly exogenous regressors $ln(age)_{-1}$ and yearly dummies and a set of GMM-style instruments drawn from lags of the endogenous and predetermined regressors. In particular, for the Arellano-Bond model, we use within the transformed equation for each time period, all available lags of the endogenous and predetermined variables in level dated $t-1$ or earlier. For the system GMM, we use the same instruments as above for the transformed equation, and for the levels equation we use contemporaneous transformed variables.

In this framework, diagnostic tests represent an important guide to select among models. In Table 4.5 we report four kinds of tests: the Arellano-Bond test for autocorrelation in disturbances, the Sargan and Hansen tests of overidentifying restrictions, and the difference-in-Hansen test of exogeneity of instrument subsets.

The Arellano-Bond statistic tests for autocorrelation in the $u_{i,t}$ by looking at first-differenced residuals $u^*_{i,t}$. The test cannot be applied to residuals of the transformed equation when orthogonal deviations are used, since in such case residuals are all interrelated by construction (Roodman (2009a)), so, regardless of the transform used, the test is always applied to first-differences. AR(1) is expected in first differences if $u_{i,t}$ are actually uncorrelated. Since the test is applied to the first difference, we have: $\Delta e_{i,t} = e_{i,t} - e_{i,t-1}$ and $\Delta e_{i,t-1} = e_{t-1} - e_{t-2}$ and both have $e_{i,t-1}$. So to check for AR(1) in levels, we must look for AR(2) in differences. In our models this test works fine since AR(1) is present as expected but AR(2) is not at a conventional threshold of 5% both for the Arellano-Bond GMM and for the system GMM

The Sargan and Hansen (see Sargan (1958) and Hansen (1982)) statistics are tests of over-identifying restrictions, *i.e.* of whether instruments, as a group, appear exogenous. The Sargan statistic is the minimized value of the one-step GMM criterion function, while the Hansen statistic is the minimized value of the two-step GMM criterion function, thus, unlike the first, the second is robust to heteroskedasticity or autocorrelation. Anyway the Hansen test, unlike the Sargan, may be weakened by instruments proliferation (see Roodman (2009a) and Roodman (2009b)). Both tests assume under the null that the instruments are jointly valid, thus rejection suggests that they are not valid. In the Arellano-Bond GMM estimates, both the Sargan and Hansen tests cannot reject the hypothesis that instruments are jointly valid. In the system GMM, while the Hansen test works fine, the Sargan test suggests that instruments are not valid, since the null hypothesis is rejected.

The difference-in-Hansen statistic tests for whether subsets of instruments are valid. We report both the test of the difference in statistics when they are included and the test when they are excluded. Rejections of these tests suggest, respectively, that instruments in the subset and the others are jointly not valid. The test is not vulnerable to instrument proliferation but it requires homoskedastic errors for consistency. In the Arellano-Bond model, GMM-type instruments turn out to be alone jointly valid as well as strictly exogenous variables. Again, in the system GMM the null hypothesis is rejected both for strictly exogenous variables and for GMM instruments in the levels equation. An overall interpretation of diagnostic tests, in conjunction with an implausible high coefficient of lagged sales for the system GMM, lead us to select the Arellano-Bond GMM as best model.

Table 4.5

DIAGNOSTICS FOR GMM ESTIMATES (ORTHOGONAL DEVIATIONS TRANSFORM)

		Arellano-Bond GMM	System GMM
1	Arellano-Bond test for AR in first differences		
	test for AR(1):	$z = -5.36$ $Pr > z = 0.000$	$z = -6.34$ $Pr > z = 0.000$
	test for AR(2):	$z = -1.93$ $Pr > z = 0.053$	$z = -1.72$ $Pr > z = 0.085$
2	Sargan test of overidentifying restrictions:	$\chi^2(323) = 364.55$ $Pr > \chi^2 = 0.055$	$\chi^2(393) = 862.68$ $Pr > \chi^2 = 0.000$
3	Hansen test of overidentifying restrictions:	$\chi^2(323) = 323.27$ $Pr > \chi^2 = 0.485$	$\chi^2(393) = 430.40$ $Pr > \chi^2 = 0.094$
4	Difference-in-Hansen tests of exogeneity of instrument subsets		
	a) i.v.: $ln(age)_{-1}$, yearly dummies		
	Hansen test excluding group:	$\chi^2(312) = 314.97$ $Pr > \chi^2 = 0.442$	$\chi^2(382) = 407.24$ $Pr > \chi^2 = 0.179$
	Difference (null H = exogenous):	$\chi^2(11) = 8.30$ $Pr > \chi^2 = 0.686$	$\chi^2(11) = 23.16$ $Pr > \chi^2 = 0.017$
	b) GMM instruments for levels		
	Hansen test excluding group:		$\chi^2(323) = 330.58$ $Pr > \chi^2 = 0.374$
	Difference (null H = exogenous):		$\chi^2(70) = 99.82$ $Pr > \chi^2 = 0.011$

Notes:

1. Diagnostics refer to Table 4.4.
2. The Arellano-Bond test for AR is performed on data transformed by first differencing regardless of the transform used for estimation. Negative first-order serial correlation is expected in first differences, thus AR(1) is uninformative about serial correlation in levels. First-order serial correlation in levels is checked looking at the AR(2) in first differences.
3. The Sargan and Hansen test assume under the null that the instruments are jointly valid, thus rejection suggests that they are not valid. The Sargan statistic is the minimized value of the one-step GMM criterion function, while the Hansen statistic is the minimized value of the two-step GMM criterion function, thus, unlike the first, the second is robust to heteroskedasticity or autocorrelation. Anyway the Hansen test, unlike the Sargan, may be weakened by instruments proliferation.
4. The difference-in-Hansen statistic tests for whether subsets of instruments are valid. We report both the test of the difference in statistics when they are included and the test when they are excluded. Rejections of these tests suggest, respectively, that instruments in the subset and the others are jointly not valid. The test is not vulnerable to instrument proliferation but it requires homoskedastic errors for consistency.

Thus, we can conclude that in our data there is evidence against the Gibrat law in the sense that smaller U.S. pharmaceutical firms grow faster. In particular (see Table 4.4), the estimated effect of lagged logarithmic sales is $\tilde{\beta} = 0.725$, which means that, *ceteris paribus*, a one-percent rise in sales of yesterday increases today's sales by 0.725%. This effect is significantly lower than 1, as the 95% confidence interval for s_{-1} displays, so the growth effect , *i.e.* β in eq. 4.1, $(\tilde{\beta} - 1) = -0.275$ is significantly less than zero. In particular, a one-percent rise in sales of yesterday reduces the growth rate $(\frac{S_t - S_{t-1}}{S_{t-1}})$ by $0.275\frac{S_t}{S_{t-1}}$ percentage points. For example, a growth rate of 5% would drop to 4.714% and a growth rate of 10% would drop to 9.698%.

Moreover, coefficients of the other regressors are plausible in sign and magnitude. Older firms in U.S. pharmaceutical sector grow slower than younger as expected (Jovanonic (1982)). In particular, one year more drops sales by 0.035% for firms aged 10 and by 0.007% for firms aged 50. Expected sales for firms in the second year of life are 0.345% lower due to the age affect. Sales of firms in the last year of life are 1.426% lower, *ceteris paribus*[18].

As regards $k^{\mathbf{p},\mathbf{q}}$ dummies, they are all positive and significant, which means that firms whose net flow of products is negative and whose ATC simultaneously does not change (the base category), have the lowest growth rates on average. Table 4.6 looks more closely at this point displaying comparisons between firms with different innovation patterns. An increase in the number of products is always an important key for higher growth. In particular we see that firms with positive net flow grow faster than firms with zero or negative flow, either when they both change the ATC or when they do not. Anyway, in the former case the difference is striking while in the latter it far lower and significant only in relation to firms which lose products. Of course, since we define the ATC code for a firm as the ATC of products with the highest share of sales out of the firm's total, it is very likely that either an increase or a decrease in the number of products is more substantial when it comes with a variation in ATC code, which may explain most of the difference in effects. When the focus is on firms with the same sign in Δk, a change in ATC comes with a substantial increase in growth (0.207) when the net flow of products is positive and with a substantial decrease (-0.298) when the flow is negative. These coefficients may pick up, as above, the effect of a larger variation in ΔK when the ATC is changed. However, these coefficients are also likely to pick up the positive growth effect of a change in ATC *per se*, unless it is the result of a relevant products phasing off which

[18] When the dummy *exit* is omitted, the negative size impact gets substantially larger suggesting that the variable plays the expected role in purging the Gibrat coefficient of the effect due to exiting firms.

translates in a shrinkage of the firm. The positive effect when there is no variation in the number of products (0.038) confirms that this is the case though the effect is modest and not significant, which suggests that it may take time (at least more than one year) to be sensible.

Table 4.6
Comparisons in Innovation Patterns

	Coeff.	*P-value*
Different pattern in ATC - Same sign in ΔK		
$k^{+,c} - k^{+,nc}$	0.207**	0.005
$k^{=,c} - k^{=,nc}$	0.038	0.631
$k^{-,c} - k^{-,nc}$	-0.298**	0.009
Different sign in ΔK - Same pattern in ATC (constant)		
$k^{+,c} - k^{=,c}$	0.224	0.057
$k^{=,c} - k^{-,c}$	0.373*	0.012
$k^{+,c} - k^{-,c}$	0.597**	0.000
Different sign in ΔK - Same pattern in ATC (varying)		
$k^{+,nc} - k^{=,nc}$	0.055	0.125
$k^{=,nc} - k^{-,nc}$	0.036	0.370
$k^{+,nc} - k^{-,nc}$	0.092*	0.038

Notes:

1. ** Significant at 1%, * significant at 5%.
2. Calculations are from Arellano-Bond GMM estimates of Table 4.4.
3. The $k^{\mathbf{p},\mathbf{q}}$ variables are dummies which indicate firms whose number of products can increase ($k^{+,\mathbf{q}}$), be constant ($k^{=,\mathbf{q}}$) or decrease ($k^{-,\mathbf{q}}$) between two consecutive years, and whose ATC code can simultaneously either change ($k^{\mathbf{p},c}$) or not ($k^{\mathbf{p},nc}$).

4.4 Results: Small versus Large Firms

Results obtained in the previous section are in line with most of the empirical literature on the Gibrat law which found that smaller firms grow faster. However, it is also interesting to investigate whether this finding applies to all firms or whether this can change with size. In fact, some early studies found that for large firms the opposite may be true[19]. Therefore we replicate the same analysis as above splitting our samples in two sub-samples of "small"

[19] Hart (1962), Samuels (1965), Prais (1974), Singh and Whittington (1975) found this result for UK manufacturing firms.

and "large" firms. In particular we identify an upper threshold of 100,000£ for the former, and a lower threshold of 400,000£ for the latter. Thresholds are chosen trading-off two requirements: a sufficient gap in size between large and small firms in order to keep groups well separated and a sufficient number of observations to obtain reliable estimates.

Tables 4.7 and 4.9 show results for small and large firms respectively. Tables 4.8 and 4.10 show the relative diagnostics. For both samples we use the Arellano-Bond GMM estimator for the purpose of comparing results with our preferred model for the whole sample[20]. Results point out that both small and large firms grow slower as the size increases, but while the size effect on growth is somewhat large for small firms, 0.49, the slope of large ones is less than the half, 0.22. Reliable intervals for consistent estimates indicated by the pooled OLS and the fixed effects are $0.20-0.51$ for small and $-0.01-0.23$ for large firms. Confidence intervals for both coefficients suggest that the difference is statistically significant. This result is also consistent with the logarithmic specification we used for lagged sales. In fact that implies a monotonically decreasing convex relation between growth and size in level so that also large firms grow slower the larger they are, but at a lower rate than small firms. These results are consistent with most of the empirical literature which compared small and large firms where it is typically found that the negative growth effect is larger for the former. However, while several studies found no relation for large firms (or even positive), our estimations support a significant negative relation also for those. This difference may be explained by the fact that data constraints force us to use a sales threshold above which some firms may not be considered actually "large".

Estimates of the effects of the covariates point out that while age is still an important determinant of growth rates for large firms, it has no effect for small firms. Small firms about to exit the market grow far less than surviving firms. The dummy *exit* is not included in the large firms analysis since the number of large firms which exit the market is negligible as one would expect[21].

Tables 4.11 and 4.12 report estimates of differences in innovation patterns. A positive net flow of products has always a positive impact on growth, but the impact is far larger for small firms.

This result has an immediate mathematic explanation as small firms have typically fewer products than large firms so when the number increases, the rise in sales should be proportionally higher for the former unless the increase

[20] Anyway, several checks were employed with the system GMM estimator but even in this case estimates are not as reliable and robust as those of the Arellano-Bond.

[21] Only 4 out of 558 in the sample of large firms exit in our sample window.

Table 4.7
SIZE GROWTH REGRESSION. GMM ESTIMATES FOR SMALL FIRMS
($< 100,000$£)

	Arellano-Bond GMM (orthogonal deviations)				
s	*Coeff.*	*Std. Err.*	*P-value*	*[95% Conf.*	*Interval]*
s_{-1}	0.513**	*0.0851*	0.000	0.346	0.680
$ln(age)_{-1}$	-0.139	*0.1467*	0.345	-0.426	0.149
exit	-1.099**	*0.1591*	0.000	-1.411	-0.787
$k^{+,c}$	1.290**	*0.3447*	0.000	0.615	1.966
$k^{=,c}$	0.663*	*0.3340*	0.047	0.008	1.318
$k^{+,nc}$	0.948**	*0.2536*	0.000	0.450	1.445
$k^{=,nc}$	0.489	*0.2588*	0.059	-0.018	0.997
$k^{-,nc}$	0.558*	*0.2519*	0.027	0.064	1.052
yearly dummies			✓		
Observations			1,619		
Firms			401		
Nr. of instruments			318		

Notes:

1. ** Significant at 1%, * significant at 5%. Notes to Table 4.4 apply here. However better diagnostics led us to exclude the first lag of lagged sales as GMM-style instrument.

Table 4.8
DIAGNOSTICS FOR GMM ESTIMATES. SMALL FIRMS ($< 100,000$£)

	Arellano-Bond GMM	
1	Arellano-Bond test for AR in first differences	
	test for AR(1):	$z = -3.17$ $Pr > z = 0.002$
	test for AR(2):	$z = 0.46$ $Pr > z = 0.649$
2	Sargan test of overidentifying restrictions:	
		$\chi^2(300) = 290.23$ $Pr > \chi^2 = 0.647$
3	Hansen test of overidentifying restrictions:	
		$\chi^2(300) = 240.82$ $Pr > \chi^2 = 0.995$
4	Difference-in-Hansen tests of exogeneity of instrument subsets	
	i.v.: $ln(age)_{-1}$, yearly dummies	
	Hansen test excluding group:	$\chi^2(289) = 229.26$ $Pr > \chi^2 = 0.996$
	Difference (null H = exogenous):	$\chi^2(11) = 11.56$ $Pr > \chi^2 = 0.398$

Notes:

1. Diagnostics refer to Table 4.7. Notes to Table 4.5 apply here.

Table 4.9

SIZE GROWTH REGRESSION. GMM ESTIMATES FOR LARGE FIRMS ($> 400,000£$)

	Arellano-Bond GMM (orthogonal deviations)				
s	*Coeff.*	*Std. Err.*	*P-value*	*[95% Conf.*	*Interval]*
s_{-1}	0.781**	*0.0401*	0.000	0.702	0.860
$ln(age)_{-1}$	-0.187**	*0.0646*	0.004	-0.314	-0.060
exit					
$k^{+,c}$	0.393**	*0.0892*	0.000	0.219	0.568
$k^{=,c}$	0.293**	*0.0931*	0.002	0.111	0.476
$k^{+,nc}$	0.336**	*0.0775*	0.000	0.184	0.488
$k^{=,nc}$	0.352**	*0.0835*	0.000	0.188	0.515
$k^{-,nc}$	0.271**	*0.0711*	0.000	0.132	0.411
yearly dummies			✓		
Observations			2,975		
Firms			445		
Nr. of instruments			331		

Notes:

1. ** Significant at 1%, * significant at 5%. Notes to Table 4.4 apply here. However better diagnostics led us to exclude the first lag of lagged sales as GMM-style instrument.

Table 4.10

DIAGNOSTICS FOR GMM ESTIMATES. LARGE FIRMS ($> 400,000£$)

Arellano-Bond GMM		
1	Arellano-Bond test for AR in first differences	
	test for AR(1):	$z = -4.74$ $Pr > z = 0.000$
	test for AR(2):	$z = -1.04$ $Pr > z = 0.296$
2	Sargan test of overidentifying restrictions:	
		$\chi^2(314) = 291.27$ $Pr > \chi^2 = 0.817$
3	Hansen test of overidentifying restrictions:	
		$\chi^2(314) = 323.30$ $Pr > \chi^2 = 0.347$
4	Difference-in-Hansen tests of exogeneity of instrument subsets	
	i.v.: $ln(age)_{-1}$, yearly dummies	
	Hansen test excluding group:	$\chi^2(303) = 310.43$ $Pr > \chi^2 = 0.372$
	Difference (null H = exogenous):	$\chi^2(11) = 12.87$ $Pr > \chi^2 = 0.302$

Notes:

1. Diagnostics refer to Table 4.9. Notes to Table 4.5 apply here.

Table 4.11
COMPARISONS IN INNOVATION PATTERNS. SMALL FIRMS ($< 100,000£$)

	Coeff.	*P-value*
Different pattern in ATC - Same sign in ΔK		
$k^{+,c} - k^{+,nc}$	0.343	0.1835
$k^{=,c} - k^{=,nc}$	0.174	0.4286
$k^{-,c} - k^{-,nc}$	-0.558*	0.0270
Different sign in ΔK *- Same pattern in ATC (constant)*		
$k^{+,c} - k^{=,c}$	0.627	0.0654
$k^{=,c} - k^{-,c}$	0.663*	0.0471
$k^{+,c} - k^{-,c}$	1.290**	0.0002
Different sign in ΔK *- Same pattern in ATC (varying)*		
$k^{+,nc} - k^{=,nc}$	0.458**	0.0038
$k^{=,nc} - k^{-,nc}$	-0.068	0.5519
$k^{+,nc} - k^{-,nc}$	0.390**	0.0076

Notes:

1. ** Significant at 1%, * significant at 5%. Notes to Table 4.6 apply here.
2. Calculations are from Arellano-Bond GMM estimates of Table 4.7.

Table 4.12
COMPARISONS IN INNOVATION PATTERNS. LARGE FIRMS ($> 400,000£$)

	Coeff.	*P-value*
Different pattern in ATC - Same sign in ΔK		
$k^{+,c} - k^{+,nc}$	0.058	0.1751
$k^{=,c} - k^{=,nc}$	-0.058	0.3414
$k^{-,c} - k^{-,nc}$	-0.271**	0.0000
Different sign in ΔK *- Same pattern in ATC (constant)*		
$k^{+,c} - k^{=,c}$	0.100	0.2050
$k^{=,c} - k^{-,c}$	0.293**	0.0016
$k^{+,c} - k^{-,c}$	0.393**	0.0000
Different sign in ΔK *- Same pattern in ATC (varying)*		
$k^{+,nc} - k^{=,nc}$	-0.016	0.4988
$k^{=,nc} - k^{-,nc}$	0.080**	0.0028
$k^{+,nc} - k^{-,nc}$	0.064**	0.0084

Notes:

1. ** Significant at 1%, * significant at 5%. Notes to Table 4.6 apply here.
2. Calculations are from Arellano-Bond GMM estimates of Table 4.9.

in K is proportionally higher for firms with more products. The latter occurrence can be reasonably ruled out, however. We provide evidence that in U.S. pharmaceutical industry a negative relation between size and growth seems to apply also to the number of product. Table 4.13 shows that this the case for the whole sample, though the negative effect is weaker than for sales, and Tables 4.15 and 4.17 show a negative dependence for both small and large firms (in terms of sales). Moreover, the negative effect is smaller for large firms as in the analysis of sales. Diagnostics reported in Tables 4.14, 4.16 and 4.18 show that these estimates are quite reliable.

Table 4.13
GMM estimates. Whole Sample

	Arellano-Bond GMM (orthogonal deviations)				
k	*Coeff.*	*Std. Err.*	*P-value*	*[95% Conf.*	*Interval]*
k_{-1}	0.841**	*0.0685*	0.000	0.707	0.975
$ln(age)_{-1}$	-0.037	*0.0329*	0.263	-0.101	0.028
exit	-0.274*	*0.1384*	0.048	-0.546	-0.003
yearly dummies			✓		
Observations			6,046		
Firms			921		
Nr. of instruments			47		

Notes:

1. ** Significant at 1%, * significant at 5%. Notes to Table 4.4 apply here. However better diagnostics led us to exclude the first and second lag of k_{-1} as GMM-style instrument. The reliable interval for consistent estimates indicated by the fixed effects and pooled OLS is $0.741 - 0.987$.

Table 4.14
Diagnostics for GMM estimates

	Arellano-Bond GMM	
1	Arellano-Bond test for AR in first differences	
	test for AR(1):	$z = -7.43$ $Pr > z = 0.000$
	test for AR(2):	$z = 0.26$ $Pr > z = 0.791$
2	Sargan test of overidentifying restrictions:	
		$\chi^2(34) = 37.98$ $Pr > \chi^2 = 0.293$
3	Hansen test of overidentifying restrictions:	
		$\chi^2(34) = 38.21$ $Pr > \chi^2 = 0.284$
4	Difference-in-Hansen tests of exogeneity of instrument subsets	
	i.v.: $ln(age)_{-1}$, yearly dummies	
	Hansen test excluding group:	$\chi^2(23) = 21.69$ $Pr > \chi^2 = 0.539$
	Difference (null H = exogenous):	$\chi^2(11) = 16.52$ $Pr > \chi^2 = 0.123$

Notes:

1. Diagnostics refer to Table 4.13. Notes to Table 4.5 apply here.

For given sign in ΔK, Tables 4.11 and 4.12 suggest that a change in ATC has a larger impact (in absolute terms) on growth for small firms. This is a genuine innovation effect since the role played by variation in the number of products should be of minor importance here. One may argue that a variation in ΔK can have a more important impact for small firms for the same reason outlined above, thus explaining also this finding, at least to some extent. But, since here we are controlling for small and large firms which do change ATC both, we can expect that the variation in products, either positive or negative, is important also for large firms. Given our firm ATC definition, what matters is the largest “ATC share” so when a large firm changes ATC

Table 4.15

GMM ESTIMATES. SMALL FIRMS ($< 100,000£$)

	Arellano-Bond GMM (orthogonal deviations)				
k	*Coeff.*	*Std. Err.*	*P-value*	*[95% Conf.*	*Interval]*
k_{-1}	0.671**	*0.1132*	0.000	0.449	0.893
$ln(age)_{-1}$	0.020	*0.0465*	0.660	-0.071	0.112
exit	-0.434**	*0.1612*	0.007	-0.750	-0.118
yearly dummies			✓		
Observations			1,691		
Firms			417		
Nr. of instruments			56		

Notes:

1. ** Significant at 1%, * significant at 5%. Notes to Table 4.4 apply here. However better diagnostics led us to exclude the first lag of k_{-1} as GMM-style instrument. The reliable interval for consistent estimates indicated by the fixed effects and pooled OLS is $0.432 - 0.808$.

Table 4.16

DIAGNOSTICS FOR GMM ESTIMATES. SMALL FIRMS ($< 100,000£$)

	Arellano-Bond GMM	
1	Arellano-Bond test for AR in first differences	
	test for AR(1):	$z = -5.15$ $Pr > z = 0.000$
	test for AR(2):	$z = 0.86$ $Pr > z = 0.388$
2	Sargan test of overidentifying restrictions:	
		$\chi^2(43) = 69.21$ $Pr > \chi^2 = 0.007$
3	Hansen test of overidentifying restrictions:	
		$\chi^2(43) = 40.92$ $Pr > \chi^2 = 0.562$
4	Difference-in-Hansen tests of exogeneity of instrument subsets	
	i.v.: $ln(age)_{-1}$, yearly dummies	
	Hansen test excluding group:	$\chi^2(32) = 23.38$ $Pr > \chi^2 = 0.866$
	Difference (null H = exogenous):	$\chi^2(11) = 17.55$ $Pr > \chi^2 = 0.093$

Notes:

1. Diagnostics refer to Table 4.15. Notes to Table 4.5 apply here.

there is no reason to expect that its variation in the *number* of products would be less relevant than its smaller counterpart. Moreover, if we focus on firms which do not change the number of products, a change in ATC yields an effect on growth of 0.17 points for small firms, and slightly negative for large. Though these effects are not always statistically significant, differences in point estimates between small and large firms are typically large enough to give at least some support on our argument that innovation is worthwhile especially for small firms.

Table 4.17
GMM ESTIMATES. LARGE FIRMS ($> 400,000£$)

	Arellano-Bond GMM (orthogonal deviations)				
k	*Coeff.*	*Std. Err.*	*P-value*	*[95% Conf.*	*Interval]*
k_{-1}	0.852**	*0.0622*	0.000	0.730	0.974
$ln(age)_{-1}$	-0.004	*0.0313*	0.890	-0.066	0.057
exit					
yearly dummies			✓		
Observations			2,987		
Firms			448		
Nr. of instruments			47		

Notes:

1. ** Significant at 1%, * significant at 5%. Notes to Table 4.4 apply here. However better diagnostics led us to exclude the first and second lag of k_{-1} as GMM-style instrument. The reliable interval for consistent estimates indicated by the fixed effects and pooled OLS is $0.803 - 0.996$.

Table 4.18
DIAGNOSTICS FOR GMM ESTIMATES. LARGE FIRMS ($> 400,000£$)

	Arellano-Bond GMM	
1	Arellano-Bond test for AR in first differences	
	test for AR(1):	$z = -6.80$ $Pr > z = 0.000$
	test for AR(2):	$z = -0.09$ $Pr > z = 0.926$
2	Sargan test of overidentifying restrictions:	
		$\chi^2(35) = 32.15$ $Pr > \chi^2 = 0.606$
3	Hansen test of overidentifying restrictions:	
		$\chi^2(35) = 39.56$ $Pr > \chi^2 = 0.274$
4	Difference-in-Hansen tests of exogeneity of instrument subsets	
	i.v.: $ln(age)_{-1}$, yearly dummies	
	Hansen test excluding group:	$\chi^2(24) = 23.29$ $Pr > \chi^2 = 0.503$
	Difference (null H = exogenous):	$\chi^2(11) = 16.27$ $Pr > \chi^2 = 0.131$

Notes:

1. Diagnostics refer to Table 4.17. Notes to Table 4.5 apply here.

4.5 Conclusions

In this chapter we investigated the relation between size and growth for a panel of US pharmaceutical firms over the period 1997-2008. Since the early 1980s the U.S. pharmaceutical sector has been characterized by a significative consolidation of large firms as well as the entry into the R&D process by small and early stage biopharmaceutical firms. From the middle 1980s the pharmaceutical industry has been experienced also significative legislative changes and market developments that have caused a decline in sales and profit particularly evident from the middle of 1990s.

Within a GMM framework, we estimate a dynamic panel growth equation to control for time-constant unobserved heterogeneity and endogeneity in the idiosyncratic error term. Estimates are also robust to heteroskedasticity and autocorrelation in the error term. Information contained in our data set allows us to use controls for the age of firm and for innovation, which are considered the most important determinants of growth. We include also a dummy which captures exit of firms. Arellano-Bond GMM estimates confirm for the U.S. pharmaceutical industry the typical finding that growth decreases both with size and age. We replicate the analysis for two sub-groups of small and large firms ($< 100,000£$ and $> 400,000£$ respectively) in order to test whether the Gibrat law may hold at least for firms above a certain threshold (as pointed aut by Hart (1962), Samuels (1965), Prais (1974), Singh and Whittington (1975)). We found that, within the U.S. pharmaceutical market, the size effect on growth is lower for the sample of large firms, but it is still significant suggesting that the Gibrat law does not hold in our sample even for large firms.

Furthermore we focused our analysis on the role of innovation on the growth. In pharmaceutical industry the cost of innovation is large, the average cost of bringing a new efficacious molecule is estimated between 800 million dollars, and risky, to produce a new molecule it can take 12-13 years and only one out of 10,000 molecules will be marketed. In last years many authors have dealt with innovation in pharmaceutical industry (see DiMasi et al. (1991), Masi et al. (2003), Ornaghi (2006), Comanor and Scherer (2011)) focusing their research on the relationship between innovation and merges, or on the relationship between innovation and size from the point of view of the existence of scale economies in pharmaceutical industry R&D (Jensen (1987), Graves and Langowitz (1993),Bobulescu and Soulas (2006), Cockburn and Henderson (2001), Miyashige et al. (2007)). According to these results our contribution here is to explore whether the innovation benefits are different for smaller and larger firms. In our dataset, the innovation, as captured by net inflows of product and change in the ATC code, plays a significant role in boosting growth rates. In U.S. pharmaceutical industry, the differences between small and large firms in point estimates of the innovation coefficients are typically large enough to give at least some support on our argument that innovation is worthwhile especially for small firms.

Chapter 5

Conclusions

The general aims of our research project can be summarized in the following points: analyze the size distributions of economic phenomena and investigate the relationship between size and growth. These two analysis correspond to two different approaches usually used in literature for testing the coherence of growth models to empirical data. As regards to the first approach we have analyzed the size distribution of wage in Italy between 1985 and 2004. The preliminary analysis of the data underlines two main changes in the sample composition between 1985 and 2004. Firstly, in 1985 women represented the 30% of the total sample, after 20 years this percentage rises at 37%. Secondly, in the twenty years of observation, an ageing of the sample occurs. Changes in the age distribution from 1985 to 2004 can be the results of a lagged entry in labour market and/or of an increase of atypical contract as main contract for young people, and/or of a delayed exit from the labour market. The analysis of the wage by gender highlights that female average daily wage growths more than the male daily wage between 1985 and 2004. Nevertheless the average daily wage perceived by women is lower than the males ones for the whole observational period.

The empirical wage distributions (by year and by gender) are fitted with 10 different models with two, three or four parameters. The distributions are compared on the basis of the likelihood ratio test (for nested models) and on the basis of other goodness-of-fit measure (such as the chi-square). For 1985 and 1995 the best fitting model is the GB2 distribution, while in 2004 data are better represented by a skewed-t distribution. For these two distributions, four inequality indices are calculated.

The analysis of inequality for the whole sample and the men sub-sample highlights two main facts. On the one hand, the dynamic of daily wage can be divided in two stages: i) an early stage (between 1985 and early 1990s) characterized by a high increase in inequality indices at the top of the distribution and a decrease at the bottom; ii) a last stage characterized by a general decreasing of the inequality indices. On the other hand, the main changes regard the top of the distribution. These results agree with the results obtained by other authors (see Brandolini et al. (2002), Manacorda (2004), Jappelli and Pistaferri (2010) and Devicienti (2003)). The same analysis on the female sub-sample reveals a different trend of the inequality. The women inequality measures seems to rise slightly also after the 1995. This fact could be a consequence of the skill based technological change (STBC). The different trends of the inequality recorded within male sub-sample and within female sub-sample suggests to investigate the dynamics of wage among the two groups defined by the gender. At a first look seems that the wage gender gap, between 1995 and 2004, was shrinking but a deeper analysis (performed by the means of a fixed effects regression model) shows that the shrinking in the difference between men wage and women wage could be due to an increasing of the percentage of women working in high-paid job and not to an "approaching" of the female wage to the male wage.

After the analysis of the size distribution we focused on the second approach proposed in literature for testing the growth models. We test the validity of the Gibrat model by testing the size-growth relation in the context of the pharmaceutical industry. The relation between size and growth is investigated for a panel of US pharmaceutical firms over the period 1997-2008. Within a GMM framework, we estimate a dynamic panel growth equation to control for time-constant unobserved heterogeneity and endogeneity in the idiosyncratic error term. Estimates are also robust to heteroskedasticity and autocorrelation in the error term. Information contained in our data set allows us to use controls for the age of firm and for innovation, which are considered the most important determinants of growth. We include also a dummy which captures exit of firms. Arellano-Bond GMM estimates confirm for the U.S. pharmaceutical industry the typical finding that growth decreases both with size and age. We replicate the analysis for two subgroups of small and large firms ($< 100,000£$ and $> 400,000£$ respectively) in order to test whether the Gibrat law may hold at least for firms above a certain threshold (as pointed aut by Hart (1962), Samuels (1965), Prais (1974), Singh and Whittington (1975)). We found that, within the U.S. pharmaceutical market, the size effect on growth is lower for the sample of large firms, but it is still significant suggesting that the Gibrat law does not hold in our sample even for large firms.

Furthermore we focused our analysis on the role of innovation on the growth. In pharmaceutical industry the cost of innovation is large, the average cost of bringing a new efficacious molecule is estimated between 800 million dollars, and risky, to produce a new molecule it can take 12-13 years and only one out of 10,000 molecules will be marketed. In last years many authors have dealt with innovation in pharmaceutical industry (see DiMasi et al. (1991), Masi et al. (2003), Ornaghi (2006), Comanor and Scherer (2011)) focusing their research on the relationship between innovation and merges, or on the relationship between innovation and size from the point of view of the existence of scale economies in pharmaceutical industry R&D (Jensen (1987), Graves and Langowitz (1993),Bobulescu and Soulas (2006), Cockburn and Henderson (2001), Miyashige et al. (2007)). According to these results our contribution here is to explore whether the innovation benefits are different for smaller and larger firms. In our dataset, the innovation, as captured by net inflows of product and change in the ATC code, plays a significant role in boosting growth rates. In U.S. pharmaceutical industry, the differences between small and large firms in point estimates of the innovation coefficients are typically large enough to give at least some support on our argument that innovation is worthwhile especially for small firms.

Bibliography

Acemoglu, D. (2002). “Technical change, Inequality, and the Labor Market”. *Journal of Economic Litterature*, 40:7–72.

Almus, M. and Nerlinger, E. (2000). “Testing Gibrat’s Law for Young Firms - Empirical Results for West Germany”. *Small Business Economics*, 15:1–12.

Ammon, O. (1898). “Some social applications of the doctrine of probability”. *Journal of Political Economiy*, 7:204–237.

Amoroso, L. (1924-1925). “Ricerche intorno alla curva dei redditi”. *Annali di Matematica pura ed Applicata*, 4:123–157.

Anderson, T. W. and Hsiao, C. (1982). “Formulation and Estimation of Dynamic Models Using Panel Data”. *Econometrica*, 18:47–82.

Arellano, M. and Bond, S. (1991). “Some Test of Specification for Panel Data: Monte Carlo Evidence and an Application to Emplyment Equations ”. *Review of Economic Studies*, 58:277–297.

Arellano, M. and Bover, O. (1995). “Another Look at the Instrumental Variables Estimation of Error Components Models”. *Journal of Econometrics*, 68:29–51.

Atkinson, A. B., Piketty, T., and Saez, E. (2011). “Top Incomes in the Long Run of History”. *Journal of Economic Literature*, 49(1):3–71.

Audretsch, D. B., Klomp, L., Santarelli, E., and Thurik, A. R. (2004). “Gibrat’s Law: are the Services Different?”. *Review of Industrial Organization*, 24:301–324.

Autor, D. A., Katz, L. F., and Kearney, M. S. (2006). “The polarization of the U.S. Labor Market ”. *American Economic Review*, 96:189–194.

Avriel, M. (2003). *"Nonlinear Programming: Analysis and Methods"*. Dover Publishing.

Azzalini, A. (1985). "A class of distributions which includes the normal ones". *Scandinavian Journal of Statistics*, 12:171–178.

Azzalini, A. and Capitanio, A. (2003). "Distribution generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution". *Journal of the Royal Statistic Society. Series B (Statistical Methodology)*, 65(2):367–389.

Azzalini, A., Cappello, T. D., and Kotz, S. (2002). "Log-Skew-Normal and Log-Skew-t Distributions as Models for Family Income Data ". *Journal of Income Distribution*, 11:12–20.

Bandourian, R., McDonald, J., and R.S.Turley (2003). "A comparison of parametric models of income distribution across countries and over time". *Estadistica*, 55:135–152.

Barabasi, A. L. and Albert, R. (2002). "Statistical mechanics of complex networks". *Reviews of Modern Physics*, 74:47–97.

Barbieri, P. (2009). "Flexible Employment and Inequality in Europe". *European Sociological Review*, 25(6):621–628.

Barron, D. N., West, E., and Hannan, M. T. (1994). "A Time to Grow and a Time to Die: Growth and Mortality of Credit Unions in New York, 1914-1990". *American Journal of Sociology*, 100(2):381–421.

Bassanini, A. and Duval, R. (2006a). "Employment Patterns in OECD Countries: Reassessing the Role of Policies and Institutions". OECD Social, Employment and Migration Working Papers 35, OECD.

Bassanini, A. and Duval, R. (2006b). "The Determinants of Unemployment across OECD Countries". Post-Print halshs-00120584, OECD.

Becchetti, L. and Trovato, G. (2002). "The Determinants of Growth for Small and Medium Sized Firms. The Role of the Availability of External Finance". *Small Business Economics*, 19(4):291–306.

Belot, M. and Ours, J. C. V. (2000). "Does the Recent Success of some OECD Countries in Lowering their Unemployment Rates lie in the Clever Design of their Labour Market Reforms?". CEPR Discussion Papers 2492, C.E.P.R. Discussion Papers.

Benini, R. (1897). "Di alcune curve descritte da fenomeni economici aventi relazione colla curva del reddito o con quella del patrimonio". *Giornale degli Economisti*, 14:177–214.

Bertola, G. and Garibaldi, P. (2003). "The Structure and History of Italian Unemployment". CESifo Working Paper Series 907, CESifo Group Munich.

Bertola, G. and Ichino, A. (1995). "Crossing the River: A Comparative Perspective on Italian Employment Dynamics". *Economic Policy*, 10(21):359–420.

Blanchard, O. and Wolfers, J. (2000). "The Role of Shocks and Institutions in the Rise of European Unemployment: The Aggregate Evidence". *Economic Journal*, 110(462):C1–33.

Blundell, R. and Bond, S. (1998). "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models". *Review of Economic Studies*, 87:115–143.

Bobulescu, R. and Soulas, C. (2006). "Innovation and firm size in the pharmaceutical industry". *Int. J. Business Environment*, 1(2):253–264.

Bond, S. (2002). "Dynamic Panel Data Models: a Guide to Micro Data Methods and Practice". Working Paper 09/02, Institute for Fiscal Studies.

Bottazzi, G., Coad, A., Jacoby, N., and Secchi, A. (2005). "Corporate Growth and Industrial Dynamics: Evidence from French Manufacturing". LEM Working Paper Series 21, Pisa, Sant'Anna School of Advanced Studies.

Bottazzi, G., Dosi, G., Lippi, M., Pammolli, F., and Riccaboni, M. (2001). "Innovation and Corporate Growth in the Evolution of the Drug Industry". *International Journal of Industrial Organization*, 19:1161–1187.

Bottazzi, G. and Secchi, A. (2003). "Common Properties and Sectorial Specifities in the Dynamics of U.S. Manufacturing Companies". *Review of Industrial Organization*, 23:217–232.

Brachmann, K., Stich, A., and Trede, M. (1996). "Evaluating parametric income distribution models". *Allgemeneis Statistiches Archiv*, vol. 80:285–298.

Brandolini, A., Cipollone, P., and Sestito, P. (2002). *"Earnings dispersion, low paid and houshold poverty on Italy, 1977-1998 "*. in: The economics of rising inequality, Oxford University Press, Oxford, New York edition.

Buldyrev, S. V., Growiec, J., Pammolli, F., Riccaboni, M., and Stanley, H. E. (2007). "The Growth of Business Firms: Facts and Theory". *Journal of the European Economic Association*, 5:574–584.

Butler, R. J. and McDonald, J. B. (1989). "Using inclompete moments to measure inequality". *Journal of Econometrics*, vol. 42:109–119.

C., K. and S., K. (2003). *"Statistical Size Distribution in Economics and Actuarial Sciences"*. Wiley.

Cabral, L. M. B. and Mata, J. (2003). "On the Evolution of the Firm Size Distribution: Facts and Theory". *American Economic Review*, 93:1075–1090.

Calvo, J. L. (2006). "Testing Gibrat's Law for Small, Young and Innovating Firms". *Small Business Economics*, 26:117–123.

Cantelli, F. P. (1929). "Sulla legge di distribuzione dei redditi". *Giornale degli Economisti e Rivista di Statistica*, 14:850–852.

Carden, S. D. (2005). "What Global Executives Think about Growth and Risk". *Mckinsey Quarterly*, 2:16–25.

Cefis, E., Ciccarelli, M., and Orsenigo, L. (2006). *"Hetereogeneity in Firms' Growth in the Pharmaceutical Industry"*, chapter 6, pages 163–207. Cambridge University Press, Cambridge, UK, in Knowledge accumulation and industry evolution: the case of pharma-biotec, M. Mazzucato and G. Dosi edition.

Champernowne, D. G. (1953). "A Model of Income Distribution". *The Economic Journal*, 63(250):318–351.

Cirillo, P. (2010). "An analisys of the size distribution of Italian firms by age". *Physica A*, vol. 389:459–466.

Clementi, F., Matteo, T. D., Gallegati, M., and Kaniadakis, G. (2008). "The k-generalized distribution: A new descriptive model for the size distribution of incomes". *Physica A: Statistical Mechanics and its Applications*, 387:3201–3208.

Coad, A. (2007). "Firm Growth: a Survay". CES Working Papers 2007.24, Centre d'Economie de la Sorbone.

Cockburn, I. M. (2004). "The Changing Structure Of The Pharmaceutical Industry". *Health Affairs*, 23(1):10–22.

Cockburn, I. M. and Henderson, R. M. (2001). "Scale and scope in drug development: unpacking the advantages of size in pharmaceutical research". *Journal of Health Economics*, 20:1033–1057.

Comanor, W. S. and Scherer, F. (2011). "Mergers and Innovation in the Pharmaceutical Market". HKS Faculty Research Working Paper Series RWP11-043, Harvard Kennedy School.

Cowell, A. F. and Kuga, K. (1981). "Additivity and the entropy concept: An axiomatic approach to inequality measurement". *Journal of Economic Theory*, 25(1):131–143.

Cressie, N., Davis, A. S., Folks, J. L., and Policello II", G. E. (1981). "The moment-generating function and negative integer moments". *The American Statistician*, 35(3):148–150.

D′Addario, R. (1932). "Intorno alla curva dei redditi". *Rivista Italiana di Statistica Economia e Finanza*, 4:89–108.

Daemmrich, A. (2011). "U.S. Healthcare reform and the Pharmaceutical industry". Workin Paper 2011-09, Harvard Business School.

Dastrup, S., Hartshorn, R., and McDonald, J. (2007). "The impact of taxes and transfer payments on the distribution of income: A parametric comparison". *Journal of Economic Inequality*, 5(3):353–369.

Dennis, J. E. and Schnabel, R. B. (1983). *"Numerical Methods for Unconstrained Optimization and Nonlinear Equations"*. Prentice Hall.

Devicienti, F. (2003). "The rise in wage inequality in Italy. Observed factors or unobservables"". Working paper, LABORatorio Revelli.

DiMasi, J., Hansen, R. W., Grabowski, H. G., and Lasagna, L. (1991). "Cost of innovation in the pharmaceutical industry". *Journal of Health Economics*, 10(2):107–142.

Draculescu, A. and Yakovenko, V. M. (2001a). "Evidence for the exponential distribution of income in the USA". *The European Physical Journal B*, 20:858–589.

Draculescu, A. and Yakovenko, V. M. (2001b). "Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States". *Physica A*, 299:213–221.

Droucopoulos, V. (1983). “International Big Business Revisited: on the Size and Growth of the World’s Larger firms”. *Managerial and Decision Economics*, 4(4):244–252.

Dunne, T. and Hughes, A. (1994). “Age, Size, Growth and Survival: UK Companies in the 1980s”. *Journal of Industrial Economics*, 42(2):115–140.

Dunne, T., Roberts, M., and Samuelson, L. (1989). “The Growth and Failure of U.S. Manifacturing Plants”. *Quartely Journal of Economics*, 104(4):671–698.

Durlauf, S. N. and Blume, L. E. (2008). *“The New Palgrave Dictionary of Economics”*. Palgrave Macmillan; 2nd edition.

Evans, D. S. (1987a). “Tests of Alternative Theories of Firm Growth”. *Journal of Political Economy*, 95(4):657–674.

Evans, D. S. (1987b). “The Relationship Between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries”. *The Journal of Industrial Economics*, 35:567–581.

Fabiani, S., Locarno, A., Oneto, G., and Sestito, P. (2000). “The sources of unemployment fluctuations: an empirical application to the Italian case”. Working Paper Series 29, European Central Bank.

F.Bordley, R., McDonald, J. B., and Mandala, A. (1996). “Something new, something old: Parametric models for the size distribution of income”. *Journal of Income distribution*, 6:91–103.

Forster, M. F. (2000). “Trends and Driving Factors in Income Distribution and Poverty in the OECD Area”. OECD Labour Market and Social Policy Occasional Papers 42, OECD Publishing.

Freedman, D. A. (2006). “On the So-Called ”Huber Sandwich Estimator” and ”Robust Standard Errors””. *The American Statistician*, vol.60:299–302.

Freel, M. S. (2000). “Do Small Innovating Firms Outperform Non-Innovators?”. *Small Business Economics*, 14:195–210.

Gabaix, X. (1999). “ Zipf ’s Law for Cities: An Explanation”. *J. Quar. Econ.*, 73:739–767.

Gabaix, X. (2009). “Power Laws in Economics and Finance”. *Annual Review of Economics*, 1:255–294.

Gabe, T. M. and Kraybill, D. S. (2002). "The Effect of State Economics Developments Incentives on Employment Growth of Establishemnts". *Journal of Regional Science*, 42(4):703–730.

Gentle, J. E. (1994). *"Computational Statistics"*. Springer-Verlag, New York.

Geroski, P. A. (2000). *"The Growth of Firms in Theory and Practice"*, pages 168–186. Oxford University Press, Oxford, UK, in Competence, governance and entrepeneuership, N. Foss and V. Mahnke (eds.) edition.

Geroski, P. A. (2005). "Understanding the Implication of Empirical Work on Corporate Growth Rates". *Managerial and Decision Economics*, 26:129–138.

Geroski, P. A. and Gugler, K. (2004). "Corporate Growth Convergence in Europe". *Oxford Economic Papers*, 56:597–620.

Geroski, P. A. and Machin, S. (1992). "Do Innovating Firms Outperform Non-Innovators?". *Business Strategy Review*, pages 79–90.

Geroski, P. A. and Toker, S. (1996). "The Turnover of Market Leaders in UK Manufacturing Industries, 1979-1986". *International Journal of Industrial Organization*, 14:141–158.

Gibrat, R. (1931). *"Les Inégalités économiques"*. Paris: Librairie du Recueil Sirey edition.

Gini, C. (1909a). "Concentration and dependency ratios (in Italian). English traslation in". *Rivista di politica economica*, 87 (1997):769–789.

Gini, C. (1909b). "Il diverso accrescimento delle classi sociali e la concentrazione della ricchezza". *Giornale degli economisti*, 38:27–83.

Gini, C. (1932). "Intorno alle curve di concentrazione". *Metron*, 9:3–76.

Goos, M. and Manning, A. (2007). "Lousy and Lovely Jobs: The Rising Polarization of Work in Britain". *Review of Economic Statistic*, 89:118–133.

Goos, M., Manning, A., and Salomons, A. (2009). "Job". *Review of Economic Statistic*, 99:58–63.

Grabowski, H. and Kyle, M. (2007). "Generic Competition and Market Exclusivity Periods in Pharmaceuticals". Open Access publications from University of Toulouse 1 Capitole http://neeo.univ-tlse1.fr, University of Toulouse 1 Capitole.

Graf, M. and Nedyalkova, D. (2010). "Variance Estimation on the Maximum Likelihood Estimates of Parameters of the GB2 Distribution of the second Kind and of the Leaken Indicators under Cluster Sampling". Working paper, SFSO.

Graves, S. B. and Langowitz, N. (1993). "Innovative productivity and returns to scale in the pharmaceutical industry". *Strategic Management Journal*, 14:593–605.

Growiec, J., Pammolli, F., Riccaboni, M., and Stanley, H. E. (2008). "On the size distribution of business firms". *Economics Letters*, 98(2):207–212.

Grubb, D. and Wells, W. (1993). " Employment Regulation and and Patterns of Work in EC Countries". Economic Studies 21, OECD.

Hall, B. H. (1987a). "The Relationship between Firm Size and Firm Growth in the U.S. Manufacturing Sector". *Journal of Industrial Economics*, 35(4):583–606.

Hall, B. H. (1987b). "The Relationship Between Firm Size and Firm Growth in the U.S. Manufacturing Sector". *Journal of Industrial Economics*, 35(4):583–600.

Hansen, L. (1982). "Large Sample Properties of Generalized Method of Moments Estimators". *Econometrica*, 50:1029–1054.

Hardwick, P. and Adams, M. (2002). "Firm Size and Growth in the United Kingdom Life Insurance Industry". *Journal of Risk and Insurance*, 69(4):577–593.

Harhoff, D., Stahl, K., and Woywode, M. (1998). "Legal Form, Growth and Exits of West German Firms - Empirical Results for Manufacturing, Construction, Trade and Service Industries". *Journal of Industrial Economics*, 46(4):453–488.

Hart, P. E. (1962). "The Size and Growth of Firms". *Economica*, 29(113):29–39.

Hart, P. E. and Oulton, N. (1996). "The Growth and Size of Firms". *Economic Journal*, 106(3):1242–1252.

Hay, M. and Kamshad, K. (1994). "Small Firm Growth: Intentions, Implementation and Impediments". *Business Strategy Review*, 5(3):49–68.

Heckman, J., Lochner, L., and Todd, P. (2003). " Fifty Years of Mincer Earnings Regressions". NBER Working Pape 9732, National Bureau of Economic Research.

Holtz-Eakin, D., Newey, W., and Rosen, H. S. (1988). "Estimating Vector Autoregressions with Panel Data". *Econometrica*, 56:1371–1395.

Ijiri, Y. and Simon, H. A. (1964). "Business Firm Growth and Size". *The American Economic Review*, 54(2):77–89.

ISTAT (1991). "Classificazione delle attivitá economiche". http://www.istat.it/.

J. Goddard, J. Wilson, P. B. (2002). "Panel Test of Gibrat's Law for Japanese Manufacturing". *Internetional Journal of Industrial Organization*, 20:415–433.

Jappelli, T. and Pistaferri, L. (2010). "Does Consumption Inequality Track Income in Italy?". *Review of Economic Dynamics*, vol. 13:133–153.

Jenkins, S. P. (2007). "Inequalityand the GB2 income distribution?". Discussion Papers 2831, IZA Institute for the Study of Labor.

Jenkins, S. P. (2009). "Distributionally-sensitive inequality indices and the GB2 income distribution". *Review of Income and Wealth*, vol. 55(No. 2):392–398.

Jensen, E. J. (1987). "Research Expenditures and the Discovery of New Drugs". *Journal of Industrial Economics*, 36:83–95.

Jones, M. C. and Faddy, M. J. (2003). "A skew extension of the t-distribution, with applications". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(1):159–174.

Jovanonic, B. (1982). "Selection and evolution of industry". *Econometrica*, 50(3):649–670.

Kalecki, M. (1945). "On the Gibrat Distribution". *Econometrica*, 13(2):pp. 161–170.

Kumar, M. S. (1985). "Growth, Acquisition Activity and Firm Size: Evidence from the United Kingdom". *Journal of Industrial Economics*, 33(3):327–338.

Layard, R., Nickell, S., and Jackman, R. (2001). *"Unemployment; Macroeconomic Performance and the Labour Market"*. Oxford: Oxford University Press.

L.C.Thurow (1970). "Analyzing the American Income Distribution". *The American Economic Review*, vol.60(No. 2):261–269.

Liu, J. M., Tsou, J., , and Hammit, J. K. (1999). "Do Small Plants Grow Faster? Evidence from the Taiwan Electronics Industry". *Economics Letters*, 65:121–129.

Lomax, K. S. (1954). "Business failures: Another example of the analysis of failure data". *Journal of the American Statistical Association*, 49:847–852.

Lotti, F., Santarelli, E., and Vivarelli, M. (2003). "Does Gibrat's Law Hold Among Young, Small Firms". *Journal of Evolutionary Economics*, 13:213–235.

Manacorda, M. (2004). "Can Scala Mobile explain the fall and the rise of earnings inequality on Italy? A semiparametric analysis, 1977-1993". *Journal of Labor Economics*, 22:585–613.

Mandelbrot, B. (1960). "The Pareto-Levy Law and the Distribution of Income". *International Economic Review*, 1:79–106.

Mansfield, E. (1962). "Gibrat's Law, Innovation, and the Growth of Firms". *The American Economic Review*, 52(5):1023–1051.

Mansfield, E., Rapoport, J., Romeo, A., Villani, E., Wagner, S., and Husic, F. (1977). *"The Production and Application of New Industrial Technology"*. Norton, New York.

Masi, J. D. and A., J. (1995). "Success rates for new drugs entering clinical testing in the United States". *Clinical Pharmacology & Therapeutics*, 58(1):1–14.

Masi, J. D., Hansen, R. W., and Grabowski, H. (2003). "The price of innovation: new estimates of drug development costs". *Journal of Health Economics*, 22:151–185.

McDonald, J. B. (1984). "The Some generalized functions for the size distribution of income". *Econometrica*, vol. 52:647–663.

McDonald, J. B. and Xu, Y. J. (1995). "A generalization of the beta distribution with applications". *Journal of Econometrics*, vol. 66:133–152.

McPherson, M. A. (1996). "Growth of Micro and Small Enterprises in Southern Africa ". *Journal of Development Economics*, 48:253–277.

Mincer, J. A. (1974). *"Schooling, Experience, and Earnings"*. Columbia University Press.

Mitzemacher, M. (2003). "A Brief History of Generative Models for Power Law and Lognormal Distributions". *Internet Mathematics*, 1:226–251.

Miyashige, T., Fujii, A., and Kimura, K. (2007). "The Scale Effect in Blockbuster Drug Development". *International Review of Business Research Papers*, 3(4):161–172.

Mowery, D. C. (1983). "Industrial Research and Firm Size, Survival, and Growth in American manufacturing, 1921-1946: an Assessment". *Journal of Economic History*, 43(4):953–980.

Newman, M. (2005). "Power laws, Pareto distributions and Zipf's law". *Contemporary Physics*, 46:323–351.

Nickell, S. (1981). "Biases in Dynamic Models with Fixed Effects". *Econometrica*, 49:1417–1426.

Nickell, S. (1997). "Unemployment and Labor Market Rigidities: Europe versus North America". *Journal of Economic Perspectives*, 11(3):55–74.

Nickell, S. and Layard, R. (1999). "Labor market institutions and economic performance". In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 3 of *Handbook of Labor Economics*, chapter 46, pages 3029–3084. Elsevier.

Oliveira, B. and Fortunato, A. (2003). "Testing Gibrat's Law: Empirical Evidence from a Panel of Portuguese Manufacturing Firms". GEMF Working Papers 2003-07, GEMF - Faculdade de Economia, Universidade de Coimbra.

Ornaghi, C. (2006). "Mergers and innovation: the case of the pharmaceutical industry". Discussion Paper Series In Economics And Econometrics 0605, Economics Division, School of Social Sciences, University of Southampton.

Pammolli, F., Buldyrev, S., Fu, D., Riccaboni, M., Matia, K., Yamasaki, K., and Stanley, H. E. (2007). "A generalized preferential attachment model for business firms growth rates". *The European Physical Journal B*, 57:127–130.

Pareto, V. (1896). *"Cours d'Economie Politique"*. Macmillan, London.

Parker, S. C. (1999). "The Generalized Beta as a model for the distribution of earnings'. *Economics Letters*, vol. 62:197–200.

Pfeffermann, D. and Sverchov, M. Y. (2003). *"Fitting Generalized Linear Models under Informative Sampling"*, pages 175–195. in Analysis of Survey data, Wiley, NY edition.

Prais, S. J. (1974). "A New Look at the Growth of Industrial Concentration". *Oxford Economic Papers*, 26(2):273–288.

Revelli, L. (2010). "Work Histories Italian Panel (WHIP)". http://www.laboratoriorevelli.it/.

R.Henderson and Cockburn, I. (1996). "Scale, Scope, and Spillovers: The Determinants of Research Productivity in Drug Discovery". *The RAND Journal of Economics*, 27(1):32–59.

Ribeiro, E. P. (2007). "The dynamics of firm size distribution". *Brazilian review of econometrics : the review of the Brazilian Econometric Society*, 27(2):199–223.

Roodman, D. (2009a). " How to Do xtabond2: An Introduction to Difference and System GMM in Stata". *Stata Journal*, 9(1):86–136.

Roodman, D. (2009b). "A Note on the Theme of Too Many Instruments". *Oxford Bulletin of Economics and Statistics*, 71:135–158.

Saint-Paul, G. (1997). "The Rise and Persistence of Rigidities". *Journal of the Royal Statistic Society. Series B (Statistical Methodology)*, 87(2):290–94.

Salem, A. B. and Mount, T. D. (1974). "A convenient descriptive model of income distribution: the gamma density". *Econometrica*, 42:1115–1127.

Samuels, J. (1965). "Size and the Growth of Firms". *Review of Economic Studies*, 32(1):105–112.

Sargan, J. D. (1958). "The Estimation of Economic Relationships Using Instrumental Variables". *Econometrica*, 26:393–415.

Scherer, F. M. (1965). "Corporate Inventive Output, Profits and Growth". *Journal of Poltical Economy*, 73(3):290–297.

Simon, H. and Bonini, C. (1958a). "The Size Distribution of Business Firms". *The American Economic Review*, 48:607–617.

Simon, H. A. (1955). "On a Class of Skew Distribution Functions". *Biometrika*, 42:425–440.

Simon, H. A. and Bonini, C. P. (1958b). " The Size Distribution of Business Firms". *The American Economic Review*, 48(4):607–617.

Singh, A. and Whittington, G. (1975). "The Size and Growth of Firms". *Review of Economic Studies*, 42(1):15–26.

Singh, S. K. and Maddala, G. S. (1976). "A function for size the size distribution of incomes". *Econometrica*, vol.44:963–970.

Soo, K. (2011). "The size and growth of state population in the United States". Technical report, Lancaster University.

Sutton, J. (1997). "Gibrat's Legacy". *Journal of Economic Literature*, 35(1):40–59.

Wagner, J. (1992). "Firm Size, Firm Growth and Persistence of Chance: Testing Gibrat's Law with Establishment Data from Lower Saxony, 1978-1989". *Small Business Economics*, 2(4):125–131.

Weiss, C. (1998). "Size, Growth, and Survival in the Upper Austrian Farm sector". *Small Business Economics*, 10:305–312.

Windmeijer, F. (2005). "A Finite Sample Correction for the Variance of Linear Efficient Two-Step GMM Estimators". *Journal of Econometrics*, 126:25–51.

Wooldridge, J. M. (2010). *"Econometric Analysis of Cross Section and Panel Data, 2nd Edition"*, volume 1 of *MIT Press Books*. The MIT Press, second edition.

Xu, K. (2004). "How Has the Literature on Gini's Index Evolved in the Past 80 Years?". Technical report, Department of Economics, Dalhousie University.

Yasuda, T. (2005). "Firm Growth, Size, Age and Behavior in Japanese Manufacturing". *Small Business Economics*, 24:1–15.

Zipf, G. (1949). *"Human Behavior and the Principle of Least Effort"*. Addison-Wesley, Cambridge MA.