Departament d'Econometria, Estadística i Economia Espanyola
Facultat de Ciències Econòmiques i Empresarials
Universitat de Barcelona

ESSAYS ON TOTAL FACTOR PRODUCTIVITY, INNOVATION, EDUCATION AND TRAINING: THE ROLE OF SIZE IN SPANISH MANUFACTURING FIRMS

Laia Castany Teixidor

Directors: Dr. Enrique López Bazo i Dra. Rosina Moreno Serrano

Tesi Doctoral per a optar al títol de Doctora per la Universitat de Barcelona
Març 2007

Programa de Doctorat en Economia, bienni 2002-2004
Especialitat en Tècniques d'Anàlisi en Economia Aplicada

Als meus pares

AGRAÏMENTS

En primer lloc, voldria donar les gràcies als directors del grup de recerca Anàlisi Quantitativa Regional, els Drs. Manuel Artís i Jordi Surinyach, per haver-me donat la possibilitat de realitzar la tesi en l'entorn d'AQR i haver donat suport a la feina de recerca que he pogut desenvolupar al llarg del doctorat. Així mateix, voldria agrair el suport que he rebut per part del Departament d'Econometria, Estadística i Economia Espanyola, i concretament, dels Caps del Departament al llarg d'aquests anys, els Drs. Manuel Artís i Miguel Ángel Sierra.

Un dels agraïments més especials va per a les dues persones amb qui segurament he estat més vinculada durant la tesi, els meus directors, el Dr. Enrique López Bazo i la Dra. Rosina Moreno. Os agradezco mucho el esfuerzo que habéis hecho para guiarme en las diferentes etapas del doctorado, así como la exigencia y el rigor que habéis mostrado en todo momento. Gracias por todo lo que he aprendido trabajando con vosotros y el respeto que habéis mostrado hacia mi trabajo. En el ámbito más personal, gracias por vuestra confianza, por darme vuestra sincera y honesta opinión y por respaldar las decisiones que ido tomando. Gracias también a los dos por las clases de licenciatura, que me abrieron una pequeña ventana a todo esto y que de alguna forma despertaron mi curiosidad.

Dec un especial agraïment al Dr. Jordi Jaumandreu per oferir-me la possibilitat de col·laborar en la realització d'algun dels seus treballs, així com per les seves idees i comentaris. I am also most thankful to Dr. Hessel Oosterbeek for his valuable comments and suggestions during my stay at the University of Amsterdam.

Estic en deute amb els companys de departament Elisabet Motellon pels seus valuosos comentaris en relació als Capítols 3 i 5 de la tesi, així com també amb Josep Lluís Carrión i Tomás del Barrio per la seva ajuda en questions estadístiques i economètriques.

Aquesta tesi s'ha beneficiat també en gran mesura dels comentaris dels participants al EARIE 2006, SAE 2005 i 2006, JEI 2005, EEA 2005, ERSA 2005, el Workshop "Competition, Innovation and Productivity" celebrat a Niça al 2005 i el Workshop "Knowledge and Regional Economic Development" del projecte COST celebrat a la Universitat de Barcelona el 2005, així com també dels comentaris de diverses persones que han pres interès en aquest treball i m'han aportat molt bones idees. En especial, estic molt agraïda amb el Drs. Francisco Requena, Jordi Jaumandreu, Ana Martín-Marcos, Carmine Ornaghi, Luis Orea, Antonio Alvarez, Emili Grifell, W. Erwin Diewert, Andries de Grip, Ada Ferrer, Néstor Duch i Joan-Ramon Borrell.

Aquesta tesi s'ha pogut dur a terme gràcies al finançament obtingut de la beca de recerca i docència concedida per la Universitat de Barcelona i gràcies als recursos materials i intel·lectuals facilitats pel grup de recerca AQR i el Departament. Així mateix, la tesi ha estat posible gràcies al suport econòmic del projecte SEJ2005-07814/ECON del Plan Nacional de I+D+i que porta per títol "Capital humano, productividad y crecimiento regional. Análisis de la heterogeneidad de los efectos".

Així mateix, m'agradaria agraïr a la Fundación Empresa Pública l'accés a les dades de la Encuesta sobre Estrategias Empresariales i a les persones del departament d'estadística de la Fundación Tripartita para la Formación en el Empleo, la seva enorme disponibilitat en oferir dades.

Estic en deute amb els companys de departament i de doctorat que he tingut de costat en el dia a dia, Elisabet Motellon, Oscar Villar, Vanessa Berenguer i Isabel Cortés. Gràcies a tots pels bons moments que hem compartit i per totes les coses que he après al vostre costat. Un record molt especial per al Jorge. Un altre agraïment molt especial és per als meus primers companys de despatx, els Drs. Salvador Torra, Paco Llorente i Oscar Clavería, que em vau acollir en el vostre espai quan vaig arribar nova. Gràcies en definitiva a totes les persones del departament i del grup de recerca amb qui he pogut intercanviar idees i que d'una manera o altre m'heu fet costat.

Als meus amics, que sempre heu estat a prop i m'heu donat tot el recolzament, no només durant aquests anys, sinó tota la vida: Marta, Adrià, Rosa, Laia, Dani i Aniol. Finalment, és als meus pares, Marga i Xavi, a qui dec un dels agraïments més grans. Gràcies per haver-me donat suport incondicional i haver-me animat sempre i en tot. Gràcies per tot el que heu lluitat i perquè de vosaltres he après algunes coses molt importants.

CONTENTS

PART III

LIST OF TABLES

LIST OF FIGURES

GENERAL INTRODUCTION

Background

The improvement of the living standards around the world strongly relies on the countries' capacity to increase productivity in the long run. Productivity is a measure of the efficiency with which a country, a firm or a worker produces goods and services. In broad terms, it refers to the quantity of output that can be obtained from a productive process by using a given quantity of inputs such as labour or physical capital. Productivity is the element that permits output increases without increasing the effort, so that theoretically everyone who contributes to the productive process could benefit from it. In contrast to the gains derived from taking a larger slice of the economic pie, productivity is expected to increase the size of the pie for everyone. As a result, productivity growth is the most important determinant of the countries living standards. More concretely, productivity increases are expected to turn into lower unitary costs, which under perfect competition results in lower prices for consumers; higher productivity also represents higher labour compensation, which enables workers to get higher wages with the same hours worked; higher productivity is also reflected in higher profits for the companies, which in some cases are dedicated to new investments; finally, at aggregate level, higher productivity leads to countries' economic growth and development. On the other hand, productivity plays a key role in shaping monetary and fiscal policy, by affecting decisions about interest rates and future tax revenues.

If our future living standards depend on productivity, how can we achieve higher productivity levels? Theoretical models and empirical evidence agree that the gains in productivity strongly rely on the technological activity and on the improvement of the quality of the labour force. On the one hand, scientific and engineering previous knowledge provides the foundation for the generation of new technologies. More specifically, private companies, universities, technological centres, centres of transference of technology, technological and scientific parks and some public administrations are important sources of generation of knowledge departing from an effort on research and development (R&D). Considering the firm-level perspective, the technological activity can take different shapes, ranging from the adoption of advanced technology, the investment in R&D or the introduction of innovations (process, product or organizational). Firms' support to R&D depends on whether the expenditure could result in new or improved products and processes. Later, firms have to transfer these innovations to commercial products that will in turn increase their profits. In reality, it is not innovative input (R&D investment) *per se* but the actual use of innovative output (i.e. innovations or use of advanced technology) that directly affects productivity. Thus, obtaining productivity increases out of an innovative effort is the result of a complex and risky process that involves: firms' consideration of technology as a strategy to improve their productivity; the organization of a multidisciplinary environment where innovations can be obtained; adaptation of workers' skills to the specificities of their technological needs which are in permanent evolution; and finally, protection of their intellectual and industrial property rights to capture the returns from their investments.

Productivity depends also on the quality of the total labour force, and not only the quality of the workers involved in R&D activities. As Keynes (1944)[1] already pointed "Economic prosperity depends not on how brilliant a few people are, but on how large a scale you are able to produce competent people in all walks of life". In this view, human capital has not only an indirect effect on productivity (through generating new technology) but also a direct positive impact on productivity. Human capital is acknowledged to be an important determinant of productivity as well as other economic outcomes, both at individual and at aggregate level, and its role is particularly crucial in today's knowledge-driven economy. At the individual level, there is clear evidence that school attainment is a primary determinant of wages. At the macroeconomic level, there is evidence that human capital positively contributes to aggregate productivity growth. Moreover, theoretical models of human capital and growth suggest that some of the benefits of a more educated labour force will "spill over" and generate macroeconomic

[1] Keynes unpublished address delivered around 1944 to the Marshall Society.

benefits beyond those in the form of higher earnings appropriated by those who undertake the relevant investment on education. Some authors comment that the existing literature has relied almost exclusively on the years of formal schooling. At firm level, though, training has special importance in considering the relation between a firm's human capital investment and its performance. In this view, firms can not only hire educated workers, but they can also provide training, which contributes to increase their productivity. The firm-provided training is expected to match better their specific skill requirements, although usually the access to training requires previous formal education. In addition, human capital is considered to have positive effects other than economic, for example, on social cohesion.

Technological activity and human capital are typically considered to have a major positive effect on productivity at employee, firm and country level. When focusing on the firm-level perspective, the literature reports evidence of high heterogeneity among firms with similar characteristics, such as belonging to the same industry or region. The heterogeneity can be found in the productivity achieved by firms or in their strategic decisions, as for example, the investment in physical capital, R&D, human capital, the geographical scope of their market, the participation of foreign capital in the firm or their export propensity. Firm size has been considered a main source of heterogeneity, implying that small and medium-sized enterprises (SMEs) and large firms would show disparity in their strategic decisions or in their efficiency levels. Actually, the evidence is compelling that large firms are more productive than SMEs. According with the literature, some of the main reasons that may explain this stylized fact are the scale economies effect; the fact that different technologies are available and that, at different production levels, some technologies would be more appropriate than others; or the industry where the firm operates, among other reasons. SMEs usually find difficulties in achieving economic results as good as in large firms and accessing the main factors that contribute to firms' productivity. However, SMEs play a very important role as employment creators, innovators and entrepreneurs. Moreover, innovation in SMEs has not only a direct contribution to their own competitiveness, but to the economy as a whole: SMEs act as initiators, catalysts and media for wider technical change. For this reason, the development of SMEs is an issue moving to the top of policy makers' agendas in industrialized countries. In this view, SMEs and large firms seem to play different roles in our economics. All in all, SMEs and large firms seem to obtain different economic results and take different strategic decisions, such as their innovative activity or the investment in human capital. For the sake of simplicity, in the remainder of the thesis, we will talk about small firms instead of SMEs.

The Spanish economy is characterized for having a smaller average firm size and a lower proportion of large firms than other advanced economies. According to the Observatory of European SMEs,[2] in 2000, only 0.1% of the Spanish firms had above 250 employees. While the average EU-15 is 0.2% and some of the most advanced economies in Europe such as Denmark, Finland, Sweden or The Netherlands reach values between 0.4% and 0.5%. Although there are few firms with more than 250 employees, they are responsible for 20% of the Spanish employment. In the average EU-15, this percentage is 34% while, in the mentioned advanced economies, they employ between 40% and 50% of the workforce. These data reflect the reduced importance of large firms in Spain compared to other advanced economies. The proportion of firms that have 10 to 250 employees in Spain is quite similar to other economies: in Spain these firms represent 5.2% and employ 32% of the workers; in the average EU-15, they are 6.6% and occupy also 32% of employees; finally, in some of the most advanced economies, these firms are between 6% and 10% and employ between 33% and 40% of the workers. As discussed in the previous paragraph, small and large firms seem to obtain different economic results and take different strategic decisions. In this view, having more small firms would be associated with lower productivity, as well as a lower innovative effort and investment in qualified workers and training. In this context, firm size, innovation and human capital could be interacting in explaining the weak productivity performance in Spain.

As in other advanced economies, the Spanish productivity has suffered a deceleration process since the mid-nineties. A common recommendation indicates that Spain should increase its efficiency and that it requires a higher investment effort in technologies and human capital (see for instance, the National Reform Program for Spain in the Lisbon Agenda). Although this economy has performed quite well in increasing the years of formal education over the last three decades, it is still far from

[2] Observatory of European SMEs
(http://www.eim.nl/Observatory_7_and_8/en/stats/2001/var2/1cou_size.html, 1st January 2007).

the EU average in terms of innovative effort. Thus, the still low levels of human capital, and, especially, innovation in Spain could in part explain its weak productivity.

In addition, it is possible that the low productivity levels are not only due to this low effort in technological activities and human capital, but also to the low impact of such effort on firms' productivity. In other words, it could be the case that it is not only necessary to invest more in the main factors driving productivity, but it is also important that this effort has a high positive impact on productivity. The emphasis here is placed on how the innovations and the skills of the workforce translate into higher productivity.

Objectives and Hypotheses

The main purpose of this PhD thesis is analysing the behaviour of different factors that contribute to increase total factor productivity (TFP) in the context of Spanish manufacturing firms: mainly innovation, formal education and training. The general underlying hypothesis is that small and large firms take different strategic decisions in relation to their investment in these factors and they may obtain different productivity.

Departing from the general hypothesis that small and large firms play different roles in our economics, obtain different economic results and take different decisions, the first question analysed in this thesis is that large firms may achieve higher productivity levels for two reasons: first, because they innovate more and employ more qualified employees than small firms, and second, because they may be able to obtain higher returns from this effort. The emphasis placed on the analysis of the effect of returns in explaining the productivity differentials between small and large firms is one of the main contributions of this study.

As commented above, human capital is generally considered to have a direct positive effect on productivity. Firms can incorporate more human capital by hiring educated workers, but they can also provide training as a way to increase the skills of their employees. Concretely, training permits adapting workers' skills to the permanent evolution of job requirements and enhances the competitive position of workers and their employers. The main purpose of continuous training is to provide knowledge and adequate skills to occupied employees so that they could adapt to the changing requirements of firms at any moment. Thus, firms' may be interested in providing training to their workers as a way of increasing their productivity.

A general finding is that large firms usually provide more training, while small firms face more difficulties in providing training. This can be seen as an additional limitation for small firms to achieve higher productivity levels. The literature suggests different reasons why large firms provide more training: scale economies, training as a way to reduce monitoring costs, access to cheaper capital or higher pay-offs from their investments, among others. According with the reasoning in the previous paragraphs that firm size is a source of heterogeneity, we depart from the idea that small and large firms behave differently in relation to their strategic decisions. We argue that this behaviour involves firms' decisions on training provision as well as other decisions that may determine such provision of training. Thus, we argue that large firms provide more training because they have certain characteristics that allow them to dedicate more efforts to training workers or that require more training. For example, large firms usually have a more qualified labour force and less temporary workers, which permit obtaining higher returns from their investment in training. Also, large firms innovate more and use more advanced technologies or operate in more competitive markets, which requires additional skills of their employees that can not be found in the labour market. These characteristics are considered relevant determinants of training by different studies in the literature. Thus, we argue that small and large firms may differ in these characteristics, which could explain in part that large firms decide to provide more training.

Spanish workers receive less training than in other countries and the Spanish firms are smaller than in other economies. Thus, if small firms provide less training, the difficulties of workers in small firms in accessing training can be considered as a limitation for the economy as a whole. The second question we analyse in this thesis is the relative contribution of training determinants in explaining the gap between small and large firms in their decisions to provide training.

In addition to the firm characteristics that determine training, those firms that receive subsidies are also expected to provide more training. Actually, there is a system of subsidies in Spain which intends to stimulate firms' provision of training. In this system, firms or groups of firms can obtain public aid to partially finance their training actions so that they make a more intense effort in training their employees. During 2001 and 2002, the eligibility conditions for a subsidy were considerably open to most firms, although the decision on whether to award the subsidy and the amount of it depended upon the evaluation of different authorities and was unknown by the firm until training actions finished.

As an extension of the second question in the thesis, we analyse the impact of subsidies on training. Following the firm-size perspective in the previous questions, we deepen in the analysis of the impact of subsidies in small and large firms. Actually, the subsidies regulation gives special importance to stimulating the provision of training in small firms.

Finally, we would like to clarify that this thesis started as an analysis of the impact of innovation and educative human capital on productivity. However, the later interest in questions related to human capital and the availability of data on training at firm level, led to the analysis of this component of human capital that is generated within the firm. Given that data on training was only available for some years, we decided not to include it as a determinant of productivity in Part II.

Methodology

To address the objectives in the thesis we follow an empirical approach based on a substantive theoretical framework, appropriate quantitative techniques and a comprehensive dataset.

In the case of TFP differences across firm size, previous to analysing such differences, we need to measure TFP at the firm level. More concretely, we are interested in a measure of TFP, which summarizes information about the relationship between output and the main inputs involved in the production process (labour, physical capital and materials). The large quantity of theoretical papers that suggest alternative ways of measuring productivity indicates that it is far from easy to suggest a unique measure of productivity. Thus, we compare the most usual TFP indices in the literature on the basis of some a priori-defined mathematical properties and on the basis of the production functions from which they are derived. According with these criteria, we select a TFP index and use it to measure TFP for a sample of Spanish manufacturing firms over the period 1990-2002.

Departing from this measure of TFP, we perform a preliminary descriptive analysis on the relationship between productivity, firm size, innovation and education and training. This analysis intends to characterize the evolution of TFP and our variables of interest for the Spanish manufacturing firms in the period of analysis. Particularly, we investigate the growth pace of productivity for small and large firms, paying special attention to the evolution of the differences in TFP between firms of different size classes. Moreover, we study whether the TFP gap by firm size is homogeneous along the distribution and whether the evolution of this gap differs for firms at different points of the distribution. Furthermore, we also analyse whether small and large firms follow different patterns in the intensity of use of technological and human capital. Finally, we provide preliminary evidence on the relationship between productivity, innovative activity, human capital and firm size. This descriptive provides further insights in the relationship between these variables, and constitutes the basis for the analysis of the differences in TFP between small and large firms.

The TFP differential between small and large firms is evaluated in the mean of the distribution using the Oaxaca-Blinder decomposition. This methodology permits decomposing the TFP differential and obtain the individual contribution of innovation and human capital in determining TFP. The decomposition departs from an auxiliary regression for each of the two groups under comparison. Every variable determining TFP may contribute to explain the TFP differential in two ways: as differences in characteristics, which means that the differences in TFP are due to the fact that one group invests more in technological or human capital, or as differences in returns, which means that the differences in TFP are due to the fact that one group obtains higher returns from these investments. One of the main contributions of this analysis is using the Oaxaca-Blinder decomposition in the field of empirical industrial organization, and concretely, in analysing the TFP differences between small and large firms.

The Oaxaca-Blinder decomposition permits evaluating the TFP differential in the mean of the distribution. However, in the presence of high heterogeneity among firms, the results for the whole distribution may differ from those at the mean of the distribution. Departing from the idea of the Oaxaca-Blinder methodology, we perform a counterfactual distribution analysis. The idea behind the counterfactual distribution analysis is transferring the Oaxaca-Blinder decomposition to the whole distribution, in the sense that we try to separate differences in firms' characteristics and returns at any point of the distribution. The counterfactual distribution analysis compares the density function of the estimated TFP of small firms with the counterfactual (or hypothetical) TFP, obtained by evaluating small firms under the returns of large firms. Thus, the difference between these density functions can be attributed to differences in returns between small and large firms. All in all, this methodology permits assessing the contribution of returns of a given variable at any point of the TFP distribution, which permits identifying the non-homogeneous behaviour of certain groups of firms.

As for the analysis of the second question raised in the thesis, the determinants of training, we consider that firms' decision on the provision of training is a double decision process, in which firms first decide whether to provide training or not and then, the quantity of it. Moreover, we argue that in the presence of fixed costs, some potential training providers may not decide to provide training and possible sample selection effects may appear. Thus, we propose estimating firms' decisions on training as a two-part model and discuss the possible sample selection problems. On the basis of these estimations, we analyse the effect of the determinants of training on firms' provision decisions: first, we estimate our specification for the subsample of small and large firms and use the results to further analyse the differences in the provision of training by firm size using the Oaxaca-Blinder decomposition; second, on the basis of this specification, we introduce a variable on subsidies on training.

As in the case of the TFP differential, we decompose the training provision differential between small and large firms in the mean of the distribution. In this case, we decompose the differential between small and large firms in the probability of providing training and the differential in the quantity of training per worker. The main objective of the Oaxaca-Blinder decomposition in this case is analysing which are the main variables that contribute to explain the gap in the provision of training between small and large firms. Both differences in characteristics and differences in the impact of these characteristics on training may contribute to explain the training gap.

The data used in this thesis are mainly drawn from the *Encuesta sobre Estrategias Empresariales* (Survey on Business Strategies, ESEE), an unbalanced panel which covers the period 1990-2002. This survey has been used in many papers on empirical industrial organization for Spain. The reference population of the ESEE is firms with 10 or more employees in the Spanish manufacturing sector. Small firms are defined as those with 10 to 200 employees and large firms, as those with more than 200. Since 1990, an average sample of 1800 firms has been surveyed yearly, on the basis of a questionnaire with more than 100 questions. According to data drawn from the Observatory of European SMEs, firms with 10 or more employees represent 5.3% of the firms in the total Spanish economy in 2000. Although this percentage is low, these firms employ 53% of the workers and produce 80% of the added value. According with data drawn from the DIRCE,[3] firms with 10 or more employees represent 18% of the firms in the Spanish manufacturing sector in 2001, indicating that the manufacturing sector concentrates larger firms than other sectors.

As for the preliminary descriptive of TFP, we obtain results for a sample of more than 13000 observations over the period 1990-2002, which means that we have around 800-1000 observations per year for more than 2000 different firms. Due to some methodological and econometric requirements, the analysis of the differences in TFP between small and large firms is based on three waves of the survey: 1994, 1998 and 2002. For the each period we count on more than 800 observations per year. Finally, the analysis on the determinants of training, firm size and subsidies is based on data for the last two years of the survey, when data on training were available. In this case, we obtain results based on data for more than 1500 firms per year. In addition, to measure the effect of subsidies on training provision, we use data from other sources. Particularly, data on subsidized training are provided by the Tripartite Foundation for Employment Training and data on the number of workers and worked hours are drawn from General Treasury of the Social Security and the Ministry of Labour and Social Affaires.

It should also be mentioned that the software used to obtain the empirical results in the thesis is Gauss v.6.0 and Stata v.9.0, specifically the commands: `heckman` and `xtprobit` and `xtreg`, in the panel data module.

Structure of the Thesis

As previously mentioned, this PhD thesis is structured in three main Parts, each one containing two Chapters that share a common empirical framework or perspective. Due

[3] *Directorio Central de Empresas* (DIRCE) database in 2001.

to this structure, an important effort has been done so as not to repeat information and specifying where it is possible to find it. Part I of the thesis includes Chapters 1 and 2. Chapter 1 presents the different methods that permit measuring productivity and, focusing on the method of index numbers, it reviews the main suggestions in the literature. Different index numbers are compared on the basis of the production functions from which they are derived and on the basis of some a priori-defined mathematical properties or axioms. Next, an index that incorporates desirable properties is chosen to develop the remainder of the analysis in Parts I and II. Chapter 2 presents the ESEE and the particularities of the variables involved in the measurement of TFP, as well as those of firm size, the innovative activity and human capital. Next, it offers a descriptive analysis of the evolution of TFP, as well as a descriptive of TFP in relation with the main variables of interest.

Part II of the thesis comprises Chapters 3 and 4 and it investigates the differences in TFP between small and large firms. Chapter 3 starts with a review of theoretical and empirical literature that relates firms' productivity, size and innovation and human capital, in their role of TFP determinants. Next, we explain the empirical specification and the methodology of estimation. This Chapter contains also a descriptive analysis and the results of the estimation. Finally, the size TFP-gap is evaluated in the mean of the TFP distribution using the Oaxaca-Blinder decomposition. Chapter 4 departs from the framework, specification and results in the previous Chapter and evaluates the TFP gap between small and large firms along the distribution using the counterfactual distribution analysis.

The last Part of the thesis contains Chapters 5 and 6 and it studies two different questions related with firms' determinants of training. Chapter 5 analyses the differences between small and large firms in their training provision decisions. This Chapter starts with a revision of the literature on the relationship between firm size and training and determinants of training. This Chapter also discusses the empirical specification and some methodological issues related with its estimation. Next, it provides a descriptive analysis and the results of the estimation. Finally, it explains the decomposition of training decisions using the Oaxaca-Blinder methodology and its results. Chapter 6 departs from the theoretical framework in the Chapter 5 and analyses the impact of subsidies on firms' provision decisions. First it describes the system of subsidies in Spain and the theoretical approach. Then, it describes the data used to measure subsidies. On the basis of the empirical model and specification in the previous Chapter, it provides the results for the estimation the effect of subsidies on training. Such effects are analysed for the total sample, as well as for the small and large firms' subsamples.

PART I

MEASUREMENT OF TOTAL FACTOR PRODUCTIVITY. DESCRIPTIVE ANALYSIS FOR THE SPANISH INDUSTRY, 1990-2002

INTRODUCTION

Productivity is considered the cornerstone that permits sustained economic growth in advanced economies. Growing faster than other countries or at least catching up with their growth rates is one of the most important objectives of macroeconomic policies nowadays. Although there is general consensus that productivity plays a central role in economic growth, it is still unclear what determines higher productivity and the way in which the effect occurs. A main strand of literature has tried to relate productivity (or productivity growth) with different variables that potentially affect it, for example, R&D, human capital or competition, among others. The positive effect of these determinants on productivity is a generally accepted finding, but a large part of what we call productivity remains still unexplained. However, before analysing what determines productivity, it is necessary to measure it, understand how it evolves and what the basic features of its distribution for the firms in an economy are.

Chapter 1 discusses different measures of total factor productivity (TFP) in order to choose an appropriate measure from the theoretical point of view. In Chapter 2, we compute a TFP measure for a sample of manufacturing firms in Spain over the period 1990-2002 using our favourite measure of productivity. In Chapter 2 we also provide a preliminary descriptive of the behaviour of the variables involved in the TFP index and of the TFP measure itself: we analyse their evolution over time and the relation between

TFP and firm size, innovative activity and the qualification of their employees, which are considered key variables in the analyses developed in this thesis.

The big quantity of theoretical works that suggest alternative ways of measuring productivity indicates that it is far from easy to suggest a unique measure of productivity. Three main methodologies, each one including a wide range of alternatives, have been proposed to deal with the question of measurement. Each methodology has its weaknesses and strengths in terms of the assumptions they make and choosing between them implies taking into account some aspects of productivity or others. Also, each methodology has its own requirements of data and the selection of one of them cannot be separated from the kind of data available (for example, the index numbers method requires availability of input prices). Moreover, the different methodologies serve different objectives (for example, the main purpose of the distance functions approach is disentangling the effect of different components of productivity). The objective of Chapter 1 is not to provide an exhaustive review of all the methods to measure productivity but to present the main arguments used to select an accurate measure of productivity. We briefly compare the three main methodologies and discuss our choice of the index numbers methodology.

Next, we go deeper into this methodology and provide an explanation of the main index numbers suggested in the literature. Although we do not pretend to review all the suggestions, we compare the characteristics of the most important indices. The main purpose of this comparison is exposing clearly the pros and cons of the main indices so as to choose one that incorporates more desirable properties. Building on the concept of TFP, we present the traditional index numbers (such as Laspeyres and Paasche, which are used to analyse many economic series) as well as TFP indices that are derived from production functions. The fact that they are derived from production functions permits establishing comparisons between them on the basis of the underlying assumptions of their corresponding production functions. Index numbers (derived from production functions or not) can also be compared on the basis of some a priori-defined mathematical properties or axioms. The main indices in the literature are compared using these two sets of tools and finally we chose the index suggested by Good et al. (1996), which incorporates more desirable properties and we argue that it constitutes an accurate measure of TFP.

In Chapter 2, we compute a TFP measure for manufacturing firms in Spain using our selected measure of productivity. Data are drawn from the *Encuesta Sobre Estrategias Empresariales* (ESEE, Survey on Business Strategies), which is an unbalanced panel that collects information of a sample of Spanish manufacturing firms over the period 1990-2002. This survey has widely been used in empirical industrial organization studies for Spain. First of all, we explain the main features of this survey and the cleaning procedure used to eliminate those observations that can be considered anomalous. A limitation of survey datasets is the fact that firms do not always respond to every question in the questionnaire and so information on all variables for every firm in the original panel is not always available. Next, we describe exactly how we calculate the output, inputs and their costs, which are used to construct the index. We pay special attention to the estimation of the stock of physical capital. The richness of this dataset permits measuring very precisely all the variables required, which is not always possible with some other firm-level datasets, for example when calculating inputs prices and costs. After all, we obtain a quite large sample of firms that permits an analysis of the main characteristics of the productivity distribution of Spanish manufacturing firms. Our final sample contains about 800-1000 firms for every year.

Later in this Chapter 2, we present a descriptive of the evolution of the output and input variables involved in the TFP index as well as the evolution of the TFP index itself. We find that TFP increases over these 13 years at an average annual increase of 1.56%. Although TFP increases almost every year, we observe a slow down during the second part of the nineties. Other studies at microeconomic level, such as Huergo and Moreno (2006), find similar results and an annual increase of TFP of 1.7% over the same period.

A main strand of literature is devoted to explain the main reasons for firms with similar characteristics achieving different TFP levels. The firm size has been considered a main source of heterogeneity and usually large firms obtain higher TFP levels. This may specially be the case of the Spanish economy, with a percentage of small and medium enterprises (SMEs) superior to other advanced economies. The substantial differences in the structure of firms' size justify our interest in the relationship between size and productivity. We present a descriptive of TFP by firm size, and according to the literature, we find that large firms achieve higher TFP levels.

On the other hand, technological and human capital have been considered two of the main factors determining productivity. Innovative firms and firms that employ more qualified workers are usually associated with being more productive. A recommendation from the National Reform Program of Lisbon Agenda indicates that Spain should invest more in them to improve its efficiency. In this view, the hypothesis is that differences in productivity across firms are closely related to the use they make of these two factors. Thus, we argue that the behaviour of these factors may be related to the low productivity levels achieved by Spain during the nineties. Actually, the innovative effort of Spanish firms is far from that in its competitors. Regarding employees' qualification, this economy has made a notorious performance over the last thirty years. Nevertheless, the educational level of the Spanish labour force is not as high as in other advanced economies. We offer a descriptive that relates firms' technological and human capital with their productivity levels and find a positive association between them.

Chapter 1

DISCUSSION ON MEASURES OF TOTAL FACTOR PRODUCTIVITY

1.1. Introduction

The objective of this Chapter is presenting different measures of total factor productivity (TFP) in order to choose an appropriate one from a theoretical point of view. Three main methodologies have been proposed: the econometric approach, the index numbers methodology and the methodology based on distance functions.

Departing from the concept of TFP, we present the traditional index numbers and the TFP indices that are derived from production functions. We compare the most usual indices in the literature on the basis of some a priori-defined mathematical properties or axioms and on the basis of the production functions from which they are derived. Finally, we chose the index suggested by Good et al. (1996), which incorporates more desirable properties and we argue that it constitutes an accurate measure of TFP. The main characteristics of this index are summarized in the conclusions of this Chapter (see Table 1.1). Basically, this index has good properties in terms of transitivity, characteristicity and it is superlative. Moreover it is derived from a translog production function, which is more general than other production functions. This index also permits a decomposition of efficiency and technological change and permits relaxing the assumption of perfect competition.

In Section 1.2 we discuss our choice of the index numbers methodology. In Section 1.3 we expose the main index numbers suggested in the literature as well as some suggestions to relax some of the assumptions that index number make. Finally, Section 1.4 concludes.

1.2. Methodologies to Measure Total Factor Productivity (TFP)

One of the most popular ways of quantifying firms' performance is measuring productivity.[4] The productivity of a firm is defined in broad terms as the ratio between a function of produced outputs and a function of inputs used in the productive process at the same moment. If we consider that only one input is involved in the productive process, the resultant measure of firm performance is called Partial Factor Productivity (PFP). If we consider more than one input, it is called Total Factor Productivity (TFP).

Labour Productivity is a PFP measure that has widely been used for measuring firms' productivity due to its simplicity and low requirements of available data. Concretely, it has been often used to measure labour productivity, where the labour input is the only input considered. Its main limitation is that, when comparing firms that use different intensities of other factors, one may obtain that firms that make a more intense use of other factors (i.e. physical capital) and a less intense use of labour are more productive. As PFP takes only one input into account, *ceteris paribus*, the less labour-intensive firm will have a higher output-labour ratio, which does not necessarily mean that it is more productive, for example in the case of using capital-intensive techniques that are more costly than labour-intensive techniques. A PFP_i measure for input i can be expressed as:

$$PFP_i = Y / X_i \tag{1.1}$$

where Y is the quantity of output and X_i is the quantity of any input, for example, labour. A TFP measure is analogue to this expression but in a more general framework, where more than one input is considered:

$$TFP = Y / \sum_{i=1}^{n} \alpha_i X_i \tag{1.2}$$

where n is the number of different inputs involved in the production function and α_i is the weight of input i. TFP measures provide more complete information on firms' performance than for example, labour productivity. Nevertheless, TFP measures are more sophisticated: they consist of an aggregation of different inputs and require information on physical capital, intermediate inputs and all the other inputs participating in the production process, which is often uneasy to obtain. Although TFP indices are

[4] Diewert (2005) comments alternative measures of firm performance such as the real rates of return and benchmarking.

more elaborated and complex to calculate, we consider that they provide a more general framework for measuring productivity and the remainder of our work will be focused on these measures.

Three different methodologies have been developed and applied to measure TFP at firm level: the econometric approach, the index numbers methodology and the methodology based on distance functions. The econometric approach consists of the estimation of production functions to obtain the contribution of inputs in the production of the output. Then the TFP measure is calculated as a "residual", that is, as the output minus a weighted sum of inputs according to their relative contribution.[5] The index numbers methodology consists of calculating TFP as in expression (1.2) by substituting X and Y by index numbers on quantities and prices of inputs and outputs.[6] Finally, the methodology based on distance functions has the objective of separating TFP in at least two components: the technical efficiency (or movements of the firms towards the production frontier) and the technical change (an outward shift in the firms' production possibility set, which is due to innovations and the diffusion of new methods of organizing production, among others). The most efficient firms, those with a higher output-input ratio, are situated on the frontier of production and the distance between any given firm and the frontier is interpreted as the "technical inefficiency" of the firm.[7] Two main methodologies are used in this third approach: the data envelopment analysis (DEA) and the stochastic frontiers method (SF).[8] The DEA methodology involves the use of mathematic programming methods to construct a non-parametric surface or frontier over the sample data. The SF method consists of an econometric estimation of parametric functions that include a random error which accounts for measurement errors and other random factors.

The objective of Chapter 1 is not to provide a full explanation of the three methodologies and their differences, but to mention their distinctive characteristics,

---

[5] The best well-known work is the seminal paper by Solow (1957), which is at macroeconomic level. At firm level, see Olley and Pakes (1996), who take into account the endogeneity problems of production functions as well as selection problems due to firm exits, or Griliches and Regev (1995), among others. For the Spanish case, see Doraszelski and Jaumandreu (2006).

[6] See Balk (1998) for a complete review of microeconomic foundations of index numbers of prices, quantities and productivity.

[7] Coelli et al. (2000, pp 2) present a framework to distinguish the different components of productivity: technical change, technical and allocative efficiency and scale economics.

[8] Farrell (1957) was the first one in suggesting the use of distance functions to measure productivity and a survey of this approach can be found in Lovell (1993). The DEA and SF methodologies are based on Malmquist indices, which are derived from distance functions.

which justify the choice of the index numbers methodology to measure productivity in our work. See Coelli et al. (2000) and Carlaw and Lipsey (2003) for a comparison of the three methodologies. We distinguish between the index number "absolute" advantages and disadvantages over the other methodologies and those that are common to other methodologies. The main "absolute" advantages of index number over the other two methodologies are the following:

- First, they only require two data points to establish a comparison, while the other methodologies require a larger number of observations.
- Second, they involve no estimation of unknown coefficients. Then, for instance, endogeneity problems associated to the demand of every input do not appear. The other methodologies need to develop different tools to deal with endogeneity problems. When they are overlooked, inconsistent coefficients and estimations of TFP are obtained.[9]
- Third, some index numbers assume that every firm uses its specific technology and thus the participation of every input in the production function is firm-specific, while the econometric approach assumes some sort of homogeneity in firms' technology which might be quite an unrealistic assumption. Actually, a change in technology can produce a variation in TFP, but it also can be reflected as a variation in the participation of the inputs in the production function.
- Prices provide additional information about firms' choices of quantities of inputs regarding prices. Index numbers use explicitly information about prices, and thus they incorporate additional information in relation to the alternative methodologies. However, information about input prices is often unavailable in micro-level datasets.

The weaknesses of index numbers derived from a production function in relation to the other two methodologies are the following:

- First, the index numbers derived from production functions assume the cost minimizing or revenue maximizing behaviour of firms, while the other methods do

[9] More concretely, endogeneity problems will appear when the inputs and the residual in the econometric approach are correlated, which is a plausible option, as probably in the estimation we may omit relevant unobservable variables that determine both the demand of inputs and the output. Olley and Pakes (1996) or Doraszelski and Jaumandreu (2006), among others, overcome this limitation in the econometric approach by the simultaneous estimation of production functions and factor demands functions.

not. However, index numbers are not always derived from production functions (using index numbers derived from a production function may be very useful as the production functions are our usual framework and thus we can select the most appropriate index number by studying the assumptions of every functional form).[10]

- Second, as shadow prices (relation of substitutability of one input for another) are unobservable and only market prices are observed, the latter are used to calculate TFP indices. But market prices are affected by market regulations and, only under perfect competition, they equal the shadow prices. When one assumes that market and shadow prices are equal for a given year, one is assuming allocative efficiency.[11]
- And third, index numbers do not permit accounting for characteristics of the market structure, which may have an influence on TFP: the capacity of utilization of inputs (they assume that the installed physical capital is optimal), the effect of regulations, the variation on the inputs quality, etc. However, some of these limitations have been overcome due to the contributions of different authors. Hall (1990) suggested that the perfect competition assumption could be relaxed by calculating the relative participation of inputs on the basis of their total cost instead of the income. This author also suggested that the constant returns to scale assumption could be relaxed by estimating a parameter of the scale of production to correct the TFP index previously calculated.

The index numbers also have advantages and disadvantages that are common to other methodologies:

- First, some index numbers and the distance functions approach permit distinguishing between technological change and technical efficiency, while the econometric approach does not permit this decomposition. When it is not possible to make this distinction, the productivity increase of firms is considered to be pure technological change, and thus we assume that firms are fully efficient (i.e. they are situated on the production frontier).

[10] Actually, the so-called economic criterion of selection of index numbers is based on choosing between the functions from which they can be derived.

[11] Balk (2001) proposes an index that permits relaxing the assumption of allocative efficiency and clearly distinguishes it from technical efficiency, technological change effects, scale efficiency and the existence of mark-ups.

- And second, the index numbers and the DEA are non-parametric methods, while the econometric approach and the SF method are parametric.

A disadvantage of the index numbers methodology and the DEA is that these methodologies do not permit accounting for noise and measurement error, while the other methodologies do. Actually, measuring productivity as a residual, that is, as the increase of output that cannot be attributed to an increase in input quantities, implies that TFP is collecting a great deal of effects that are unknown to the analyst.[12]

1.3. Index Numbers to Measure TFP

In this analysis, TFP is measured using an index derived from a production function. In this Section we present a revision of the most important indices, highlighting their strengths and weaknesses so as to establish a comparison that permits choosing an appropriate index for the empirical analysis. We start by the simplest index numbers and discuss the improvements incorporated in the indices suggested over the years. We have organized this section following basically the works by Diewert (2005), Carlaw and Lipsey (2003), Coelli et al. (2000), Balk (1998) and Good et al. (1996). At the end of the discussion, we present Table 1.1, which summarizes the main features of the index numbers introduced before to facilitate their comparison. An exhaustive revision is beyond the scope of this Chapter, as there is a wide range of index numbers, each of them introducing slight variations in relation to others.

1.3.1. The basic structure of productivity indices

We start with the simplest case of comparison of TFP in a firm in moments 0 and 1, for the one output and one input case. This constitutes the basis of the following explanation on the use of index numbers to measure TFP at firm level. Assume that X_1/X_0 is an index of input quantities and Y_1/Y_0 is an index of output quantities. This way, the change in productivity between periods 0 and 1 is expressed as:

$$gTFP = (Y_1/Y_0)/(X_1/X_0) \tag{1.3}$$

If, instead of quantifying inputs (outputs) using physical quantities, we can quantify them using monetary units, it is possible to aggregate the inputs (outputs),

[12] Abramowitz (1956) considered the Solow Residual as "a measure of our ignorance".

overcoming one of the limitations of a multi-input (multi-output) index of productivity. Thus, data on prices of output and inputs are required, which are not always available in standard firm-level microeconomic datasets. In the case of multiple outputs and inputs, X and Y have to be replaced by indices that consider more than one output or input. An input (output) quantity index number is a function of quantities of input (output) and its prices for a given period, $F(w_0,w_1,X_0,X_1)$ and $Q(p_0,p_1,Y_0,Y_1)$, being w_1 and w_0 the prices of input in periods 1 and 0, and p_1 and p_0 the prices of output. The most popular quantity indices are those proposed by Paasche, Laspeyres, Fisher and Törnqvist.[13] This way, we can calculate TFP as a ratio between the output quantity index and the input quantity index. For example, using Laspeyres indices:

TFP index for multiple outputs and inputs: $TFP^L = \dfrac{Q^L}{F^L}$ (1.4)

Index of outputs: $$Q^L(p_0,p_1,Y_0,Y_1) = \sum_{j=1}^{m} \frac{Y_{j1}}{Y_{j0}} S_{j0} = \sum_{j=1}^{m} \frac{Y_{j1}}{Y_{j0}} \left(\frac{p_{j0}Y_{j0}}{\sum_{j=1}^{m} p_{j0}Y_{j0}} \right)$$

Index of inputs: $$F^L(w_0,w_1,X_0,X_1) = \sum_{i=1}^{n} \frac{X_{i1}}{X_{i0}} S_{i0} = \sum_{i=1}^{n} \frac{X_{i1}}{X_{i0}} \left(\frac{w_{i0}X_{i0}}{\sum_{i=1}^{n} w_{i0}X_{i0}} \right)$$

where $j=1,...m$ are the different outputs and $i=1,...n$ the different inputs. In the case of the Törnqvist index:[14]

TFP index for multiple outputs and inputs: $TFP^T = \dfrac{Q^T}{F^T}$ (1.5)

Index of outputs: $$Q^T(p_0,p_1,Y_0,Y_1) = \prod_{j=1}^{m} \left(\frac{Y_{j1}}{Y_{j0}} \right)^{\left(\frac{S_{j0}+S_{j1}}{2} \right)}$$

[13] See for example Carlaw and Lipsey (2003). The Laspeyres index is defined as the value of input (output) in period 1 measured at prices in moment 0 and divided by the value of input (output) in period 0 measured at prices in moment 0. The Paasche index uses the prices in period 1 to weight the input (output) instead of using the prices in 0. The Fisher index is the square root of the product of the Laspeyres and Paasche indices.

[14] The Törnqvist input (output) index is the geometric average of input (output) over two periods weighted by an average of the two periods weights. In the case above these weights are the value of every input (output). The Törnqvist also constitutes the basis of the most used TFP indices.

$$where \quad s_{j0} = \frac{p_{j0}Y_{j0}}{\sum_{j=1}^{m} p_{j0}Y_{j0}} \quad and \quad s_{j1} = \frac{p_{j1}Y_{j1}}{\sum_{j=1}^{m} p_{j1}Y_{j1}}$$

Index of inputs: $$F^T(w_0, w_1, X_0, X_1) = \prod_{i=1}^{n}\left(\frac{X_{i1}}{X_{i0}}\right)^{\left(\frac{s_{i0}+s_{i1}}{2}\right)}$$

$$where \quad s_{i0} = \frac{w_{i0}X_{i0}}{\sum_{i=1}^{n} w_{i0}X_{i0}} \quad and \quad s_{i1} = \frac{w_{i1}X_{i1}}{\sum_{i=1}^{n} w_{i1}X_{i1}}$$

Notice that in the case of one output and one input, these expressions simplify to (1.3). Another comment is in order: at this point, one has to make a decision on which of the indices is most appropriate, which is a strong assumption. However, we can go one step further and instead of substituting X and Y by a Laspeyres, Paasche, Fisher or Törnqvist index, substitute them by a production function (Cobb-Douglas, CES, translogarithmic, etc). Now, $Y_0=f_0(X_0)$ and $Y_1=f_1(X_1)$ are production functions, from which it is possible to derive index numbers.[15] Every production function imposes different restrictions and every index number is derived from a production function. Then, the different index numbers impose different restrictions and can be defined according to their underlying assumptions. For example, it can be proved that the Törnqvist index numbers are derived from translogarithmic production functions. One of the main advantages of using index numbers derived from production functions is that, as economists, we are familiar with them and it seems reasonable to assume that production is driven by one of these usual functions. The problem is that we are still making assumptions on the functional form. However, as explained above, the econometric approach has the same disadvantage.

To summarize, there are two main groups of index numbers: those that are derived from underlying production functions and those that do not. We select the former as they permit establishing comparisons on the basis of the underlying assumptions of their corresponding production functions. In Section 1.3.3 we expose in more detail the most common indices derived from production functions.

[15] See Diewert (2005) for an introduction to index numbers using production functions.

1.3.2. Criteria to select index numbers

There are two main criteria to compare index numbers. The axiomatic approach consists of several mathematical properties called tests or axioms, which are based on some a priori reasoning and that index numbers may satisfy or not. The economic approach consists of choosing an index number according to the underlying production function from which it is derived and, thus, according to the economic assumptions that it makes. The axiomatic criterion is applicable to any index number, while the economic approach is only applicable to indices that can be derived from production functions.

1.3.2.1. Axiomatic approach

Fisher (1922) and Diewert (1992b) proposed exhaustive lists of the different axioms. Some of the basic and commonly used axioms are listed below:

1. Positivity test: the index should be everywhere positive.
2. Continuity test: The index is a continuous function of the quantities.
3. Proportionality test: if inputs are scaled up by some constant, the value of the index is also multiplied by this constant.
4. Commensurability test: the index is not sensitive to the units of measurement of prices and quantities.
5. Point reversal test: $I_0^1 = 1/I_1^0$, where I can refer to any index of inputs, outputs or productivity and subscripts may refer to different firms, time periods or combinations of both.
6. Mean-value test: The quantity index must lie between the respective minimum and maximum changes at the commodity level.
7. Transitivity or circularity test: $I_0^1 = I_0^2 I_2^1$. Transitivity is important in cross-section comparisons (comparisons between observations that do not follow any natural order).
8. Characteristicity test: is the degree in which the shares of inputs or outputs are specific of the specific firms under comparison. The weights of every input in the production function may be specific of every pair of firms under comparison or equal for the whole sample, which is a quite restrictive assumption.[16]

[16] For axioms 7 and 8, see Caves et al. (1982b).

9. Exact index number: index number directly derived from an underlying function, which can be either a production, cost, income or utility function.
10. Superlative index number: index number directly derived from a flexible underlying function (that is, the index number provides a second order local approximation to an arbitrary functional form). Indices that are derived from a translogarithmic production function can be superlative.[17]

The Fisher and Törnqvist indices satisfy properties 1 to 6, 9 and 10. These two indices are not transitive, that is, they do not permit cross-section comparisons, but they can be transformed in a way that these comparisons are feasible.[18]

1.3.2.1. Economic approach

The economic approximation is based on the production function from which an index is derived. The main characteristics of these indices are listed as follows:

- The most important characteristic of these indices is the specific functional from which they are derived. The different production functions (Cobb-Douglas, CES, quadratic, translogarithmic, etc) originate different index numbers, and as long as some of them are more restrictive than others, their corresponding indices will also be more or less restrictive. The Cobb-Douglas is easy to mathematically manipulate but it imposes fixed returns to scale and an elasticity of substitution equal to unity. The translog does not impose these restrictions, but it is more difficult to mathematically manipulate. There are a number of functional forms that lie between these and that impose some restrictions. For example, the CES, which relaxes the assumption of unitary elasticity. The translog and quadratic are called flexible functional forms, as they provide a second order local approximation.
- The production function may also permit taking into account changes in the technology over time and thus, allow factor prices to change over time. This

[17] Diewert (1976) shows that the Fisher and Törnqvist indices are exact and superlative.
[18] This transformation is performed using the EKS method (proposed by Eltetö and Köves, 1964, and Szulc, 1964), as explained in Coelli et al. (2000, pp 84-87). Caves et al. (1982a) proposed a Törnqvist index that allows multilateral comparisons.

characteristic becomes especially desirable when long periods are considered or when the prices change considerably over time.[19]

- Moreover, some indices do not depend on the sample: when adding data for a new period, the values for the index in the previous periods remain unaltered. This is desirable as it allows updating the TFP measurement.

1.3.3. Index numbers derived from production functions

In this Section, we present some of the most commonly used indices in the literature. We follow Coelli et al. (2000) and Good et al. (1996) to organize our exposition of the index numbers according to the production function from which they are derived. Moreover, their axiomatic properties and other particularities are also discussed. Building on the simplest indices, we expose the improvements incorporated to the indices that have been suggested over the years. We end up with the index suggested by Good et al. (1996), which incorporates most of the desirable properties of the previous indices, according with both the economic and axiomatic approach. The comparison in this Section justifies our choice of this index to calculate TFP in the empirical analysis. Table 1.1 presented at the conclusions summarizes the properties of the indices commented here and intends to clarify our reasons for selecting this index. Although all the indices presented in this Section permit accounting for multiple outputs and inputs, for simplicity we consider only the case of a single output and multiple inputs, as in our database we cannot distinguish different outputs. However, notice that the indices of quantities of output in the case of multiple outputs are analogue to those of inputs.

1.3.3.1. The TFP according to Solow. The Cobb-Douglas production function

In his seminal work, Solow (1957) used a Cobb-Douglas production function assuming constant returns to scale to measure TFP: $Y = AL^{\alpha}K^{1-\alpha}$, where A represents the idea of total factor productivity and L and K the quantities of labour and capital inputs respectively. Even though Solow measured productivity using the growth accounting method, the literature has often used an index number that can be derived from the

[19] As Good et al. (1996) argue.

Cobb-Douglas production function, using Solow as a point of departure. We also use it as a point of departure for the discussion in this Section.[20]

To aggregate the different inputs, this index uses a geometric weighted average of the inputs. The weights α_i are calculated using input expenditure shares, *S*, across all the observations in the sample, and thus, they are constant across all firms *(e=1, ...E)* and time periods *(t=1, ...T)*. As Hall (1990) suggested, if this geometric average is calculated using the participation of the cost of every input in the total cost of inputs, the assumption of perfect competition can be relaxed. The TFP increase for a given firm, *e*, between two time periods, 0 and 1, is:

$$gTFP_e = \ln\left(\frac{Y_{e1}}{Y_{e0}}\right) - \sum_{i=1}^{n} \alpha_i \ln\left(\frac{X_{ei1}}{X_{ei0}}\right)$$

$$with \quad \alpha_i = \frac{1}{ET}\sum_{e=1}^{E}\sum_{t=1}^{T} S_{eit} = \frac{1}{ET}\sum_{e=1}^{E}\sum_{t=1}^{T}\left(\frac{w_{eit}X_{eit}}{\sum_{i=1}^{n} w_{eit}X_{eit}}\right) \qquad (1.6)$$

where Y_{et} is the quantity of output, X_{eit} is the quantity of the *n* types of *i*-inputs and w_{eit} their prices. As commented before, the different empirical applications introduce slight variations in the same index. For example, in Foster et al. (1998), the weights of the inputs are not specific for every firm, but they are calculated using the average of the industry where every firm operates. Bailey et al. (1992) express the quantities of outputs and inputs in differences to the mean for every industry. Bernard and Jones (1996), in a macroeconomic work, suggest an improvement to this index where the weights of the factors are specific of every economic entity. Coe and Helpman (1995), also at aggregate level, calculate the weights of a given country as the average over all the years in the sample for this country. These are only four examples of the wide variety of possibilities that index numbers permit. Of course, the different possibilities are very often conditioned on the available data and so index numbers have to be adapted.

1.3.3.2. The TFP index by Kendrick. The CES production function

The index suggested by Kendrick (1961) is based on a Constant Elasticity of Substitution production function (CES), which is more flexible than the Cobb-Douglas

[20] See Diewert and Lawrence (1999, pp 7).

(the Cobb-Douglas assumes elasticity of substitution equal to one, while the CES assumes that the elasticity of substitution between inputs is any constant). The expression of the linearly homogeneous CES production function is $Y = A\,L\,K\,(cL^{\rho} + dK^{\rho})^{-1/\rho}$, where ρ is the parameter that guarantees constant elasticity of substitution. This author used arithmetic weighted average of input prices to aggregate the inputs in the TFP index. The expression in terms of TFP growth rates for a firm between two time periods is:

$$gTFP_e = \frac{(Y_{e1}/Y_{e0})}{\sum_{i=1}^{n} \alpha_{i0}(X_{ei1}/X_{eio})} \quad with \quad \alpha_{i0} = \frac{1}{E}\sum_{e=1}^{E}\left(\frac{w_{ei0}}{\sum_{i=1}^{n} w_{ei0}}\right) \tag{1.7}$$

1.3.3.3. The Fisher Ideal TFP index. A superlative index

The Fisher Ideal and Törnqvist indices are presented in what follows. They make fewer assumptions than indices derived from the Cobb-Douglas and CES and they are also superlative, one of the desirable properties of index numbers according to the axiomatic criterion. The Fisher Ideal index was suggested by Fisher (1922) and it was the index that better satisfied the desirable properties of index numbers suggested by this author himself. This index is known as the geometric mean of the indices Laspeyres and Paasche, and so, apparently, it is not derived from a production function. However Diewert (1976) proved that this index is directly derived from a flexible function: the function of quadratic mean of order two, and thus, it is a superlative index.[21]

The main weakness of this index is that it only allows comparisons over time. The comparison of observations over time follows a natural order and this index makes binary comparisons between consecutive observations, and so transitivity is guaranteed. In cross-section comparisons, the index would not satisfy the property of transitivity, considered as one of the most desirables by Fisher himself.[22] The main advantage of this index is the characteristicity test, that is, the degree of specificity of the shares used to calculate the weights. When the shares are specific of the two specific firms under comparison, this property is completely satisfied, as in the case of this index. The

[21] See also Caves et al. (1982b) and Diewert (1992b).

[22] Coelli et al. (2000, pp 87) suggest a transformation of the Fisher index using the EKS method that permits transitive comparisons across firms.

properties of transitivity and characteristicity are both desirable; however, Fisher discovered that they suffer a trade-off so that it is difficult to have a transitive index that maintains a high degree of characteristicity. This is known as the Fisher dilemma. Other transitive indices showed a low degree of characteristicity which led Fisher to choose this index as the "Fisher Ideal", and later on, to discard the property of transitivity as a desirable property for index numbers. The Fisher Ideal TFP index is expressed as:

$$\text{TFP index: } TFP^F = \frac{Q^F}{F^F} \tag{1.8}$$

$$\text{Index of output: } Q^F(p_0, p_1, Y_0, Y_1) = \left(Q^L(p_0, p_1, Y_0, Y_1)Q^P(p_0, p_1, Y_0, Y_1)\right)^{1/2}$$

$$\text{Index of inputs: } F^F(w_0, w_1, X_0, X_1) = \left(F^L(w_0, w_1, X_0, X_1)F^P(w_0, w_1, X_0, X_1)\right)^{1/2}$$

where Q^P, Q^L, F^P, F^L are the indices of quantity of output and inputs of Paasche and Laspeyres.

1.3.3.4. The Törnqvist TFP index. The translog production function

The most widely used superlative index is the Törnqvist-Theil-translog. Diewert (1976) proved that this index is directly derived from a flexible function, the multiproduct and multifactor homogeneous translogarithmic function, and then, it is a superlative index. The expression of the translogarithmic function, in the case of two inputs (capital and labour) and one output, where the second order parameters are equal for all firms and imposing constant returns to scale is:

$$\ln Y = \alpha_0 + \alpha_L \ln L + \alpha_K \ln K + \alpha_{LL}(\ln L)^2 + \alpha_{KK}(\ln K)^2 + \alpha_{LK} \ln L \ln K \tag{1.9}$$

where $\alpha_L + \alpha_K = 1$ *and* $2\alpha_{LL} + \alpha_{LK} = 2\alpha_{KK} + \alpha_{LK} = 0$ and subscripts L and K refer to labour and capital inputs, respectively.

In the next sub-sections, we detail the properties of the most important Törnqvist indices, which are derived from translog production functions. Concretely, we explain the particularities of the divisia "chaining", the TFP index by Caves, et al. (1982a) and the "chained" multilateral index by Good et al. (1996).

a) The TFP index by Jorgenson and Griliches. The divisia "chaining"

The Divisia discrete index (also called Divisia "chaining") belongs to the family of Törnqvist indices and it was developed by Jorgenson and Griliches (1972). This index establishes binary comparisons between consecutive observations and these comparisons are chain-linked so that it is possible to compare observations that are not necessarily consecutive: the transitivity property permits comparing each time period with the first one (normalized). This index cannot be used in the case of comparisons between observations that do not follow a natural order, and thus it would not be transitive for cross-section comparisons. An advantage of this index is that the chain-linking of binary indices implies using different shares that minimize the cost of the inputs at every moment, and then, they approximate to the production technology that would be the most appropriate. This characteristic is especially useful when the shares suffer a great variation over time or when very long time series are available. Moreover the shares are not sample dependent and they remain fixed even when the sample is extended.

This TFP index is obtained from subtracting an input index (*lnX*) from an output index (*lnY*), being both exact and allowing comparisons between two time periods, *0* and *1*:

TFP index: $$gTFP_e = \ln Y - \ln X \qquad (1.10)$$

Index of output: $$\ln Y = \ln Y_{e1} - \ln Y_{e0}$$

Index of inputs: $$\ln X = \sum_{i=1}^{n} \alpha_{ei}(\ln X_{ei1} - \ln X_{ei0})$$

Weights of the inputs:

$$\alpha_{ei} = \frac{S_{ei1} + S_{ei0}}{2} = \frac{1}{2}\left[\left(\frac{w_{ei1}X_{ei1}}{\sum_{i=1}^{n} w_{ei1}X_{ei1}}\right) + \left(\frac{w_{ei0}X_{ei0}}{\sum_{i=1}^{n} w_{ei0}X_{ei0}}\right)\right]$$

The reference point is time 0 and it indicates how to normalize (usually in time series the normalization is in relation to the first time period). The choice of the reference point is not irrelevant. In this index, the input weights are specific of every firm *e*.

b) The TFP index by Caves, Christensen and Diewert. Multilateral comparisons.

The multilateral indices overcome the limitation of binary comparisons in the Divisia index and permit transitive comparisons both for time series and cross-section data.

Caves et al. (1982a) suggested an index that is transitive for all kind of comparisons, while maintaining a high degree of characteristicity. Moreover, given that it can be derived from a flexible production function, it is a superlative index. In spite of these advantages, this index still suffers from the same problems as all the indices derived from production functions: it assumes the price-acceptant cost-minimizing behaviour of firms. Moreover, it assumes constant returns to scale and perfect competition, although these assumptions can be relaxed as we explain in Section 1.3.4.[23] Another weakness of the Caves et al. (1982a) index is that it is sample dependent and does not take technological change into account.

To calculate the Caves et al. (1982a) index of output, one has to measure the output of any firm in relation to the output of a hypothetical firm (considered a reference point), that is, make a bilateral comparison. Then, the same calculation is done for any other firm in relation to the hypothetical firm. This way, a multilateral comparison between the outputs of any pair of firms in the sample can be established, and this index is transitive for temporal and cross-section comparisons. By introducing the hypothetical firm, one has an unequivocal basis to make comparisons between observations that do not follow a natural order. The same reasoning holds for the inputs index when more than one input is considered. Finally, TFP is obtained by subtracting the inputs index from the output index. The quantities of output and inputs in the hypothetical firm are the geometric average of output and inputs for all the firms in the sample. The shares of the hypothetical firm are the arithmetic average of the shares in all the other firms. To establish a comparison with the Jorgenson and Griliches (1972) index, S_{i0}, lnX_{i0}, lnY_{j0}, would become, respectively, $\overline{S_i}, \overline{\ln X_i}, \overline{\ln Y_j}$ where the variables with the bar are calculated as a mean across all the firms $(e=1,...E)$ and all the time periods $(t=1,...T)$. The expression of the TFP index for any given firm in a given time period is:

TFP Index: $\ln TFP_{et} = \ln Y - \ln X$ (1.11)

Index of output: $\ln Y = \ln Y_{et} - \overline{\ln Y}$

[23] Caves et al. (1982a) obtain a Törnqvist index from the average of two Malmqvist indices when their underlying production function is a translog but imposing much fewer restrictions than in the Caves et al. (1982a). Actually, in some cases, this index permits relaxing the assumption of constant returns to scale, so that it is more general.

Index of inputs: $$\ln X = \sum_{i=1}^{n} \alpha_{eit} (\ln X_{eit} - \ln X_{i})$$

Weights of inputs:

$$\alpha_{eit} = \frac{S_{eit} + \bar{S}_{i}}{2} = \frac{1}{2}\left[\left(\frac{w_{eit} X_{eit}}{\sum_{i=1}^{n} w_{eit} X_{eit}}\right) + \frac{1}{ET}\sum_{e=1}^{E}\sum_{t=1}^{T}\left(\frac{w_{eit} X_{eit}}{\sum_{i=1}^{n} w_{eit} X_{eit}}\right)\right]$$

Harrigan (1997) considered the Caves et al. (1982a) index for R different types of labour input. If the weight of the labour input is an aggregation of all of them, then the share of the type of labour r is: $S_r = w_r L_r / \sum_{r=1}^{R} w_r L_r$. To apply this variation of the original index, it is necessary to have information on the wage differential between the different labour categories. Griffith et al. (2004) suggest another variation that differs from the original index in that the TFP level for every individual is measured in relation to the TFP level of a frontier individual (defined as the individual with higher productivity growth in every industry and time period).

c) The TFP index by Good, Nadiri and Sickles. "Chained" multilateral comparisons

This index was suggested by Good in his doctoral thesis and it appears in Good et al. (1996). This index incorporates the desirable characteristics of the Divisia "chaining" indices and of the multilateral Caves at al. (1982a). On the one hand, it approximates the most adequate technology available at any time period and, on the other hand, it is transitive for observations that do not follow a natural order. Additionally it is not sample dependent.

This index can be calculated as the Caves at al. (1982a) but with a hypothetical firm that is not common to all the observations, but specific of every time period, which guarantees transitivity for all the observations in a given time period. The hypothetical firms are chain-linked over time, which guarantees transitivity across all the observations in the sample. The expression of the TFP index in levels for a given firm e in a given time period t is:

TFP Index:

$$\ln TFP_{et} = \ln Y - \ln X$$

$$= (\ln Y_{et} - \ln \overline{Y_t}) - \frac{1}{2}\sum_{i=1}^{n}(S_{iet} + \overline{S_{it}})(\ln X_{iet} - \overline{\ln X_{it}}) \tag{1.12}$$

$$+ \sum_{s=2}^{t}(\overline{\ln Y_s} - \overline{\ln Y_{s-1}}) - \frac{1}{2}\sum_{s=2}^{t}\sum_{i=1}^{n}(\overline{S_{is}} + \overline{S_{i,s-1}})(\overline{\ln X_{is}} - \overline{\ln X_{i,s-1}})$$

Index of output: $\ln Y = \left(\ln Y_{et} - \overline{\ln Y_t}\right) + \sum_{s=2}^{t}(\overline{\ln Y_s} - \overline{\ln Y_{s-1}})$

Index of inputs: $\ln X = \sum_{i=1}^{n}\alpha_{eit}(\ln X_{eit} - \overline{\ln X_{it}}) + \sum_{s=2}^{t}\alpha_{is}(\overline{\ln X_{is}} - \overline{\ln X_{i,s-1}})$

Weights of inputs:

$$\alpha_{eit} = \frac{S_{eit} + S_{it}}{2} = \frac{1}{2}\left[\left(\frac{w_{eit}X_{eit}}{\sum_{i=1}^{n} w_{eit}X_{eit}}\right) + \frac{1}{E}\sum_{e=1}^{E}\left(\frac{w_{eit}X_{eit}}{\sum_{i=1}^{n} w_{eit}X_{eit}}\right)\right]$$

$$\alpha_{is} = \frac{\overline{S_{is}} + \overline{S_{i,s-1}}}{2} = \frac{1}{2}\left[\frac{1}{E}\sum_{e=1}^{E}\left(\frac{w_{eis}X_{eis}}{\sum_{i=1}^{n} w_{eis}X_{eis}}\right) + \frac{1}{E}\sum_{e=1}^{E}\left(\frac{w_{ei,s-1}X_{ei,s-1}}{\sum_{i=1}^{n} w_{ei,s-1}X_{ei,s-1}}\right)\right]$$

The bar refers to the hypothetical firm for every year. The two first terms on the right hand side of the TFP index in expression (1.12) establish the comparison between any observation and the hypothetical firm in its time period. The other two terms of this expression establish the comparison between all the hypothetical firms over time. This way, a productivity differential is added to the index every year. This differential permits relaxing the assumption of constant technology over time. By introducing these yearly variations in the shares and quantities of the inputs one assumes that the technology is not constant, as the shares and quantities of inputs will vary over time. The result is an index that permits establishing comparisons of all the firms in the sample in relation to the hypothetical firm in the base time period (usually the first period is considered the point of reference).

Moreover, this index permits decomposing the TFP in productive efficiency (the two first terms on the right hand side, which describe the change in TFP of any firm in relation with the hypothetical firm) and technological change (the other two terms, which describe the change in productivity of a firm representative of its period, considered as a frontier).

According with the discussion in this Section, the different indices have different desirable properties. Table 1.1 in Section 1.4 summarizes the properties of each index. On the basis of this discussion, we select the index suggested by Good et al. (1996) to perform the remaining of our analysis. This index seems to incorporate more desirable characteristics than the previous suggestions, which justifies our choice of this index. The index by Good et al. (1996) has been applied for instance by Aw et al. (2003) or Lim and Hahn (2003) among others and, in the Spanish case, by Delgado et al. (2002) and Máñez et al. (2005).

1.3.4. Some improvements to index numbers. Relaxing the assumptions

Many efforts have been focused on overcoming some of the limitations of index numbers by relaxing their restrictive assumptions. Using more flexible production functions or directly estimating the influence of these assumptions on productivity are two possible ways of relaxing these assumptions. In what follows, we comment some of the suggestions that have appeared in the literature to relax assumptions (some of the assumptions have already been mentioned in Section 1.3.3). Maintaining these assumptions when they do not hold makes the TFP measure depart from its true value.

1.3.4.1. The existence of market power

Hall (1990) suggests a modified Solow residual that permits taking market power into account. The particularity of this modified Solow residual is that the shares are based on the cost of production instead of income. Under market power, prices equal marginal costs plus a profit margin (mark-up) and thus the value of the production is higher than the total cost of the factors involved in the production process. If one uses income-based shares instead of cost-based shares, in the presence of mark-ups, the inputs participation is being underestimated. Although some indices did not originally incorporate this modification, since the suggestion by Hall, most authors have been using cost-based shares in empirical applications. Thus, we have incorporated this modification in expressions (1.6), (1.10), (1.11) and (1.12). The cost-based shares are expressed as follows:

$$S_{eit} = \frac{w_{eit} X_{eit}}{\sum_{i=1}^{n} w_{eit} X_{eit}} \tag{1.13}$$

where X is the quantity of every input i and w, their unitary prices. In the case of the income-based shares, the denominator of this expression would have been $p_{et}Y_{et}$.

1.3.4.2. Non-constant returns to scale

Under market power it is possible that some firms show increasing returns to scale. In the absence of market power, it is not viable that a firm operates with increasing returns to scale, as it would not have enough revenue to pay for its inputs. Under non-constant returns to scale, the suggested TFP measures cannot distinguish between scale efficiency (an increase in the scale of production) and technical efficiency. Most indices do not permit relaxing the assumption of constant returns to scale, however in some cases it is possible.[24] Hall (1990) suggests estimating a returns-to-scale index, γ (or a parameter of the scale of production). This parameter is the inverse of the elasticity of output with respect to inputs:[25]

$$\gamma^{-1} = 1 \Big/ \left(\frac{\partial Y}{\partial X} \frac{X}{Y} \right) \tag{1.14}$$

Under constant returns to scale, the numerator and denominator increase at the same rate so that $\gamma = 1$. Under increasing returns to scale, the elasticity of output with respect to inputs will be $\gamma > 1$. This author suggested a modified Solow residual (MRS) that permits taking non constant returns to scale into account, otherwise the obtained Solow residual, and thus the productivity measure, would be biased. The MSR can be expressed as:

$$MSR = \Delta Y - \sum\nolimits_{i=1}^{n} \alpha_i \Delta X_i = (\gamma - 1)(\sum\nolimits_{i=1}^{n} \alpha_i \Delta X_i) + \theta \tag{1.15}$$

where θ is equal to the Solow residual under constant returns to scale.

[24] Balk (2001).

[25] As Hall (1990, pp 92) comments, it is better to estimate the inverse rather than the elasticity of output with respect to inputs directly because when this parameter is large, the standard error becomes also very large, and thus there is greater uncertainty on how much larger than one this parameter is (estimating the inverse of the elasticity, the parameter will take values between 0 and 1).

1.3.4.3. Measurement errors in output, physical capital and labour

Finally, some authors highlight the existence of measurement errors as a source of deviation of TFP from the real firm productivity that we try to measure (see for example, Carlaw and Lipsey, 2003). For instance, in the declining part of the economic cycle, firms may occupy part of their idle workers in tasks related to the construction or repairing the physical capital (*labour hoarding effect*). Then, even though the output does not increase in the current time period, the improvements of physical capital are probably translated into later output increases and thus in higher productivity. If this measurement error is overlooked, the TFP measure will deviate from the true productivity. Another source of measurement errors is related to labour input and it is due to the fact that wages do not depend on the effort of employees. Workers that make a more intense effort will be more productive, and not accounting for this effort will underestimate TFP. Another possible source of errors in the labour input is due to the fact that in general there is only data available on the monthly or yearly wage; however it would be more appropriate to have wages by hours (Bernard and Jones, 1996). The third source of error in relation to labour input is due to the fact that actually firms contract workers with some anticipation before the production is made and this lag is never reflected in the TFP indices, where labour and output are usually contemporaneous. The measurement errors related to physical capital originate in the impossibility of observing the true cost of capital and the period between the investment and the utilization of physical capital. Finally, another source of error in the measurement of capital is that most models incorporate depreciation of capital as a function of time and not of the actual utilization.

1.4. Conclusions

As a way of summarizing the discussion on the different indices presented before, Table 1.1 collects the most outstanding characteristics of these indices, according to both the economic and axiomatic approach and with the suggestions that permit relaxing some of the assumptions of index numbers. The columns in this table show the different indices commented in Section 1.3.3 and the rows, the characteristics that permit establishing comparisons between them. The first part of the Table refers to the characteristics related to the axiomatic approach and the second, to those related to the economic approach. The third part of the Table mentions the main contributions to relax the assumptions of index numbers in general and whether they apply to each index. As one moves from the left to the right part of the Table, the indices present more desirable characteristics.

The discussion in this Section shows that the different indices incorporate different desirable properties. The last column of this Table shows that the Good et al. (1996) index incorporates more desirable characteristics than the previous suggestions in the literature. For this reason, we choose this index to perform the empirical analysis in the remainder of Part I and Part II.

Table 1.1. Comparison of index numbers properties: economic and axiomatic approach

	Description	***Solow (1957)***	***Kendrick (1961)***	***Fisher Ideal (1922)***	***Jorgenson & Griliches (1972)***	***Caves, Christensen & Diewert (1982a)***	***Good, Nadiri & Sickles (1996)***
Axiomatic Approach:							
Transitivity	*Permits cross-section comparisons*	No	No	No	No	Yes	Yes
Characteristicity	*Specificity of weights*	No	No	Yes	Yes	Yes	Yes
Superlative indices	*Exact and directly derived from a flexible function*	No	No	Yes	Yes	Yes	Yes
Economic Approach:							
Production function	*Degree of restrictions*	Cobb-Douglas	CES	Quadratic mean of order two	Translog	Translog	Translog
Accounting for technological change	*Weights are specific of every time period*	No	No	No	Yes	No	Yes
Sample independent	*Index values do not change when adding new observations*	No	No	No	Yes	No	Yes
Shares	*Built as costs or prices*	Costs	Prices	Costs	Costs	Costs	Costs
Relaxing the Assumptions:							
Non-constant returns to scale	*Requires estimating a returns- to-scale index*	Yes	Yes	Yes	Yes	Yes	Yes
Imperfect competition	*Cost based shares*	Yes	No	Yes	Yes	Yes	Yes

Source: Own elaboration based on previous discussion

Chapter 2

COMPUTATION AND DESCRIPTIVE ANALYISIS OF TOTAL FACTOR PRODUCTIVITY FOR SPANISH MANUFACTURING FIRMS, 1990-2002

2.1. Introduction

In this Chapter, we calculate a total factor productivity (TFP) index for a sample of Spanish manufacturing firms between 1990 and 2002, using our selected measure of productivity, the index by Good et al. (1996). Data are drawn from the *Encuesta Sobre Estrategias Empresariales* (ESEE, Survey on Business Strategies). In the following Section, we introduce the technical particularities of this survey and the cleaning procedure used to eliminate possible anomalous observations. Our dataset is an unbalanced panel with around 800-1000 observations in each year, and which collects information of 2104 different firms over the 13 years.

Section 2.3 describes the estimation of the stock of physical capital and the other measures of variables on output, inputs and their costs, involved in the TFP index. In Section 2.4 we define and explain the particularities associated with the measurement of our main variables of interest apart from TFP: firm size, the innovative activity and human capital. Section 2.5 offers a description of the evolution of the output and input variables that intervene in the TFP index and of the evolution of the TFP index for Spanish manufactures. According with other studies, we find that TFP increases at an average annual rate of 1.56% between 1990 and 2002 and it slows down during the second part of the nineties. Given the importance of small firms in Spain in relation to other advanced economies and the fact that large firms are usually more productive, in the same section, we show a descriptive of TFP by firm size. Also, we show that

Spanish manufacturing firms that incorporate more technological and human capital are more productive. Finally, in Section 2.6 we conclude.

2.2. Dataset: the "*Encuesta sobre Estrategias Empresariales*"

In this Section we expose the basic characteristics of the dataset used for the empirical analysis in this work: the ESEE. This survey has been used in a great number of papers on empirical industrial organization for Spain. We also explain the cleaning procedure used to remove some anomalous observations.

2.2.1. Description of the dataset

We use a sample of Spanish manufacturing firms drawn from the ESEE.[26] The ESEE has its origin in an agreement subscribed in 1990 between the Ministry of Science and Technology (at the time, the Ministry of Industry and Energy), and the *Fundación SEPI* (formerly the *Fundación Empresa Pública*); the later has been the responsible for its design and control of its execution through the Program of Economic Research. Since that year, an average sample of 1800 firms has been surveyed yearly, on the basis of a questionnaire with more than 100 questions. Nowadays, the ESEE is partly financed by the Ministry of Industry, Tourism and Trade and *Fundación ICO.*

The ESEE was designed with two main purposes: first, it permits analysing in depth the evolution over time of the manufacturing industry through a large number of data regarding the activity and the decisions taken by manufacturing firms; second, the design of the ESEE intends to provide panel microeconomic information adapted to the specification and contrast of the econometric models resulting from the economic theory. Regarding its informative content, the sample is aimed at capturing information about the firms' strategies, that is to say, about the decisions they take regarding the competition tools at their disposal. These instruments include flexible instruments that vary in the short run (such as prices) and those which require a longer term to be effective (such as R&D expenditure) as well as information on the markets where firms operate and some accounting data. One of the main virtues of this dataset is the richness of its information as there are plenty of variables that permit analysing the heterogeneous behaviour of the productive units. Regarding the specific content of the

[26] See Fariñas and Jaumandreu (1999) for further details.

questionnaire, it includes the following sections: activity, products and manufacturing processes; customers and suppliers; costs and prices; markets covered; technological activities; foreign trade and finally, employment. The panel structure of this dataset is also a very valuable characteristic.

This annual survey covers the period 1990-2002. Every four years, firms answer a complete questionnaire; for the other years, they only answer a reduced form of it (with those issues that are supposed to change yearly), so that nowadays full information is available in 1990, 1994, 1998 and 2002. The reference population of the ESEE is firms with 10 or more employees dedicated to one of the activities corresponding to divisions 15 to 37 from the NACE-93, excluding division 23 (activities related to refinement of oil and fuel treatment).[27] In the base time period, all the firms with more than 200 employees were required to participate (and so 70% of them did). The firms with 10 to 200 employees were sampled randomly by industry and four size strata, retaining about 5%, so that representativity for every industry and firm size class was guaranteed.[28]

The ESEE is designed to change as the industry composition evolves. Newly created firms enter the sample using the original selection criteria. There are also exits in the survey (due to death and attrition) and these firms have been replaced with other firms in the same industry and size group so as to maintain representativity. In the initial year, information was available for 2188 firms. In later years, an important effort has been made to avoid the deterioration of the initial sample so as to obtain the above-mentioned panel structure. So, the ESEE is an unbalanced panel that attempts to capture the entry and exit of manufacturing firms over the sample period. This is an important characteristic given our interest in establishing comparisons of TFP distributions in different time periods. The data contained in this survey will permit calculating the output and input quantities and prices required to measure TFP. The variables on input

[27] The sectors following NACE-93 are grouped in the following 20 categories: Meat-processing industry; foodstuffs and tobacco; drinks; textiles; leather and footwear; wood industry; paper; editing and printing; chemical industry; rubber and plastics; non-metallic minerals products; iron and steel; metallic products; machinery and mechanical goods; office machinery, computers, processing, optical and similar; electrical and electronic machinery and material; motor vehicles; other transport material; furniture and a final group called other manufacturing industries.

[28] The random sampling scheme is stratified: different strata were defined by combining the sector with four size groups (according the number of workers: 10-20, 21-50, 51-100 and 101-200).

and output quantities and prices are not directly obtained from the ESEE and so they have been calculated according to the particularities explained in the next Section.

2.2.2. The cleaning procedure

Between 1990 and 2002, this survey has 37141 observations, for 3462 different firms. However it has not been possible to include them all in the TFP measurement because of the following reasons: first, some firms do not respond to some of the fields in the questionnaire that are necessary in our analysis, which implies that the sample is reduced to 15078 observations. For these observations, data are available to calculate labour, capital, intermediate inputs, output and their prices, as defined below. Although this loss of observations seems quite large, it is in line with other works that calculate these variables with the same dataset. Moreover, we consider that representativity is maintained, as a priori, productivity does not seem to be related to those firms responding to some of the questions in the questionnaire and thus these observations are randomly deleted. Second, before computing the TFP index we need to clean the sample in a way that removes anomalous observations according to the criteria described below.[29]

The cleaning procedure intends to eliminate the typing errors or anomalous observations, although it is not possible to treat other sources of bias such as the measurement errors explained in Section 1.3.4.3. We have followed these steps:

i. We remove 119 observations with negative value added.
ii. We drop all observations where the growth rate of output is higher than 1 but the growth rate of some of the inputs is lower than 0.5. We also drop the observations where the growth rate of output is lower than 0.5 but the growth rate of some of the inputs is higher than 1. We remove all observations where the growth rate of output is lower than -0.5 but the growth rate of some of the inputs is higher than -0.25. Finally, we remove all observations where the growth rate of output is higher than -0.25 but the growth rate of some of the inputs is lower than -0.5. In total, 1717 observations are removed by this procedure.

[29] Ornaghi (2006) cleans the same dataset with similar criteria.

iii. We remove 152 observations where the share of labour input or materials is higher than 0.95 or lower than 0.05.

iv. Due to the requirements of the subsequently applied techniques, we remove 55 additional observations corresponding to firms for which the TFP index is available in a given year, but not in the previous or the next one. This way we can obtain a TFP index for those firms that have observations for at least two consecutive years.

At the end, we obtain a sample of 13035 observations over 13 years (1990-2002), for 2104 different firms. Given that we have an unbalanced panel, this means that we have around 800-1000 observations per year.

2.3. Measurement of the Variables Involved in the TFP Index

In this Section we explain the methodology and data used to measure the quantities of output and inputs required for the TFP index, as well as their cost. Given the complexity associated to the measurement of the stock of physical capital, we pay special attention to this issue in Section 2.3.1. In Section 2.3.2 we comment the remaining inputs and output.

2.3.1. The stock of physical capital

According to OECD (1993), and following the System of National Accounts, the physical capital of firms can be defined as "the value, at a given point in time, of the durable, tangible, reproducible and fixed capital assets that are installed in producers' establishment and which constitute one of the factors that intervene in the production of other goods and services". The durable assets are those that have a duration of more than one year; they specify that these assets have to be tangible to exclude intangible assets such as patents, copyrights and financial assets; they also have to be reproducible in the sense that natural forests, land or mineral deposits are excluded; finally, the fixed assets implies that inventories and work in progress are excluded.

The difficulty involved in measuring the actual stock of physical capital of firms is considerable. As Martín-Marcos and Suárez-Gálvez (1997) comment, in the first waves of the ESEE, this question was directly asked in the questionnaire, however it had very few respondents and the answers did not use to be consistent. Thus, it becomes necessary to estimate the stock of physical capital from the balance sheets from the companies' accounts in the ESEE. We have built two different series: the stock of equipment and the stock of constructions, which constitute the stock of physical capital of the firm.[30]

To calculate TFP, it is necessary to obtain the net capital stock valuated at replacement cost in thousands of constant pesetas of 1990 (*KNR*). To obtain an estimation of *KNR*, we use the methodology of the permanent inventory, which has been widely used in the literature. The basic idea of the permanent inventory method is calculating the *KNR* for an initial year and, for the subsequent years, subtracting the depreciation, adjusting the prices to take inflation into account, and finally, adding the flows of gross fixed capital formation that have taken place over the year under consideration. This stock is computed by means of an iterative formula that permits adding the yearly investment to the initial stock of capital, obtaining series of net capital stock valuated at replacement cost for every year. Investments are considered to take place at the middle of every year, while the stock of capital refers to 31st of December.

To construct the stock of physical capital for the Spanish manufacturing firms in the ESEE, we follow Martín-Marcos and Suárez-Gálvez (1997), who obtained series for the period 1990-1994, which have subsequently been extended to 1999. First of all, one has to calculate the net stock of capital in the initial year (KNR_0). Although the survey starts on 1990, in that year, the questionnaire did not include questions on all the variables required to calculate the stock, and thus 1990 cannot be taken as the initial year. In 1991, the questionnaire included new variables that permit measuring the stock of physical capital; however, there were very few respondents to these new fields, so that 1991 is taken as the initial year only for those firms who provided the necessary information. For the firms that did not, 1992 is the base time period, even if they were already in the sample in 1991. The constant replacement reflects the value of the stock of capital assets supposing that all assets were purchased new in 1990. Our dataset contains data from the balance sheets on gross fixed assets valuated at acquisition cost for equipment goods and constructions (KBH_0). The acquisition cost (or historic cost) means that each asset is valued at the prices prevailing at the time the assets were

[30] Other studies do not include the stock of constructions as part of the stock of physical capital. However, we do it because otherwise we would be overestimating the TFP of those firms that have their own buildings and do not have to rent them (the rents are counted as intermediate inputs).

purchased, and thus the assets are aggregated using a variety of different prices. To obtain the net stock of capital it is necessary to subtract the depreciation of these assets since the moment they were incorporated in the fixed assets. With this operation, the net capital stock valued at acquisition cost for the base time period (KNH_0) is obtained.

More concretely, KBH_0 has to be multiplied by $(1-d)^{al}$, where d is the depreciation rate and al is the average asset life. To obtain the net capital at replacement cost, one has to take inflation into account. So, the net capital stock valued at acquisition cost in the base time period is multiplied by a price index. Table 2.1 summarizes these steps:

Table 2.1. Construction of the stock of physical capital

$$KBH_0 \longrightarrow KNH_0 = KBH_0\,(1-d)^{al} \longrightarrow KNR_0 = KNH_0(P_0/P_{0-al'})$$

(fixed assets in balance sheet)

where:
d = depreciation rate
al = average asset life
P_0 = Price index in the base time period
$P_{0-al'}$ = Price index according to:

$$al' = \begin{cases} 0 - regularization\ year & if\ (0 - regularization\ year) \le al \\ al & if\ (0 - regularization\ year) > al \end{cases}$$

Source: Own elaboration based on Martín-Marcos and Suárez-Gálvez (1997)

The depreciation rate is assumed to be constant over time but specific of every sector. The depreciation is the inexorable decrease in the flow of income from the capital assets expected in the future. The average asset life of the fixed assets of the firm (al) is calculated as the average age of the fixed assets on every year. The variable al' is also the average asset life but with the following particularity: as firms may regularize the value of their assets, this value does not always correspond to the year of acquisition, but to the year of the last regularization. Thus, if the difference between the base time period and the last regularization year is smaller than the average age of the assets, al' is calculated as the difference between these years. Otherwise, if the difference is larger than the average age, al' is calculated as the average age itself, as the assets have not been regularized before. The ratio of prices permits taking inflation into account and then obtaining KNR_0.

Once KNR_0 is computed, it is possible to use the iterative formula to obtain the series of stock of capital for the following years. As commented above, although the survey starts on 1990, the base time period is 1992. It should be noticed that in the case of firms that were in the sample in 1992, but which did not respond to the questions necessary to calculate the stock of capital, we have considered another base time period. If the firm was already in the sample in 1991 and offered appropriate data, 1991 is taken as the base time period. If it did not exist in 1991 or the required variables are not available, the base time period will be 1993. If the same problem occurs in 1993, then the base time period will be 1994 and so on. The same procedure is used for firms that were incorporated in the sample after 1992, taking as the base time period the first year for which data are available. This way, every firm has a specific base time period and the iterative formula is applied from this year on. The expression of the permanent inventory method is:

$$KNR_t = KNR_{t-1}(1-d)\frac{P_t}{P_{t-1}} + I_t \qquad (1.16)$$

This formula provides the net stock of capital at replacement cost in t, on the basis of the initial capital in $t-1$, which has suffered some depreciation over the period t. Moreover a price adjustment has been done so that the net stock of capital of the previous year is valued at replacement cost. After actualizing prices and accounting for depreciation in the stock of capital of the previous year, the investment done over period t is added to the stock of the previous year. This way, we obtain the net stock of capital at replacement cost for every year in the series. The permanent inventory method can be applied "forward" as in (1.16) and used to construct the series for years after the base time period or "backward" as we explain next. The "backward" version is the reverse of the "forward" version and is used to construct the series for the years before the base time period. It is obtained by simply isolating KNR_{t-1} in equation (1.16):

$$KNR_{t-1} = (KNR_t - I_t)\left(\frac{1}{1-d}\right)\left(\frac{P_{t-1}}{P_t}\right) \qquad (1.17)$$

However, OECD (1993) acknowledges the following limitations of the permanent inventory method. First, this method lies on some simplifying assumptions on the average life of assets: this methodology provides estimations on the "capacity stock" rather than on the "utilized stock", because we assume that assets depreciate over time and not due to the intensity of use. However, during some periods, the assets remain in the firms but they are idle or withdrawn from production. Second, the depreciation rate in this work is assumed to be exponential and constant over time, which is only one of the multiple possibilities than can be considered. Finally, the estimation of net capital stock on the initial year has also been criticized due to the difficulty that its measurement involves. Crucial information on the useful life of the stock of capital is seldom available. However, as explained in OECD (1993), the initial estimation has a decreasing impact on the reliability of the stock of capital estimations as years go by. They argue that after 25 years, most part of the assets in the initial stock has been retired from production.

The ESEE variables used to measure the stock of capital by the permanent inventory methodology are explained below:

- *d:* the depreciation rate is constant over time, but specific for every sector. As in Alonso and Collado (1999), the depreciation rate for each firm corresponds to the depreciation rate for the main activity of the firm according to the NACE-93 classification. We need a depreciation rate for equipment and constructions and they are obtained from Martín-Marcos (1990). In her work, the depreciation rates are calculated as the inverse of the useful life and they are expressed according to the classification of activities in the *Encuesta Industrial* and according to NACE-74, so that the equivalences between NACE-74 and NACE-93 have been considered.
- *al:* the average asset life in a given year is calculated as the average age of the assets in the firm, distinguishing between equipment and constructions. For those firms whose assets are older than 37 years, we have established the age at 37 years, as the OECD considers that assets older than 25 years are totally depreciated and because series of prices before 1954 are not available.
- *year of the last regularization* (if done), also distinguishing between equipment and constructions.
- *KBH*: the gross capital at acquisition cost is obtained from the value of the fixed assets in the balance sheets of the firms.
- The *capital deflator* that we use is the one in Martín-Marcos and Suarez-Gálvez (1997). As for constructions, between 1970 and 1994, these authors use the Index of

the Cost of Construction provided by the SEOPAN[31] with basis 1989; they use for the previous years, the deflator of the gross fixed capital formation from the National Accounts. As for equipment, between 1975 and 1994, these authors suggest using the component of equipment of the Index of Industrial Prices with basis 1990 (by the National Institute of Statistics, INE); for the previous years, they use the deflator of the gross fixed capital formation from the National Accounts. Our limitation is that these series are only available until 1994,[32] and thus from 1995 on, we have calculated it as we explain next. As for constructions, a price increase of constructions (from the implicit deflator) has been used to extend the available series. To calculate this implicit deflator, we have used series of the gross fixed capital formation in constructions by the INE (dividing series in real terms by series in nominal terms). As these series are expressed with basis 1995, we have had to express the deflator in pesetas of 1990. Finally, we calculate the increases of this deflator to obtain a measure of the price increase of constructions. The price increase is applied to the series from 1994 on, so that we obtain a price index of constructions until the end of the period. For equipments, the price index from 1995 on has been obtained from the series on Price Indices of Equipment Goods by the INE, which provides annual data with basis 1990.[33]

- I: is the net nominal investment in fixed assets, valued in the middle of the year and as if the whole expenditure was done at once. It includes the acquisition of fixed

---

[31] SEOPAN is the Association of Nationwide Construction Firms (*Asociación de Empresas Constructoras de Ámbito Nacional*).

[32] Martín-Marcos and Suárez-Gálvez (1997) calculate the series of the stock of physical capital and constructions for the period 1990-1994. The *Fundación Empresa Pública* (FUNEP) has extended these series covering the period 1995-1999. So series of the stock of capital are available for the period 1990-1999, but we need it for the period 1990-2002, so we have had to recalculate the whole series. Some of the data we use are provided in the original work by Martín-Marcos and Suárez-Gálvez (1997), but this data is only available until 1994. In Appendix 2.1 we compare our series and the series provided by the FUNEP to check the robustness of our results between 1990 and 1999.

[33] In absence of longer price series in the paper by Martín-Marcos and Suárez-Gálvez (1997), we need to use other series of prices or extend their series using the price increases in other similar series. To check whether they are appropriate, we perform the following comparisons. As for the prices of constructions, we compare the series of prices for constructions over the period 1980-2002 using the implicit deflator by the INE with those by Martín-Marcos and Suárez-Gálvez, which have been extended using this deflator between 1995 and 2002. Given that the variation between the two series is very small, we use the latter to estimate the stock of constructions. As for the prices of equipment, we compare the series of Price Indices of Equipment Goods by the INE over the period 1975-2002 with those by Martín-Marcos and Suárez-Gálvez (1997), which have been extended using the former series between 1995 and 2002. We obtain that the variation between the two series is minimal and use the latter to estimate the stock of equipment. As the variation is minimal for both constructions and equipment price series, we consider that the series used here to estimate the stock of physical capital are appropriate.

assets (constructions, computers, machinery, transport material and furniture) minus the sales of these assets. Given that in the database we cannot distinguish the proportion of the sales that can be attributed to equipment and constructions we estimate a percentage according to investment in these two concepts and impute them in each series.[34]

The difficulty involved in the estimation of the stock of physical capital and the assumptions required for this calculation, require a validation of the estimation. Thus we compare the series obtained by Martín-Marcos and Suárez-Gálvez between 1990 and 1999 with ours (see Appendix 2.1).

2.3.2. The output and inputs in the TFP index

In this Section we explain the particularities associated with the measurement of the output and inputs quantities required for the computation of the TFP index, as well as their prices. The quantities are expressed in thousands of constant pesetas of 1990, except for labour, which is measured as the number of hours worked in each firm. The TFP index has been computed only for those observations that had data available for all these variables and later on we have cleaned the sample, as explained in Section 2.2.2.

- *Output.* The output is defined as the production of the firm, but this variable is not directly available at the ESEE, so that we have calculated it as sales plus the variation of stocks for sale. The correction by stocks for sale permits taking into account that, in a given year, firms may have produced more than they manage to sell (which is accumulated in the inventories) or they may have produced less than they actually sell (the difference might have been produced in previous years being accumulated as stocks for sale to be sold in the following years). This way, we obtain the production in nominal terms. To obtain the real production in pesetas of 1990, we have divided the nominal production by a price index specific for every firm.

In order to obtain a firm-specific price index of output, the ESEE provides information on the price increases in the five main markets where firms operate. The sales in these five markets in total constitute at least half of the firms' sales. For every market, we have data on the price increase in a given year and the percentage of sales

[34] Martín-Marcos and Suárez-Gálvez only consider the acquisition of fixed assets, but we consider that sales of fixed assets should also be taken into account to avoid an overestimation of the stock of physical capital.

that this market represents on the total sales of the firm. To calculate a price increase specific of every firm, we calculate a weighted sum of the price increases in the different markets where the firms operate. The weights are the percentage of sales in each market. When calculating the weighted sum, we take into account that, first, not all the firms operated in five different markets. Second, firms offer data on the percentage that every market represent on their total sales, but the sum of all the percentages does not necessarily cover their total sales. In other words, we assume that the sum of the percentages reported by firms is equal to their total sales; otherwise the price increase would be underestimated. The limitation of this assumption is that the total sales are unknown. However, after some validations, the sum of the percentages reported by firms is close to 100%.[35]

The firm-specific price increase permits calculating a firm-specific price index. For each firm, we assume that the base time period is 1990, so that the price index takes value 1 in this year. For the following years, we calculate the price index by adding the price increases calculated before. The main limitation is that the ESEE is an unbalanced panel so that price increases are not available for every firm since 1990. The price increase for firms that enter the survey after 1990 is unknown until they enter the survey, and so it is not possible to construct a firm-specific series of prices. The assumption in these cases has been adopted in the literature[36] and consists of supposing that every firm has a price increase as if it had been in the survey between 1990 and the year it enters the survey. The hypothetical firm-specific price increase is calculated as the mean of the price increase for all firms in the same sector for any year. This way we obtain a firm-specific series of prices that covers the whole period and that is formed by: a hypothetical series of prices for those years before firms entered the survey and by a firm-specific series of prices reported by the firm itself when it enters the survey. These series of prices permit building a firm-specific price index for every year with basis 1990. The nominal production is deflated using this price index and we obtain the production in real terms at prices of 1990 for every firm on every year.

- *Labour input.* The amount of labour input used by the firm is calculated as the total effective hours of work. This measure is more precise than the number of employees.

[35] The percentage of sales revealed by the firms in the main five markets where they operate is quite high: in 1990, 89.74%; in 1994, 91.56%; in 1998, 92.45%; and in 2002, 92.05%.

[36] As suggested by Ana Martín-Marcos in personal communication.

The number of effective hours is obtained by multiplying the total number of employees by the effective hours worked during the year. The total number of employees is the number of full time employees plus the number of part time employees divided by two (both on December 31st) plus the number of temporary employees.[37] We calculate the number of temporary workers following this criterion: if the firm reports that the number of temporary workers has changed significantly over the year, we consider the arithmetic mean of the number of temporary workers at the end of every quarter. Otherwise, we simply consider the number of temporary workers on December 31st. The number of effective hours is calculated as the normal hours (work-time by law, collective agreement or labour contract for the majority of the personnel) plus overtime minus lost hours (hours paid but not worked).

- *Capital Input.* The amount of this factor is measured by the net stock of capital at replacement cost in real terms, calculated as explained in Section 2.3.1.

- *Intermediate Inputs or Materials.* The amount of intermediate inputs in real terms is obtained by deflating the nominal amount of intermediate inputs using a firm-specific price index. The amount of intermediate inputs in nominal terms includes: the purchases (acquisition of raw materials purchases, energy, etc.) and external services minus the variation in stocks of purchases. The correction by the stocks of purchases takes into account that the firm may use intermediate inputs bought in previous years and stored to produce the output in a given year. This correction also accounts for the fact that, in the current year, the firm may have bought intermediate inputs but without using them for the production.

The ESEE permits calculating a firm-specific price index to deflate the amount of intermediate inputs. The price index is calculated on the basis of the price increases of the intermediate inputs in the firm. Data on price increases of raw materials, energy and external services are available in the survey and we build a price index for each series. The price index in 1990 takes value 1 and, for the following years, we add the price increase in every year for each series. At this point we face the same limitation as in the case of the output price index: for firms entering the sample later than 1990, it is not possible to construct the series of the price index and thus we assume that the price increase of these firms is the arithmetic mean of the price increase of all firms in the

[37] See for example Suárez-Gálvez (2001)

same sector for any year. Next, we obtain the price index for the three series. The deflator of intermediate inputs as a whole is constructed as a weighted sum of the external services price index plus the raw materials price index plus the energy price index. On the one hand, the external services price index is weighted by the share of the cost in external services on the total cost of intermediate inputs. On the other hand, a global index for raw materials and energy, weighted by the share of the cost of purchases minus the variation of the stock of purchases, is constructed. The global index for raw materials and energy is obtained as the geometric mean of the two price indices where the two factors are weighed 0.95 and 0.05 respectively.[38] The expression of the intermediate inputs firm-specific price index is:

$$IP_{II} = \left(\frac{V_{RME}}{V_{II}}\right) * IP_{RME} + \left(\frac{V_S}{V_{II}}\right) * IP_S \tag{1.18}$$

where IP_{II} is the intermediate inputs price index for a firm in a given year; V_{RME} is the value of the purchases; V_S is the value of the external services; V_{II} is the value of the intermediate inputs and it is calculated as V_{RME} plus V_S; IP_{RME} is the price index of raw materials and energy and it is equal to $IP_{RM}^{0.95} * IP_E^{0.05}$; IP_{RM} is the price index of raw materials and IP_E is the price index of energy; finally, IP_S is the price index of external services.

- *Cost of Labour Input.* To calculate the shares of the inputs we use the percentage of their cost on the total cost of inputs. The ESEE provides the personnel costs of firms (including the employees' salaries, payments to the Social Security System and other labour costs paid by the firm). To obtain the expenditure on labour input in real terms we deflate the variable using the consumer price index, which in many industries is the reference price index used to update employees' wages every year. We use the series on the year-to-year price increase on December, so that it collects the price increase over the whole year. Then, we normalize the series by imposing that the value in 1990 takes value 1.

- *Cost of Capital.* The cost of capital is calculated as the user cost of capital, that is, the price of every unit of capital multiplied by the units of capital. The price of the capital input is defined as the interest rate minus the price increase (real interest rate)

[38] Martín-Pliego et al. (2001) use this methodology and weights.

plus the depreciation rate.[39] We construct series of the cost of capital for equipment and for constructions. First of all, the nominal interest rate is calculated as the average cost of the amounts falling due within more than one year (which is the interest rate paid by the firm to banks or other creditors). The ESEE has data on the interest rate paid for bank loans and to other creditors as well as the value of the balance liabilities. Then we calculate a weighted average of the two interest rates, where the weights are the volume of funding obtained from each type of creditors. Data on the interest rate and value of the liabilities for 1990 are not available in the survey and we assume that, for every firm, they are equal to the average firm in the same sector in 1991. Second, the depreciation rates and the price increase of capital goods are obtained as explained in Section 2.3.1. With these three elements, interest rates, depreciation rates and price increases, we calculate the price of one unit of capital, and later we multiply it by the quantity of capital used by the firm. This way we obtain the cost of physical capital as the sum of the cost of equipment and the cost of constructions. The cost of physical capital will be used to calculate the shares of this input in the production function.

- *Cost of intermediate inputs.* As usual in the literature, the expenditure on intermediate inputs is calculated as explained in *Intermediate Inputs and Materials*. See for example, Aw et al. (2003). Then, the weight of intermediate inputs is calculated as the ratio of the cost of intermediate inputs over the total cost of inputs.

2.4. Measurement of Firm Size, Innovative Activity and Human Capital

As commented in the Introduction of the thesis, we aim at investigating different questions in relation with two of the main determinants of firms' productivity, innovative activity and human capital, paying special attention to the differences in their effects for small and large firms. In the next section, we offer some preliminary descriptive on the relationship between these variables and TFP, which motivates the analysis in the following Parts of the thesis. In this Section we introduce and explain the particularities associated with the measurement of our variables of interest.

[39] Delgado et al. (2002) use this methodology to calculate the user cost of capital. However, according to Hall and Jorgenson (1967), the price of input capital is the real interest rate multiplied by the depreciation rate. The main weakness of their suggestion is that, in presence of high inflation, (for example in constructions) one may obtain very small or even negative prices of the capital input. The negative prices could imply a negative cost of capital and thus a negative weight of this input in the TFP index. We follow the methodology of the former paper as we do not expect to have negative returns to physical capital (negative weight); otherwise firms would not make any use of it.

As for the firm size, this variable is defined as the total number of employees and measured as the number of full time employees plus the number of part time employees divided by two (both on December 31st) plus the number of temporary employees.[40] We have defined small firms as those with 200 or less employees. Some exercises for other countries consider that small and medium enterprises (SMEs) are those firms under 250 employees.[41] As commented in Section 2.2.1, the ESEE makes the distinction at 200 employees and it uses different sampling schemes for the two groups. We consider it is appropriate to use the same criterion to guarantee representativity by size strata. Moreover, an outstanding characteristic of the Spanish industry is the reduced size of its firms in comparison with other advanced economics. So, it seems quite natural to consider firms with more than 200 employees, rather than 250, as large firms. Barrios (2000), Fariñas and Martín-Marcos (2001), Delgado, et al. (2002); Fariñas and Ruano (2004), Máñez et al. (2004), Fariñas and Martín-Marcos (2005) and Ornaghi (2006) use the same criterion when using data from the ESEE.

In relation with the measurement of innovative activity, the literature suggests a wide variety of variables to measure it at the firm level. On the one hand, the expenditure on R&D is a measure of innovative inputs or the effort of firms in R&D. The expenditure on R&D used here is expressed in thousands of 1990 constant pesetas per worker and measured as explained in the questionnaire: "Report the R&D expenses made by the company". The relationship between productivity and R&D expenditures embodies two different processes: the production of innovations starting from R&D activities and the incorporation of these innovations into production the production process. Thus, the innovative capacity can also be measured by process and product innovations. These variables are a measure of innovative output, the innovative effort that effectively turns into innovation. In our analysis, process innovation is a dicothomic variable that takes value 1 if the firm responds affirmatively to the following request: "Indicate if your firm introduced some significant modification in the production process (process innovation). If the answer is yes, please indicate the way: (a) introduction of new machines; (b) introduction of new methods of organization; (c)

[40] The number of temporary workers is calculated as explained in Section 2.3.2.
[41] Studies for other European countries often consider that small firms have 10 to 50 employees; medium firms have 50 to 250; and large firms have more than 250 employees. Moreover, some studies do not only classify the firms by size according with the number of employees, but also according with their annual turnover, their total balance sheet and the percentage of their capital that is participated by other firms.

both". Product innovation is a dicothomic variable that takes value 1 if the firm responds affirmatively to: "State whether the company has obtained product innovations (completely new products, or with such modifications that they are different from those produced earlier). If so, state how many and type of novelty which they entail".

In Part II of the thesis, we are interested in the relationship between innovation and productivity, and so the measures of innovative output are preferred over the measures of innovative input.[42] These measures of innovation have been used in other empirical works such as Huergo and Jaumandreu (2004a), Huergo and Moreno (2006) or Máñez et al. (2006). Still, in the descriptive analysis in Section 2.5.4 we offer results for the three different measures of innovative activity. Notice also that, for requirements of the econometric procedures used in Part II, the innovative activity is lagged one period, as commented in Section 3.3. In addition, there are economic reasons to believe that the effect of innovation on productivity takes place some time after the innovation is obtained. For coherence, these variables are also defined with lags all over Parts I and II.

As for the measurement of human capital, two different components of human capital are measured in this thesis. The first is related with schooling and it is called education or skilled labour. The second is related with life-long learning of occupied workers and it is called continuous training. The skilled labour is measured as the proportion of qualified workers according to their level of education. The category of qualified workers includes: engineers, graduates, middle level engineers, experts and qualified assistants. As in the case of innovation, for requirements of the econometric methods used in Part II and the economic reasoning, the variable on skilled labour is lagged one period. For coherence, these variables are also defined with lags all over Parts I and II. Our dataset only provides information about this variable in 1990, 1994, 1998 and 2002 because it is not considered to change yearly. Thus, to obtain the variable lagged one period, we have interpolated its values for 1993, 1997 and 2001, assuming that it increases linearly.

Continuous training is measured as the external expenses on training per worker, including fives different types of training: computation and information technologies,

[42] In addition, in Section 3.2.2, we argue that process innovation have a direct effect on productivity through the production function, while product innovations have an effect through the demand function. For all these reasons, in Part II the innovative activity is measured as process innovations.

foreign languages, sales and marketing, engineering and technical training and other issues (and expressed in 2001 real euros). The data on training are only available for 2001 and 2002.

2.5. Descriptive Analysis

This section contains a preliminary descriptive analysis of different variables related with TFP. First, we offer a description of the variables that involved in the TFP index to have a first idea of their evolution and to compare them with those described in other papers, especially at the macroeconomic level. Second, we characterize the evolution of TFP using both synthetic measures and an analysis based on the entire distribution of TFP. Finally, we describe the behaviour of TFP in relation with our main variables of interest: firm size, the innovative activity and human capital. This descriptive analysis shows interesting results for the Spanish manufacturing firms over the period 1990 to 2002. It provides preliminary evidence of certain patterns that will be further analysed in Parts II and III.

2.5.1. The variables involved in the TFP index

In this Section we characterize the most relevant aspects of the variables that intervene in the TFP index calculation. First, we offer a descriptive for the total sample, and next we analyse whether the results for the subsamples of small and large firms differ. We also compare the most outstanding findings on the behaviour of these variables in our firm-level dataset with the conclusions for the Spanish economy obtained by previous studies. Similar magnitudes and trends obtained from this comparison provide some guarantee on the construction of these variables and their adequacy for the computation of a representative TFP index for the Spanish manufacturing firms.

Table 2.2 shows the evolution of the average values of firms' production, number of effective hours of work, stock of physical capital and intermediate inputs for the total sample of firms over the period 1990-2002. It also shows the evolution of the average shares of every input in total cost of production and the evolution of some interesting ratios: the capital-labour ratio, indicating the degree of capitalization in the manufacturing industry, and the output-labour ratio as a measure of labour productivity, using both production and value added to define output. Tables 2.3 and 2.4 show the same descriptive, but for the subsamples of small and large firms. Tables 2.2 to 2.4 express the quantities of inputs and output in millions of constant 1990 pesetas, except for labour input, which is expressed in thousands of worked hours. The above-mentioned ratios are expressed in thousands of constant 1990 pesetas by worked hours, while the input cost shares are expressed in proportions over the total cost.

In Table 2.2, we observe that, in average, firms in the sample have slightly increased their production and stock of physical capital over the period of analysis. However, these variables show a cyclical behaviour with decreases in the first half of the nineties, some recovery in the second half, and a decrease again in the slowdown of the early 2000s.[43] The number of effective hours of work has diminished progressively during the period of analysis, with only a partial increase in the last two years of the nineties. This finding is also in line with those reported at aggregate level, which suggest that employment was not the main driver economic growth in Spain during the nineties.[44] On the other hand, the consumption of intermediate inputs has consistently followed the business cycle.

As for the participation of inputs in the total cost, the contribution of labour is approximately one third and it has slightly decreased over time, while the contribution of capital has increased from 3.8% in 1990 to 5.4% in 2002. This might be reflecting the increasing capitalization of manufacturing firms, although it can also be affected by the evolution of relative factor prices.[45] Finally, intermediate inputs contribute almost two thirds in the total cost without much variation over the period.

The analysis of the ratios confirms the existence of an intense process of capitalization and an improvement in the level of labour productivity, which is consistent with the above-mentioned evolution of production, labour and capital. The capital-labour ratio experienced an increase over the entire period of more than 70%. A higher capitalization of the industry is expected to carry on higher productivity. And so we observe in the output-labour ratio, which has substantially increased over the period. This trend is even more intense when value added is used to proxy for labour

[43] Some authors have shown the procyclical behaviour of capital in Spain, although with some lag (see for instance Pérez et al., 1998, pp 49).
[44] See for example, Pérez et al. (1998, pp 36) and Myro (2001, pp 56), who comment that, in Spain, as in other advanced economies (except USA and Japan), the economic growth is more due to the increase in labour productivity, rather than to the increase of employment.
[45] A deeper analysis of such issue is beyond the objectives of this work.

productivity: in this case, the level of productivity at the end of the period more than doubles that in 1990. However, the increase in both the capital-labour ratio and labour productivity is clearly more intense in the first half of the nineties.

All in all, figures from our sample of firms are consistent with previous evidence that reveals the process of capitalization and the improvement in the level of labour productivity that took place basically up to the mid nineties. Afterward, a slowdown in the pace of growth is clearly detected (Myro, 2001; Goerlich et al., 2002; Gual et al., 2006; Huergo and Moreno, 2006).

In what follows, we perform a similar analysis for the subsamples of small and large firms to show some evidence on the differences by firm size in the variables that characterize the process of production. Average values for the samples of small and large firms are shown in Tables 2.3 and 2.4 respectively. As for the participation of inputs in the total cost, the contribution of labour in small firms is approximately one third and it remains quite stable over time. However, in the case of large firms it decreases progressively from 29% in 1990 to 21% in 2002. Interestingly, the contribution of capital in small firms has increased from 3.3% to 5.4%, while in the case of large firms it has increased in a smaller magnitude (from 4.7% to 5.5%). Finally, the contribution of intermediate inputs has increased throughout the period in the case of large firms, while it has remained stable or even decreased for small firms. This effect can be related to a more intense process of outsourcing in large than in small firms.

The process of capitalization observed in the evolution of the capital-labour ratio for the entire sample of firms, is again observed for both small and large firms. But in magnitude, the intensification of capital seems to be more important in small than in large firms, which may be due to the fact that they depart from lower levels of capital and tend to converge to large firms. In 1990, the capital-labour ratio for large firms doubled that for small firms, and despite a somewhat better evolution in the case of small firms, differences in capitalization at the end of the period are substantially more favourable to large firms. As for the output-labour ratio, it has also increased progressively over the period for the two subsamples. In fact, it is reasonable to expect that a more intense use of capital leads to higher labour productivity. And so has happened to large firms, where higher labour productivity was achieved.

Table 2.2. Average of the variables involved in the TFP index, total sample.

year	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
No of obs	799	1084	1169	1052	996	934	910	1060	1076	1049	1056	986	864
Y	3617.4829	3367.0041	3329.3581	3085.3647	3425.488	3346.2728	3602.5036	3417.9827	4457.3894	4804.3743	4834.3583	3764.5106	3784.8471
L	391.3876	361.3084	330.4893	286.3844	279.018	270.5506	257.7949	244.8653	271.4894	276.2414	273.6829	252.9438	250.4076
K	1446.1877	1473.5941	1449.5525	1428.8272	1446.6438	1432.2571	1309.4486	1268.7032	1571.5961	1633.6172	1708.9285	1498.5405	1587.1438
M	2418.9435	2217.6356	2189.8923	1986.3062	2202.9263	2029.3367	2028.6316	1972.6409	2712.1154	2995.068	2894.4738	2213.8603	2250.4145
α_L	0.3013	0.3136	0.3188	0.3263	0.3091	0.299	0.2973	0.2896	0.2872	0.2915	0.2856	0.2869	0.2885
α_K	0.0381	0.0466	0.054	0.056	0.0527	0.0485	0.0516	0.0498	0.0494	0.0516	0.0492	0.0502	0.0542
α_M	0.6605	0.6398	0.6273	0.6177	0.6382	0.6525	0.6511	0.6605	0.6634	0.6569	0.6652	0.6629	0.6573
K/L	2.4121	2.6184	2.7138	3.0458	3.1268	3.3281	3.4755	3.2909	3.4211	3.478	3.7814	3.8561	4.135
Y/L	6.9577	7.1193	7.3017	7.6485	8.1394	8.8496	9.2526	9.2215	9.6277	9.6443	9.9993	10.194	10.0965
VA/L	2.3958	2.5919	2.7718	3.0392	3.3161	3.707	4.0812	3.9268	4.1792	4.2573	4.524	4.7226	4.6863

Note: Y is the quantity of output and L, K and M, the quantities of inputs. Y, K and M are expressed in millions of pesetas of 1990. L is expressed in thousands of hours worked.
α_L, α_K and α_M represent the participation (weights) of the cost of labour, capital and materials in the total cost of production.
K/L is the capital-labour ratio; Y/L is the product-labour ratio; and VA/L is the value added-labour ratio (in thousands of constant 1990 pesetas by worked hour).

Table 2.3. Average of the variables involved in the TFP index, small firms subsample

year	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
No of obs	507	704	812	762	731	699	694	833	847	831	822	769	678
Y	495.6177	448.6821	489.4915	564.7856	575.1999	668.5704	723.6461	708.1423	795.8902	766.9482	765.4694	756.5643	710.1372
L	74.1144	68.1981	71.8661	72.6514	71.6185	71.9693	71.7286	71.9418	76.889	75.5057	74.9947	73.9077	71.8516
K	170.1769	154.4851	184.7135	224.0782	247.0547	260.967	290.0653	270.9162	295.7583	287.8395	308.5072	296.9207	302.4855
M	325.966	283.3721	296.1663	342.2958	342.0782	401.752	410.7202	411.7007	440.9257	415.9877	402.252	383.9449	369.6003
α_L	0.3073	0.328	0.3353	0.3422	0.3286	0.3143	0.3121	0.3035	0.3031	0.3087	0.3043	0.3063	0.3096
α_K	0.0332	0.0388	0.0479	0.0511	0.0484	0.045	0.0488	0.0475	0.0477	0.0499	0.0483	0.0496	0.054
α_M	0.6595	0.6331	0.6168	0.6067	0.6231	0.6407	0.6392	0.6489	0.6492	0.6413	0.6473	0.6441	0.6363
K/L	1.7574	1.8608	1.9297	2.1694	2.3585	2.5595	2.7896	2.6433	2.7338	2.7887	3.1053	3.2113	3.436
Y/L	5.9054	6.1268	6.2405	6.5122	6.7664	7.5591	7.9016	7.993	8.2894	8.2263	8.4363	8.798	8.6364
VA/L	2.0312	2.2786	2.4377	2.6648	2.8442	3.2123	3.5356	3.449	3.7094	3.7836	3.9445	4.2759	4.2162

Note: see Table 2.2

Table 2.4. Average of the variables involved in the TFP index, large firms subsample

year	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
No of obs	292	380	357	290	265	235	216	227	229	218	234	217	186
Y	9037.9817	8773.5797	9788.6626	9708.4035	11287.981	11311.013	12852.166	13362.023	18000.14	20194.746	19127.635	14424.007	14992.661
L	942.2695	904.3338	918.7302	847.9864	851.1278	861.2244	855.6193	879.4257	991.2562	1041.4315	971.6388	887.4081	901.2731
K	3661.7269	3917.4171	4326.4413	4594.409	4755.699	4916.2222	4584.6892	4930.1861	6290.5243	6763.6234	6628.3571	5756.8152	6269.9306
M	6052.9832	5801.1132	6497.1908	6306.0855	7336.0584	6870.5356	7226.9211	7700.6722	11112.542	12826.333	11649.201	8698.6755	9106.2859
α_L	0.291	0.2869	0.2812	0.2845	0.2555	0.2533	0.2498	0.2386	0.2285	0.226	0.2199	0.2183	0.2114
α_K	0.0467	0.061	0.0678	0.0689	0.0646	0.0591	0.0606	0.0583	0.0558	0.0579	0.0521	0.0521	0.0547
α_M	0.6623	0.652	0.651	0.6466	0.6799	0.6876	0.6897	0.7031	0.7158	0.7161	0.728	0.7297	0.7339
K/L	3.5489	4.022	4.4971	5.3486	5.2461	5.6144	5.6793	5.6674	5.9633	6.1054	6.1564	6.1413	6.6831
Y/L	8.7848	8.9581	9.7154	10.6342	11.927	12.6879	13.5935	13.7295	14.5778	15.0497	15.49	15.1411	15.4186
VA/L	3.029	3.1725	3.5317	4.023	4.618	5.1785	5.8342	5.6799	5.9168	6.0631	6.5599	6.3059	6.4

Note: see Table 2.2

2.5.2. Evolution of firms' TFP over the period 1990-2002

Using the variables described in the previous Sections and the index suggested by Good et al. (1996) —expression 1.12—, we calculate a TFP index for a sample of Spanish manufacturing firms over the period 1990-2002. Table 2.5 shows some descriptive measures of TFP for all firms in the sample. Specifically, for each year in the period under analysis, Table 2.5 shows the average level of TFP, the standard deviation, the values for the most and less productive firms and those in selected percentiles of the distribution. Additionally, it provides information on the yearly growth rate observed in the index for an average firm in the sample.

The average TFP increases by more than 18% over the whole period, which corresponds to an average annual increase of 1.56%. However, most of the TFP growth corresponds to the evolution in the first half of the nineties, where growth was around 2.3% per year. There is a clear slowdown in productivity growth starting in the mid nineties, with a modest yearly average growth of 0.78%. These findings are in line with the evidence reported by other studies at the firm level for Spain, which have obtained similar TFP growth in that period (see for example, Huergo and Moreno, 2006). Different studies on aggregate productivity have reported a similar pace of growth (Myro, 2001; Goerlich et al., 2002; Pérez et al, 2006; Gual et al., 2006).

Another interesting feature has to do with the degree of dispersion in the firms' TFP distribution. By comparing the standard deviation with the average TFP on every year, we can state that there exists high degree of heterogeneity in the productivity levels across firms. Moreover, the standard deviation increases over time, reflecting that firms' heterogeneity increases as well. Actually, both the ratio between firms showing the minimum and maximum values of TFP on every year, and the distance between the TFP levels at percentiles 10 and 90, and 25 and 75, increase over the period. The analysis of the TFP figures at different percentiles shows that productivity increased at any initial TFP level, according with the results discussed above for the mean of the distribution. The increase associated to the percentile 10 is around 10%, while at percentile 90, the increase is around 30%. Similar conclusions are obtained when comparing changes in TFP levels at other extreme points of the distribution. Thus, the general picture is that TFP growth was not homogeneous over the whole distribution of firms, being more intense for the already most productive firms.

Table 2.5. Evolution of the TFP index (1990-2002)

Year	No of obs	Mean	Std dev	Minimum	Maximum	Growth rate	10%	25%	50%	75%	90%
1990	799	-0.0682	0.1859	-0.69	0.6317		-0.2968	-0.1907	-0.0726	0.0418	0.1681
1991	1084	-0.049	0.2058	-1.2269	0.9201	0.0192	-0.2987	-0.1751	-0.0506	0.0698	0.1915
1992	1169	-0.0338	0.218	-1.4085	0.8033	0.0152	-0.2968	-0.1622	-0.0318	0.0961	0.2218
1993	1052	-0.013	0.2359	-1.1763	1.0418	0.0208	-0.2826	-0.1568	-0.0151	0.1339	0.2517
1994	996	0.0149	0.2294	-0.9287	0.8966	0.0279	-0.2671	-0.1167	0.0106	0.1556	0.2943
1995	934	0.0429	0.2439	-1.3614	1.3002	0.028	-0.2451	-0.1089	0.0398	0.1957	0.3406
1996	910	0.0714	0.2528	-1.1243	1.266	0.0285	-0.2423	-0.0881	0.0653	0.2332	0.3755
1997	1060	0.0598	0.247	-1.1401	1.231	-0.0116	-0.2279	-0.0862	0.0477	0.2077	0.359
1998	1076	0.0747	0.2559	-1.1804	1.1318	0.0149	-0.2355	-0.0692	0.0696	0.2283	0.3887
1999	1049	0.082	0.2672	-1.1255	1.4117	0.0073	-0.2315	-0.0791	0.0756	0.2302	0.4029
2000	1056	0.0931	0.2701	-1.1024	1.4261	0.0111	-0.2183	-0.0611	0.0872	0.2391	0.4158
2001	986	0.1223	0.2629	-0.8044	1.4838	0.0292	-0.1898	-0.0397	0.1089	0.273	0.4364
2002	864	0.1184	0.2676	-0.8573	1.4501	-0.0039	-0.1924	-0.0366	0.1034	0.272	0.4648

Evidence on heterogeneity in the size TFP-gap over the distribution suggests the necessity to analyse TFP considering the entire distribution, instead of just some synthetic measures such as the average. In so doing, firstly we estimate non-parametrically the density function associated to the TFP firms' distribution for every year.[46] To save space, only the density functions for 1990, 1994, 1998 and 2002 are depicted in Figure 2.1. It is clearly observed that the distribution of TFP shifts to the right, which is interpreted as a generalized increase of firms' TFP levels over time. However, the shift is not neutral, as there is an increasing larger proportion of firms in the range of high and low TFP levels. The comparison of densities for each year also confirms the different pattern of evolution of productivity in the first and in the second half of the period under analysis, as well as the fact that the evolution is not homogeneous along the distribution. For instance, the formation of a mass of probability on the right tail (high TFP levels) is very active in the second half of the nineties, while only minor changes are observed in the middle part of the distribution in that subperiod.

[46] A brief description of the method used to estimate density functions can be found in the Appendix 2.2.

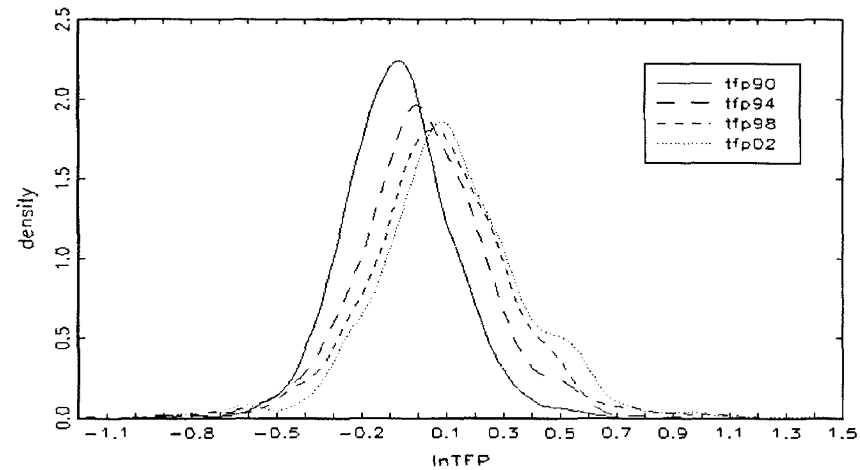


Figure 2.1. Estimated density functions for TFP in the total sample of Spanish manufacturing firms

The density function summarizes the external shape of the distribution of TFP, and the comparison of densities for different years allows assessing changes in the external shape. But this analysis says nothing about distribution dynamics, that is, how the different types of firms evolve within the distribution. This is relevant, as the same density for two years may be consistent with different patterns of firms' dynamics. For instance, a pattern of perfect persistency (firms with high TFP at the beginning of the period are the ones with high TFP at the end; and the same for low-TFP firms) or a pattern with a lot of churning (high TFP firms at the beginning of the period show low productivity at the end; and the other way round). To analyse the probability of transition of firms from one TFP level in a given period to any other level in the next period, we estimate a stochastic kernel (Stokey and Lucas, 1989).[47] The stochastic kernel permits evaluating the characteristics of the dynamics within the entire distribution. Our analysis is inspired in the analysis of the dynamics of the distribution of personal income, which has been used in recent analyses on income convergence at macroeconomic level (see for instance Quah, 1996a, and Fingleton and López-Bazo, 2003).

Figure 2.2 shows the estimation of the stochastic kernel from the dynamics observed within the firms' TFP distribution in two consecutive years. The two

[47] The stochastic kernel has been estimated non-parametrically by the kernel method. Further details can be found in Johnson (2000).

horizontal axes in the three-dimensional plot refer to TFP levels in two consecutive periods, and the vertical axis measures the probability of transition (joint density for any pair of TFP values in two consecutive years, conditional on the density of the TFP level at the initial year). The mass of probability following the positive diagonal indicates strong persistence (low mobility) of firms in their productivity levels. When this mass of probability twists, it indicates mobility of firms. If it twists until becoming parallel to axis *t*, it indicates that firms tend to converge to a common TFP level: all firms will achieve a similar TFP level, regardless of their initial level. To facilitate the interpretation, Figure 2.2 also includes a bidimensional graph, which is a contour plot of the three-dimensional graph.

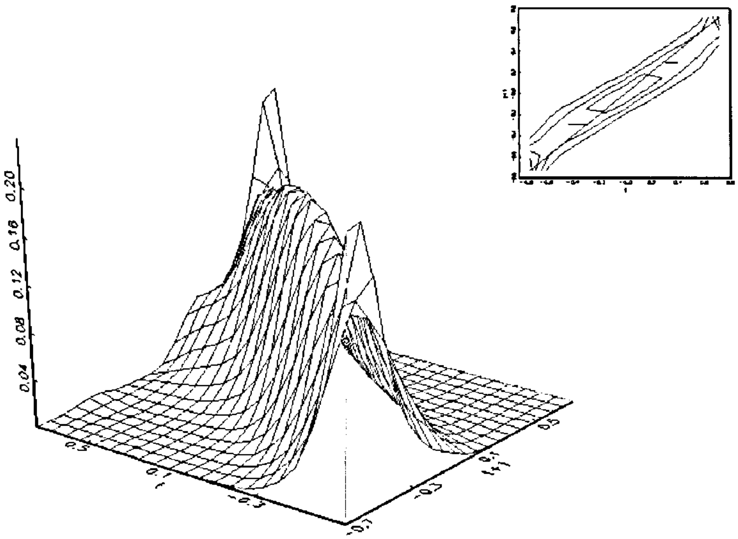


Figure 2.2. Estimated stochastic kernel for TFP in the total sample of Spanish manufacturing firms

The most relevant aspect of the estimated stochastic kernel in Figure 2.2 is that the mass of probability follows the positive diagonal, indicating a high degree of persistence of firms in their initial productivity levels. This implies that firms with low (high) TFP levels in any given year have a high probability of achieving low (high) TFP in the following year.[48] The most relevant movements in the distribution seem to be related with the improvement of firms that start below, but not far, from the average,

[48] The conclusion on the degree of persistence is robust to the consideration of longer time spans (2 and 3 years). However, it should be kept in mind that the longer the time span being considered the lower the number of firms in the sample from which data is available.

and with the formation of the group of highly productive firms already detected when analysing the evolution of the external shape of the distribution.

To summarize, TFP has increased between 1990 and 2002 in the Spanish manufacturing firms. Nevertheless the increases are smaller during the second half of the nineties, indicating that the productivity growth has slowed down. Increases over the period of analysis are larger for those firms with higher TFP levels, leading to larger dispersion in the TFP distribution at the end of the period. The amount of dispersion in TFP levels and its increase over time suggests that firms are heterogeneous. A strand of literature has focused in explaining the different sources of heterogeneity in firms' productivity. Concretely, firm size appears to be a main source of heterogeneity in TFP among firms with similar characteristics. The reduced size of Spanish firms appears as a relevant characteristic of our economy in relation to other advanced economies.[49] In what follows, we present a descriptive of TFP by firm size to analyse whether firms of different sizes present different TFP levels, and thus if size can be explaining part of the TFP heterogeneity, a phenomenon that seems to become more important in the last years in the case of Spain.

2.5.3. Descriptive of TFP by firm size

In this Section we provide a descriptive analysis of TFP by firm size. In so doing, we consider the groups of small (10 to 200 employees) and large firms (more than 200 employees) as defined in Section 2.4. Results confirm that there are sharp differences in TFP levels for small and large firms and these differences are not homogeneous over the distribution and evolve over time.

Tables 2.6 and 2.7 report the same descriptive measures as in Table 2.5 but for the two subsamples of firms. In addition, these tables show the average firm size in each subsample. In the group of small firms the average number of employees is around 40 and it remains quite stable over the period, while for large firms it is around 500 employees.

[49] According with the data from the Observatory of European SMEs, Italy, Spain and Portugal are, respectively, the countries with a smaller percentage of large firms in the EU-15 (for further details, see http://www.eim.nl/Observatory_7_and_8/en/stats/2001/var2/1cou_size.html, last time visited on 1st January 2007).

As for the average TFP, figures clearly confirm that productivity in large firms is higher than in small ones, with differences being statistically significant on every year (the t-test of equality of means in the last column of Table 2.7 rejects the null hypothesis that small and large firms have the same average TFP). However, differences in TFP between small and large firms tend to reduce over time and the gap is narrower at the end of the period under analysis. This is caused by a higher pace of productivity growth of small firms since the mid nineties. The general evolution of TFP for small and large firms is similar to the one described in the previous section for the total sample: it increases over time although there is a slow down during the second half of the nineties. In contrast with the first half of the nineties, in which growth rates in small and large firms were quite similar (yearly average of 2.44% and 2.66% respectively), since the mid nineties small firms become more dynamic (an annual TFP growth rate of 0.9% versus a 0.4% in large firms). Thus the slow down in productivity growth was much more severe in the case of large firms.

Differences in TFP levels between small and large firms are not only observed in the mean, but also in other points of the distribution. The level of TFP corresponding to selected percentiles of the distribution for small and large firms is reported in the last set of columns in Tables 2.6 and 2.7. Differences in TFP are more severe for firms at the lowest part of the distribution and they diminish as we move up to higher TFP levels: the TFP gap is around 10% at percentile 10, while there seem to be no significant differences at the upper part of the distributions. Actually, a closer look at the evolution of TFP values at the different percentiles reveal non-homogeneous trends in the size gap over the whole distribution. This is confirmed by the inspection of the estimated density functions for the TFP levels in large and small firms in Figure 2.3.

Table 2.6. Evolution of the TFP index. Subsample of small firms (1990-2002).

year	No of obs	Avg size (workers)	Mean	Std dev	Min	Max	Growth rate TFP	10%	25%	50%	75%	90%
1990	507	41.1	-0.097	0.1884	-0.69	0.6317		-0.3252	-0.2223	-0.1022	0.017	0.1519
1991	704	37.9	-0.0679	0.2152	-1.2269	0.9201	0.0291	-0.3284	-0.1933	-0.0798	0.06	0.1902
1992	812	40.42	-0.0529	0.226	-1.4085	0.8033	0.015	-0.3135	-0.187	-0.0641	0.0828	0.2158
1993	762	41.27	-0.0355	0.247	-1.1763	1.0418	0.0174	-0.2934	-0.1764	-0.0498	0.1212	0.248
1994	731	40.08	-0.0095	0.2366	-0.9287	0.8966	0.026	-0.303	-0.1517	-0.0103	0.1336	0.2772
1995	699	40.19	0.0193	0.2515	-1.3614	1.3002	0.0288	-0.2718	-0.1405	0.0062	0.1755	0.3257
1996	694	39.96	0.0496	0.2578	-1.1243	1.266	0.0303	-0.2677	-0.1105	0.0446	0.2109	0.3694
1997	833	40.01	0.0401	0.2468	-1.1401	1.231	-0.0095	-0.2416	-0.1044	0.0275	0.1817	0.3403
1998	847	42.78	0.0594	0.2603	-1.1804	1.1318	0.0193	-0.2467	-0.091	0.0542	0.2046	0.3679
1999	831	42.27	0.0684	0.2729	-1.1255	1.4117	0.009	-0.2516	-0.0965	0.0548	0.2101	0.4012
2000	822	42.08	0.0739	0.2771	-1.1024	1.4261	0.0055	-0.2432	-0.0892	0.066	0.2164	0.4059
2001	769	41.52	0.1099	0.2719	-0.8044	1.4838	0.036	-0.2059	-0.0597	0.0969	0.2545	0.4429
2002	678	40.6	0.1054	0.2762	-0.8573	1.4501	-0.0045	-0.2183	-0.0508	0.0954	0.2515	0.459

Table 2.7. Evolution of the TFP index. Subsample of large firms (1990-2002).

year	No of obs	Avg size (workers)	Mean	Std dev	Min	Max	Growth rate TFP	10%	25%	50%	75%	90%	Test Eq means
1990	292	545.38	-0.0182	0.1705	-0.5493	0.5552		-0.2117	-0.1212	-0.0192	0.0756	0.1814	6.0481***
1991	380	522.53	-0.0139	0.1823	-0.6084	0.7276	0.0043	-0.2476	-0.1129	-0.0066	0.0889	0.1905	4.3635***
1992	357	534.68	0.0095	0.1919	-0.6271	0.615	0.0234	-0.2191	-0.0866	0.0092	0.1245	0.2255	4.8429***
1993	290	499.71	0.0461	0.1919	-0.5829	0.6009	0.0366	-0.1895	-0.063	0.0528	0.1592	0.2582	5.674***
1994	265	491.53	0.0823	0.1931	-0.6042	0.6966	0.0362	-0.1458	-0.0369	0.075	0.2011	0.3012	6.2288***
1995	235	486.22	0.113	0.2044	-0.3615	0.6891	0.0307	-0.1318	-0.0186	0.0943	0.2357	0.3823	5.7145***
1996	216	484.22	0.1415	0.2224	-0.5234	1.071	0.0285	-0.1413	0.0002	0.1309	0.2749	0.3958	5.1007***
1997	227	497.46	0.1322	0.2345	-0.54	1.1308	-0.0093	-0.1373	0.0028	0.1116	0.2518	0.406	5.1824***
1998	229	561.66	0.1314	0.2309	-0.7116	1.0397	-0.0008	-0.1374	-0.017	0.1306	0.269	0.4018	4.0736***
1999	218	591.74	0.1339	0.2376	-0.6843	1.0414	0.0025	-0.1415	-0.0062	0.1199	0.2745	0.4081	3.5088***
2000	234	551.81	0.1605	0.2321	-0.4633	1.168	0.0266	-0.1177	0.0124	0.132	0.2965	0.4257	4.8146***
2001	217	503.1	0.1661	0.2234	-0.4161	0.8096	0.0056	-0.109	0.0112	0.1472	0.3159	0.4309	3.1118***
2002	186	509.85	0.1657	0.2281	-0.3073	0.8682	-0.0004	-0.1282	-0.0064	0.1375	0.3397	0.4679	3.0423***

Note: the test of equality of means compares the values in Tables 2.6 and 2.7; (***) denotes significant at 1%.

At the beginning of the period, the TFP distribution of large firms was clearly at the right of the distribution of small firms. Actually, a (first order) stochastic dominance test clearly indicates that the distribution of large firms stochastically dominates that of small firms in 1990 (see results in Table 2.8).[50] However, the visual inspection of the density functions that year reveals that differences in the distribution of small and large firms were more pronounced in the range of TFP values below the average. For the higher TFP levels (right tail), both distributions are quite similar.

As time goes by, the two densities seem to be more alike, indicating a reduction in the gap in whole distribution, particularly in the second half of the period. This is confirmed by the decrease in the values of the two-sided statistic of the stochastic dominance test. This causes a concentration of a a large mass of probability around a similar mode in both distributions in 2002. As time goes by, we also observe an increasing mass of probability in the left tail in the case of small firms, denoting the existence of a greater number of small firms with TFP levels well below the average. Correspondingly, a larger mass of probability in values above the average appears in the distribution of large firms.

In conclusion, it seems clear that large firms are in general more productive; however a more careful analysis shows a heterogeneous behaviour along the productivity distribution: the most productive small firms are as productive as the most productive large firms. The analysis in the mean shows that the size TFP-gap diminishes over time, however the analysis in the distribution shows that this reduction is more important in the central and in the very top values. Finally, it should be noted that dispersion in both distributions increases over time, which can be read as an indication of boosting firms' heterogeneity in the level of TFP.

[50] A description of the test of stochastic dominance is provided in Appendix 2.3.

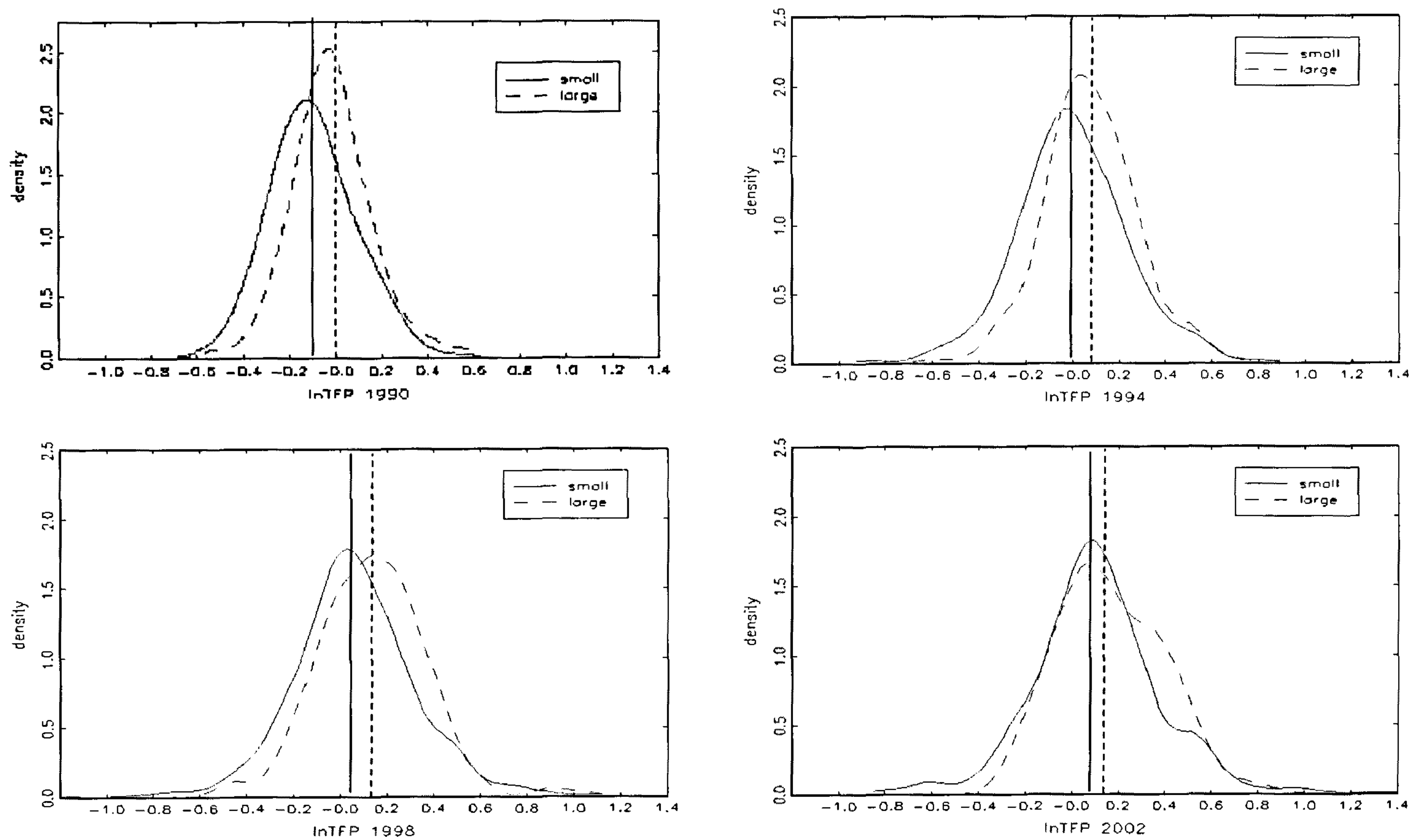


Figure 2.3. Estimated density functions for TFP in small and large firms

Table 2.8. Tests of stochastic dominance in TFP distributions of large and small firms

	Two-sided test		**One-sided test**	
	KS Statistic	P-value	KS Statistic	P-value
1990	44.694***	0	0.0029	0.9986
1994	23.811***	0	0.084	0.9589
1998	20.266***	0	0.1891	0.9098
2002	7.7784**	0.0205	0.2	0.9048

Note: KS denotes Kolmogorov-Smirnov. See the description of the test in Appendix 2.3.

2.5.4. Descriptive of TFP by innovative activity and human capital

As mentioned in the introduction of Part I, the innovative activity and human capital are considered two of the main determinants of productivity: firms that make a more intense innovative effort and firms that obtain more product and process innovations are expected to be more productive. Also, those firms that employ a more qualified labour force or that provide more training are expected to reach higher TFP levels. A main difference between these two sources of economic growth in the Spanish case is that firms' human capital has improved more than their technological level in the last decades.[51] In this section, we provide some descriptive evidence on the connection between innovation and human capital and the level of TFP in the Spanish manufacturing firms, stressing the differences observed for large and small firms. According with the definitions in Section 2.4,[52] the variables on the innovative activity and proportion or skilled workers are lagged one period. So the following Tables show results for TFP in 1994, 1998 and 2002 in relation with these variables in 1993, 1997 and 2001. Results for 1990 are not offered because data for innovative activity and skilled workers were not available in 1989.

[51] De la Fuente et al. (2003) show that the average years of schooling of the Spanish population above 25 years old have by over 60% during the period 1960-2000. López-Bazo and Moreno (2007) show evidence of a continuous increase of educational human capital over the period 1964-2000: concretely, the average years of schooling of the employees in the private productive sector has increased from around 4 to 10 years. Gual et al. (2006) show that Spain has increased the stock of human capital since 1992 and that, in 2003, it takes values around 80% the average of the EU-15. The stock of technological capital has improved substantially, however it departed from very low levels in relation with the average EU-15 and it is still far from it: in 2003, it is around 50% of it.

[52] As we argue in Section 2.4, there are economic reasons to believe that the impact of these variables on productivity takes place after some time. Moreover, due to econometric requirements explained in Section 3.3, the lagged specification is preferred over the contemporaneous one.

TFP by innovative activity

Tables 2.9 to 2.11 show the differences in TFP between innovative and non-innovative firms according with the definitions in Section 2.4: a firm is considered to be innovative if it has introduced at least one product/process innovation or has made some expenses in R&D.

Table 2.9 reports the results for the first measure of innovative activity: process innovations. The upper part of this Table shows that around one third of the firms in our sample obtain new processes and that this proportion does not seem to increases over time. According with the theoretical arguments and previous empirical evidence, firms that obtain process innovations are more productive. Actually, the t-tests of equality of means strongly reject the null that innovating and non-innovating firms have equal TFP levels, indicating that innovative firms are significantly more productive. The median TFP of innovative firms is also higher than for non-innovative ones and the dispersion in TFP is quite important. Moreover, the productivity levels increase as time goes by.

Notice that there are differences in the innovative activity by firm size: around half of large firms obtain process innovations, while only one quarter of small firms do. This result is consistent with the general finding that large firms are more innovative.[53] Thus, it is possible that the higher innovative propensity in large firms is reflected in higher productivity levels for these firms. We start by comparing TFP in small and large innovative firms. Large innovative firms are significantly more productive in average than their smaller counterparts. The differences in TFP between small and large innovative firms are statistically significant over the whole period, but they seem to decrease over time (in 2002, only at 10%).[54] We are also interested in analysing if a similar effect occurs between small and large non-innovative firms. As for the group of non-innovative firms, large firms are also significantly more productive than small ones. In addition, the differences in TFP

[53] Bucsa and Molero (2001, pp 141) find similar results. They comment that the industrial sector is the most innovative and that the probability of innovating is much higher in large firms. Huergo and Jaumandreu (2004b) find that process innovations are strongly associated with firm size.

[54] Notice that although large firms are more innovative, there are more small than large firms in the innovative group. However, this result is explained by the fact that small firms have a very important participation in the Spanish industry and so, in the sample. Actually, there is also a much larger participation of small firms in the group of non-innovative firms for the same reason.

associated with size are more important in the group of non-innovative firms than in the group of innovative ones. In this view, innovation seems to mitigate the differences in TFP between small and large firms. Or in other words, performance of small innovative firms is more similar to their larger counterparts, while the differences are more severe between small and large firms that do not innovate.

We are also interested in investigating the effect of process innovations on TFP controlling by firm size: so we compare small firms that obtain new processes and those that do not. Table 2.9 shows that the average TFP of small innovative firms is much higher than for small non-innovative ones and that the differences are significant all over the period. Thus, small firms obtaining a process innovation are strongly associated with higher TFP levels. However, in the case of large firms, the TFP gains derived from obtaining a process innovation are more modest than in the case of small firms. Actually, the differences in TFP between large innovative and non-innovative firms are not significant, except for 1998. Since the gains in productivity associated with process innovation are more important in small than in large firms, obtaining process innovations may be the key for small firms to increase productivity and become more competitive.

Table 2.9. TFP, process innovations and size

					Innovative			Non-innovative					
Year	#obs	#Innovative	#Small Inn	#Large Inn	Mean	Std dev	Median	Mean	Std dev	Median	Eq mean Total ($)	Eq mean Small ($)	Eq mean Large ($)
1994	852	35.45%	28.34%	54.55%	0.0624	0.2218	0.0601	-0.003	0.2345	-0.011	4.01***	2.99***	0.8593
1998	968	34.40%	29.92%	50.97%	0.1326	0.2369	0.1119	0.0446	0.2622	0.0329	5.75***	4.30***	1.82**
2002	864	30.32%	25.36%	48.39%	0.1503	0.2301	0.1282	0.1045	0.2814	0.094	2.74***	1.93**	0.63
					Small innovative			Large innovative					
Year	#obs		#Small	#Large	Mean	Std dev	Median	Mean	Std dev	Median	Eq mean (&)		
1994	302		58.28%	41.72%	0.0438	0.2339	0.0275	0.0884	0.2018	0.0688	1.77**		
1998	333		68.47%	31.53%	0.1184	0.2443	0.1	0.1637	0.218	0.1514	1.81**		
2002	262		65.65%	34.35%	0.1364	0.2256	0.1198	0.1768	0.2375	0.1339	1.45*		
					Small non-innovative			Large non-innovative					
Year	#obs		#Small	#Large	Mean	Std dev	Median	Mean	Std dev	Median	Eq mean (&)		
1994	550		80.91%	19.09%	-0.0189	0.2409	-0.0215	0.0661	0.1917	0.0631	3.88***		
1998	635		84.09%	15.91%	0.0331	0.2649	0.0294	0.1056	0.2395	0.1199	2.74***		
2002	602		84.05%	15.95%	0.0949	0.2909	0.0852	0.1553	0.2198	0.1493	2.33***		

Note: test of equality of mean: (***), (**) and (*) denotes significant at 1%, 5% and 10%. ($) compares TFP in innovative and non-innovative firms; (&) compares TFP in small and large firms. Results correspond to TFP levels in 1994, 1998 and 2002 in relation with process innovations in 1993, 1997 and 2001.

Table 2.10 reports the results for a similar analysis using another measure of innovative output: product innovations. The upper part of this Table shows that less than one quarter of firms obtains product innovations and that this proportion decreases over time. Compared with Table 2.9, there are more firms obtaining process than product innovations and the same result is obtained for small and large firms. However, there are clear differences in innovative activity, measured as product innovations, between them. Around 22% of small firms obtain product innovations in 1993, while almost 40% of large firms do. These percentages decrease over time until 14% and 35% respectively in 2001.

Table 2.10. TFP, product innovations and size

					Innovative			**Non-innovative**					
Year	#obs	#Innovative	#Small Inn	#Large Inn	Mean	Std dev	Median	Mean	Std dev	Median	Eq mean Total ($)	Eq mean Small ($)	Eq mean Large ($)
1994	852	26.64%	22.22%	38.53%	0.0269	0.2152	0.0047	0.0181	0.238	0.0165	0.51	0.21	0.15
1998	968	23.86%	20.34%	36.90%	0.092	0.2576	0.068	0.0695	0.2569	0.0662	1.16	0.19	0.83
2002	864	19.10%	14.60%	35.48%	0.1195	0.2222	0.1018	0.1181	0.2774	0.1036	0.07	0.97	0.25
					Small innovative			**Large innovative**					
Year	#obs		#Small	#Large	Mean	Std dev	Median	Mean	Std dev	Median	Eq mean (&)		
1994	227		60.79%	39.21%	-0.0047	0.2322	-0.0166	0.0758	0.176	0.0387	2.96***		
1998	231		67.10%	32.90%	0.0621	0.2626	0.0447	0.1529	0.2372	0.1251	2.64***		
2002	165		60%	40%	0.0849	0.2149	0.083	0.1713	0.2245	0.1304	2.46***		
					Small non-innovative			**Large non-innovative**					
Year	#obs		#Small	#Large	Mean	Std dev	Median	Mean	Std dev	Median	Eq mean (&)		
1994	625		77.28%	22.72%	-0.0001	0.2429	-0.007	0.0798	0.2099	0.0802	3.84***		
1998	737		82.36%	17.64%	0.0577	0.2617	0.0539	0.1248	0.226	0.1325	2.99***		
2002	699		82.83%	17.17%	0.1089	0.2854	0.0982	0.1626	0.231	0.1493	2.22***		

Note: test of equality of mean: (***), (**) and (*) denotes significant at 1%, 5% and 10%. ($) Compares TFP in innovative and non-innovative firms. (&) Compares TFP in small and large firms. Results correspond to TFP levels in 1994, 1998 and 2002 in relation with product innovations in 1993, 1997 and 2001.

An important result is that firms that obtain product innovations show TFP levels similar to non-innovative firms. Actually, the differences in TFP are not significantly different from zero for the years considered in this analysis. As before, we are interested in analysing whether the differences in the innovative activity by firm size are reflected in TFP differences. So, we analyse differences in TFP between small and large firms controlling for the innovative activity. First, we consider only those firms that obtain new products and we find that TFP is higher in large firms and that the differences are significant at 1%. Second, we consider only the firms that do not obtain new products, and we find a similar result. Thus, TFP appears to be higher in large firms independently of their innovative activity, measured as product innovations. This result seems to be confirmed by the fact that small (large) innovative firms are not significantly more productive than small (large) non-innovative firms.

All in all, the firms that obtain product innovations do not seem to achieve higher productivity levels regardless of their size. This is in sharp contrast with the higher productivity levels observed in the case of firms innovating in processes. Thus, at least for the sample of Spanish manufacturing firms, productivity seems to be associated to process rather than to product innovations.

As for the input measure of innovative activity, Table 2.11 offers the results for a similar analysis using the expenditure in R&D per worker, as defined in Section 2.4. It shows that around one third of the Spanish manufacturing firms in our sample spend money on R&D, although this proportion tends to decrease over time.[55] Firms that spend money in R&D are significantly more productive than those that do not do any expenditure. Although productivity increases over time for both groups, the TFP differences become less important over time.

As for the differences in innovative effort by firm size, around 70% of large firms and 22% of small do a positive expenditure in R&D in 1993. These percentages decrease over time until 68% and 19% respectively in 2001. The average expenditure in R&D for the total sample is around 600 euros per worker. The amount is much higher in the case of large firms (almost 1200 euros) than for small firms (400 euros). However, when only the firms that do a positive expenditure are considered, the average is slightly higher for small firms (2000 euros in the case of small firms and 1700 euros in the case of large firms).[56]

[55] Buesa and Molero (2001, pp 140) comment that firms play a key role in the innovative system but they have performed worse than other agents that participate in R&D activities. The authors also highlight the wide gap between Spain and other economies in terms of R&D expenditure. Actually, between 1990 and 2000, the R&D expenditure suffered a decrease with a later recuperation, although in general terms Spain still needs to progress considerably during several decades to achieve an innovative capacity similar to other advanced economies.

[56] However, it should be mentioned that further analysis shows that, among those firms that spend money on R&D, 59% obtained process innovations in 1993 and decreasing until 53% in 2001. By firm size, the percentage of large firms that obtain process innovations is higher than that of small firms: around 60% of

Table 2.11. TFP, R&D expenditure and size

					Innovative effort			No innovative effort					
Year	#obs	#Innovative	#Small Inn	#Large Inn	Mean	Std dev	Median	Mean	Std dev	Median	Eq mean Total ($)	Eq mean Small ($)	Eq mean Large ($)
1994	852	35.21%	22.22%	70.13%	0.0736	0.2141	0.0713	-0.009	0.2365	-0.012	5.15***	3.14***	1.23
1998	968	29.86%	18.63%	71.36%	0.1333	0.2386	0.1216	0.05	0.2608	0.0442	4.84***	3.16***	1.10
2002	864	29.17%	18.58%	67.74%	0.1651	0.2194	0.1517	0.0992	0.2831	0.0902	3.67***	2.5***	1.19
					Small innovative			Large innovative					
Year	#obs		#Small	#Large	Mean	Std dev	Median	Mean	Std dev	Median	Eq mean (&)		
1994	300		46%	54%	0.0553	0.2393	0.0362	0.0891	0.1895	0.0877	1.34*		
1998	289		49.13%	50.87%	0.1194	0.2527	0.107	0.1468	0.2241	0.134	0.974		
2002	252		50%	50%	0.1508	0.2087	0.1452	0.1794	0.2295	0.1639	1.0348		
					Small non-innovative			Large non-innovative					
Year	#obs		#Small	#Large	Mean	Std dev	Median	Mean	Std dev	Median	Eq mean (&)		
1994	552		87.50%	12.50%	-0.0172	0.2385	-0.0159	0.0527	0.2134	0.0215	2.5***		
1998	679		91.31%	8.69%	0.0447	0.2619	0.039	0.1062	0.2439	0.128	1.84**		
2002	612		90.20%	9.80%	0.0951	0.2886	0.0867	0.137	0.2244	0.124	1.33*		

Note: test of equality of mean: (***), (**) and (*) denotes significant at 1%, 5% and 10%. ($) Compares TFP in innovative and non-innovative firms. (&) Compares TFP in small and large firms. Results correspond to TFP levels in 1994, 1998 and 2002 in relation with R&D expenditure in 1993, 1997 and 2001.

As with the previous measures of innovation, we compare whether there are differences in TFP by firm size after controlling for their expenditure in R&D. Considering only those firms that do a positive expenditure in R&D, we find that TFP is higher in large firms although we cannot reject the null hypothesis that the differences between the two groups are zero. Considering only the non-innovative firms, we find significant differences in TFP by firm size although these differences seem to decrease over time.

Next, we analyse whether there are differences in productivity between firms of the same size class that do a positive expenditure and those that do not. First, we find that small innovative firms are significantly more productive than small non-innovative ones. Second, we find that large innovative firms are as productive as large non-innovative firms. Thus, spending money on R&D does not seem to be associated with higher productivity in the

large firms obtained process innovations over the period, while 56% of small firms did in 1993 and decreasing until 45% in 2001. As for product innovations, 48% of firms investing in R&D obtained process innovations in 1993 and decreasing until 44% in 2001. By firm size, in 1993 the percentage of small firms that obtain product innovations (50%) is higher than that of large firms (46%). However, in 1997 and 2001, the percentage of product innovators remains stable in the case of large firms and it decreases in the case of small firms (until 43%). Thus, among firms with a positive R&D expenditure, small firms spend more than large firms per worker. However in terms of innovative output, large firms are more likely to obtain at least one new process or product (except for product innovations in 1993). This interesting result can be interpreted as a reflection of small firms having more difficulties in obtaining innovations out of their innovative effort.

case of large firms, while in the case of small ones, it seems to be a key element fostering their productivity.

To summarize, we obtain that the percentage of large innovative firms is superior to the percentage of small ones, according with our three definitions of innovative activity. Firms that obtain new processes are more productive, while product innovations do not seem to be associated with higher TFP levels. Moreover, process innovations seem to be a key element for small firms to achieve higher TFP levels. Thus, obtaining new processes appears to determine higher TFP levels and it seems to contribute to explain the differences in TFP between small and large firms: not only through a direct effect on productivity, but also through an indirect effect associated with firm size. Similar results are obtained when we use the expenditure in R&D measure of innovative input: spending money on R&D seems to be a key element fostering productivity in the case of small firms.

These results suggest that a further analysis is required to identify the channels through which innovation determines firms' productivity, with special attention to firm size. In addition, results for the descriptive analysis support the reasoning in Section 2.4, in the sense of using the variable on process innovations to define the firms' innovative activity in Part II.

TFP by human capital

Tables 2.12 and 2.13 offer a descriptive of TFP in relation with the intensity of use of human capital for the sample of Spanish manufacturing firms. Two different components of human capital are considered in this analysis: the formal education of employees, which has been defined in Section 2.4 as the proportion of white collars, and the continuous training, defined as expenditure on training per worker.

Table 2.12 shows the differences in TFP between firms that have a high proportion of skilled workers (above the median) or low (below the median).[57] The average percentage of qualified workers for the total sample is around 8% in 1993 and increasing over time until 10% in 2001. For small firms, the percentage increases from 7% to 9%, and for large firms from 10% to 12%. This result is in line with the general finding that large firms

[57] The median is specific of each period and common for the group of small and large firms.

employ more qualified employees. As expected, firms with a high proportion of qualified employees are significantly more productive in average. The t-tests of equality of means reject the null that firms with a proportion of white collars above and below the median are equally productive.

Table 2.12. TFP, workers' qualification and size

					High % of qualified			Low % of qualified					
Year	#obs	% of High Qualified	% of High Qualified (Small)	% of High Qualified (Large)	Mean	Std dev	Median	Mean	Std dev	Median	Eq mean Total ($)	Eq mean Small ($)	Eq mean Large ($)
1994	852	7.99%	7.04%	10.53%	0.0701	0.2216	0.0638	-0.03	0.2318	-0.024	6.44***	4.74***	3.55***
1998	968	9.08%	8.19%	12.36%	0.1338	0.2346	0.1132	0.016	0.2652	0.0034	7.32***	6.10***	3.03***
2002	864	9.74%	8.96%	12.58%	0.1559	0.2394	0.132	0.0807	0.2886	0.0701	4.17***	3.48***	1.70**
					Small - high % of qualified			Large - high % of qualified					
Year	#obs		#Small	#Large	Mean	Std dev	Median	Mean	Std dev	Median	Eq mean (&)		
1994	429		65.27%	34.73%	0.0483	0.2315	0.0309	0.1111	0.1961	0.108	2.96***		
1998	484		72.31%	27.69%	0.1194	0.2407	0.0973	0.1714	0.2139	0.1614	2.31***		
2002	433		72.29%	27.71%	0.1444	0.2408	0.1208	0.1861	0.2341	0.1747	1.64**		
					Small - low % of qualified			Large - low % of qualified					
Year	#obs		#Small	#Large	Mean	Std dev	Median	Mean	Std dev	Median	Eq mean (&)		
1994	423		80.61%	19.38%	-0.0417	0.2403	-0.0411	0.0186	0.1858	0.0026	2.48***		
1998	484		85.12%	14.88%	0.0069	0.2679	-0.0088	0.0677	0.2449	0.0334	1.92**		
2002	431		84.69%	15.31%	0.072	0.2996	0.0628	0.1285	0.2135	0.1152	1.84**		

Note: test of equality of mean: (***), (**) and (*) denotes significant at 1%, 5% and 10%. ($) Compares TFP in firms with a ratio of qualified workers above and below the median. (&) Compares TFP in small and large firms. Results correspond to TFP levels in 1994, 1998 and 2002 in relation with the percentage of qualified workers in 1993, 1997 and 2001.

But, is the higher participation of white collars in large firms translated into higher productivity? Among firms that have a high proportion of qualified workers,[58] large firms are significantly more productive than their smaller counterparts. However, the differences in TFP tend to decrease over time. Considering only the group of firms with a low proportion of qualified workers, large firms are also significantly more productive. The differences in TFP associated with size are quite similar in magnitude for the group of firms that employ a high and low proportion of white collars. In contrast to what we obtained for process innovations, incorporating more human capital does not seem to mitigate the

[58] There is a higher proportion of small firms both in the group of firms with a high percentage of white collars and in the group of firms with a low percentage of white collars. As mentioned before, it is due to the higher participation of small firms in the industry and thus, in the sample.

differences in TFP between small and large firms. The fact that after controlling for human capital, large firms are still significantly more productive, suggests that they might be obtaining higher returns from their investment in human capital than small firms. In any case, small and large firms that employ a high proportion of white collars are significantly more productive than those firms in the same size group employing a low proportion of qualified workers.

All in all, making a more intense use of a qualified labour force seems to be strongly associated with higher productivity levels. Moreover, the differences in TFP between small and large firms (among the groups that use a high and low proportion of qualified workers) seem to indicate that these two groups obtain different returns from their investment in human capital. This relevant result will be further explored in the analysis in Part II.

The second component of human capital considered in the thesis is the firm-provided continuous training. Table 2.13 shows the differences in TFP between firms that provide continuous training and those that do not for years 2001 and 2002, as well as the percentage of firms providing training. First of all, we observe that 30% of firms in the sample provide training. As the previous evidence in the literature points out,[59] firms that provide training are significantly more productive than those firms that do not.

As before, we are interested in the differences between small and large firms. We can state with no doubt that training is more frequent among large firms: 70% of large firms versus only 20% of small firms do training. Considering the total sample, the average expenditure on training is around 40 euros per worker; 25 euros in the case of small firms and 100 in the case of large ones (expressed in constant euros of 2001). When considering only the subsample of firms providing training, the average expenditure is 130 euros per worker for the total sample and 115 euros and 140 euros in the small and large firms' subsamples respectively. This result is thus consistent with the general finding that large firms provide more training, and motivates the extensive analysis performed in Chapter 5.[60]

Among firms that provide training, the differences in TFP by firm size are not statistically different from zero, and similar results are obtained when considering the firms

[59] See for example, Alba-Ramírez (1994) or Barrett and O'Connell (2001) for an analysis of the impact of training on productivity.

[60] See for example, Black et al. (1999).

that do not provide training. Although there are very important differences between small and large firms in the provision of training per worker, these differences do not seem to turn into productivity differences between the two groups.

Table 2.13. TFP, continuous training and size

					Provide training			No training					
Year	#obs	#Provide Training	#Small	#Large	Mean	Std dev	Median	Mean	Std dev	Median	Eq mean Total ($)	Eq mean Small ($)	Eq mean Large ($)
2001	952	32.35%	22.53%	68.32%	0.1874	0.2402	0.1814	0.087	0.2657	0.0765	5.82***	4.57***	2.44***
2002	857	31.74%	21.24%	71.51%	0.1948	0.2342	0.1766	0.0828	0.2757	0.0755	6.15***	4.88***	3.07***
					Small - provide training			**Large - provide training**					
Year	#obs		#Small	#Large	Mean	Std dev	Median	Mean	Std dev	Median	Eq mean (&)		
2001	307		55.05%	44.95%	0.1873	0.2529	0.18	0.1875	0.2247	0.1857	0.01		
2002	272		52.94%	47.06%	0.193	0.2305	0.1469	0.1967	0.2393	0.2009	0.13		
					Small - no training			**Large - no training**					
Year	#obs		#Small	#Large	Mean	Std dev	Median	Mean	Std dev	Median	Eq mean (&)		
2001	645		90.08%	9.92%	0.0846	0.2713	0.0764	0.1084	0.209	0.0809	0.84		
2002	585		91.28%	8.72%	0.0818	0.2829	0.0769	0.0934	0.1865	0.0537	0.40		

Note: test of equality of mean: (***), (**) and (*) denotes significant at 1%, 5% and 10%. ($) Compares TFP in firms that provide and do not provide training. (&) Compares TFP in small and large firms. Results correspond to TFP levels and expenditure on training in 1994, 1998 and 2002.

Next, we compare TFP in small firms that provide and do not provide training. Table 2.13 shows that productivity is much higher in firms that provide training than in those that do not. Similar results are obtained in the case of large firms. Then, results suggest that training seems to be strongly associated with productivity both in small and in large firms. Actually, once we control by the provision of training the TFP-size gap turns to be non-significant (the tests do not reject the null hypothesis of equal TFP averages for small and large firms providing training and for small and large firms not providing training).

The results obtained in this Section seem to indicate that a more intense innovative activity and use of human capital are strongly related with higher TFP in firms. Concretely, firms that obtain new processes and that invest in R&D are more productive. However product innovations do not seem to be associated with TFP levels. Moreover, process innovations and the R&D effort seem to be a key element for small firms to achieve higher TFP levels rather than for large firms. On the other hand, using more white collars is strongly associated with higher productivity levels. And the differences in TFP between small and large firms suggest that these two groups might be obtaining different returns from their investment in human capital. In relation with training, we find that it seems to be strongly associated with productivity.

However, the causality relationship between the variables considered in each case might be questioned. On the basis of this analysis, we cannot extract conclusions on, for instance, whether R&D increases firms' productivity or the more productive firms spend more on R&D. Moreover, this descriptive is an unconditional analysis as we are not considering the effect that other firm characteristics may have on TFP. In Part II we further the analysis by investigating whether the use of innovative activity and human capital has an impact on firms' productivity, when conditioning to other possible determinants of productivity. We also check if the only differences in the effect of innovations and human capital on TFP between small and large firms are caused by the different intensity with which they use these factors, or if there are also differences in the return obtained by both types of firms.

2.6. Conclusions

Departing from the index suggested by Good et al. (1996) discussed in Chapter 1, in this Chapter we calculated a TFP measure for a sample of manufacturing firms in Spain over the period 1990-2002. This measure of TFP constitutes one of our variables of interest in Parts I and II. The variables involved in the TFP index as well as the variables that measure firms' size, innovative activity and human capital are drawn from the ESEE. In this Chapter we described the particularities and justified the use of these variables, not only for the descriptive analysis in this Chapter, but also for the remainder of the thesis.

The ESEE is an unbalanced panel that has been used in many studies on empirical industrial organization for Spain. This annual survey contains information for firms with 10 or more employees over the period 1990-2002 and it is representative by size and industry. After a cleaning procedure, we obtain a TFP measure for an unbalanced panel with 800-1000 observations each year (13035 observations in total, for 2104 different firms).

To measure TFP, we need to calculate the variables involved in this index: output, labour, capital, intermediate inputs as well as the cost of these three inputs. We paid particular attention to the measurement of the stock of physical capital. After describing the specific methodology to measure the output and input variables that intervene in the TFP index, we presented a descriptive analysis of the evolution of these variables. We obtained evidence about the intensification in the use of physical capital together with an increase of the production per hour worked over the period and, specifically, during the second half of the nineties.

The TFP measure calculated in this analysis shows that firms' productivity increases almost every year, however since the second half of the nineties there is a slow down in the productivity growth. This growth rate corresponds to a representative Spanish firm in the period under analysis; however the TFP increases are not homogeneous along the distribution: the most productive firms are more capable of increasing their TFP levels and the density functions show evidence of the formation of a group of highly productive firms during the second half of the nineties. The stochastic kernels show a highly persistent behaviour of firms in their TFP levels. The dispersion in firms' productivity tends to increase over time and thus firms become more and more heterogeneous in TFP over time.

Firm size is considered a main source of heterogeneity in TFP among firms. Actually we found that large firms are significantly more productive than their smaller counterparts, although differences tend to reduce over time. This is due to a higher productivity growth of small firms since the mid-nineties. The differences in TFP are not homogeneous along the distribution: they are more severe in the lower part of the distribution, while there is a group of very productive small firms which are as productive as the most productive large firms. Although the TFP gap between small and large firms decreases at any initial TFP level, it becomes narrower for firms at the central and upper part of the distribution.

Given that it is generally accepted that innovation and human capital play a crucial role in improving firms' performance, we provided a descriptive of TFP in relation with these relevant variables. A main difference between these two sources of economic growth is that Spain has performed better in increasing its human capital rather than its technological capital in the last decades. Large firms in the Spanish manufacturing sector innovate more than their smaller counterparts, according with our three definitions of innovative activity. Firms that obtain new processes are more productive, while product innovations do not seem to be associated with higher TFP levels. Moreover, process innovations appear to be a key element for small firms to achieve higher TFP levels. Thus, the relationship between innovation and productivity seems to be conditioned by firm size, suggesting a possible indirect effect of innovation on productivity. Similar results are obtained when using the expenditure on R&D as a measure of innovative input.

Large firms in the Spanish manufacturing sector employ more qualified employees. Firms that employ a high proportion of white collars are more productive than those that employ a low proportion, which is obtained for both small and large firms. The differences in TFP between small and large firms (within the groups that use a high and low proportion of qualified workers) seem to indicate that these two groups obtain different returns from their investment in human capital. As for continuous training, another component of human capital, we obtain that productivity is much higher in firms that provide training than in those that do not (for both small and large firms).

This descriptive analysis supports the hypothesis that the TFP differences between small and large Spanish manufactures is not only due to differences in the level of use of knowledge capital, but also to differences in the effect that this capital has on TFP. This constitutes the basis for the analysis in Part II, where we perform a causal analysis and we control for the effect of other variables.

Appendix 2.1. Validation of the stock of physical capital

In this Appendix, we compare the estimation of the stock of physical capital obtained here with that provided by the *Fundación SEPI* over the period 1990-1999, departing from the work by Martín-Marcos and Suárez-Gálvez (1997), who use the same dataset. As, the stock of physical capital provided by this organization was only available for this period, we have estimated our own stock of physical capital following the methodology of these authors, so that the series are expected to be very similar. The stock of physical capital has been calculated by the permanent inventory method and it is expressed in thousands of 1990 constant pesetas. Tables A2.1 and A2.2 show the average stock of physical capital, the standard deviation and the median on every year, for both the equipment and constructions respectively. The average stock of equipment has slightly increased over the period of analysis, showing a cyclical behaviour with decreases in the first half of the nineties and some recovery in the second half. On the last four columns, we have performed tests of equality of means and the one-sided Kolmogorov-Smirnov tests of stochastic dominance to compare our stock with that by Martín-Marcos and Suárez-Gálvez (1997) on every year. In all the cases, we cannot reject the null that in average the stock of equipment is equal in the two estimations. The same happens with the KS test, as we cannot reject the null that the two distributions are equal. The estimation of stock of constructions shows some differences between the two studies. The stock by Martín-Marcos and Suárez-Gálvez (1997) shows a slight decrease between 1990 and 1999, while ours shows a slight increase. More concretely, the stock of constructions by these authors decreases between 1990 and 1996 (with the exception of 1994), and increases between 1996 and 2002. In our series, there is also a decreasing tendency between 1991 and 1996 (with the exception of 1993 and 1994) and an increase between 1996 and 2002. However, the tests of equality of means show that in general the two series do not have significant differences, except for 1996 and 1997, where the null is rejected. The KS test does not reject the null of equality of distributions on any year.

Table A2.1. Comparison of the stock of physical capital between our calculation and the work by Martín-Marcos and Suárez-Gálvez (1997) (MMSG): Equipment

Year	No of obs	Mean MMSG	Std dev MMSG	Median MMSG	Mean	Std dev	Median	Equality mean	p-value	KS statistic One-sided test	p-value
1990	721	794672.26	2447505.24	96552.00	856674.82	2717119.66	100878.25	0.481	0.315	0.024	0.988
1991	1021	837511.15	2460834.93	76351.50	876402.88	2567344.74	78653.70	0.357	0.361	0.022	0.972
1992	1150	864668.75	2564079.58	66520.35	897733.24	2637819.80	71165.54	0.309	0.379	0.019	0.984
1993	1033	840273.13	2519101.49	63554.20	883254.57	2636297.72	67080.12	0.388	0.349	0.022	0.960
1994	970	818796.48	2568113.94	61655.70	851462.36	2500952.65	65211.54	0.280	0.390	0.020	0.992
1995	912	799640.66	2522032.42	60706.95	850624.80	2624481.04	64596.17	0.432	0.333	0.035	0.628
1996	888	758541.81	2074080.70	63604.00	805286.18	2049526.49	71872.05	0.475	0.317	0.036	0.611
1997	1037	699863.85	2195183.79	57307.40	754349.48	2223665.56	67233.42	0.565	0.286	0.040	0.392
1998	1052	841636.93	4912459.19	75306.10	963544.97	6058519.77	83459.15	0.569	0.285	0.038	0.432
1999	1023	930896.65	5734506.63	73546.40	1056939.73	6915647.04	80122.01	0.497	0.310	0.035	0.551

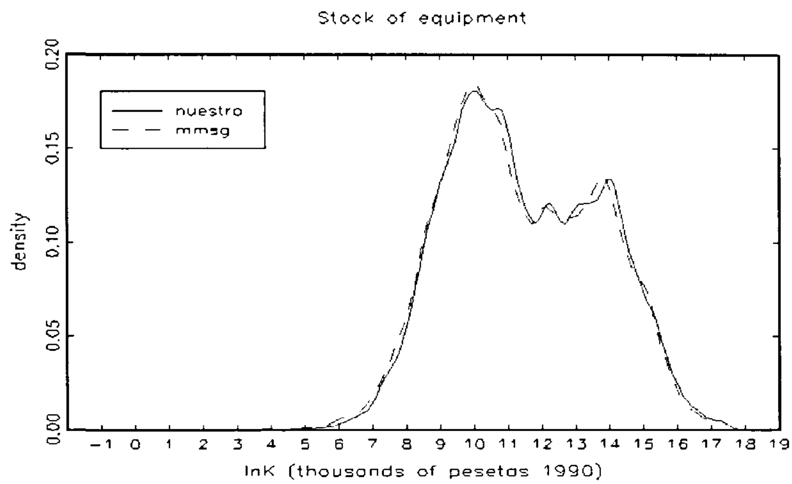
Note: test of equality of mean: (***), (**) and (*) denotes significant at 1%, 5% and 10%. The stock of capital is expressed in constant thousands of pesetas of 1990

Table A2.2. Comparison of the stock of physical capital between our calculation and the work by Martín-Marcos and Suárez-Gálvez (1997) (MMSG): Constructions

Year	No of obs	Mean MMSG	Std dev MMSG	Median MMSG	Mean	Std dev	Median	Equality mean	p-value	KS statistic One-sided test	p-value
1990	651	548173.27	2737022.32	27457.00	556512.40	2683048.13	29905.58	0.056	0.478	0.028	0.965
1991	951	513432.88	2325202.12	23713.70	564347.41	2462202.78	25854.62	0.464	0.321	0.019	0.996
1992	1125	467241.91	2124863.96	18437.70	523390.64	2282806.54	20914.77	0.604	0.273	0.044	0.236
1993	1030	435086.26	2055051.18	17888.30	524465.65	2368092.14	21412.71	0.915	0.180	0.033	0.629
1994	972	447962.37	2107360.65	17460.50	568341.18	2542499.86	21999.43	1.136	0.128	0.039	0.447
1995	915	436884.82	2085760.08	17526.40	554581.22	2538877.21	23727.45	1.084	0.139	0.040	0.443
1996	892	354723.90	986813.78	17205.40	479537.47	1385442.65	24584.43	2.192***	0.014	0.045	0.331
1997	1040	360245.37	1841762.73	16629.65	481185.25	2353497.20	22886.62	1.305*	0.096	0.046	0.218
1998	1062	440110.81	2647085.08	21052.05	587395.50	3456152.85	29722.19	1.103	0.135	0.052	0.116
1999	1029	482078.15	3409918.88	19189.30	581189.40	3535046.85	27296.21	0.647	0.259	0.047	0.213

Note: idem Table A2.1.

Finally, Figure A2.1 compares the external shape of the stock of physical capital in this Chapter with data provided by the *Fundación SEPI* and the two series appear to be very similar, although in the case of the constructions our series provide slightly higher values than in the case of Martín-Marcos and Suárez-Gálvez (1997). These differences may be related to the use of different price series after 1995.



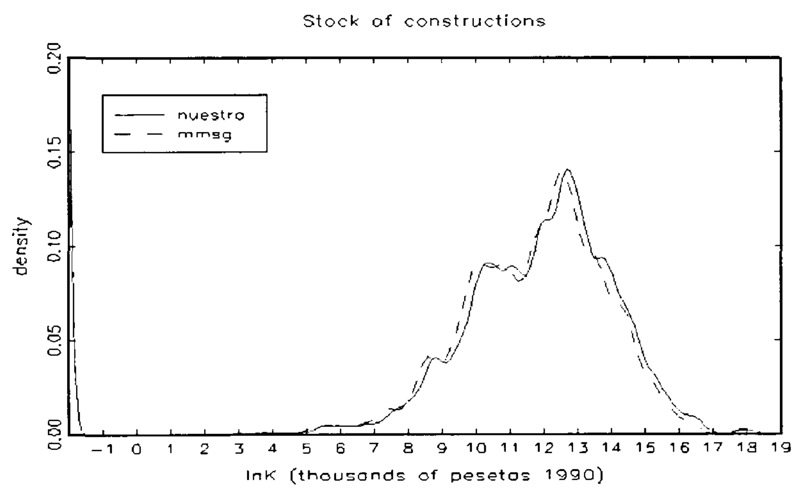


Figure A2.1. Comparison of the stock of physical capital between our calculation and the work by Martín-Marcos and Suárez-Gálvez (MMSG)

Appendix 2.2. Estimation of the density function

We estimate density functions (the external shape of the distribution) using a nonparametric methodology, as it does not assume TFP to follow any known distribution. The expression of the Rosenblatt-Parzen kernel density estimator is expressed as follows (Silverman, 1986):

$$\hat{f}(x) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{h}K\left(\frac{x - X_i}{h}\right)$$

We use the Gaussian kernel for the K(·) function and a bandwidth that is estimated by applying the plug-in method suggested by Sheather and Jones (1991). In order to compare density functions of different distributions, the bandwidth has been settled down to: 0.05148001 in the case of Figure 2.1 (which is the arithmetic mean of bandwidths for the four distributions under comparison: TFP in 1990, 1994, 1998 and 2002); 0.19450392 in the case of Figure A2.1 (which is the arithmetic mean of bandwidths for the four distributions under comparison: stock of equipment and constructions in the two studies); and to 0.06683227 in the case of Figure 2.3 and in all the Figures in Chapter 4 (which is the arithmetic mean of bandwidths for the six distributions: TFP in small and large firms in 1994, 1998 and 2002).

We have also considered the possibility of reporting estimates of the firms' size weighted densities, as two firms equally productive may have different impact on the whole distribution according to their size. However, when comparing the external shape of the weighted and unweighted distributions, not much difference was observed. Then, for simplicity, we report results for the unweighted densities.

Appendix 2.3. Test of stochastic dominance

The tests of (first order) stochastic dominance allow comparing productivity distributions of different groups of firms and establishing a ranking between them.[61] Let's suppose two independent random samples of size n and m. Let $Z_1, \ldots, Z_n$, be a random sample corresponding to a group of firms from the cumulative distribution function F, and $Z_{n+1}, \ldots, Z_{n+m}$, from the cumulative distribution function G; z_i is the productivity level of firm i. Then, the condition of first order stochastic dominance of F relative to G is: $F(z)-G(z)\leq 0 \ \forall z \in \Re$, with strict inequality for at least one z. The hypotheses we are testing are (i) that the null in the two sided test

$$H_0 : F(z) - G(z) = 0 \ all \ z \in \Re \quad vs \quad H_1 : F(z) - G(z) \neq 0 \ some \ z \in \Re$$

can be rejected and (ii) that the null in the one-sided test

$$H_0 : F(z) - G(z) \leq 0 \ all \ z \in \Re \quad vs \quad H_1 : F(z) - G(z) > 0 \ some \ z \in \Re$$

cannot be rejected.

This test can alternatively be formulated as:

(i) Two sided test

$$H_0 : \sup_{z \in \Re} |F(z) - G(z)| = 0 \quad vs \quad H_1 : \sup_{z \in \Re} |F(z) - G(z)| \neq 0$$

(ii) One-sided test

$$H_0 : \sup_{z \in \Re} \{F(z) - G(z)\} = 0 \quad vs \quad H_1 : \sup_{z \in \Re} \{F(z) - G(z)\} > 0$$

The two-sided test will determine whether there exist significant differences between the two TFP distributions. The one-sided test will determine whether $F(z)$ stochastically dominates $G(z)$. Then, in case we cannot reject the null in the two-sided test, or in case we reject the null in both tests, $F(z)$ will not stochastically dominate $G(z)$. When the two-sided test is rejected and the one-sided test cannot be rejected, $F(z)$ is on the right of $G(z)$, and we conclude that F dominates G.

The Kolmogorov–Smirnov test statistics for the one and two-sided tests are respectively:

$$\delta_n = \sqrt{\frac{n * m}{N}} \max_{1 \leq i \leq N} |T_N(Z_i)| \quad \text{and} \quad \eta_n = \sqrt{\frac{n * m}{N}} \max_{1 \leq i \leq N} \{T_N(Z_i)\}$$

[61] This strategy has been recently applied in Delgado et al. (2002) to check for higher productivity among exporting firms. Here we follow the description of the test strategy in their paper.

where $T_N(Z_i)=F_n(Z_i)-G_m(Z_i)$ and $N=n+m$. F_n and G_m represent the empirical distribution functions for F and G, respectively. The limiting distributions of both test statistics, δ_N and η_N, are known under independence[62].

In Section 2.5.3, F_n and G_m represent the empirical distributions of TFP for large and small firms respectively.

[62] Kolmogorov (1933) and Smirov (1939) showed that, under the assumption that all observations are independent, the limiting distributions of δ_S and η_n under H_0 are given by $\lim_{n\to\infty} P(\delta_n > v) = -2\sum_{k=1}^{\infty}(-1)^k \exp(-2k^2v^2)$ and, $\lim_{n\to\infty} P(\eta_n > v) = \exp(-2v^2)$, respectively.

CONCLUSIONS

In Part I of the thesis, we intended to establish a suitable framework to measure firm productivity and then we used the resulting measure of TFP to obtain preliminary evidence on its relationship with firm size, innovative activity and human capital.

Chapter 1 discusses different measures of TFP with the objective of selecting a measure that has more desirable properties. Concretely, we focused on index numbers and chose the index by Good et al. (1996). This index is derived from a translog production function, which is more general than other production functions, and so it is a superlative index. Moreover, it is transitive, which permits cross-section comparisons, and it has a high degree of characteristicity. This index also permits a decomposition of efficiency and technological change and it permits relaxing the assumption of perfect competition. In addition, it is sample independent, which permits extending the series of TFP as new data is available.

In Chapter 2 we measured TFP for a sample of manufacturing firms in Spain over the period 1990-2002 on the basis of the index selected in the previous Chapter. After describing the ESEE and the cleaning procedure, we explain in detail the variables used to measure TFP. Based on this TFP measure, we perform a descriptive analysis to obtain preliminary insights on the behaviour of TFP in relation to firm size, innovative activity and human capital.

We confirm the previous evidence that large firms are more productive, innovate more and have a more qualified labour force. An outstanding result is that small innovative firms achieve TFP levels close to large innovative ones. In this view, process innovations seem to a key element for small firms to achieve higher TFP levels and similar results are obtained for R&D. Another interesting result is that productivity in large firms with a high proportion of qualified workers is higher than in small firms, after controlling for this characteristic. Finally, productivity is higher in firms that provide training than in those that do not.

Thus, there seem to be interesting relations between these variables which are further explored in Part II. Specifically, firm size seems to play a central role in explaining firms' productivity and to condition the effect of innovation and human capital on firms' productivity. The results presented in this Part I constitute the basis for the remaining analysis in Parts II and III.

PART II

DIFFERENCES IN TOTAL FACTOR PRODUCTIVITY BETWEEN SMALL AND LARGE FIRMS

INTRODUCTION

The evolution of productivity has been one of the issues of major concern among economists, especially in the last years, when productivity growth has slowed down in many advanced economies. As commented in OECD (2003), "Productivity has accelerated in some of the most affluent economies, most notably the United States, and slowed down substantially in others, such as continental Europe and Japan".[63] The Spanish economy has also suffered a deceleration process during the nineties (Gual et al., 2006).[64] As stated by the Lisbon Agenda in March 2000, Spain has the objective of achieving convergence in income per capita and an employment rate higher than the average in the European Union. In 1990, the GDP per capita in Spain was about 87.6% of the European average and it increased in the following years until 92.7% in 2000. The main purpose of the Spanish National Reform Program in the Lisbon Agenda is that this convergence process is achieved in 2010.

This objective requires increasing productivity. Although total factor productivity (TFP) has increased in the period of our analysis, it has slowed down

[63] OECD (2003, pp 51) shows that some OECD countries have accelerated their TFP growth during the 1980s and 1990s (New Zealand, United States, Canada, Australia, Sweden, Denmark, Norway, Finland or Ireland) while other countries have suffered a slow down (Spain, Germany, France, Japan, Italy, Belgium, Austria or the Netherlands).

[64] Gual et al. (2006) explain that the gap between the EU-15 and US labour productivity has been widening since mid eighties. Moreover, Spain occupies a very unsatisfactory position in relation to the other EU-15 countries: while the average growth rate of EU-15 during the nineties is 1.36%, Spain only increases 1.2%. And the situation becomes worse since 2001.

during the nineties, especially during the second half. Actually Goerlich et al. (2002) find that during the nineties, the TFP decreases its contribution to economic growth, which affects sustained economic growth. Some aggregate studies attribute this decrease in productivity to the behaviour of the manufacturing sector.[65] A generalized recommendation indicates that Spain should increase its efficiency and that this requires a higher investment effort in technologies and human capital. The economic policies included in this National Reform Program are directed to deal with the weaknesses of the Spanish Economy and thus to achieve a more modern production system. Actually two of the key elements of the Program are the "Increase and Improvement of Human Capital" and the "R&D&I Strategy".

But the effort in technological and human capital may be explaining only part of the story. In the Introduction of the thesis, we highlighted the reduced average firm size and the high percentage of small firms in Spain in relation to other advanced economies. In this Part II, we argue that the structure of the Spanish industry may also play a role in explaining the lower TFP levels in Spain in relation to other advanced economies. Concretely, the fact that the Spanish economy is characterised by the predominance of small firms could also explain this low productivity. See for example, Bartelsman and Doms (2000) or Ruano (2002), who explain that smaller firms tend to be less efficient. Small firms have certain characteristics that can be seen as limitations: they are usually considered to innovate less than large firms and to employ less qualified employees. The difficulties of small firms in accessing levels of innovation and human capital close to the levels of large firms may constitute a limitation for them to achieve higher productivity levels. Thus, the predominance of small firms in the Spanish economy can be seen as a limitation for the economy as a whole.

In this Part II, we investigate the TFP differences between small and large firms. More specifically, our hypothesis is that the higher productivity in large firms may be associated with two of the main determinants of firms' performance: the human and technological capital that firms incorporate. In addition, the contribution of these factors in explaining the TFP differences between small and large firms may be due to two different effects: first, the fact that large firms have a higher percentage of qualified employees and obtain more innovations; and second, the fact that large firms obtain

[65] See Estrada and López-Salido (2001b).

higher returns from their investment in human and technological capital. In other words, every innovation or every additional qualified worker incorporated in a large firm could provide higher returns (higher impact on productivity) than in a small firm. Thus, the higher returns of these factors may also explain why large firms are more productive and why they have more incentives to use them. In Chapter 3, we analyse the contribution of these factors at the mean of the distribution using the Oaxaca-Blinder decomposition, while Chapter 4 explores the TFP differential at every point along the distribution by means of a counterfactual distribution analysis.

The analysis of the effect of returns in explaining the productivity differentials between small and large firms is the main contribution of Part II by placing special emphasis on the idea that firm size conditions the effect of innovations and employees' qualification on productivity, so that size *indirectly* affects productivity. Moreover, this analysis adds to the previous empirical evidence that the innovative activity and the use of skilled labour have a positive impact on firms' productivity. It also contributes to the literature that considers firm size as a main source of heterogeneity in firms' productivity.

In Chapter 3 we present empirical evidence and theoretical reasons in favour that small and large firms follow different patterns of behaviour in relation to productivity, innovation and human capital: large firms are more productive, innovate more and use more qualified labour. Departing from the descriptive in Sections 2.5.3 and 2.5.4, we obtain similar result for the case of Spanish manufacturing firms. Using the TFP index defined in Chapter 1 and data from the *Encuesta sobre Estrategias Empresariales* (ESEE) to measure human and technological capital, we obtain that innovation and human capital have a positive and significant effect on productivity. In order to analyse differences in productivity by firm size, we estimate the impact of the knowledge variables for the subsamples of small and large firms and we find considerable differences between them. The results in this Chapter indicate that large firms obtain higher returns from their investments in these factors, whereas small firms present smaller coefficients and in some cases they are not significant. Finally, we decompose the TFP differential between small and large firms in differences in the use of human and technological capital and differences in returns to these factors in the mean of the distribution.

The decomposition based on the mean uses the Oaxaca-Blinder methodology to analyse the individual contribution of our variables of interest. This methodology has extensively been used in labour economics to decompose the wage gap among different groups of workers. In the case of TFP, this decomposition permits studying the relative importance of technological and human capital, as well as their returns, in explaining the productivity differences between small and large firms.

The Oaxaca-Blinder decomposition shows that our variables of interest explain part of the average TFP gap between small and large firms. Regarding human capital, it explains quite a large part of the gap —both as differences in the level of qualified workers between small and large firms and differences in their returns. Thus, we come to the conclusion that large firms seem to obtain higher returns from their skilled labour, together with the fact that they invest more than small firms in human capital. With respect to innovation, it explains a smaller part of the differential and it is basically due to differences in this characteristic: once innovations are obtained, the effect on TFP seems to be the same, regardless of the size of the firm.

This decomposition evaluates the contribution of innovation and human capital in the mean of the distribution, but these effects are not necessarily homogeneous along the TFP distribution. Using the counterfactual distribution analysis, inspired in Jenkins (1994), we propose transferring the idea of the Oaxaca-Blinder decomposition to the entire distribution by studying the contribution of differences in characteristics and returns at any point of the distribution.

In brief, results confirm that the contribution of differences in returns and differences in endowments are not homogeneous over the firms' TFP distribution. With regards to innovation and human capital, differences in returns to these factors can only explain a modest part of the TFP differential. This effect is quite heterogeneous along the distribution and small firms with high TFP levels would improve their productivity if they had the same returns as large firms. In other words, if small firms had the returns to human and technological capital of large firms, some of them would increase their TFP, becoming as productive as the most productive large firms.

Chapter 3

THE TFP DIFFERENTIAL BETWEEN SMALL AND LARGE FIRMS: ANALYSIS IN THE MEAN OF THE DISTRIBUTION

3.1. Introduction

The objective of this chapter is assessing the relative contribution of human and technological capital to explain the total factor productivity (TFP) differential between small and large firms. Concretely, the TFP gap is evaluated in the mean of the distribution using the Oaxaca-Blinder methodology, which permits decomposing this gap in differences in the endowments of technological and human capital and differences in their returns.

Chapter 3 is structured as follows. In the next Section, we discuss the role of firm size as a source of heterogeneity in productivity as well as the impact of innovation and human capital on productivity, considering the role of firm size. Section 3.3 presents our empirical specifications and a brief discussion on the estimation method. In Section 3.4 we describe the variables used in this analysis and provide a descriptive analysis, showing that large firms in the Spanish manufacturing sector are more innovative and employ a more qualified labour force. Section 3.5 offers the results of the OLS and random effects estimation, which show the positive relationship between productivity and human and technological capital, although with differences by firm size. In Section 3.6 we present the results of the Oaxaca-Blinder decomposition and, finally, Section 3.7 concludes.

3.2. Factors Determining Firms' Productivity

3.2.1. Productivity, heterogeneity and firm size

The vast majority of firm-level studies on productivity recognize the existence of high heterogeneity among firms with common characteristics (heterogeneity in terms of size, age, technologies, productivity levels, entry-exit patterns, and so on). Such heterogeneity cannot be appreciated under the macroeconomic approach as it aggregates different firms that have different characteristics and they are all supposed to be affected by economic forces in a similar way. Thus, such models may not explain the observed differences in firms' productivity adequately, while the microeconomic approach permits a deeper analysis of the characteristics that may explain such differences in productivity. "The evolutionary literature recognizes the large amount of heterogeneity across firms regarding their productivity and seeks to explore the factors behind this heterogeneity within the framework of firm behaviour" (Bartelsman and Doms 2000, pp. 570).

A strand of microeconomic literature that analyses the heterogeneity of productivity behaviour in Spain has emerged with the appearance of the micro-level dataset *Encuesta de Estrategias Empresariales* (ESEE). This literature has focused mainly on the effects of firm dynamics, exports and innovative activity.[66]

Some models in industrial organization have provided a framework that permits analysing the heterogeneity in productivity among firms. In these models, industries are not composed by representative firms anymore. Lucas (1978) proposes a model of the size distribution of business firms. It consists of a distribution of people by managerial talent which underlies the distribution of businesses by firm size. The individuals may become either employees (working for someone else and earning a salary) or managers (taking managerial decisions and obtaining their returns). One implication of this model is that, as capital per capita increases, workers become more productive, their wages increase, and they will prefer working for someone else, so that firm size will increase. Thus, greater capital instensity might be associted to a larger firm size. Jovanovich (1982) proposed a model in which, as firms gain experience, they learn about their level of costs and about their efficiency level. If they learn they are efficient, they decide to

[66] See Fariñas and Ruano (2004) for an analysis of firm dynamics; Delgado et al. (2002) for an analysis of exports; Huergo and Jaumandreu (2004a,b), Máñez et al. (2004, 2005) and Ornaghi (2006) among others for the innovative activity.

expand, or otherwise, they decide to contract or even exit the market. Thus, the most efficient firms are expected to survive and grow in size, while the most inefficient are expected to fail. This process suggests a relationship between size and productivity in a framework of heterogeneity, even between firms in the same industry or with similar characteristics. Hopenhayn (1992) proposes a model in which firms are heterogeneous in their productivity levels, and thus on their flows of expected future benefits. On the one hand, when firms enter a market, they face sunk costs and they only decide to enter when the future expected benefits are higher than sunk costs. In this sense, sunk costs act as an entry barrier. On the other hand, firms decide to exit the market when this carries lower costs than continuing in it. In absence of sunk costs, the less productive firms would exit the market. However, with sunk costs, firms would continue in the market to try to compensate them and in order to avoid higher losses. In this sense, sunk costs are also exit barriers and the implication is that, in their presence, some non-efficient firms will continue in the market. Ericson and Pakes (1995) propose a model in which firms invest in R&D to improve their productivity levels. This way, their productivity is a function of their own R&D investment, of the productivity of their competitors and of the pressure of firms entering the market. If a firm succeeds and is productive enough, it will grow, or otherwise it will fail and contract or even exit the market. Finally, in the model by Olley and Pakes (1996), firms decide whether to continue in the market and demand a certain amount of inputs, or otherwise, exit the market. This decision depends on whether firms expect to achieve a certain (unobservable) efficiency level or not. As long as firms continue in the market, firm size and productivity are also related in this model.

Both the theoretical and empirical literature on productivity at firm level agrees in considering size as a main source of heterogeneity in firms performance. A first explanation of the productivity dispersion by size is the difference in the available technologies. Even if all the available tecnologies were equally efficient, at different production levels, some technologies would be more appropiate than others. So that different firm sizes correspond to different appropiate technologies. In addition, in the presence of scale economies, firms can produce larger quantities with lower unitary costs. Other theoretical arguments that explain why large firms are more productive include: the scope economies effect, the experience effect and the organization effect.[67] The heterogeneity of firms can also be due to the absence of perfect competition in prices. If perfect competition exists, any loss of competition, would make the firm disappear. However, if there is not competition in prices, firms can grow with objectives other than competitiveness and so it can also explain firms' dimension.[68] Finally, the industrial effect is recognized to capture a great deal of heterogeneity in productivity and firm size, especially in relation with the innovative activity.[69]

Although large firms are considered to be more productive, small firms are often seen as "the engines of growth" because of their role as employment creators, innovators and entrepreneurs (Audretsch, 2002). As Schumpeter pointed out, they play a main role in the economy, not only because their innovation has a direct contribution to their competitiveness, but also because they act as initiators, catalysts and media for wider technical change. They operate in a very competitive environment, so that they make a great innovative effort to be able to survive; they are very flexible, which allows them to adopt technologies developed in other environments; and they act as catalysts, because their closeness to the market allows them to appreciate the opportunities and develop a technological response.

Geroski (1998) argues that controling for firm size in regressions can be even considered as a routine. Thus, when we introduce a firm size variable in a regression we are accounting for different technologies associated to a certain firm size, which, in terms of Geroski is called the *direct effect* of size on productivity, that is, as a variable that *ceteris paribus* improves efficiency. This author claims that size may have also an *indirect effect* on productivity, that is, conditioning the effect of other variables on productivity as they will show different patterns of behaviour for small and large firms. This author suggests controlling for the indirect effect through analysing separately the coefficients of small and large firms and evaluating to what extent they differ. Differences in the returns of firms' endowments between small and large firms indicate

[67] See Audretsch et al. (1998).

[68] For example, when managers and owners are different agents, the decisions of the managers affect the owners (stakeholders). Under imperfect competition in prices, managers may decide to expand the firms for reasons other than efficiency. Sargant (1943) suggested that many owner-managed companies adopt "satisficing" rather than maximizing policies. In such case, the firm does not disappear from the industry, even if its size is not competitive. Thus, differences between firms in costs, size and market share arise.

[69] Rajan et al. (2001) review the different sources of firm size dispersion according with different theories. See also Martin (2002).

that, an additional unit of innovation or qualified employees hired in one of the two groups of firms would obtain higher returns to these factors than in the other group. In this sense, size is exerting an indirect effect on firm productivity, as it conditions the impact of other factors on productivity. Building on this idea, one of the main contributions of the present analysis is using the Oaxaca-Blinder decomposition to assess the relative importance of differences in firms' endowments and in their returns to explain the productivity differential between small and large firms.

3.2.2. Productivity, innovation, labour qualification and firm size

The technological and human capital endowments of firms have traditionally been considered two factors fostering productivity. Griliches (1979) is a pioneer work in assessing the contribution of R&D on productivity growth. Most literature on the innovative activity estimates the elasticity or the rates of return to a stock of knowledge (calculated on the basis of the R&D effort) on productivity. Studies using firm level data show a wide range of estimates, and some of them have found weaker correlations than at sectoral or country level, especially when including industry dummies (see Mairesse and Sassenou, 1991, for a survey). However, the relationship between productivity and R&D expenditures embodies two different processes: the production of innovations starting from R&D activities and the incorporation of these innovations to production.[70] Firms invest in R&D in order to develop process and product innovations, which in turn may contribute to their productivity and economic performance. Crépon, et al. (1998) emphasize that it is not innovative input (R&D) but innovative output that increases firms' productivity. This measure of innovative activities permits measuring those changes that firms consider relevant for their production process. Moreover, this measure avoids having to distinguish between formal and informal R&D activities (Huergo and Jaumandreu, 2004a). Also, product and process innovations play different roles on firms' performance. Process innovations reduce the unit cost of production of the good, and then productivity increases, because the knowledge capital acquired by a firm improves the mechanism by which input is transformed into output. Product

[70] Griffith et al. (2004) highlight that the effort of R&D of a firm increases its productivity not only because of the fact that the firm has a higher probability of introducing an innovation, but also because it rises its absorptive capacity, that is, it becomes more flexible and adaptable to benefit from spillovers than its rivals.

innovations (improvements in the quality of existing products and the introduction of new goods) are supposed to increase output as it gives the innovator transitory power over a group of new buyers. However this effect is achieved through a demand function, rather than through a cost function.[71]

Other studies have related the innovative capacity with firm characteristics, such as size, finding a positive relation between them. Schumpeter (1942) hypothesized that large firms have an advantage over small companies as their financial situation allows them to be the most capable innovators. Acs et al. (1994) find that large firms invest more in R&D and innovate more, however, small firms appear to have higher innovative productivity. Huergo and Jaumandreu (2004b) in a study for Spain obtain that "innovation is strikingly related to size". In these studies, the underlying hypothesis is that firms show different patterns in their innovative activity according to their size. However, small and large firms can differ not only in their innovative intensity, but also on the returns to such innovative activity. In other words, once an innovation has been done, are there differences in returns between small and large firms?

Klepper (1996) proposed a theoretical model where firm size plays a crucial role in firms' appropiation of the returns to innovation and in firms engaging in R&D activities. The larger the firm, the more output over which process R&D fix costs can be averaged, then returns to process innovations are higher, which encourages additional innovative effort. Cohen and Klepper (1996) corroborate the hypothesis of easier apropiability of returns to innovation in the case of large firms for the US case. These papers suggest the existence of differences in returns to innovation between small and large firms. Máñez et al. (2006) estimate the impact of innovations on productivity growth for the Spanish case and obtain that implementing process innovations leads to an extra productivity growth both for large and small firms. However, the persistence of this extra productivity growth is longer for large than for small firms. Parisi, Schiantarelli and Sembenelli (2002) analyse the impact of innovations on productivity by different firm sizes for the Italian case and obtain that large firms have a large impact of innovation on productivity. What we intend to analyse in this Chapter is the contribution of innovations in explaining differences in TFP between small and large

[71] See Ornaghi (2006) and Huergo and Jaumandreu (2004a).

firms, both as differences in the technological levels and in the impact of firms' technology in productivity.

On the other hand, the literature that analyses the effects of human capital on productivity argues that those workers with better skills in solving problems and better communication skills, will do any task that requires something else than simple workforce in a more efficient way (see for example, De la Fuente 2004). Bolton and Dewatripont (1994) argue that communication in firms is costly as agents have to learn the information sent by others. Agents can reduce these costs by specializing in processing some kinds of information. When the returns to specialization outweigh the costs of communication, agents will collaborate in a firm. Then, if education translates into higher learning capacity to solve problems and to communicate, those workers with better education would be more productive, and so will the firms where they work. The microeconomic literature, concerned about the impact of investing in human capital on productivity levels, has typically estimated mincerian equations (see the survey by Harmon et al., 2002). In relation to firm level studies, a strand of literature is dedicated to analyse complementarities between innovations and human capital. This strand of literature analyses the so called skill biased technological change and, although the results are not conclusive, many papers find positive complementarities between these two factors (Manasse and Stanca, 2003; Aguirregabiria and Alonso-Borrego, 2001). Nevertheless, only a few microeconomic studies have considered studying the effect of human capital at firm level. Griliches and Regev (1995) estimate a production function including R&D capital services as well as a measure of quality of labour as a proxy for human capital for the Israeli industry. They find a coefficient of qualified labour of about 0.4 for the total sample and 0.5 for large firms in the pooled regressions, taking differences to control for individual heterogeneity, which they acknowledge to be quite large. Haltiwanger et al. (1999), using a matched employer-employee dataset, obtain that labour productivity is associated with certain characteristics of the workforce, such as the proportion of educated workers. Their result is consistent with a human capital model where more-skilled workers make the firm more productive. Other papers at firm level have considered the effect of training, another component of human capital that differs from formal education (Black and Lynch, 1996; Dearden et al., 2000).

Evans and Leighton (1989) found evidence of some sorting on observed and unobserved ability characteristics across firm sizes, and so, better educated workers are employed in large firms. Zábojník and Bernhardt (2001) propose a model in which workers in larger firms and large industries acquire more human capital. The general finding is that large firms employ more educated workers. In addition to the fact that large firms employ a more qualified labour force, it is also possible that returns to human capital are higher in large firms. Actually, in the literature that analyses the positive relationship between firm size and wages, Oosterbeek and Van Praag (1995) or El-Attar and López-Bazo (2007) obtain that large firms pay higher wages because they obtain higher returns to human capital. In this work we are interested in analysing if large firms are more productive due to the fact that they employ more qualified workers. But we are also interested in analysing whether small and large firms obtain different returns to any additional qualified employee hired.

The empirical evidence and the theoretical arguments suggest that technological and human capital affect productivity. There are also reasons to believe that larger firms have better endowments of these factors. As already pointed in Chapter 2, we argue that technological and human capital play different roles in determining productivity for small and large firms: first, because large firms are usually more innovative and employ more qualified workers; second, because the returns of these endowments on productivity may be larger in the case of large firms. This chapter analyses to what extent the productivity differentials by firm size, evaluated in the mean of the distribution, are due to firms' endowments in technological and human capital or to the returns to such endowments.

3.3. Empirical Specification and Estimation

Our empirical framework relates the TFP index obtained in Chapter 2 to innovation and skilled labour, our variables of interest, as well as to several control variables. Our approach is quite close to Griliches and Regev (1995). These authors estimate a production function at firm level including measures of human and technological capital. Instead of the production function, we use the estimate of a measure of TFP as our dependent variable and innovation and skilled labour as the explanatory variables, whose effects on productivity we want to assess. Hence, the empirical model can be expressed as follows:

$$TFP_{it} = \beta_0 + \beta_1 INN_{it-1} + \beta_2 HK_{it-1} + Z'\gamma_{it} + u_{it} \tag{3.1a}$$

where *TFP* is the logarithm of the total factor productivity index in firm *i* in year *t*, *INN* is a dummy variable that takes value one if the firm *i* reports to have made an innovation in year *t-1*, *HK* is the proportion of skilled labour for firm *i* in year *t-1*, *Z* is a set of standard control variables: firms' size,[72] age, industry and year effects, and *u* is an error term. We estimate two different specifications: our main specification is defined as (3.1a) and our second specification (labelled 3.1b) includes some additional variables in *Z* so as to check for robustness of the results after controlling for some other firms' characteristics. The variables included in *Z* for the robustness analysis are basically controlling for the ownership structure, for the degree of competition faced by the firm and its market orientation, for the region where it is located and for the economic cycle.

The possible endogeneity problems in labour, capital and materials that appear in production functions estimations are avoided when calculating a TFP index and using input prices instead of estimating their returns to calculate the participation of each input in the production function. Endogeneity problems associated to the demand of labour, capital and intermediate inputs when estimating a production functions are well known: the demands of inputs are not only determining firms' productivity, but they also depend on the productivity they obtain. Then, the residual is correlated with the part of the inputs that is endogenously determined, producing biased coefficients for the inputs and thus an inconsistent estimation of the production function parameters. Also, if some relevant variable is omitted in the estimation of the production function and this variable is also relevant determining the demand of inputs, the error term in the production function and the demand of inputs will be correlated, producing biased coefficients.[73]

However, innovative activity and human capital may suffer the same limitation. Thus, it would be appropriate to find some variable correlated with these variables but uncorrelated with the residual, so that it could be used as an instrument, but the

[72] The variable on firms' size controls for the existence of a possible scale economies effect (for which we are not controlling in the TFP index itself), the effect of firm size on TFP as well as other effects associated with size that are not controlled by the other variables in the equation.

[73] Olley and Pakes (1996) or Doraszelski and Jaumandreu (2006) among others have developed different methodologies to deal with endogeneity when estimating production functions.

literature highlights the absence of appropriate instruments to approximate these variables. Often, lags of the variables themselves are introduced in the regressions to reduce endogeneity problems, although the high persistence of the variables may render this method ineffective.[74] We have introduced the innovative activity and the percentage of skilled workers in the empirical specification with one lag.

Another source of bias is the existence of high heterogeneity among firms. As we argued in Section 3.2.1, even firms that share similar characteristics may present high heterogeneity in TFP. We recognize the existence of unobservable factors that determine TFP but which escape our control. If there are unobserved firm-specific effects, the simple pooled regression may produce biased and inconsistent estimates. To deal with this problem we estimate a random effects model, which assumes that the individual heterogeneity is part of a compound error term and that it is uncorrelated with the regressors. In the case of micro-databases, where firms in the sample are selected randomly from a larger population, it is quite common to estimate a random effects model, rather than a fixed effects model.[75] In addition, notice that we also control for specific effects as for instance region and sector. Finally, to take possible heterosckedasticity problems into account, we estimate robust standard errors in both cases. The obtained coefficients will be used in the methods explained in Section 3.6.1 to decompose the productivity gap between small and large firms.

3.4. Variables and Descriptive Analysis

3.4.1. Description of the variables

In this section we define the variables used in the empirical model of Part II. The TFP index is calculated as explained in Part I. Basically, it is measured using the index suggested by Good et al. (1996) in logs. Its most relevant properties are transitivity (it permits cross-section comparisons) and it approximates the most adequate technology available at any time period.

[74] Hall and Mairesse (1995) explain the likely endogeneity of the R&D stocks in the production function and Sianesi and Van Reenen (2003) explain the endogeneity of human capital accumulation in the economic growth context. They suggest using lags of these variables as instruments, however, they may have long delayed effects, and thus lags will not be good instruments.

[75] For example, Groot and Maassen van den Brink (2003) estimate a random effects probit model to analyse the frequency of training in Dutch firms. Barrios et al. (2003), Máñez et al. (2004), Licandro et al. (2004) among others also estimate a random effects model using the ESEE.

The variables of interest, firm size, innovative activity and human capital, have already been defined in detail in Section 2.4. The firm size is defined as the log of the total number of employees. The innovative activity of the firm is defined through a measure of innovative output. Concretely, it is a dicothomic variable that takes value 1 if the firm has obtained a process innovation. Human capital is measured in terms of formal education of the labour force. This variable is defined as the proportion of qualified workers according to their education level. The category of qualified workers includes: engineers, graduates, middle level engineers, experts and qualified assistants.

As for the control variables, they are defined as follows:

- Firms' age is the number of years since the constitution of the firm.
- The sector of the firm is defined through a set of 20 dummy variables according to the National Classification of Economic Activities (NACE-93). The omitted category is "Other manufacturing industries".
- We also include time dummies.

The variables included for the robustness analysis are defined as follows:

- The productive capacity used by the firm is a question directly asked to firms in the survey.
- As variables related to the ownership structure of the firm we introduce the proportion of foreign-owned capital of the firm, the proportion of publicly-owned capital of the firm and a dummy on whether the firm belongs to a group of firms.
- To approximate the competition faced by the firm, we include a set of dummy variables on the geographical scope of the firm's main market. It considers whether the market is local, provincial (NUTS III), regional (NUTS II), national, international and a category that includes all the previous categories, which is the omitted category.
- The exports are measured as the log of the value of exports expressed in constants pesetas of 1990.
- The region of the firm is a set of 17 dummy variables for the NUTS II regions. The omitted category is "La Rioja".

3.4.2. Descriptive analysis

As stated in Section 2.5.2, the TFP of Spanish manufacturing firms has increased between 1990 and 2002 although these increases are smaller during the second half of the nineties. The TFP growth is not homogeneous over the whole distribution of firms, being more intense for the most productive firms.

The analysis in Section 2.5.3 confirms that TFP in large firms is higher than in small ones, with differences being statistically significant on every year. However, the TFP gap between small and large firms is heterogeneous: differences in TFP are more severe for firms at the lowest part of the distribution and they diminish as we move up to higher TFP levels. In other words, the group of most productive small firms are as productive as the most productive large firms. As time goes by, differences in TFP tend to reduce due to a higher pace of productivity growth in small firms during the second half of the nineties. This reduction is more important in the central and in the very top values. Finally, the dispersion in both distributions increases over time and so heterogeneity becomes more and more important.

In Section 2.5.4, we state that large firms are more innovative than small ones in terms of process innovations and employ a more qualified labour force. In that section, we also mention that different studies at aggregate level report a notorious improvement in human and technological capital, although Spain is still far from the average EU in terms of technological level.

Tables 3.1 and 3.2 show a descriptive analysis of the variables on process innovations and the proportion of white collars that complements the descriptive in Tables 2.9 and 2.12. As in Chapter 2, this analysis corresponds to years 1993, 1997 and 2001 because these variables are lagged in our empirical specification. As before, we obtain that firms in the Spanish manufacturing industry increase their percentage of skilled workers over time, but they do not report increasing their innovative output between 1993 and 2001.

Focusing on differences by firm size, Tables 3.1 and 3.2 confirm our reasoning that large firms are associated with innovating more and having a more qualified labour force. Table 3.1 shows that the proportion of large innovative firms almost doubles the proportion of small innovative firms. The test of equality of proportions rejects the null that small and large firms report doing innovations in the same proportion. This table also shows the quantiles of the distribution of process innovations. Notice that small firms in the two lower quartiles do not report having innovated, while it only happens in the first quartile in the case of large firms. For the most innovative firms, the differences between small and large are larger than in the average.

Table 3.2 shows that large firms report having a higher proportion of skilled labour force than small firms (0.1 in 1993 and increasing to 0.12 in 2001, while in the case of small firms it increases from 0.07 to 0.9). The t-test of equality of means rejects the null that in average they have the same ratio of white collars. Notice also that small firms in the first quartile do not report having qualified workers. Differences in innovations and skilled labour endowments between small and large firms remain quite stable over time, which is reflected later in the decomposition of the TFP gap. This descriptive analysis shows basically that large Spanish manufacturing firms innovate and invest more in human capital than their smaller counterparts.

Table 3.1. Proportion of firms reporting process innovations

	1993			1997			2001		
	Total sample	**Small**	**Large**	**Total sample**	**Small**	**Large**	**Total sample**	**Small**	**Large**
Mean	0.3545	0.2834	0.5455	0.344	0.2992	0.5097	0.3032	0.2537	0.4839
Var	0.2291	0.2034	0.249	0.2259	0.21	0.2511	0.2115	0.1896	0.2511
Test eq prop		7.1081***			5.6427***			6.0499***	
0-25%	0	0	0	0	0	0	0	0	0
25-50%	0	0	0.0870	0	0	0.0194	0	0	0
50-75%	0.1393	0.0430	0.3931	0.1253	0.0648	0.3442	0.0710	0.0039	0.3094
No of firms	852	621	231	968	762	206	864	678	186

Note: test of equality of proportions: (***) denotes significant at 1%.

Table 3.2. Proportion of skilled workers over the total workers for the Spanish manufacturing firms

	1993			1997			2001		
	Total sample	**Small**	**Large**	**Total sample**	**Small**	**Large**	**Total sample**	**Small**	**Large**
Mean	0.0799	0.0704	0.1053	0.0908	0.0819	0.1236	0.0974	0.0896	0.1258
Var	0.0088	0.0080	0.0102	0.0123	0.0114	0.0144	0.0143	0.0145	0.0122
Test eq mean		4.6147***			4.5297***			3.8759***	
Percentiles									
25%	0	0	0.0414	0	0	0.052	0.0156	0	0.0548
50%	0.0556	0.0487	0.0734	0.0646	0.0538	0.0867	0.0685	0.0584	0.0909
75%	0.1082	0.0976	0.1317	0.1281	0.1211	0.152	0.1332	0.124	0.1582
No of firms	852	621	231	968	762	206	864	678	186

Note: test of equality of mean: (***) denotes significant at 1%.

In Section 2.5.4. we obtained large firms are more productive. We also found that process innovators and firms that employ a high proportion of qualified workers are more productive. Moreover, we obtained that the gains in productivity derived from obtaining new processes are significant in the case of small firms but not in the case of large firms. Then, innovation can be seen as a key element for small firms to improve their productivity. This result implies that innovation does not only have a direct effect increasing firms' productivity, but its effect is also conditioned by firm size (innovation is associated with different gains in productivity for small and large firms). In that section we also obtained that large firms employing a high proportion of qualified workers are more productive than small ones. And a similar result is obtained for firms that employ a low proportion of white collars. However, the TFP gap between small and large firms is not much different between firms that employ a high proportion of qualified workers and the ones that do not. That is, once we control for the fact of having high endowment of human capital, the TFP gap does not reduce. This could suggest that this gap may be due to something else than the different levels in the use of white collars. We argue that differences in TFP between small and large firms may indicate that these two groups obtain different returns from their investment in human capital.

Table 3.3 offers the average TFP for innovative and non-innovative firms by their use of qualified workers, considering small and large firms separately. Basically, it complements the results in Section 2.5.4 but it adds to the previous results in the sense that it permits controlling for innovation and human capital. Moreover, it defines four intervals for the variable on the proportion of white collars instead of two so that further information on the distribution of this variable is available.

After controlling for firms' level of qualified workers, this table confirms in general terms the previous result that innovative firms are more productive than non-innovative. However, changing from being a non-innovative firm to an innovative one is associated to higher TFP increases in the case of small than in the case of large firms, corroborating the results in Section 2.5.4. This result suggests the possibility of an indirect effect of innovation on productivity through firm size for different levels of use of human capital. Table 3.3 also confirms the result that those firms that make a more intense use of white collars are more productive, after controlling for the innovative activity. When changing from a given interval in the distribution of white collars to another, TFP increases in most cases.

The TFP differential between innovative and non-innovative firms is smaller than the TFP increase when changing from a proportion of white collars in the first to the fourth quartile. This result suggests that TFP increases are more associated with an increase in human capital than with firms innovative activity.

When comparing small innovative and large innovative firms with a similar proportion of qualified workers (same interval), we find that large firms are more productive than their smaller counterparts. Large non-innovative firms are also more productive than small non-innovative ones in the same interval. All in all, after conditioning for innovative activity and human capital, large firms appear to be more productive. Among other reasons, we could think that innovation and human capital may not only have a direct relation with productivity (firms that innovate and those that use qualified workers are more productive) but they also could have an indirect effect through firm size (the increases in productivity when innovating or increasing the proportion of white collars are different for small and large firms). This is to be checked in the nect Section

Table 3.3. TFP index for the Spanish manufacturing firms by innovative activity, human capital and size

		Small firms				Large firms			
		1st HK	2nd HK	3rd HK	4rt HK	1st HK	2nd HK	3rd HK	4rt HK
Innovative firms	1993	0.0164	0.0090	0.0442	0.0828	0.0375	0.0708	0.1378	0.1011
	1997	0.0539	0.0900	0.1279	0.1846	0.0949	0.1052	0.1711	0.2721
	2001	-0.0014	0.0952	0.1604	0.2005	0.1136	0.1739	0.1619	0.2465
Non-innovative firms	1993	-0.0716	-0.0175	-0.0120	0.0621	-0.0109	-0.0148	0.1199	0.1709
	1997	-0.0582	0.0390	0.0982	0.1036	0.0032	0.1503	0.1457	0.1325
	2001	0.0355	0.1064	0.1170	0.1427	0.0856	0.1755	0.1567	0.2173

Note: 1st, 2nd, 3rd and 4th HK are the observations corresponding to the different quartiles of the distribution of the proportion of skilled workers, which are used to define four intervals of human capital.

3.5. Estimation

As a first step in our analysis, we estimate the empirical specification in (3.1a) for the total sample of firms and for the small and large subsamples separately. Results for the pooled OLS and random effects estimation for this specification are summarized in columns 1 to 3 of Tables 3.4 and 3.5 respectively. A robustness analysis is performed by including additional control variables as described above. Results are summarized in columns 4 to 6 under the label of Specification (3.1b).[76]

3.5.1. The ordinary least squares estimation

In the OLS estimation for the total sample (Table 3.4), the coefficients of our two variables of interest, innovative activity and workers' qualification, are positive and significant at 1%. This suggests that the knowledge capital acquired by a firm improves the mechanism by which inputs are transformed into output. Process innovations reduce the unitary cost of production, and then productivity increases. However the effect seems to be modest: changing from being a non-innovative to innovative firm increases the TFP by 3%. The positive and significant coefficient for human capital confirms that a higher qualification of the labour force increases productivity because workers can do any task that requires something else than simple workforce in a more efficient manner. Actually, a 10 points increase in the ratio of skilled workers, increases TFP of the average Spanish manufacturing firms a 2.1%. As for the control variables, the coefficients of firms' size and age are significant and with the expected sign (although the coefficient of firms' age is very small in magnitude). The coefficients for the sets of dummies on sectors and years are jointly significant. Actually, differences in TFP levels are usually found to be strongly related to the industry in which firms operate.

The last set of columns in Table 3.4 shows that the estimates of the effects of innovation and human capital are quite robust to the inclusion of additional control variables. The major change is observed for the returns to human capital, which decreases its magnitude around one third. The coefficient of the control variable on firms' size is significant in specification (3.1a), but not in specification (3.1b). This indicates that, after controlling for all the additional variables in the robustness analysis, the size effect disappears. The coefficient of firms' age and the sector and year dummies remain significant. Even after controlling for all these variables, innovation and skilled labour still remain positive and significant, suggesting that our results are robust to various specifications.

[76] Tables 3.4 and 3.5 show the most relevant results of the estimations. For more detailed results, see Tables A3.1 and A3.2 at the Appendix 3.1.

Table 3.4. Results for the OLS estimation. Dependent variable: lnTFP

	Specification (3.1a)			Specification (3.1b)		
	Total	**Small**	**Large**	**Total**	**Small**	**Large**
Innovation	0.0344***	0.0355***	0.0374***	0.0321***	0.0306***	0.0389***
	(0.0092)	(0.0113)	(0.0158)	(0.0091)	(0.0112)	(0.0159)
Qualified workers	0.2112***	0.1793***	0.3028***	0.1418***	0.1068**	0.2333***
	(0.0458)	(0.0532)	(0.0897)	(0.0462)	(0.0543)	(0.0927)
Controls						
Size	0.0085**	0.0115*	0.0025	-0.0055	-0.0036	-0.0061
	(0.0035)	(0.0063)	(0.0133)	(0.0049)	(0.0075)	(0.0137)
Age	0.0010***	0.0013***	0.0006*	0.0007***	0.0009***	0.0005
	(0.0002)	(0.0003)	(0.0004)	(0.0002)	(0.0003)	(0.0004)
Sector dummies	yes	yes	yes	yes	yes	yes
Year dummies	yes	yes	yes	yes	yes	yes
Robustness analysis						
Productive capacity				0.0637**	0.06*	0.1115**
				(0.0323)	(0.0369)	(0.0585)
Foreign capital				0.0347**	0.0039	0.0515***
				(0.0154)	(0.0263)	(0.0209)
Group				0.0076	0.0306	-0.0158
				(0.0139)	(0.0199)	(0.0206)
Public capital				-0.0401	-0.0747	-0.0434
				(0.0520)	(0.0964)	(0.0655)
Exports				0.0013**	0.0016**	0.0011
				(0.0006)	(0.0007)	(0.0017)
Market dummies				yes	yes	yes
Region dummies				yes	yes	yes
constant	0.0652**	0.0445	0.1652*	0.1261**	0.0888	0.2514**
	(0.0291)	(0.0352)	(0.0949)	(0.0563)	(0.0649)	(0.1265)
No of obs	2684	2061	623	2684	2061	623
R^2 (adj)	0.2053	0.1868	0.2887	0.248	0.2361	0.3488
H_0: $Sector_t=0$	19.89***	14.44***	53.34***	16.12***	12.37***	9.12***
H_0: $Year_t=0$	38.72***	27.05***	13***	34.66***	26.11***	10.46***
H_0: $Market_t=0$				2.23**	1.77	0.98
H_0: $Region_t=0$				5.55***	5.16***	2.89***

Note: robust standard deviation in parentheses; (***), (**) and (*) denote significant at 1%, 5% and 10%.

Since we are interested in investigating whether innovations and human capital play different roles in their contribution to enhance productivity in small and large firms, we estimate our specifications for the subsamples of small and large firms separately. As shown in Table 3.4, the coefficients for innovation are positive and significant at 1% in both cases. More concretely, in specification (3.1a), being an innovative firm increases TFP a 3.5% in small firms, and a 3.7% in large firms. The estimate of this effect is 3% and 3.9% respectively when additional control variables are included. As for human capital, its effect is positive and significant at 1% in both types of firms. But the most interesting feature is that the effect of using human capital seems to be much larger in the case of large firms. Concretely, a 10 points increase in the ratio of skilled workers, increases TFP by 3% in large firms, while it only increases productivity by 1.8% in the case of small firms. A decrease in the effects of human capital is observed for both types of firms when additional control variables are included. Even in such case, the gap in the return to the use of human capital between firms of different groups is quite important (1% for small firms and 2.3% for large firms). The coefficients for control variables on firms' age and the sector and year dummies are significant for small and large firms, while firms' size is only significant for the subsample of small firms.

The general picture is that innovating and having more skilled workers enhance productivity for both small and large firms. Furthermore, the impact of these variables on TFP is more important for large than for small firms. The fact that the impact of human capital differs considerably for the two subsamples points out that small and large firms do not only have different human capital endowments, but they also have different returns to this factor, suggesting differences in their behaviour. Therefore, the incentive to hire qualified workers as a way to increase productivity seems to be stronger in large firms. In the case of innovative activity, the differences in the returns to these endowments are not as large as with human capital.

3.5.2. The random effects model

The empirical evidence highlights the existence of considerable heterogeneity among firms with similar characteristics, thus, including only some of the observed characteristics as regressors may not be sufficient to account for such heterogeneity. The random effects model permits taking unobservable characteristics of the firms into account. This unobservable heterogeneity is considered as a component of the disturbance term. Table 3.5 shows the estimation of specification (3.1a) under a random effects model. As for the total sample, the coefficients of our two variables of interest, innovative activity and workers' qualification, remain positive and significant once we consider the unobservable firm-specific effects. Concretely, a 10 points increase in the ratio of skilled workers, increases TFP a 1.5%.[77] Changing from being a non-innovative to an innovative firm increases TFP by almost 2%.[78] The coefficients of the control variables on firms' size and age are significant and have the expected sign. The coefficients of the sector and year dummies are also jointly significant. More importantly, the LM test for the random effects rejects the null that panel-level variance component is not statistically significant.

The last set of columns in Table 3.5 shows the estimation of specification (3.1b) under a random effects model. Our variables of interest remain positive and significant after controlling for firm-specific heterogeneity and a set of additional control variables, which supports our findings. However the two coefficients are smaller in magnitude than in specification (3.1a). As with the OLS estimation, the coefficient of firms' size is not significant in specification (3.1b). The coefficients for firms' age and the sector and year dummies remain significant. As before, LM test suggests the necessity to control for firm-specific effects.

As we did with the pooled data, we estimate specification (3.1a) for the subsample of small and large firms separately. Now, the coefficient of innovation is not significant for small firms, however it is still positive and significant at 1% in large firms and the magnitude of its effect does not seem to be affected by considering the firm-specific effects. Large innovative firms have TFP levels almost 4% higher than those large firms that do not report having innovated. This effect is quite close to the one obtained from the OLS estimation. The coefficients for human capital are positive and significant for both subsamples and larger in magnitude in the case of large firms. In the case of human capital, the values are also a bit smaller in magnitude than in the OLS estimation. A 10 points increase in the ratio of skilled workers, increases TFP a 1.2% and a 2% in small and large firms respectively. After controlling for firm-specific effects, the coefficient of firms' size is not significant for any subsample. The coefficient of firms' age remains significant and with the expected sign, although, very small in magnitude. The coefficients of the sets of dummies for sector and region are jointly significant.

[77] Our results are smaller than those obtained by Griliches and Regev (1995) for a pooled sample taking differences to control for firm heterogeneity. However they argue that their coefficients are too high, probably due to the omission of other relevant variables.

[78] Our results are quite close to those by Huergo and Jaumandreu (2004a), who find an impact of process innovations on productivity growth around 0.015.

Table 3.5. Results for the random effects estimation. Dependent variable: lnTFP

	Specification (3.1a)			Specification (3.1b)		
	Total	**Small**	**Large**	**Total**	**Small**	**Large**
Innovation	0.0190**	0.0132	0.0379***	0.0162**	0.0102	0.0355***
	(0.0081)	(0.0101)	(0.0134)	(0.0080)	(0.0099)	(0.0129)
Qualified workers	0.1475***	0.1203***	0.2051***	0.0912**	0.0569	0.1626*
	(0.0430)	(0.0489)	(0.0891)	(0.0424)	(0.0494)	(0.0865)
Controls						
Size	0.0090**	0.0063	-0.0032	-0.0050	-0.0102	-0.0065
	(0.0040)	(0.0070)	(0.0155)	(0.0053)	(0.0081)	(0.0156)
Age	0.0009***	0.0009***	0.0009**	0.0006**	0.0005	0.0008**
	(0.0003)	(0.0004)	(0.0004)	(0.0003)	(0.0004)	(0.0004)
Sector dummies	yes	yes	yes	Yes	yes	yes
Year dummies	yes	yes	yes	yes	yes	yes
Robustness analysis						
Productive capacity				0.1159***	0.1224***	0.1166**
				(0.0312)	(0.0360)	(0.0534)
Foreign capital				0.0506***	0.0391	0.0518***
				(0.0154)	(0.0278)	(0.0190)
Group				-0.0101	0.0137	-0.0373**
				(0.0130)	(0.0192)	(0.0184)
Public capital				-0.0866*	-0.1546	-0.0897
				(0.0486)	(0.0761)	(0.0625)
Exports				0.0014**	0.0016***	0.0010
				(0.0006)	(0.0007)	(0.0013)
Market dummies				Yes	yes	yes
Region dummies				yes	yes	yes
constant	0.0686*	0.0726	0.1572	0.0782	0.0644	0.2412
	(0.0374)	(0.0442)	(0.1156)	(0.0713)	(0.0817)	(0.1518)
No of obs	2684	2061	623	2684	2061	623
No of firms	1585	1211	415	1585	1211	415
H_0: Random effects$_i$=0	616.81***	476.95***	70.85***	571.58***	423.35***	58.99***
H_0: Sector$_t$=0	254.02***	185.27***	984.9***	205.54***	156.51***	145.11***
H_0: Year$_t$=0	134.41***	91.93***	44.05***	115.77***	81.86***	30.59***
H_0: Market$_t$=0				12.05**	10.8**	5.46
H_0: Region$_t$=0				62.54***	50.67***	33.98***

Note: robust standard deviation in parentheses; (***), (**) and (*) denote significant at 1%, 5% and 10%.

As we did with the pooled data, we estimate specification (3.1a) for the subsample of small and large firms separately. Now, the coefficient of innovation is not significant for small firms, however it is still positive and significant at 1% in large firms and the magnitude of its effect does not seem to be affected by considering the firm-specific effects. Large innovative firms have TFP levels almost 4% higher than those large firms that do not report having innovated. This effect is quite close to the one obtained from the OLS estimation. The coefficients for human capital are positive and significant for both subsamples and larger in magnitude in the case of large firms. In the case of human capital, the values are also a bit smaller in magnitude than in the OLS estimation. A 10 points increase in the ratio of skilled workers, increases TFP a 1.2% and a 2% in small and large firms respectively. After controlling for firm-specific effects, the coefficient of firms' size is not significant for any subsample. The coefficient of firms' age remains significant and with the expected sign, although, very small in magnitude. The coefficients of the sets of dummies for sector and region are jointly significant.

Results for the random effects estimation of specification (3.1b) are shown in the last set of columns in Table 3.5. The magnitude of the coefficients is a bit smaller than in the OLS estimation. After the inclusion of additional control variables, the coefficient of human capital is not significant in the case of small firms, while it is still positive and significant in the case of large firms. In specification (3.1b), a 10 points increase in the ratio of skilled workers, increases TFP a 0.6% and 1.6% in small and large firms respectively. The coefficient of innovation is not significant for small firms, however it is still positive and significant at 1% in large, although a bit smaller in magnitude than for specification (3.1a). Firms' size is not significant, age is only significant for the large firms' subsample and the sector and year dummies are jointly significant.

The general result is that after taking firm heterogeneity into account, innovating and having more skilled workers enhance productivity for large firms. However, in the case of small firms, innovation is not significant, and human capital is only significant in specification (3.1a). The fact that these variables are not significant in specification (3.2b) may be due to possible collinearity between the additional control variables and human capital. In that case, the control variables may capture the effects of human capital. In all the cases the LM test for the random effects rejects the null that panel-level variance component is not statistically significant.

To summarize, both innovation and human capital seem to play a role in enhancing firms' productivity, thought the evidence suggests that the magnitude of their effect is dramatically related to firm size. Actually, after controlling for a large set of conditioning variables and accounting for firm heterogeneity, the effect of the innovation and human capital is only marginal and it is not statistically significant in the case of the group of small firms. This suggests the existence of a threshold in the effect of innovation and human capital related to the dimension of the firm.

Thus, small and large firms show different patterns of behaviour in relation to the knowledge variables: they do not only have different endowments, but the returns to these endowments are quite different between them. In this framework we argue that TFP differences between small and large firms are associated to differences in firms' endowments as well as differences in the returns to these endowments. In the next section we analyse the relative contribution of these effects in explaining the productivity differences by firm size.

3.6. Differences in Total Factor Productivity by Firm Size Evaluated in the Mean of the Distribution

3.6.1. The Oaxaca-Blinder decomposition

The Oaxaca-Blinder decomposition methodology has widely been used to study wage gaps associated to differences in workers characteristics and to discrimination by gender or race. Following Oaxaca (1973) and Blinder (1973), the differential between the average wages between two groups, for example men and women, in period *t* can be decomposed into an explained or predicted difference due to disparities in observed or measured characteristics between the two groups, and an unexplained or residual difference attributable to both wage discrimination and unmeasured disparities in characteristics.

We apply the Oaxaca-Blinder decomposition to analyse differences in TFP between small and large firms.[79] This methodology permits analysing how much of the TFP differential between small and large firms can be explained by either, differences in firms' *endowment* of human and technological capital, and differences in the *returns* to these endowments. The standard methodology decomposes the TFP differential between small and large firms departing from the coefficients estimated in auxiliary regressions for each type of firms specified as in (3.1a). From such regressions, the average TFP in the sample of small and large firms is obtained as:

[79] To the best of our knowledge, Smith et al. (2004) is the only paper that uses this decomposition to analyse differences between firms. More concretely, they compare firms that make an R&D effort and firms that do not.

$$\overline{T\hat{F}P_S}' = \overline{X}'_S \hat{\beta}_S$$
$$\overline{T\hat{F}P_L} = \overline{X}'_L \hat{\beta}_L \tag{3.2}$$

where $\overline{T\hat{F}P}$ denotes the mean of the logarithm of total factor productivity,[80,81] $\overline{X}$ is the vector of the mean values of the regressors in specification (3.1a) or (3.1b), $\hat{\beta}$ is the conforming vector of estimated coefficients and subscripts L and S refer to large and small firms respectively. Then, the differences in TFP between small and large firms can be decomposed as:

$$\overline{T\hat{F}P_L} - \overline{T\hat{F}P_S} = \left(\overline{X}_L' - \overline{X}_S'\right)\hat{\beta}_L \quad + \quad \overline{X}_S'\left(\hat{\beta}_L - \hat{\beta}_S\right) \tag{3.3}$$

where the first term on the right hand side is the part of the TFP gap due to differences in characteristics between the representative small and large firms and the second term on the right hand side is the contribution of differences in returns between both types of firms.

The first term on the right hand side in expression (3.3), assumes that all the firms have the returns of large firms, $\hat{\beta}_L$. The second term, assumes that all firms have the endowments of small firms, $\overline{X}_S$. However we could write a symmetric equation where these values were replaced by $\hat{\beta}_S$ and $\overline{X}_L$ respectively. The standard version of the Oaxaca-Blinder decomposition builds on the assumption that one of the two equations is the "natural" model. For example, in the case of wage differentials by gender, it may appear quite natural to impose that women are the "discriminated" group: we would impose that women have the same returns as men in the first term, and the second term could be interpreted as discrimination by gender. In our case there is no compelling reason to calculate the differences in firms' endowments assuming that all the firms had the returns of either large or small firms. A strand of literature considers that it is not always easy to establish which the natural model is and often results may differ considerably. This literature suggests a variation of the standard decomposition that avoids assuming which is the natural model. According to this perspective, there exists a "non-discriminatory structure of returns" in relation to which one group is "discriminated" while the other is "favoured". The TFP differential without assuming that any of the two equations is the natural model can be expressed as:

[80] Notice that the OLS estimation method guarantees that $\overline{T\hat{F}P} = \overline{T\hat{F}P}$ as the average of the errors is zero. That is, the decomposition is exact.
[81] In what follows, we use TFP to denote lnTFP calculated as in expression (1.12).

$$\overline{TFP}_L - \overline{TFP}_S = (\bar{X}_L' - \bar{X}_S')\hat{\beta}^* + \bar{X}_L'(\hat{\beta}_L - \hat{\beta}^*) + \bar{X}_S'(\hat{\beta}^* - \hat{\beta}_S) \tag{3.4}$$

where $\hat{\beta}^*$ is the estimated nondiscriminatory returns structure. The first term on the right hand-side of (3.4) is an estimate of the productivity differential in absence of differences in returns between small and large firms, reflecting productivity differences due to differences in firms' endowments. The second and third terms are estimates of the large firms' advantage and small firms' disadvantage in relation to the non-discriminatory returns structure. The two terms together are considered differences in TFP by firm size associated with differences in returns without imposing a discriminated group. Since we are not interested in distinguishing the advantage and disadvantage effects, but in evaluating the differences in returns as a whole without imposing a discriminated returns structure, here we report these two terms together. To implement this decomposition, one has to make an assumption on what the non-discriminatory returns structure would be. Oaxaca and Ransom (1994) deal with a proper choice of the non-discriminatory structure and propose estimating it as a weighted average of the two returns structure, $\hat{\beta}^* = \Omega\hat{\beta}_L + (I - \Omega)\hat{\beta}_S$, where the weights Ω are calculated as $(X'X)^{-1}(X'_L X'_L)$ $(X'X)^{-1}(X'_L X'_L)$, X being the matrix of regressors for the entire sample of firms.[82]

3.6.2. Results of the Oaxaca-Blinder decomposition

The Oaxaca-Blinder decomposition permits analysing the relative contribution of firm characteristics and returns to these characteristics to explain the differential in productivity between small and large firms. Moreover, it allows decomposing the individual effect of each variable. The individual decomposition is especially useful in our case, since we are interested in the effect of human and technological capital.

When more than one set of dummy variables is included in the empirical specification, as in our case, their individual contribution to the productivity differential is not identified. Gardeazábal and Ugidos (2004) suggest a transformation of the dummy variables that permits analysing each set of dummies individually and which makes the results invariant to the omitted category of the dummies. In absence of this transformation it is not possible to distinguish the part of the effect due to a set of

[82] It can be easily proved that a consistent estimate of β^* can be obtained by OLS in the whole sample of firms (see Oaxaca and Ransom, 1994).

dummies from another. Given that innovation is defined as a dummy variable, this transformation will be much useful in this case. The transformation basically consists of imposing the constraint that the summation of all the coefficients for a given set of dummies is equal to one. This constraint can be interpreted as a normalizing constraint on the coefficients of the dummy variables. Then, one has to express all the categories in a set of dummies in differences with respect to the left-out category.

Tables 3.6 and 3.7 show the results for the decomposition as suggested by Oaxaca and Ransom (1994) for 1994, 1998 and 2002 based on the OLS estimation with the identification constraint.[83] The TFP differential between small and large firms takes values around 0.08 in 1994 and decreases until 0.06 in 2002. Table 3.6 shows the results for the decomposition based on specification (3.1a). The decomposition for all the variables together shows that almost 100% of the differences are explained by differences in firms' endowments. More concretely, in 1994, differences in endowments explain about 97% of the differential and differences in returns, 3%. Between 1994 and 2002, the contribution of firms' endowments increases, reaching a 109% in 2002. The contribution of differences in endowments higher than 100% implies that differences in returns have the opposite sign. In aggregate, the slight reduction in the TFP gap between small and large firms can be attributed to higher returns that favour small firms.

Table 3.6. TFP gap decomposition. OLS estimation. Specification (3.1a)

	1994		1998		2002	
$\overline{TFP_L} - \overline{TFP_S}$	**0.0794**		**0.0766**		**0.0603**	
	Charact	Returns	Charact	Returns	Charact	Returns
Total	0.0775	0.0018	0.0799	-0.0033	0.0654	-0.0052
	97.69%	2.31%	104.33%	-4.33%	108.60%	-8.60%
Innovation	0.0090	0.0004	0.0072	0.0003	0.0079	0.0002
	11.36%	0.48%	9.45%	0.33%	13.14%	0.38%
Qualified workers	0.0074	0.0119	0.0088	0.0139	0.0076	0.0144
	9.28%	14.99%	11.50%	18.19%	12.68%	23.86%
Innovation and	0.0164	0.0123	0.0160	0.0142	0.0156	0.0146
qualified workers	20.64%	15.46%	20.95%	18.52%	25.82%	24.24%

But this general result is an aggregation of the individual effects of all the variables in the specification and so positive and negative contributions might somehow

[83] Tables 3.6 and 3.7 show the most relevant results of the decomposition. For complete results, see Tables A3.3 and A3.4 at the Appendix 3.2.

be compensated. We are especially interested in the contribution of innovation and human capital. These two variables together, including differences in endowments and differences in returns, explain a great part of the differential in productivity in our three periods. In 1994, the two variables explain 36% of the TFP gap. Their contribution increases over time and they explain about 50% of the gap in 2002. In 1994, innovation individually explains about 12% of the differential, and it is mostly due to the fact that large firms innovate more than small firms (differences in characteristics). In 1998, the contribution of innovation decreases slightly but, in 2002, innovation explains about 13% of the differential (differences in characteristics).

The contribution of the differences associated to the ratio of skilled workers is even higher and increasing over time: it ranges from 24% in 1994, 30% in 1998 to 36% in 2002. In the case of this variable, both differences in endowments and differences in returns contribute to explain the TFP gap. Interestingly, differences in returns to human capital between small and large firms explain a more important part of the TFP gap than differences in the amount used of this factor. And the contribution of differences in returns increases over the period up to almost one quarter of the TFP gap in 2002.

Table 3.7 shows results of the decomposition for the OLS estimation of specification (3.1b). Results are quite similar to those in Table 3.6. The major difference between the two tables has to do with the lower estimated contribution of human capital, and particularly, with the smaller contribution of the differences in characteristics effect. This reduction is due to the decrease in the coefficients when including additional control variables. The contribution of differences in returns in the case of innovation remains very small in magnitude, although with negative sign. Nevertheless, the individual and joint effect of our knowledge variables remains clearly important and increasing over the period.

As we argued in Section 3.3, firms present a high degree of heterogeneity and, to account for it, we estimate the TFP equations including random effects. However, the Oaxaca-Blinder decomposition is not exact for the random effects model (GLS estimation) as the mean of the error term is different from zero, causing that $\overline{\hat{TFP}} \neq \overline{TFP}$.[84] Considering the previous results regarding the sensitivity of the

[84] In the RE model the transformed residuals have zero mean, but not the residuals from the original specification. This prevents obtaining an exact decomposition of the TFP gap based on the RE estimates of the coefficients.

coefficients in the TFP equation to observed and unobserved firm heterogeneity and although only the coefficients based on the OLS estimation provide an exact decomposition, we next offer the results of the decomposition based on the random effects estimation.

Table 3.7. TFP gap decomposition. OLS estimation. Specification (3.1b)

	1994		1998		2002	
$TFP_L - TFP_S$	**0.0794**		**0.0766**		**0.0603**	
	Charact	returns	Charact	Returns	Charact	Returns
Total	0.0752	0.0041	0.0804	-0.0038	0.0690	-0.0087
	94.79%	5.21%	104.92%	-4.92%	114.49%	-14.49%
Innovation	0.0084	-0.0002	0.0068	-0.0001	0.0074	-0.0005
	10.60%	-0.03%	8.82%	-0.31%	12.27%	-0.81%
Qualified workers	0.0050	0.0121	0.0059	0.0142	0.0051	0.0146
	6.23%	15.25%	7.72%	18.50%	8.52%	24.30%
Innovation and	0.0134	0.0121	0.0127	0.0139	0.0125	0.0142
qualified workers	16.84%	15.22%	16.54%	18.19%	20.78%	23.49%

Tables 3.8 and 3.9 show the decomposition for the random effects model for specifications (3.1a) and (3.1b) respectively.[85] In general, the results of the decomposition using the pooled dataset (Tables 3.6 and 3.7) and using the random effects model are quite similar.

The first row in Table 3.8 shows the observed TFP gap between small and large firms. In agreement with the lower magnitude of the coefficients based on the RE estimation, the contribution of the knowledge variables is in this case is slightly lower. However, their impact on the TFP gap seems to be far from negligible, and it increases over time (22.31% in 1994, 24.16% in 1998, and 29.29% in 2002). The major difference has to do with the lower contribution of innovation, which is caused by a lower impact of differences in its endowment (as the RE estimation assigns a lower weight —estimated coefficient— to differences in innovation between large and small firms). As for human capital, the inclusion of unobserved firm heterogeneity seems to maintain the main conclusions about this factor: this variable alone is able to explain a

[85] Tables 3.8 and 3.9 show the most relevant results of the Oaxaca-Blinder decomposition based on the random effects model. For more detailed and complete results, see Tables A3.5 and A3.6 at the Appendix 3.2.

large portion of the TFP gap, its contribution increases over time and it is mostly caused by the different returns observed in large and small firms.

Table 3.8. TFP gap decomposition. Random effects estimation. Specification (3.1a)

	1994		1998		2002	
$\overline{TFP_L} - \overline{TFP_S}$	**0.0794**		**0.0766**		**0.0603**	
	Charact	Returns	Charact	Returns	Charact	Returns
Total	0.0722	0.0105	0.0755	0.0240	0.0597	-0.0095
	90.94%	13.17%	98.63%	31.35%	99.04%	-15.79%
Innovation	0.0050	-0.0004	0.0040	-0.0010	0.0044	-0.0017
	6.27%	-0.51%	5.21%	-1.29%	7.25%	-2.89%
Qualified workers	0.0051	0.0080	0.0061	0.0094	0.0053	0.0097
	6.48%	10.06%	8.03%	12.21%	8.86%	16.07%
Innovation and	0.0101	0.0076	0.0101	0.0084	0.0097	0.0079
qualified workers	12.75%	9.56%	13.24%	10.92%	16.10%	13.19%

Note: given that the decomposition is not exact in the case of using the RE estimates, the sum of the shares of the components does not equal 100%.

Table 3.9. TFP gap decomposition. Random effects estimation. Specification (3.1b)

	1994		1998		2002	
$\overline{TFP_L} - \overline{TFP_S}$	**0.0794**		**0.0766**		**0.0603**	
	Charact	Returns	Charact	Returns	Charact	Returns
Total	0.0708	0.0111	0.0781	0.0196	0.0637	-0.0117
	89.24%	13.96%	101.94%	25.64%	105.77%	-19.34%
Innovation	0.0043	-0.0004	0.0034	-0.0010	0.0037	-0.0018
	5.36%	-0.54%	4.46%	-1.33%	6.20%	-2.98%
Qualified workers	0.0032	0.0099	0.0038	0.0116	0.0033	0.0120
	4.01%	12.52%	4.97%	15.18%	5.48%	20.00%
Innovation and	0.0074	0.0095	0.0072	0.0106	0.0070	0.0103
qualified workers	9.37%	11.98%	9.43%	13.85%	11.68%	17.02

Note: given that the decomposition is not exact in the case of using the RE estimates, the sum of the shares of the components does not equal 100%.

Similar results are obtained for the decomposition based on random effects after the inclusion of additional control variables (Table 3.9). The contribution of the knowledge variables to explain the TFP gap between small and large firms is quite large in magnitude and increases over time (21.35% in 1994, 23.28% in 1998, and 28.7% in 2002). Notice that their contribution is smaller than in Table 3.8 due to the fact that other variables are included in the empirical specification. As before, when including firm-specific effects the contribution of innovation is smaller and it is caused by the smaller effect of differences in characteristics. The results for the contribution of human capital are close to those explained in the previous paragraph.

On the one hand, results in Tables 3.6 and 3.7 are quite similar, but they differ in some aspects with the results in Tables 3.8 and 3.9 as explained above. Taking firm heterogeneity into account provides somehow different results for the estimation and thus for the Oaxaca-Blinder decomposition. The results do not appear to differ so much in relation with the specification, suggesting the robustness of the results for the two variables of interest. Considering all the variables in our empirical specification, the general picture is that the TFP differential between small and large firms is mostly due to differences in characteristics.

The contribution of our variables of interest is smaller under the random effects model than in the OLS estimation. Innovation and human capital together seem to explain between 30% and 50% of the TFP gap for the pooled data, and between 20% and 30% for the random effects model. Innovation individually explains a modest part of the TFP gap: in the case of the OLS, its contribution is a bit more than 10%, whereas in the in the case of the RE estimation, its contribution is a bit less than 10%. The contribution of innovation is mainly due to differences in firms' endowments, although with a small part of the effect due to differences in returns in favour of small firms (except in the case of the decomposition based on OLS for specification (3.1a)). Human capital alone explains quite a large part of the TFP differential and increases over time: as for the decomposition based on pooled data, human capital explains between 21% of the gap in 1994 and 36% in 2002; as for that based on panel data, the contribution of this variable increases between 16% to 25% over this period. Out of these percentages, around 1/3 is due to differences in firms' endowments and 2/3 is due to differences in returns.

All the results are quite similar over time, except the part of the gap due to differences in returns to human capital. The increasing contribution of differences in returns to human capital to the TFP gap has important implications for small firms: they have fewer incentives to hire higher qualified workers because they obtain lower returns, and this increases their TFP gap in relation with large firms.

3.7. Conclusions

In Chapter 3, we analysed the relative contribution of technological and human capital in explaining the average TFP gap between small and large firms, both as differences in the levels of these endowments and differences in returns.

First of all, the descriptive analysis here corroborates the results in Chapter 2: large Spanish manufacturing firms are more productive than small ones; they innovate more and have a more qualified labour force. The results also confirm that the most innovative firms and those with a higher proportion of qualified workers are more productive.

After controlling for a large set of conditioning variables and accounting for firm heterogeneity, both innovation and human capital seem to play a role in enhancing firms' productivity. However, small and large firms follow different patterns of behaviour in relation to innovation and human capital: large firms obtain positive and significant returns to their investments in these factors, which are higher than for small firms. In the case of small firms, the effects of innovation are only marginal and not statistically significant in all our specifications.

Finally, we decomposed the TFP differential between small and large firms in differences in the use of human and technological capital and differences in returns to these factors using the Oaxaca-Blinder methodology. The decomposition for all the variables indicates that differences in characteristics explain most of the TFP gap. But a more detailed analysis reveals that innovation and human capital together explain quite a large part of the TFP differential: innovation individually explains around 10% and it is mostly due to differences in this characteristic; human capital alone explains quite a large part of the differential: ranging between 16% to 36%, out of which around 1/3 is due to differences in the level of workers' qualification in the firm and 2/3 is due to differences in returns to it. Moreover, the contribution of differences in returns to human capital in the TFP gap tends to increase over time.

In Chapter 2, we obtained that the TFP gap between small and large firms is more severe for firms with lower TFP levels, where small firms perform much worse than large firms. Thus, an improvement of TFP for this group of low-productive firms would *ceteris paribus* represent an improvement of TFP for the industry as a whole.

The lower productivity for small firms is associated with the fact that they do not innovate as much as large firms. Actually, those that innovate achieve productivity levels close to large firms. In the present Chapter, we obtained that the returns to innovation are quite similar in small and large firms, and so they have a similar incentive to innovate. In this view, economic policies should focus in increasing the innovative activity for small firms. Given that the returns to innovation are similar in small and large firms, small innovative firms would achieve TFP levels close to large innovative firms, increasing TFP in the industry as a whole.

The lower productivity in small firms is also associated with the fact that they employ less qualified workers than large firms. However, the returns derived from employing qualified workers were larger in the case of large than in small firms. The higher returns to human capital in large firms can be explained by the fact that the costs of communication related to the absorption of new information can be somehow attenuated by specialization, and large firms are more likely to specialize (Bolton and Dewatripont, 1994). This suggests that differences in productivity may derive from something else than the levels of use of qualified workers. Actually, the Oaxaca-Blinder decomposition suggests that a large part of the TFP gap between the two groups is due to the fact that the returns to human capital in small firms are much smaller than in large firms, implying that they have fewer incentives to invest in it. Moreover, this effect tends to increase over time. Thus, policy implication focused on stimulating the more intense use of qualified labour force in small firms would only make sense if these firms improved their returns to human capital, that is, if they could take more advantage of their investment in human capital.

These results add to the previous literature that analyses the role of technological and human capital in improving productivity in placing special emphasis in the role of the returns derived from investments in these two factors. In agreement with the literature that considers that Spain is still far from the average EU in terms of technological capital and that it still has to make an effort in increasing this investment so as to improve productivity, we obtain that, particularly, small firms may play a role and increase their technological levels so that productivity improves for the industry as a whole. In contrast with some studies that consider that human capital is close to the average EU and thus it is not necessary to make a more intense effort, we find that increasing human capital in small firms can improve their productivity and the productivity in the whole industry. In this view we agree with the general recommendation of the National Reform Program that Spain should increase its human capital levels. However, we add to this recommendation in emphasizing that small firms have lower returns to investment in human capital. Thus, increasing the proportion of qualified workers in small firms would only have a positive impact on productivity if the returns increased. Otherwise, the effort on hiring more qualified workers would have a limited impact on small firms' productivity and thus in the industry as a whole.

Appendix 3.1. Estimation of specifications (3.1a) and (3.1b). Complete results

Table A3.1. Results for the OLS. Dependent variable: lnTFP (corresponding to Table 3.4)

	Specification (3.1a)			Specification (3.1b)		
	Total	**Small**	**Large**	**Total**	**Small**	**Large**
Innovation	0.0344***	0.0355***	0.0374***	0.0321***	0.0306***	0.0389***
	(0.0092)	(0.0113)	(0.0158)	(0.0091)	(0.0112)	(0.0159)
Qualified workers	0.2112***	0.1793***	0.3028***	0.1418***	0.1068**	0.2333***
	(0.0458)	(0.0532)	(0.0897)	(0.0462)	(0.0543)	(0.0927)
Controls						
Size	0.0085**	0.0115*	0.0025	-0.0055	-0.0036	-0.0061
	(0.0035)	(0.0063)	(0.0133)	(0.0049)	(0.0075)	(0.0137)
Age	0.0010***	0.0013***	0.0006*	0.0007***	0.0009***	0.0005
	(0.0002)	(0.0003)	(0.0004)	(0.0002)	(0.0003)	(0.0004)
Sector 1	-0.0402	-0.0094	-0.1630**	-0.0263	0.0145	-0.2204***
	(0.0372)	(0.0418)	(0.0735)	(0.0394)	(0.0452)	(0.0756)
Sector 2	-0.1579***	-0.1543***	-0.2109***	-0.1139***	-0.1008***	-0.2809***
	(0.0298)	(0.0346)	(0.0539)	(0.0312)	(0.0360)	(0.0623)
Sector 3	-0.0568	-0.0762	-0.0774	-0.0025	-0.0245	-0.1341*
	(0.0473)	(0.0644)	(0.0635)	(0.0543)	(0.0720)	(0.0791)
Sector 4	-0.1722***	-0.1670***	-0.2157***	-0.16***	-0.1495***	-0.2905***
	(0.0298)	(0.0336)	(0.0577)	(0.0308)	(0.0345)	(0.0620)
Sector 5	-0.1140***	-0.1023***	-0.5212***	-0.1008***	-0.0919***	-0.5274***
	(0.0328)	(0.0353)	(0.0501)	(0.0336)	(0.0365)	(0.0769)
Sector 6	-0.0401	-0.0128	-0.1833***	-0.0113	0.0218	-0.2271***
	(0.0317)	(0.0358)	(0.0585)	(0.0344)	(0.0390)	(0.0675)
Sector 7	-0.0335	-0.0510	-0.0203	-0.0366	-0.0511	-0.1146
	(0.0358)	(0.0397)	(0.0769)	(0.0353)	(0.0393)	(0.0815)
Sector 8	0.0311	0.0438	-0.0355	0.0323	0.0468	-0.0936
	(0.0350)	(0.0397)	(0.0612)	(0.0364)	(0.0411)	(0.0648)
Sector 9	0.0911***	0.0875**	0.0504	0.0834***	0.0879**	-0.0475
	(0.0323)	(0.0380)	(0.0597)	(0.0332)	(0.0391)	(0.0624)
Sector 10	0.0389	0.0381	-0.0027	0.0290	0.0439	-0.1216**
	(0.0301)	(0.0343)	(0.0580)	(0.0310)	(0.0352)	(0.0629)
Sector 11	0.0241	0.0515	-0.0962*	0.0442	0.0814**	-0.1745***
	(0.0326)	(0.0377)	(0.0591)	(0.0344)	(0.0404)	(0.0643)
Sector 12	0.1590***	0.1299***	0.1607***	0.1489***	0.1191***	0.0661
	(0.0361)	(0.0417)	(0.0679)	(0.0379)	(0.0421)	(0.0777)
Sector 13	0.0343	0.0494	-0.0623	0.0369	0.0587*	-0.1460***
	(0.0284)	(0.0319)	(0.0556)	(0.0298)	(0.0337)	(0.0609)
Sector 14	0.0038	0.0275	-0.1006*	-0.0081	0.0117	-0.1790***
	(0.0286)	(0.0323)	(0.0562)	(0.0296)	(0.0338)	(0.0576)
Sector 15	0.0054	0.0292	-0.1027	-0.0097	0.0138	-0.1762*
	(0.0413)	(0.0429)	(0.0945)	(0.0425)	(0.0444)	(0.1024)
Sector 16	0.0115	0.0016	-0.0100	0.0148	0.0224	-0.1330**
	(0.0303)	(0.0350)	(0.0563)	(0.0318)	(0.0370)	(0.0629)
Sector 17	-0.0059	0.0059	-0.0736	-0.0197	0.0058	-0.1895***
	(0.0300)	(0.0366)	(0.0541)	(0.0317)	(0.0385)	(0.0596)
Sector 18	-0.0724*	-0.0589	-0.1414**	-0.0360	-0.0163	-0.1749**
	(0.0395)	(0.0451)	(0.0737)	(0.0410)	(0.0486)	(0.0758)
Sector 19	-0.1646***	-0.1548***	-0.2537***	-0.1507***	-0.1306***	-0.3730***

	(0.0312)	(0.0344)	(0.0656)	(0.0324)	(0.0359)	(0.0741)
Year 94	-0.0979***	-0.0980***	-0.0978	-0.0939***	-0.0974***	-0.0921***
	(0.0112)	(0.0134)	(0.0194)	(0.0113)	(0.0135)	(0.0201)
Year 98	-0.0439	-0.0433***	-0.0470	-0.0465***	-0.0469***	-0.0560***
	(0.0111)	(0.0130)	(0.0205)	(0.0110)	(0.0128)	(0.0205)
Robustness analysis						
Productive capacity				0.0637**	0.06*	0.1115**
				(0.0323)	(0.0369)	(0.0585)
Foreign capital				0.0347**	0.0039	0.0515***
				(0.0154)	(0.0263)	(0.0209)
Group				0.0076	0.0306	-0.0158
				(0.0139)	(0.0199)	(0.0206)
Public capital				-0.0401	-0.0747	-0.0434
				(0.0520)	(0.0964)	(0.0655)
Local market				-0.0662***	-0.0554**	-0.0797
				(0.0219)	(0.0245)	(0.0892)
Province market				-0.0079	0.0014	-0.0196
				(0.0194)	(0.0226)	(0.0531)
Region market				-0.0067	0.0001	-0.0123
				(0.0199)	(0.0232)	(0.0441)
National market				-0.0182	-0.0049	-0.0365**
				(0.0118)	(0.0157)	(0.0192)
International market				-0.0063	0.0152	-0.0408
				(0.0198)	(0.0259)	(0.0335)
Exports				0.0013**	0.0016**	0.0011
				(0.0006)	(0.0007)	(0.0017)
Region 1				-0.0659*	-0.0714	0.0392
				(0.0406)	(0.0449)	(0.0533)
Region 2				-0.0547	-0.0543	0.0170
				(0.0419)	(0.0475)	(0.0578)
Region 3				-0.0857*	-0.0614	-0.0927
				(0.0477)	(0.0553)	(0.0608)
Region 4				-0.0855	-0.0512	-0.2818***
				(0.0543)	(0.0598)	(0.0791)
Region 5				-0.0775	-0.0882	0.0114
				(0.0701)	(0.0808)	(0.1050)
Region 6				-0.0078	-0.0062	0.0052
				(0.0503)	(0.0791)	(0.0602)
Region 7				-0.0689	-0.0425	-0.1097**
				(0.0432)	(0.0481)	(0.0535)
Region 8				-0.0614	-0.0595	-0.0185
				(0.0421)	(0.0493)	(0.0476)
Region 9				-0.0113	0.0117	-0.0330
				(0.0370)	(0.0413)	(0.0434)
Region 10				-0.0776**	-0.0651	-0.0633
				(0.0373)	(0.0416)	(0.0488)
Region 11				-0.3922***	-0.4502***	-0.1671*
				(0.0785)	(0.0913)	(0.0973)
Region 12				-0.1141***	-0.1133***	-0.0935*
				(0.0399)	(0.0450)	(0.0531)
Region 13				-0.0130	-0.0088	0.0208
				(0.0373)	(0.0416)	(0.0474)

Region 14				-0.1188***	-0.1200***	-0.0906
				(0.0448)	(0.0496)	(0.0690)
Region 15				-0.0724	-0.0466	-0.0874
				(0.0459)	(0.0536)	(0.0577)
Region 16				-0.0131	0.0012	-0.0161
				(0.0388)	(0.0441)	(0.0469)
constant	0.0652**	0.0445	0.1652*	0.1261**	0.0888	0.2514**
	(0.0291)	(0.0352)	(0.0949)	(0.0563)	(0.0649)	(0.1265)
H_0: $Sector_t$=0	19.89***	14.44***	53.34***	16.12***	12.37***	9.12***
H_0: $Year_t$=0	38.72***	27.05***	13***	34.66***	26.11***	10.46***
H_0: $Market_t$=0				2.23**	1.77	0.98
H_0: $Region_t$=0				5.55***	5.16***	2.89***
No of obs	2684	2061	623	2684	2061	623
R^2 (adj)	0.2053	0.1868	0.2887	0.248	0.2361	0.3488

Note: robust standard deviation in parentheses; (***), (**) and (*) denote significant at 1%, 5% and 10%.

Table A3.2. Results for the random effects estimation. Dependent variable: lnTFP (corresponding to Table 3.5)

	Specification (3.1a)			Specification (3.1b)		
	Total	**Small**	**Large**	**Total**	**Small**	**Large**
Innovation	0.0190**	0.0132	0.0379***	0.0162**	0.0102	0.0355***
	(0.0081)	(0.0101)	(0.0134)	(0.0080)	(0.0099)	(0.0129)
Qualified workers	0.1475***	0.1203***	0.2051***	0.0912**	0.0569	0.1626*
	(0.0430)	(0.0489)	(0.0891)	(0.0424)	(0.0494)	(0.0865)
Controls						
Size	0.0090**	0.0063	-0.0032	-0.0050	-0.0102	-0.0065
	(0.0040)	(0.0070)	(0.0155)	(0.0053)	(0.0081)	(0.0156)
Age	0.0009***	0.0009***	0.0009**	0.0006**	0.0005	0.0008**
	(0.0003)	(0.0004)	(0.0004)	(0.0003)	(0.0004)	(0.0004)
Sector 1	-0.0290	-0.0049	-0.0978	-0.0037	0.0303	-0.1569
	(0.0457)	(0.0524)	(0.0891)	(0.0481)	(0.0550)	(0.1125)
Sector 2	-0.1493***	-0.1396****	-0.1866***	-0.0988***	-0.0766*	-0.2643***
	(0.0388)	(0.0451)	(0.0703)	(0.0403)	(0.0458)	(0.1002)
Sector 3	-0.0516	-0.0703	-0.0334	0.0131	-0.0037	-0.1078
	(0.0617)	(0.0834)	(0.0822)	(0.0654)	(0.0872)	(0.1159)
Sector 4	-0.1563***	-0.1512***	-0.1778***	-0.1412***	-0.13***	-0.2602***
	(0.0390)	(0.0441)	(0.0758)	(0.0406)	(0.0447)	(0.1022)
Sector 5	-0.1133***	-0.1039**	-0.4931***	-0.0967**	-0.0885*	-0.5129***
	(0.0431)	(0.0466)	(0.0643)	(0.0438)	(0.0472)	(0.1099)
Sector 6	-0.0533	-0.0299	-0.1461*	-0.0143	0.0127	-0.1833*
	(0.0404)	(0.0458)	(0.0777)	(0.0433)	(0.0482)	(0.1073)
Sector 7	-0.0046	-0.0142	0.0358	-0.0045	-0.0103	-0.0523
	(0.0451)	(0.0505)	(0.0967)	(0.0455)	(0.0500)	(0.1204)
Sector 8	0.0490	0.0572	0.0175	0.0590	0.0722	-0.0548
	(0.0434)	(0.0492)	(0.0836)	(0.0450)	(0.0501)	(0.1063)
Sector 9	0.1074***	0.1049**	0.0937	0.1062***	0.11**	-0.0046
	(0.0410)	(0.0487)	(0.0748)	(0.0417)	(0.0486)	(0.1002)
Sector 10	0.0378	0.0436	0.0242	0.0330	0.0523	-0.0904
	(0.0387)	(0.0445)	(0.0756)	(0.0397)	(0.0446)	(0.1026)
Sector 11	0.0257	0.0545	-0.0671	0.0569	0.0969**	-0.1339
	(0.0418)	(0.0487)	(0.0770)	(0.0436)	(0.0502)	(0.1037)
Sector 12	0.1712***	0.1488***	0.1982**	0.1658***	0.1399***	0.1096
	(0.0475)	(0.0541)	(0.0914)	(0.0493)	(0.0537)	(0.1195)
Sector 13	0.0322	0.0485	-0.0337	0.0427	0.0636	-0.1091
	(0.0371)	(0.0420)	(0.0712)	(0.0386)	(0.0428)	(0.0999)
Sector 14	0.0251	0.0486	-0.0635	0.0171	0.0382	-0.1370
	(0.0379)	(0.0431)	(0.0725)	(0.0394)	(0.0441)	(0.0993)
Sector 15	0.0412	0.0598	-0.0246	0.0356	0.0501	-0.1014
	(0.0517)	(0.0547)	(0.1078)	(0.0527)	(0.0558)	(0.1284)
Sector 16	0.0079	-0.0014	0.0015	0.0126	0.0190	-0.1130
	(0.0391)	(0.0452)	(0.0725)	(0.0405)	(0.0461)	(0.1005)
Sector 17	0.0080	0.0143	-0.0292	-0.0005	0.0128	-0.1402
	(0.0388)	(0.0461)	(0.0703)	(0.0406)	(0.0476)	(0.0992)
Sector 18	-0.0469	-0.0251	-0.1099	0.0016	0.0299	-0.1374
	(0.0510)	(0.0583)	(0.1030)	(0.0518)	(0.0603)	(0.1188)
Sector 19	-0.1626***	-0.1563***	-0.2147***	-0.1450***	-0.1281***	-0.3211***
	(0.0411)	(0.0459)	(0.0830)	(0.0422)	(0.0464)	(0.1096)

Year 94	-0.0969***	-0.0978***	-0.0849***	-0.0923***	-0.0946***	-0.0790***
	(0.0085)	(0.0102)	(0.0143)	(0.0086)	(0.0105)	(0.0147)
Year 98	-0.0434***	-0.0474***	-0.0192	-0.0461***	-0.0501***	-0.028**
	(0.0078)	(0.0091)	(0.0131)	(0.0077)	(0.0091)	(0.0131)
Robustness analysis						
Productive capacity				0.1159***	0.1224***	0.1166**
				(0.0312)	(0.0360)	(0.0534)
Foreign Capital				0.0506***	0.0391	0.0518***
				(0.0154)	(0.0278)	(0.0190)
Group				-0.0101	0.0137	-0.0373**
				(0.0130)	(0.0192)	(0.0184)
Public Capital				-0.0866*	-0.1546	-0.0897
				(0.0486)	(0.0761)	(0.0625)
Local market				-0.0677***	-0.0637	-0.0676
				(0.0209)	(0.0234)	(0.0683)
Province market				-0.0235	-0.0196	0.0039
				(0.0188)	(0.0214)	(0.0636)
Region market				-0.0084	-0.0043	-0.0081
				(0.0198)	(0.0227)	(0.0438)
National market				-0.0197*	-0.0104	-0.0342**
				(0.0113)	(0.0150)	(0.0164)
International market				-0.0010	0.0216	-0.0292
				(0.0197)	(0.0258)	(0.0310)
Exports				0.0014**	0.0016***	0.0010
				(0.0006)	(0.0007)	(0.0013)
Region 1				-0.0621	-0.0647	0.0074
				(0.0563)	(0.0629)	(0.0495)
Region 2				-0.0493	-0.0506	-0.0396
				(0.0571)	(0.0648)	(0.0597)
Region 3				-0.0716	-0.0435	-0.0936
				(0.0620)	(0.0717)	(0.0655)
Region 4				-0.0540	-0.0242	-0.2939**
				(0.0753)	(0.0829)	(0.1476)
Region 5				-0.0819	-0.0859	0.0134
				(0.0868)	(0.0978)	(0.1261)
Region 6				0.0071	-0.0060	-0.0140
				(0.0709)	(0.1186)	(0.0653)
Region 7				-0.0848	-0.0629	-0.1365***
				(0.0593)	(0.0672)	(0.0520)
Region 8				-0.0641	-0.0670	-0.0254
				(0.0580)	(0.0677)	(0.0458)
Region 9				-0.0022	0.0129	-0.0442
				(0.0529)	(0.0595)	(0.0354)
Region 10				-0.0676	-0.0565	-0.0869**
				(0.0534)	(0.0600)	(0.0422)
Region 11				-0.3442***	-0.3778***	-0.2104*
				(0.0912)	(0.1084)	(0.1161)
Region 12				-0.1090**	-0.1088*	-0.1222**
				(0.0568)	(0.0643)	(0.0541)
Region 13				-0.0128	-0.0075	-0.0040
				(0.0531)	(0.0596)	(0.0421)
Region 14				-0.1011*	-0.1021	-0.0802

				(0.0608)	(0.0672)	(0.0751)
Region 15				-0.0983	-0.0670	-0.1656***
				(0.0622)	(0.0738)	(0.0550)
Region 16				0.0068	0.0201	-0.0324
				(0.0554)	(0.0627)	(0.0452)
constant	0.0686*	0.0726	0.1572	0.0782	0.0644	0.2412
	(0.0374)	(0.0442)	(0.1156)	(0.0713)	(0.0817)	(0.1518)
H_0: Random effects$_t$=0	616.81***	476.95***	70.85***	571.58***	423.35***	58.99***
H_0: Sector$_t$=0	254.02***	185.27***	984.9***	205.54***	156.51***	145.11***
H_0: Years$_t$=0	134.41***	91.93***	44.05***	115.77***	81.86***	30.59***
H_0: Market$_t$=0				12.05**	10.8**	5.46
H_0: Region$_t$=0				62.54***	50.67***	33.98***
No of obs	2684	2061	623	2684	2061	623
No of firms	1585	1211	415	1585	1211	415

Note: robust standard deviation in parentheses; (***), (**) and (*) denote significant at 1%, 5% and 10%.

Appendix 3.2. Oaxaca-Blinder decomposition for specifications (3.1a) and (3.1b). Complete results

Table A3.3. TFP gap decomposition. OLS estimation. Specification (3.1a) (corresponding to Table 3.6)

	1994			1998			2002		
$\overline{TFP_L} - \overline{TFP_S}$		**0.0794**			**0.0766**			**0.0603**	
	Charact	**RetsS**	**RetsL**	**Charact**	**RetsS**	**RetsL**	**Charact**	**RetsS**	**RetsL**
Total	0.07752	0.00050	0.00133	0.07992	-0.00070	-0.00261	0.06544	-0.00112	-0.00407
	97.69%	0.63%	1.68%	104.33%	-0.92%	-3.41%	108.60%	-1.85%	-6.75%
Innovation	0.00901	0.00024	0.00014	0.00724	0.00022	0.00003	0.00792	0.00028	-0.00005
	11.36%	0.30%	0.17%	9.45%	0.29%	0.04%	13.14%	0.46%	-0.08%
Qualified workers	0.00737	0.00225	0.00965	0.00881	0.00261	0.01132	0.00764	0.00286	0.01152
	9.28%	2.83%	12.16%	11.50%	3.41%	14.78%	12.68%	4.74%	19.12%
Size	0.02329	-0.00955	-0.03590	0.02268	-0.00985	-0.03607	0.02309	-0.00979	-0.03625
	29.35%	-12.03%	-45.24%	29.61%	-12.85%	-47.09%	38.31%	-16.25%	-60.15%
Age	0.01819	-0.00470	-0.01408	0.01545	-0.00494	-0.01335	0.01250	-0.00522	-0.01260
	22.92%	-5.92%	-17.74%	20.16%	-6.45%	-17.43%	20.75%	-8.66%	-20.91%
Sector	0.01966	-0.00119	0.01992	0.02575	-0.00150	0.01719	0.01430	-0.00261	0.01187
	24.78%	-1.50%	25.10%	33.61%	-1.96%	22.45%	23.72%	-4.34%	19.70%

Note: "retsS" corresponds to the small firms' disadvantage in relation with the non-discriminatory returns structure; "retsL" corresponds to large firms' advantage in relation with the non-discriminatory returns structure.

Table A3.4. TFP gap decomposition. OLS estimation. Specification (3.1b) (corresponding to Table 3.7)

	1994			1998			2002		
$\overline{TFP_L} - \overline{TFP_S}$		**0.0794**			**0.0766**			**0.0603**	
	Charact	**RetsS**	**RetsL**	**Charact**	**RetsS**	**RetsL**	**Charact**	**RetsS**	**RetsL**
Total	0.07522	0.00112	0.00301	0.08036	-0.00080	-0.00296	0.06899	-0.00188	-0.00685
	94.79%	1.41%	3.80%	104.92%	-1.05%	-3.87%	114.49%	-3.12%	-11.37%
Innovation	0.00841	-0.00033	0.00031	0.00676	-0.00031	0.00007	0.00739	-0.00038	-0.00011
	10.60%	-0.42%	0.39%	8.82%	-0.40%	0.09%	12.27%	-0.62%	-0.18%
Qualified workers	0.00495	0.00247	0.00963	0.00591	0.00287	0.01130	0.00513	0.00314	0.01150
	6.23%	3.11%	12.13%	7.72%	3.75%	14.75%	8.52%	5.21%	19.08%
Size	-0.01514	-0.00626	-0.00340	-0.01474	-0.00646	-0.00342	-0.01501	-0.00642	-0.00343
	-19.08%	-7.89%	-4.28%	-19.25%	-8.43%	-4.46%	-24.91%	-10.65%	-5.70%
Age	0.01305	-0.00233	-0.00707	0.01108	-0.00245	-0.00670	0.00897	-0.00259	-0.00633
	16.45%	-2.94%	-8.91%	14.47%	-3.20%	-8.75%	14.89%	-4.30%	-10.50%
Sector	0.01928	-0.00194	0.00927	0.02204	-0.00215	0.00805	0.01329	-0.00343	0.00104
	24.30%	-2.44%	11.69%	28.78%	-2.81%	10.51%	22.05%	-5.70%	1.73%

Note: see Table A3.3.

Table A3.5. TFP gap decomposition. Random effects estimation. Specification (3.1a) (corresponding to Table 3.8)

	1994			1998			2002		
$\overline{TFP_L} - \overline{TFP_S}$	**0.0794**			**0.0766**			**0.0603**		
	Charact	**RetsS**	**RetsL**	**Charact**	**RetsS**	**RetsL**	**Charact**	**RetsS**	**RetsL**
Total	0.07216	0.00083	0.00963	0.07555	0.00438	0.01964	0.05968	-0.00092	-0.00859
	90.94%	1.04%	12.13%	98.63%	5.72%	25.64%	99.04%	-1.53%	-14.26%
Innovation	0.00497	-0.00126	0.00086	0.00399	-0.00117	0.00018	0.00437	-0.00143	-0.00031
	6.27%	-1.59%	1.08%	5.21%	-1.53%	0.24%	7.25%	-2.38%	-0.51%
Qualified workers	0.00514	0.00191	0.00607	0.00615	0.00222	0.00713	0.00534	0.00243	0.00725
	6.48%	2.41%	7.65%	8.03%	2.90%	9.30%	8.86%	4.04%	12.04%
Size	0.02459	0.00859	-0.07249	0.02395	0.00886	-0.07284	0.02438	0.00881	-0.07320
	30.99%	10.83%	-91.35%	31.26%	11.57%	-95.09%	40.45%	14.62%	-121.47%
Age	0.01583	0.00027	0.00099	0.01344	0.00028	0.00094	0.01088	0.00030	0.00089
	19.95%	0.34%	1.25%	17.55%	0.37%	1.23%	18.06%	0.49%	1.48%
Sector	0.02162	-0.00125	0.01638	0.02802	-0.00158	0.01431	0.01472	-0.00274	0.01099
	27.25%	-1.57%	20.64%	36.58%	-2.06%	18.68%	24.42%	-4.54%	18.24%

Note: see Table A3.3. Given that the decomposition is not exact in the case of using the RE estimates, the sum of the shares of the components does not equal 100%.

Table A3.6. TFP gap decomposition. Random effects estimation. Specification (3.1b) (corresponding to Table 3.9)

	1994			1998			2002		
$\overline{TFP_L} - \overline{TFP_S}$	**0.0794**			**0.0766**			**0.0603**		
	Charact	**RetsS**	**RetsL**	**Charact**	**RetsS**	**RetsL**	**Charact**	**RetsS**	**RetsL**
Total	0.07081	0.00060	0.01047	0.07808	0.00348	0.01616	0.06373	-0.00128	-0.01038
	89.24%	0.75%	13.20%	101.94%	4.54%	21.09%	105.77%	-2.12%	-17.22%
Innovation	0.00425	-0.00130	0.00088	0.00342	-0.00121	0.00019	0.00374	-0.00148	-0.00031
	5.36%	-1.64%	1.10%	4.46%	-1.58%	0.24%	6.20%	-2.46%	-0.52%
Qualified workers	0.00318	0.00242	0.00752	0.00380	0.00281	0.00882	0.00330	0.00307	0.00898
	4.01%	3.04%	9.47%	4.97%	3.67%	11.52%	5.48%	5.10%	14.90%
Size	-0.01379	0.01659	-0.00882	-0.01343	0.01711	-0.00886	-0.01367	0.01701	-0.00890
	-17.38%	20.90%	-11.11%	-17.53%	22.33%	-11.56%	-22.68%	28.23%	-14.77%
Age	0.01093	0.00178	0.00525	0.00928	0.00188	0.00498	0.00751	0.00198	0.00470
	13.78%	2.25%	6.62%	12.12%	2.45%	6.50%	12.47%	3.29%	7.80%
Sector	0.02157	-0.00232	0.00649	0.02456	-0.00256	0.00708	0.01391	-0.00382	0.00204
	27.19%	-2.93%	8.18%	32.07%	-3.34%	9.25%	23.08%	-6.35%	3.38%

Note: see Table A3.3. Given that the decomposition is not exact in the case of using the RE estimates, the sum of the shares of the components does not equal 100%.

Chapter 4

THE TFP DIFFERENTIAL BETWEEN SMALL AND LARGE FIRMS: A DISTRIBUTIONAL ANALYSIS

4.1. Introduction

The objective of this Chapter is extending the decomposition analysis of the total factor productivity (TFP) gap to the entire distribution. In Chapter 3 we focused on the analysis in the mean of the distribution. Although it is attractive to use a synthetic measure, the analysis in the whole distribution permits identifying patterns of behaviour that differ from the mean and that may uncover heterogeneous behaviour of firms. In this regard, the analysis in the present Chapter is in the spirit of Jenkins (1994) and Juhn et al. (1993).

In Chapter 4 we depart from the same theoretical framework, empirical specifications and estimation methods exposed in Chapter 3. There, we argued that large firms are more productive and that they innovate more and have a more qualified labour force. Estimates of the coefficients of our empirical specifications pointed to a positive relationship between productivity and human and technological capital, in the case of both large and small firms. As a result, we obtained that these types of capital contribute to explain the TFP gap between the *representative* (average) small and large firms.

Furthermore, in the present Chapter we intend to uncover heterogeneous patterns that cannot be identified using synthetic measures. In so doing, we firstly sketch the method used to decompose the TFP gap between small and large firms using a counterfactual distribution analysis in Section 4.2. As the most relevant features of the difference between the TFP distributions for large and small firms have already been described in Section 2.5.3, in Section 4.3 we directly present and discuss the results of the counterfactual distribution analysis. Section 4.4 summarizes and concludes.

4.2. Methodology: The Counterfactual Distribution Analysis

The Oaxaca-Blinder decomposition carried on in Chapter 3 permits evaluating the relative contribution of characteristics and returns of our variables of interest to explain the productivity differentials in the mean of the distribution. Although it is very attractive to summarize the TFP differentials within a single number, for example the mean, Jenkins (1994) argues that synthetic measures represent a loss of information because they do not allow evaluating such differences along the whole distribution. In other words, the same statistical may be consistent with very different distributions. Using an argument similar to that used by Jenkins, if innovation explains 10% of TFP differences, one could extract different policy implications depending on whether it occurred in the whole distribution or only in the first quantiles (small firms with the lowest TFP). For example, in the second case, policies that try to increase the innovative capacity of small firms should not focus on all the firms as for firms in some TFP ranges they would be unnecessary.

Jenkins (1994) suggested a modification of the traditional counterfactual analysis in the mean of the distribution by applying methods developed in the study of income distributions. Here, we take the idea of the counterfactual distribution analysis but propose using the estimated density functions to compare the distributions of the predicted and counterfactual TFP levels in large and small firms. This permits studying TFP differentials in the complete distribution. In addition, we estimate a conditional bivariate density for the predicted and counterfactual distributions to detect those firms with larger probabilities of changing their TFP levels under the scenario of equal returns regardless of firm size. The idea behind the counterfactual distribution analysis is transferring the Oaxaca-Blinder decomposition to the whole distribution, in the sense that we try to separate the contribution of differences in endowments and returns at any point of the distribution. The counterfactual analysis permits comparing the distribution of the predicted TFP of small firms with the counterfactual (or hypothetical) TFP, obtained by evaluating small firms under the returns of large firms.

Computing the predicted and counterfactual TFP values

The predicted TFP values for small and large firms are obtained as:

$$\hat{TFP}_S = X'_S \hat{\beta}_S$$
$$\hat{TFP}_L = X'_L \hat{\beta}_L \tag{4.1}$$

while the counterfactual distribution for small firms is obtained as:

$$\hat{TFP}_S^L = X'_S \hat{\beta}_L \tag{4.2}$$

In other words, $\hat{TFP}_S^L$ would be the TFP of small firms if they had the same returns of large firms. Both $\hat{TFP}_S$ and $\hat{TFP}_S^L$ use the characteristics of small firms. The difference between them is that the former uses the estimated returns of small firms, $\hat{\beta}_S$, and the latter, those of large firms $\hat{\beta}_L$. Thus, the difference between $\hat{TFP}_S$ and $\hat{TFP}_S^L$ can be attributed to differences in returns. The same reasoning can be applied for the difference between $\hat{TFP}_S^L$ and $\hat{TFP}_L$, which use the estimated returns of large firms. In this case, the former uses the characteristics of small firms, X_S, and the latter the characteristics of large firms X_L. Thus, the gap between $\hat{TFP}_S^L$ and $\hat{TFP}_L$ can be attributed to differences in characteristics.

Comparing the (external) shape of the predicted and the counterfactual distributions

Let $F(\hat{TFP}_S)$, $H(\hat{TFP}_L)$ and $G(\hat{TFP}_S^L)$ denote the cumulative distribution functions that correspond to the estimated TFP for small and large firms and the counterfactual TFP, respectively. The difference between $F(\hat{TFP}_S)$ and $G(\hat{TFP}_S^L)$ is then associated to differences in returns. When $G(\hat{TFP}_S^L)$ shifts to the right of $F(\hat{TFP}_S)$, it indicates that small firms have a higher probability of achieving higher TFP levels if they have the returns of large firms. Following the same reasoning, the difference between $G(\hat{TFP}_S^L)$ and $H(\hat{TFP}_L)$ is associated to differences in characteristics. So this methodology permits assessing the contribution of differences in characteristics and differences in returns to the TFP gap between small and large firms in the whole distribution.

Finally, it should be noticed that, instead of using the estimate of the distribution functions to perform the above-mentioned analysis, we will estimate and compare the corresponding density functions, $f(\hat{TFP}_S)$, $h(\hat{TFP}_L)$ and $g(\hat{TFP}_S^L)$. And given that we are interested in the differences between the distributions, we will report the difference in the estimated densities in each TFP value, as suggested in DiNardo et al. (1996).[86, 87]

Movements between the predicted and the counterfactual distributions

The analysis of changes in the shape of the TFP distribution of small firms when their characteristics are evaluated as in large firms says nothing about which small firms will be more affected by a change in returns. In other words, the same density for the counterfactual distribution may be consistent with very different transition patterns of small firms from the predicted to the counterfactual TFP levels. Following the example above, densities for the predicted and counterfactual TFP distributions showing heterogeneous differences along the distribution indicate that policy actions aiming to equalize returns in small firms to those in large firms will be more effective for firms departing from some specific TFP ranges. However, this analysis does not permit identifying those specific ranges. To get some insight on such issue, we need to estimate the probabilities of transition of every TFP level from the predicted to the counterfactual distribution. In this way, we will have a tool that permits assessing which firms (regarding their TFP levels) could benefit more from the improvement in the return to their characteristics.

As in Section 2.5.2, to analyse the probabilities of transition from two distributions we estimate a stochastic kernel. The stochastic kernel permits evaluating the dynamics within the distribution in two periods of time (as in Section 2.5.2) or between the predicted and the counterfactual distributions (see, for instance, Fingleton and López-Bazo, 2003). The stochastic kernel is interpreted in a way similar to a first order probability of transitions matrix where the number of states tends to infinite. That is, in a continuous framework instead of a discrete framework, as in the first-order Markov probability of transitions matrix. Following Johnson (2000), let $f_S(T\hat{F}P = a)$ and $g_S^L(T\hat{F}P = a)$ be the probability of $T\hat{F}P = a$ for small firms and for small firms evaluated under the returns of large firms respectively. Assuming the existence of marginal and conditional density functions for the TFP distribution, the relationship between the two distributions can be expressed as:

[86] DiNardo et al. (1996) suggest a similar analysis based on the comparison of actual and counterfactual distributions but from a different methodological approach.
[87] Appendix 4.1 provides further explanation on the interpretation of this analysis.

$$g_S^L(T\hat{F}P = a) = \int_0^1 l(T\hat{F}P = b \mid T\hat{F}P = a) f_S(T\hat{F}P = a) \tag{4.3}$$

where $l(T\hat{F}P = b \mid T\hat{F}P = a)$ is the probability of $T\hat{F}P = b$ for small firms when their returns are those of large firms, conditional on a level of $T\hat{F}P = a$ in the predicted distribution of small firms. The conditional density function $l(\cdot)$ summarizes information on transitions between the two distributions. It is computed by first estimating the joint density for the distributions of $T\hat{F}P_S$ and $T\hat{F}P_S^L$ by the kernel method[88] and then, dividing it by the marginal density of $T\hat{F}P_S$, which is obtained by integrating the joint density over $T\hat{F}P_S^L$.

4.3. Results

4.3.1. Comparing the (external) shape of the predicted and the counterfactual distributions

To start with, we briefly describe the distance between the densities of the predicted TFP levels for small and large firms, that is the difference between $f(T\hat{F}P_S)$ and $h(T\hat{F}P_L)$. For each year under analysis, this difference is represented by the continuous line in Figure 4.1. The picture we can get from the differences between densities is similar to the one in Section 2.5.3, when we compared the densities for the actual TFP values for small and large firms. In 1994 and 1998, the relative number of small firms with low TFP levels is much higher than the number of large firms. Correspondingly, large firms are more abundant in the range of high TFP levels. Changes in the distribution during the second half of the period under analysis imply that, in 2002, the two distributions are more similar than in previous years. In this year, the relative number of small firms with very low TFP levels is much lower; and there seems to be more small firms at medium levels of TFP, although once again they are underrepresented in the range of high TFP levels.

As described in Appendix 4.1, the distance between the counterfactual —$g(T\hat{F}P_S^L)$— and the predicted —$f(T\hat{F}P_S)$— density for small firms in relation to the

[88] The non-parametric kernel methodology used to estimate these density functions is explained in Appendix 2.2.

distance between the two predicted densities, permits an analysis of the contribution of differences in returns to explain the productivity gap at every TFP level along the distribution. The dashed line in Figure 4.1 represents the difference between the counterfactual and the predicted densities for small firms using OLS estimates of the coefficients from specification (3.1a). In this case, the counterfactual distribution assumes that all the characteristics of small firms are paid as in the case of large firms. That is to say, Figure 4.1 permits assessing the effect of differences in returns to the whole set of characteristics.

The first interesting issue is that the effect of differences in returns is far from homogeneous: it clearly depends on the range of the TFP level. In addition, in the three years under analysis, differences in returns account for a non-negligible part of the gap at high TFP levels. Concretely, they make small firms with medium and low TFP levels shift to the right. Moreover, differences in returns cause that a mass of probability at the very low TFP levels in the counterfactual distribution emerges (which is not present in the predicted distribution). This result suggests that, if small firms were evaluated under the returns of large firms, some of them would be even less productive than they are. This is a signal of heterogeneous behaviour that cannot be identified using an analysis in the mean of the distribution.

Therefore, we can conclude that evaluating the endowment of small firms as in large firms will improve productivity in some of them but, simultaneously, will provoke a decrease in the TFP levels of some others. The contribution of differences in returns seems to be more intense in 2002 than in previous years, where apart from the mass of probability in the counterfactual distribution at very low TFP levels, the dashed line is quite close to the continuous one. This suggests that a big deal of the gap in a wide range of TFP levels is due to differences in returns between large and small firms.

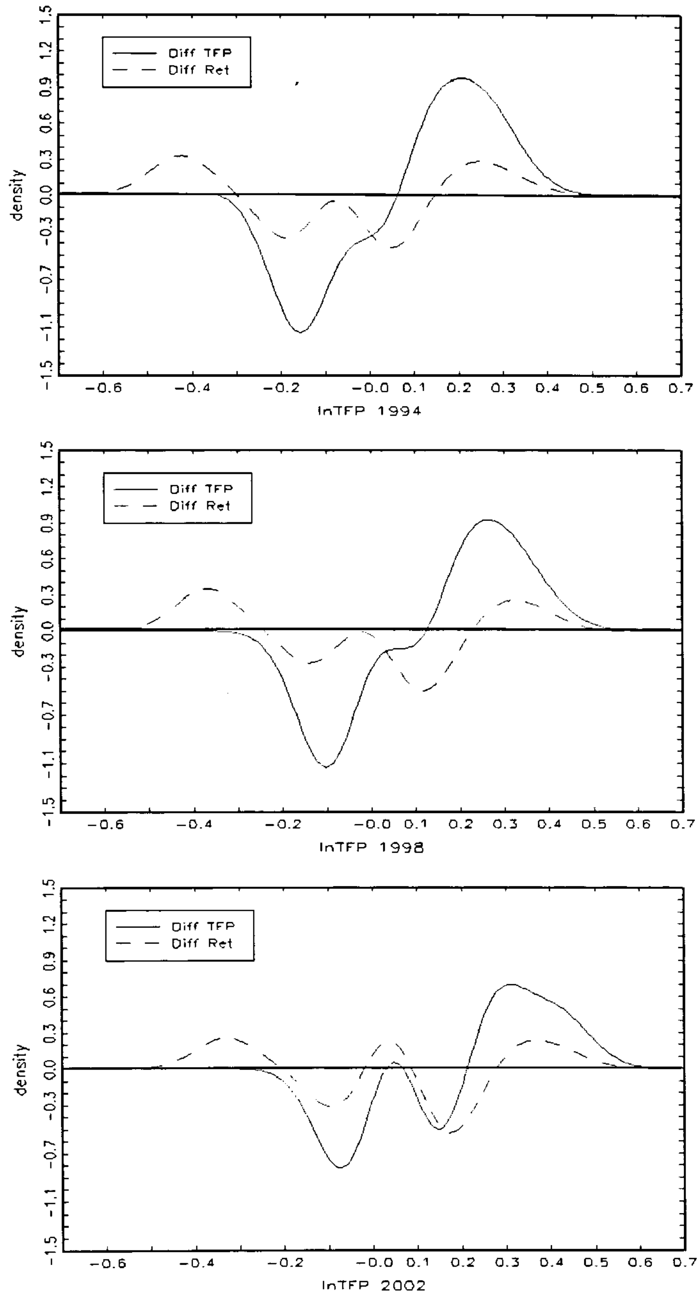


Figure 4.1. Differences between density functions of predicted and counterfactual TFP levels based on OLS estimations of specification (3.1a). Effects of all variables.

In Chapter 3, we showed that both observable and unobservable firm heterogeneity is present in the empirical specification of TFP. When considering the two sources of heterogeneity in the estimation of the coefficients, we observed relevant differences between the estimations that include unobservable firm specific effects (random effects) and those that do not. For this reason, we also report here the results based on the estimation of the random effects model of specification (3.1a). Results based on the OLS and the RE estimations of the specification that include the additional control variables (specification 3.1b) are reproduced in the Appendix 4.2, as only minor differences are observed.

Figure 4.2 displays the difference between densities based on the RE estimations of specification (3.1a). The general picture is quite similar to the one derived from results based on the OLS estimations. The most striking feature is that in this case differences in returns seem to explain even a larger portion of the TFP gap, as the dashed line is closer to the continuous line. In addition, controlling for unobservable heterogeneity does not prevent obtaining a mass of probability at the very low TFP levels in the counterfactual distribution that is not observed in the predicted distribution.

Given our interest in the particular effect of innovation and human capital, we have obtained a counterfactual distribution for small firms under the assumption that only these two factors were evaluated as in large firms. That is, in this case, $\hat{TFP}_S^L$ is computed using all the estimated parameters in the sample of small firms except those associated to innovation and human capital, for which we use the estimated parameters of large firms. The differences between the counterfactual and predicted densities when using the OLS and the RE estimates are depicted in Figures 4.3 and 4.4, respectively.

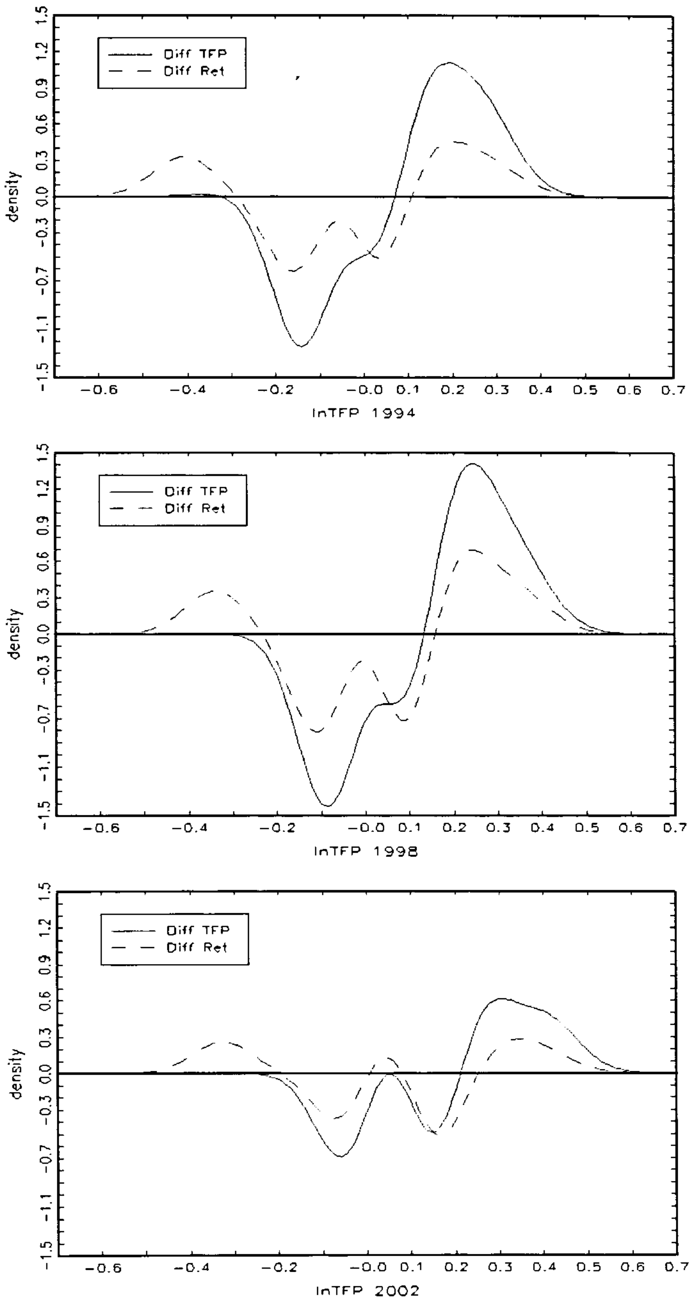


Figure 4.2. Differences between density functions of predicted and counterfactual TFP levels, based on RE estimations of specification (3.1a). Effects of all variables.

The first thing to say is that differences in returns to innovation and human capital show a modest contribution in explaining the TFP gap between small and large firms. The remaining portion of the gap is explained by differences in firms' endowment of technological and human capital, as well as differences in returns and characteristics of the other variables in our specification. But still in this case, we can state that the effect is not homogeneous. The contribution of differences in returns to these types of capital is more intense for firms with TFP values above the average. However, they do not seem to help explaining the situation of small firms with the lowest TFP levels. Or, in other words, small firms with low TFP levels would not improve their productivity if they had the same returns as large firms.

Two final remarks are in order. First, the difference between the counterfactual and the predicted densities for small firms is quite similar in the three years under analysis, even when the difference between the predicted densities for small and large firms varies, particularly in 2002. This is caused by persistence in the endowments of qualified labour and innovation in small firms over the period under analysis.[89] Second, in agreement with the results provided in Chapter 3, the effect attributed to differences in returns to these factors is lower when using the coefficients from the RE estimation.

[89] Notice that the difference between the counterfactual and the predicted distribution has to do with the difference in the coefficients associated to innovation and human capital in large and small firms, weighted by the endowment of these factors in small firms. Thus, persistence in endowments causes stability in the contribution of differences in returns, as we are imposing stability over time in the estimated coefficients.

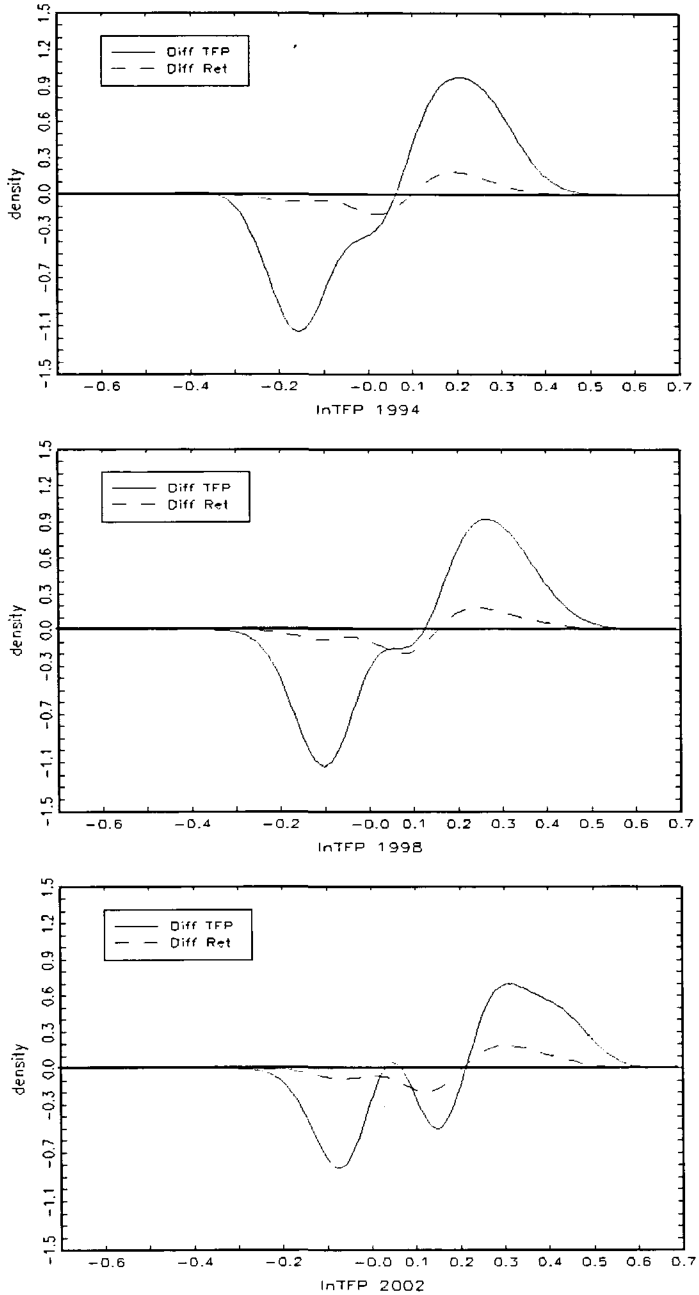


Figure 4.3. Differences between density functions of predicted and counterfactual TFP levels, based on OLS estimations of specification (3.1a). Effect of innovation and human capital

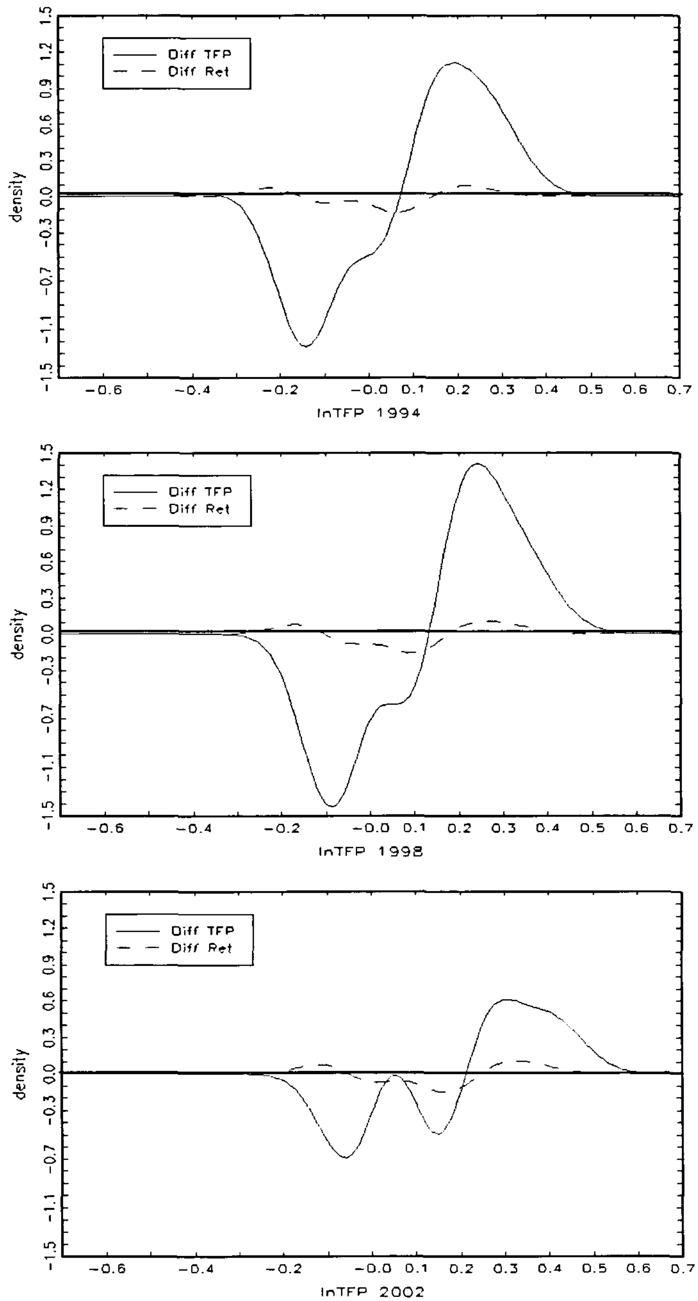


Figure 4.4. Differences between density functions of predicted and counterfactual TFP levels, based on RE estimations of specification (3.1a). Effect of innovation and human capital.

4.3.2. Movements between the predicted and the counterfactual distributions

The analysis of differences in the external shape of the predicted and counterfactual distributions for small firms reveals a non-negligible degree of heterogeneity in the contribution of returns to the TFP gap. Actually, results in the previous section can be read as a signal that the less productive small firms might react quite differently from the other small firms if their returns were equalized to those in large firms. However, robust evidence on such issue demands an analysis of the movements or transitions from the predicted to the counterfactual distribution. As described in Section 4.2, we propose estimating a stochastic kernel for these distributions to obtain evidence on this question.

Figure 4.5 shows the contour plots of the conditional density function in (4.3), $l(T\hat{F}P = b \mid T\hat{F}P = a)$, which summarize information on transitions between the predicted and counterfactual distributions of small firms, when using estimates of the coefficients from the OLS estimation. Correspondingly, Figure 4.6 shows the same information in the case of using estimates from the RE model. In both cases, the estimates correspond to specification (3.1a).[90] For a hypothetical small firm, contour plots graph the probability of achieving each of the TFP levels in the counterfactual distribution when departing from any TFP level in the predicted distribution. The contour lines represent pairs of predicted and counterfactual TFP values with the same probability: the most external (internal) lines correspond to pairs of values with low (high) probability. When the mass of probability lies on the positive diagonal, it indicates a high degree of persistence, that is, small firms have a high probability of reaching a similar TFP level when they are evaluated either under the returns of large or small firms. On the contrary, if the mass of probability lies parallel to the horizontal axis, it indicates a high degree of mobility, as all the small firms would obtain the same TFP level if they were evaluated under the returns of large firms, regardless of their predicted TFP level. When the mass of probability shifts above the positive diagonal (upward and to the left), it indicates a tendency of small firms to reach higher TFP levels when they are evaluated under the returns of large firms.

[90] The results based on estimates from the specification including additional control variables are shown in Appendix 4.3.

Results from Figures 4.5 and 4.6 offer a similar picture: small firms would experience some improvement in their TFP levels if their characteristics were evaluated as in large firms, but the reaction if far from homogeneous. Some small firms above the average TFP will benefit from convergence to returns in large firms, but those with the highest TFP will remain unaffected. This suggests that returns of those small firms at the top of the TFP distribution might not differ from returns in large firms. For the small firms below the average, the reaction is heterogeneous as well. The clockwise twist in the mass of probability corresponding to the lowest TFP levels points to an improvement in TFP that is more intense the lower the level of productivity in the predicted distribution. In contrast, for the small firms with TFP not far below the average the effect of changing returns seems to be negligible (the mass of probability basically follows the diagonal). Finally, it is clearly observed that small firms with predicted TFP below the average have some chances to end up with even much lower values in the counterfactual distribution (*island* of probability at the bottom left of the graphs). Therefore, the analysis of movements between both distributions confirms that the appearance of the mass of probability at very low TFP levels in the counterfactual distribution is caused by some of the low-TFP small firms.

As with the comparison of the external shape of the distribution, and given our interest in the effects of human capital and innovation, we have repeated the exercise by only changing the returns to these two factors. The contour plots of the corresponding estimated stochastic kernels are depicted in Figures 4.7 and 4.8, for the OLS and the RE estimates of specification (3.1a) respectively. In this case the mass of probability is much more concentrated over the diagonal indicating that both the predicted and the counterfactual distributions are quite similar. Thus, the effect of differences in returns to human capital and innovation accounts for only a modest portion of the gap in TFP. Actually, the analysis reveals that only small firms with TFP above the average will somehow improve their TFP if their characteristics are evaluated as in large firms. The absence of any twist in the mass of probability additionally indicates that this latter effect would be homogeneous in those types of firms.

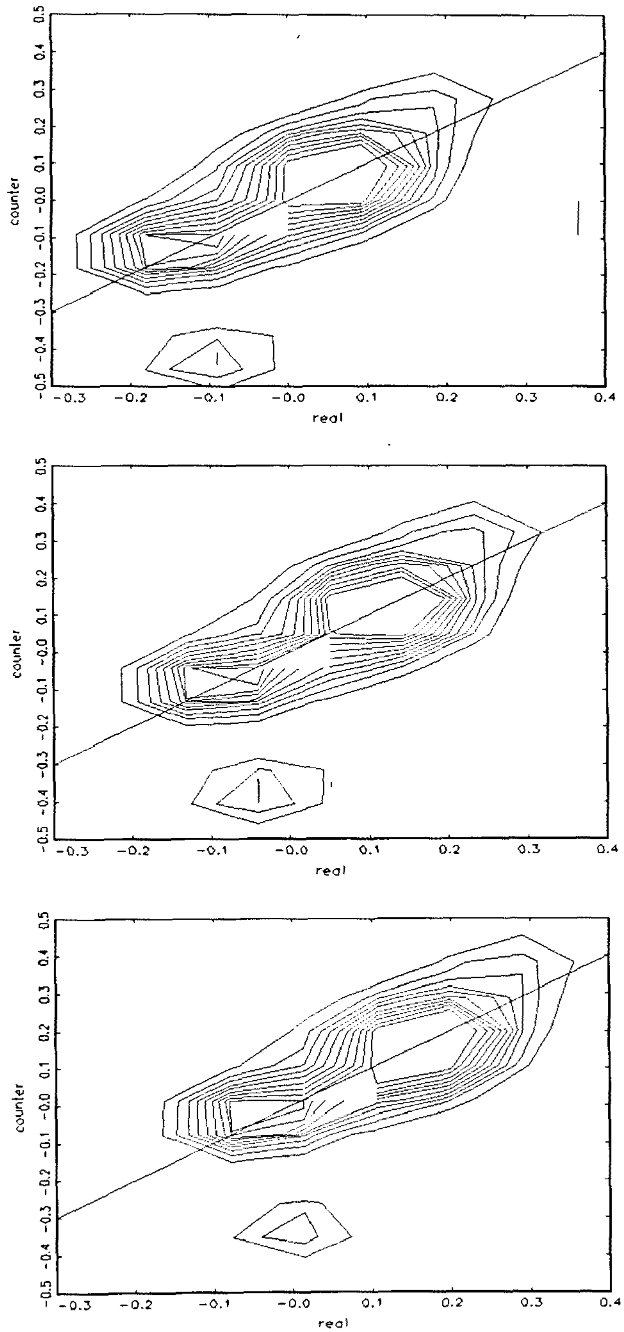


Figure 4.5. Stochastic kernel for the predicted and counterfactual TFP levels based on OLS estimations of specification (3.1a). Effects of all variables.

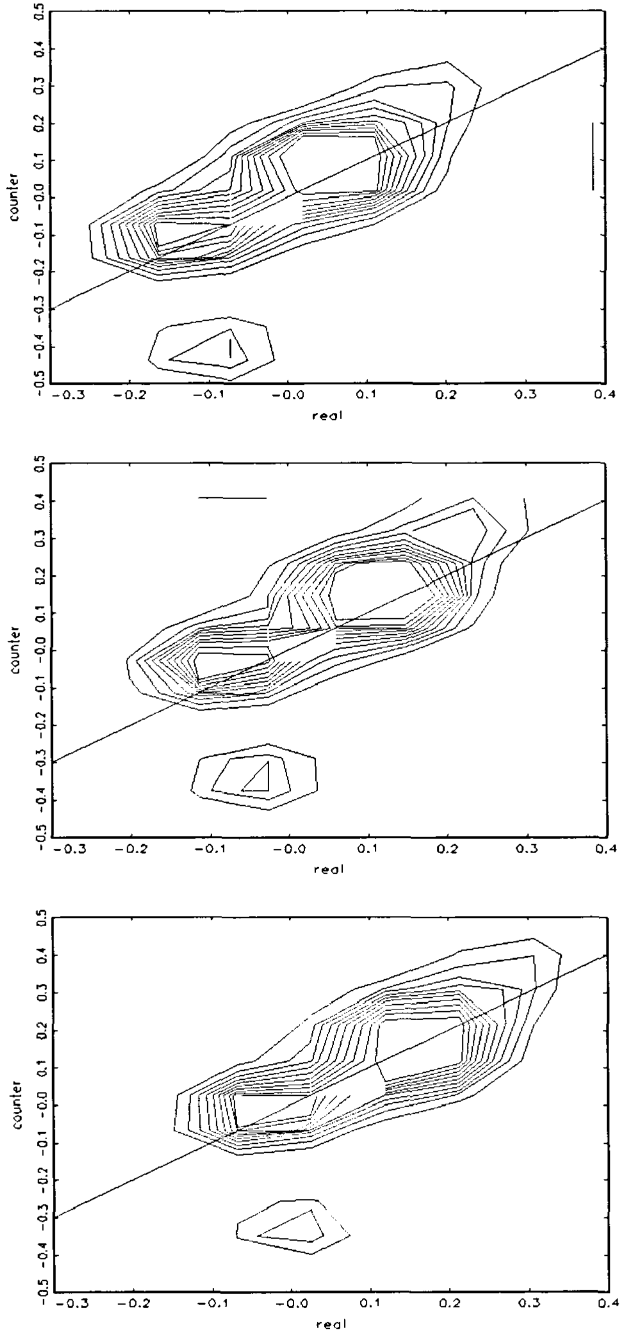


Figure 4.6. Stochastic kernel for the predicted and counterfactual TFP levels based on RE estimations of specification (3.1a). Effects of all variables.

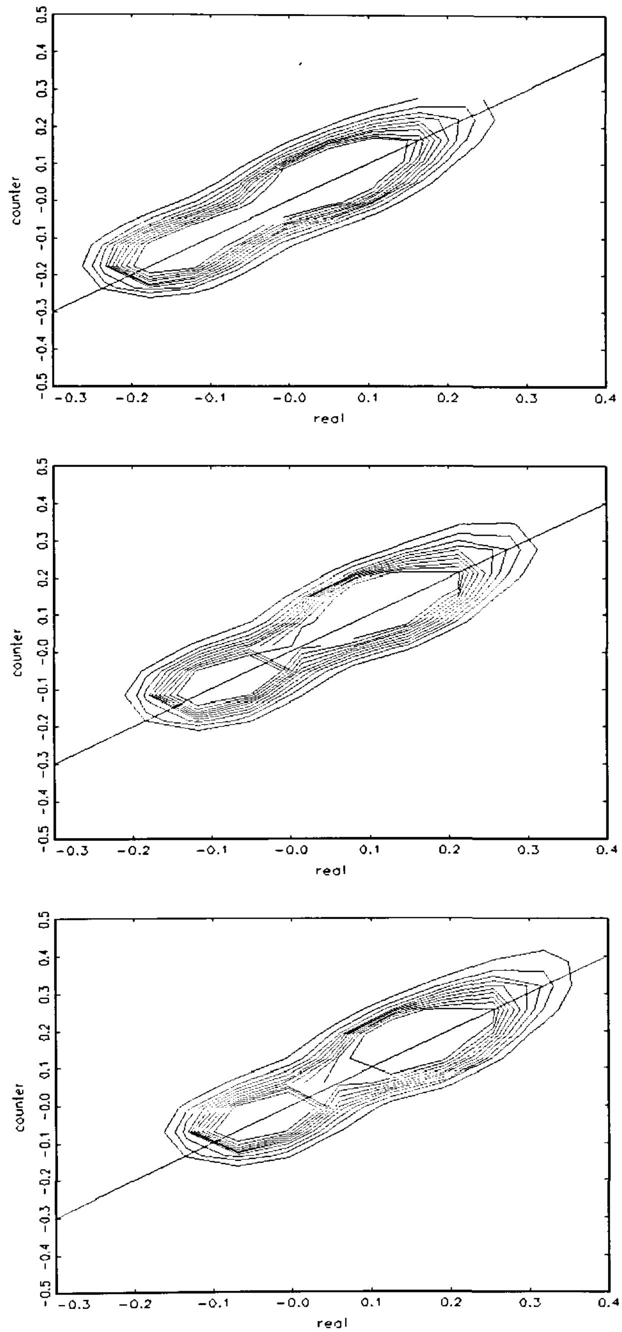


Figure 4. 7. Stochastic kernel for the predicted and counterfactual TFP levels based on OLS estimations of specification (3.1a). Effects of innovation and human capital

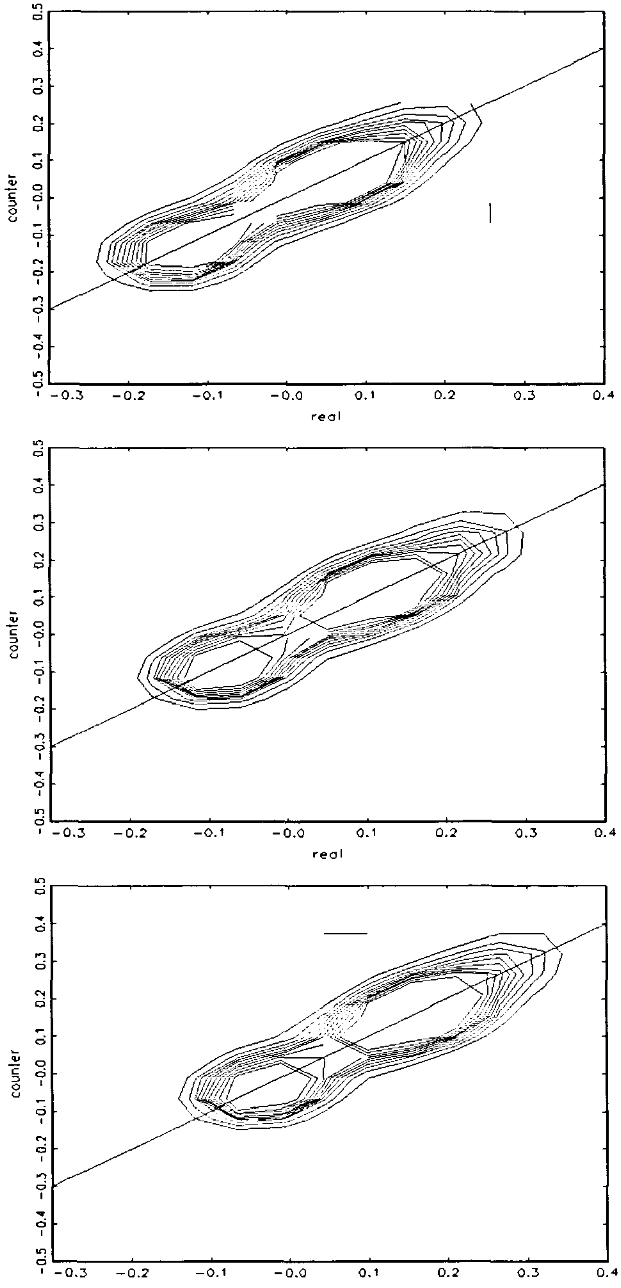


Figure 4.8. Stochastic kernel for the predicted and counterfactual TFP levels based on RE estimations of specification (3.1a). Effects of innovation and human capital.

4.4. Conclusions

Building on the idea of the Oaxaca-Blinder decomposition, this Chapter analyses the contribution of differences in returns to firms' characteristics in explaining the TFP gap between small and large firms along the distribution. The objective of this analysis is uncovering heterogeneous patterns that cannot be identified using synthetic measures.

The counterfactual distribution analysis permits comparing the distribution of the predicted TFP of small firms with the counterfactual (or hypothetical) TFP, obtained by evaluating small firms under the returns of large firms. More specifically, we analyse the contribution of all the factors included in our empirical specification and the contribution of technological and human capital. First, we compare the predicted and the counterfactual density functions, which permit analysing the external shape of these distributions. Next, we compare the conditional bivariate density functions, which permit studying the probability of transition of small firms from one productivity level to another after imposing the returns of large firms, conditional to their initial productivity level.

First of all, we found that large firms are more abundant in the range of high TFP levels, confirming the previous result that large firms are more productive. In addition, we confirmed the previous result that differences in TFP between small and large firms tend to decrease over our period of analysis.

Next, the counterfactual distribution analysis for all the variables in our specification indicates that differences in returns account for a considerable part of the TFP gap between large and small firms. Moreover, the contribution of differences in returns is far from homogeneous as small firms at different TFP levels react differently to changes in their returns. We obtained that a group of some small firms with low productivity would perform even worse if they were evaluated under the returns of large firms. However, the small firms with the lowest TFP levels would improve. This indicates that there are some opportunities to improve productivity for some low-productivity firms and this requires obtaining returns to their characteristics as high as in large firms. Another relevant result is that small firms with the highest TFP levels would not improve their productivity when they are evaluated under the returns of large firms, suggesting that they might have returns to firm characteristics close to those of large firms, which permit achieving TFP levels close to them.

Returns to innovation and human capital, our variables of interest, can only explain a modest part of the TFP differential, however their effect is not homogeneous. Only small firms with TFP above the average will somehow improve their TFP if their characteristics are evaluated as in large firms. However, they do not seem to help explaining the situation of small firms with the lowest TFP levels.

Thus in the case of small firms with TFP levels above the average, the economic policies dedicated to increase the levels of innovation and human capital would have a more positive effect on reducing the productivity gap if small firms could extract higher returns from the investment in these factors. As in the present analysis differences in returns to these factors play a minor role in explaining the TFP gap, the effectiveness of such policies would be modest. Moreover, small firms would not change their position in the ranking and the less productive would generally keep their relatively low productivity levels. Thus, if an economic policy was applied, we should not expect that it provokes changes within the distribution that increased or decreased the relative TFP levels of small firms in relation to their actual values.

Appendix 4.1. Interpretation of the counterfactual distributional analysis

Let $f(T\hat{F}P_S)$ and $h(T\hat{F}P_L)$ represent density functions of hypothetical TFP distributions for small firms and large firms respectively, and $g(T\hat{F}P_S^L)$ the density function for the counterfactual levels of TFP for small firms under the returns of large firms. They have been simulated for the purpose to illustrate the method used in this Chapter. The upper part of Figure A4.1 shows the density functions $f(T\hat{F}P_S)$, $g(T\hat{F}P_S^L)$ and $h(T\hat{F}P_L)$. The lower part of this Figure shows the differences of the density at each TFP level for large and small firms (continuous line), and the difference of densities for the counterfactual and predicted TFP for small firms (dashed line). The distance between $h(T\hat{F}P_L)$ and $f(T\hat{F}P_S)$ is the difference in the density (probability) of observing a large and a small firm in a given TFP level. The distance between $g(T\hat{F}P_S^L)$ and $f(T\hat{F}P_S)$corresponds to the change in the probability of observing a small firm with a particular TFP level when its endowments are evaluated with the returns of large firms. The difference can thus be assigned to differences in returns between small and large firms.

The magnitude of the later distance relative to the distance observed between the two predicted densities permits an analysis of the contribution of differences in returns to explain the productivity gap at every TFP level along the distribution. For example, in Figure A4.1, the difference between the densities of small and large firms at a TFP level of 0.05 (vertical light line) is decomposed in differences in returns (vertical dark line) and differences in characteristics (the remaining of the vertical light line). Notice that a flat dashed line at a value of 0 would mean no differences between the counterfactual and the predicted distribution, and thus that differences in returns do not help explaining the TFP gap at any point of the distribution. On the contrary, the dashed line overlapping the continuous line should be read as the gap being fully explained by differences in returns. In other words, had the small firms had the returns of large firms, they would have achieved similar TFP levels over the entire distribution. Intermediate cases indicate that differences in returns explain a portion of the gap, while the rest should be attributed to differences in endowments.

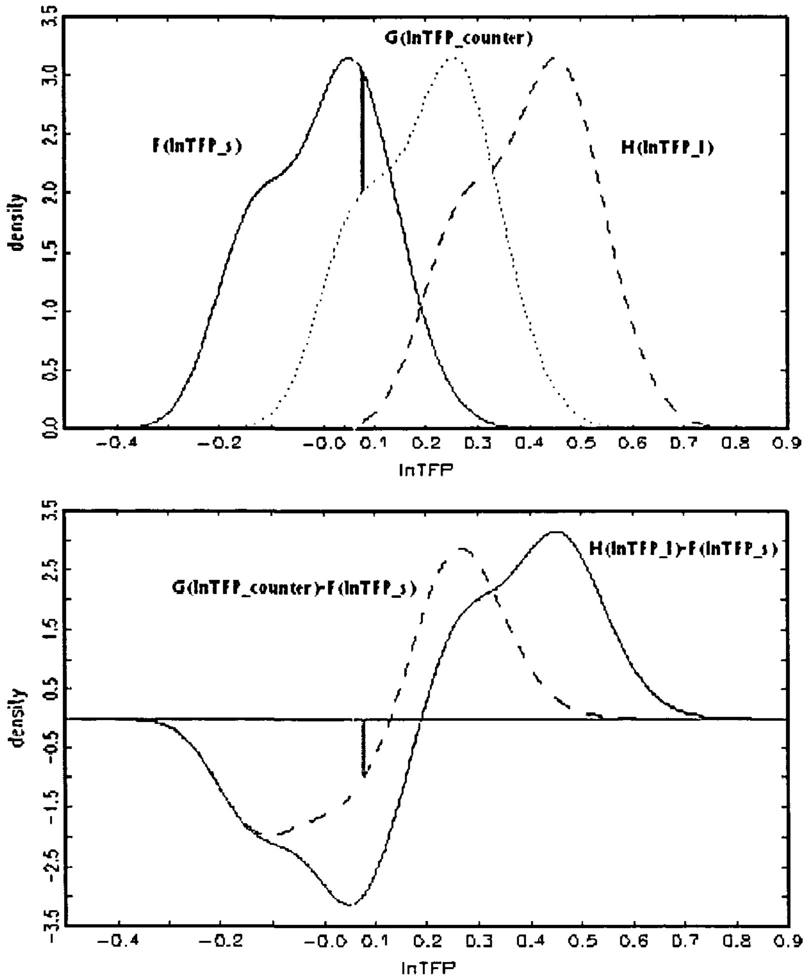
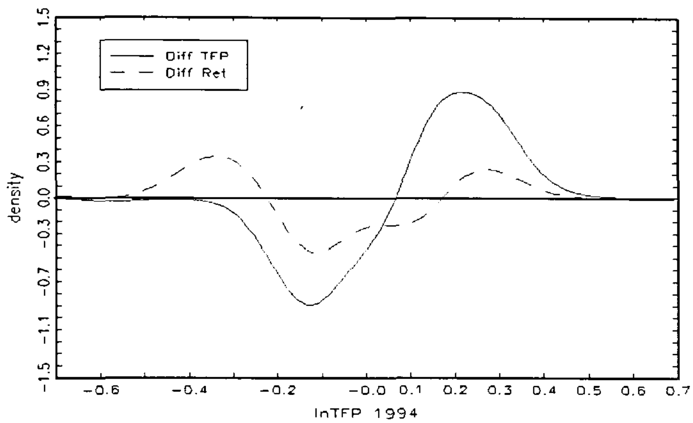
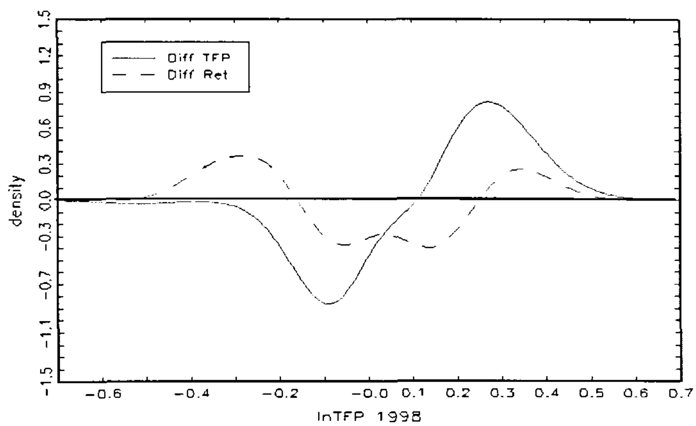


Figure A4.1. (Simulated) Density functions and their differences

Appendix 4.2. Results of the counterfactual analysis based on coefficients from specification (3.1b). External shape of the distributions





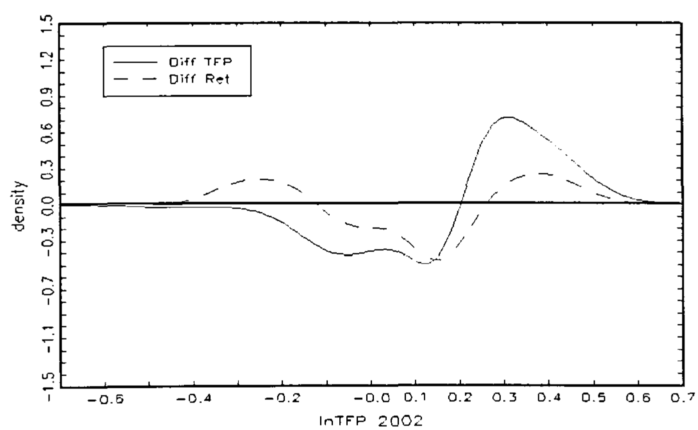


Figure A4.2. Differences between density functions of predicted and counterfactual TFP levels, based on OLS estimations of specification (3.1b). Effects of all variables

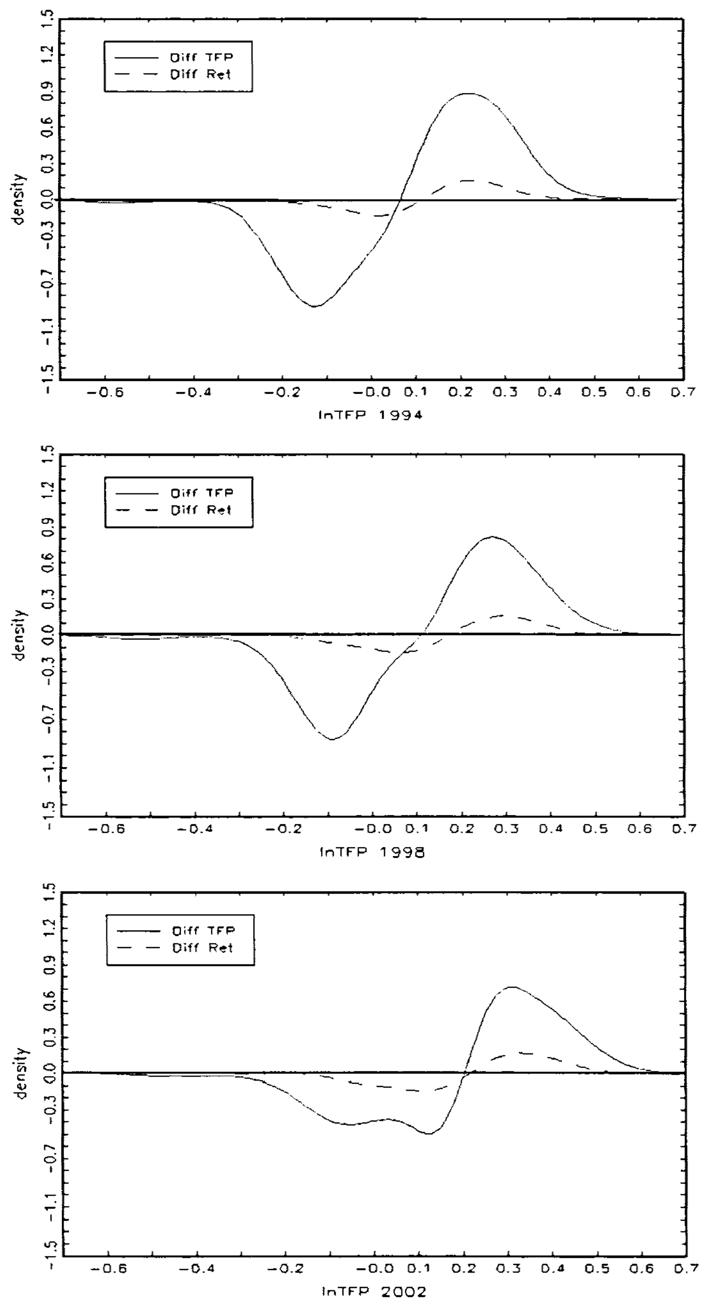


Figure A4.3. Differences between density functions of predicted and counterfactual TFP levels, based on OLS estimations of specification (3.1b) Effect of innovation and human capital.

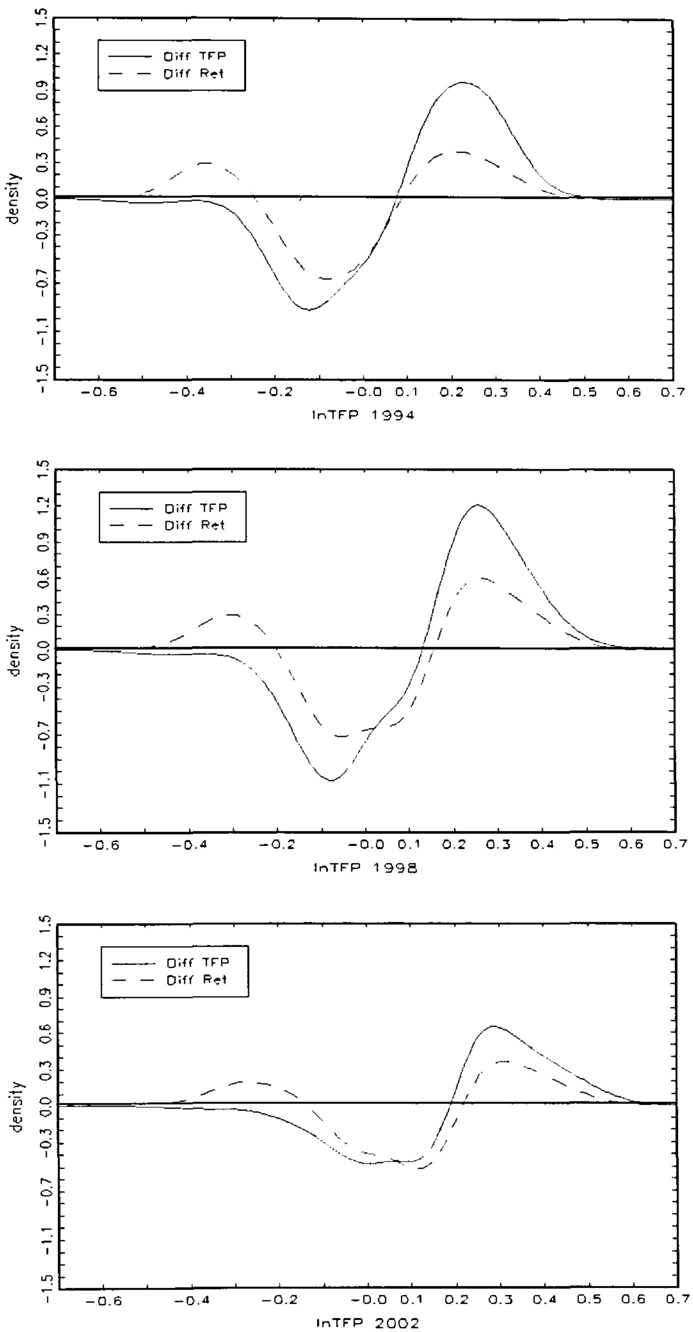


Figure A4.4. Differences between density functions of predicted and counterfactual TFP levels, based on RE estimations of specification (3.1b). Effects of all variables.

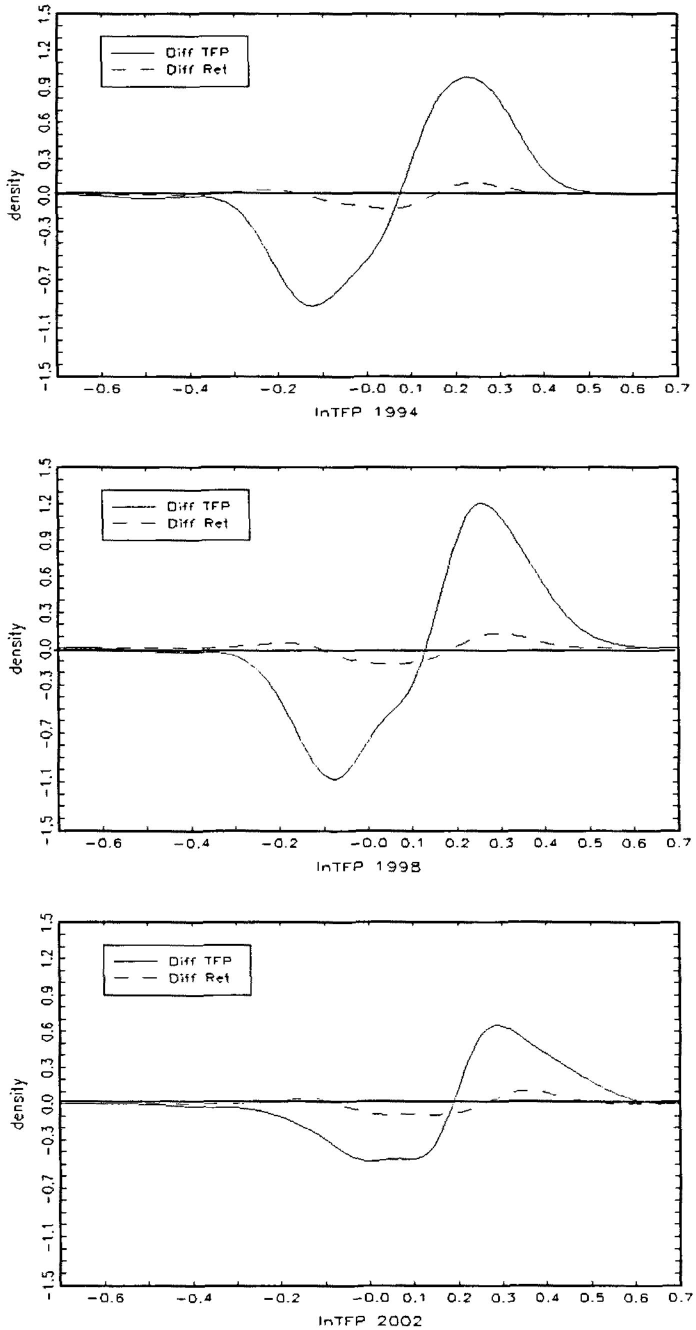


Figure A4.5. Differences between density functions of predicted and counterfactual TFP levels, based on RE estimations of specification (3.1b) Effect of innovation and human capital.

Appendix 4.3. Results of the counterfactual analysis based on coefficients from specification (3.1b). Movements within the distributions

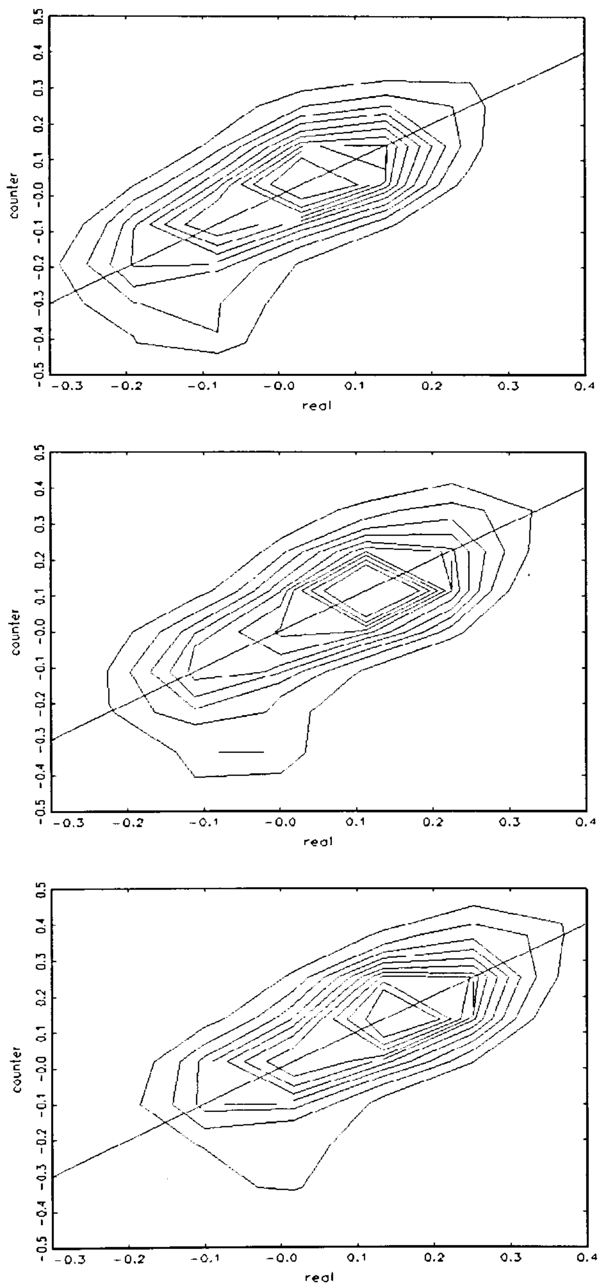


Figure A4.6. Stochastic kernel for the predicted and counterfactual TFP levels, based on OLS estimations of specification (3.1b). Effects of all variables.

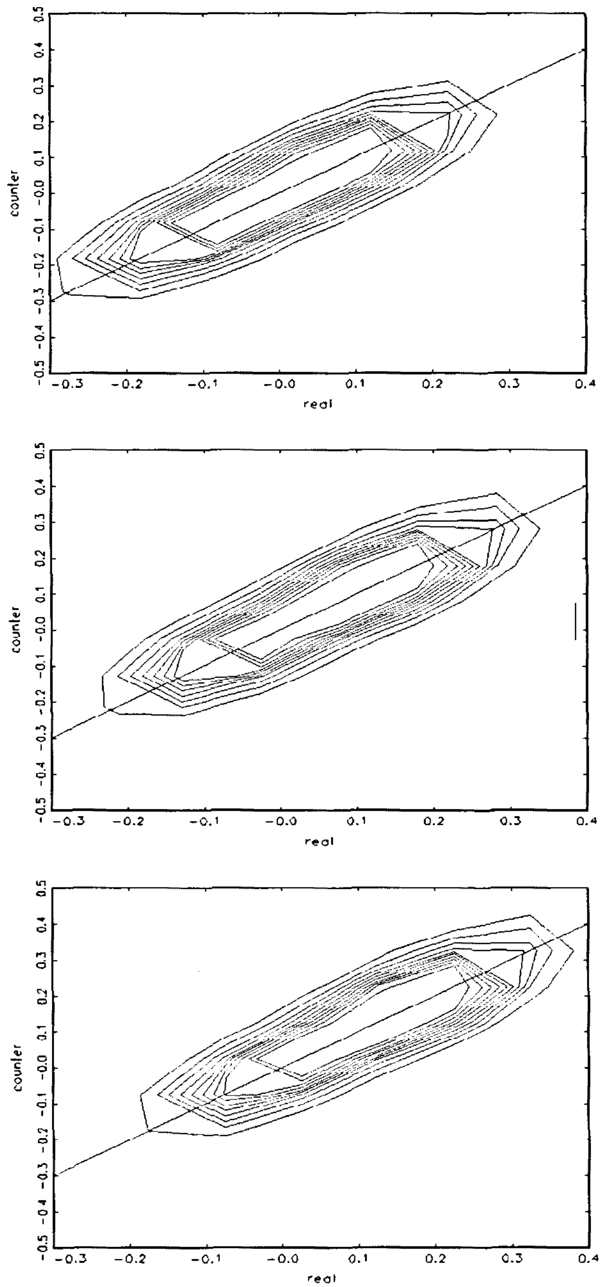


Figure A4.7. Stochastic kernel for the predicted and counterfactual TFP levels, based on OLS estimations of specification (3.1b). Effects of innovation and human capital

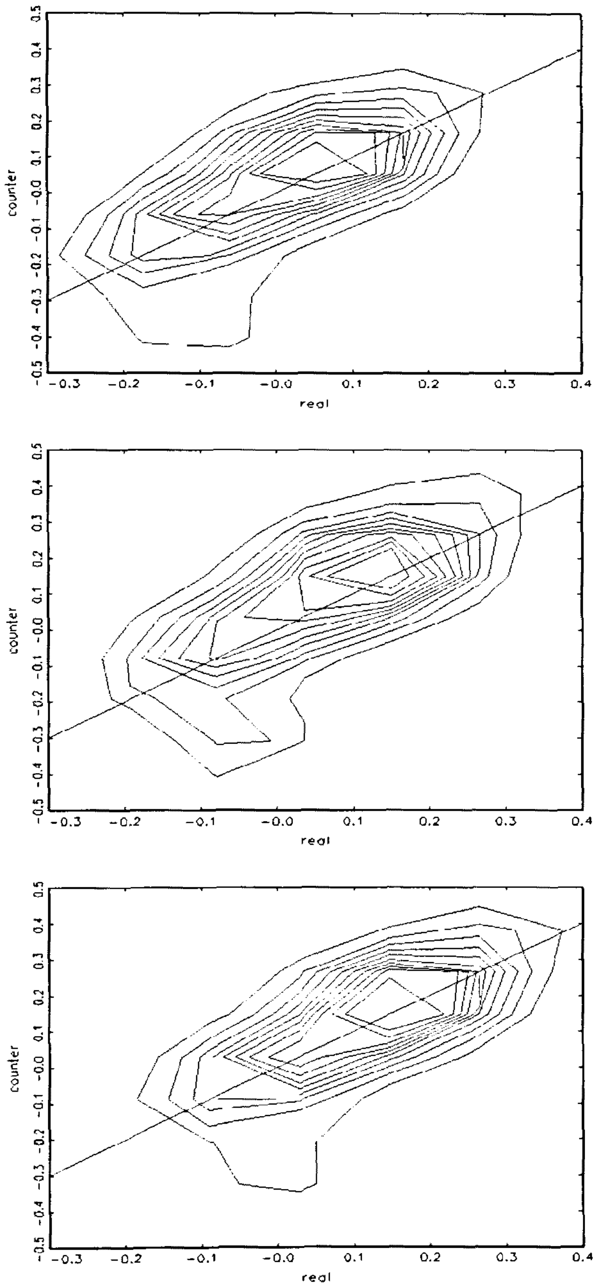


Figure A4.8. Stochastic kernel for the predicted and counterfactual TFP levels, based on RE estimations of specification (3.1b). Effects of all variables.

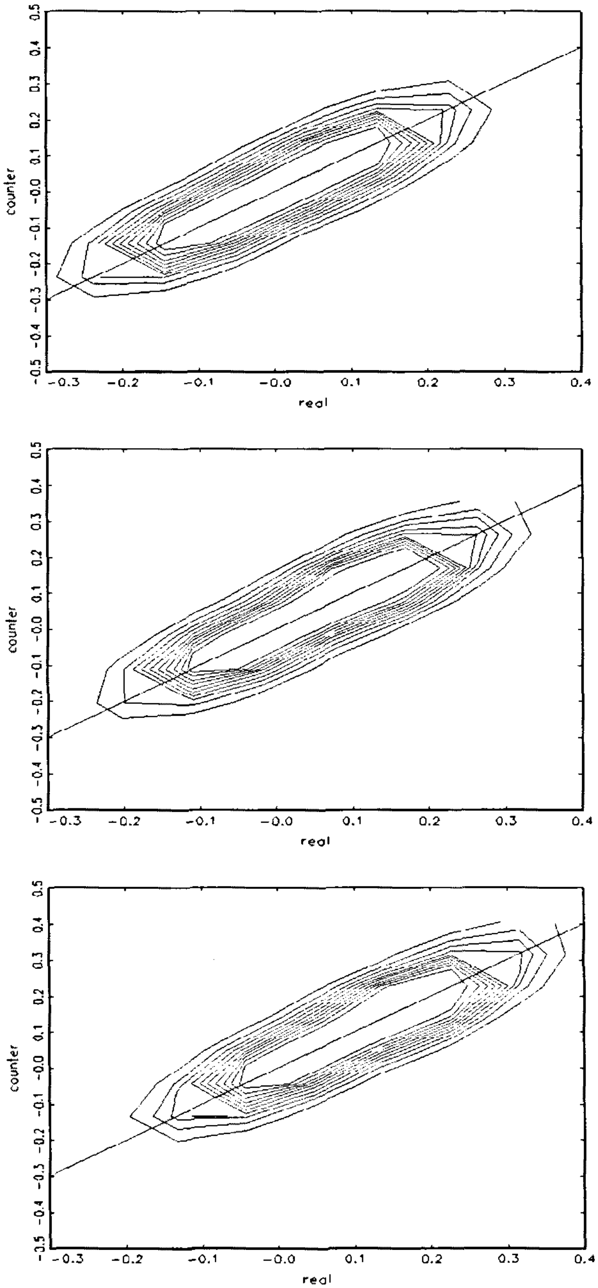


Figure A4.9. Stochastic kernel for the predicted and counterfactual TFP levels, based on RE estimations of specification (3.1b). Effects of innovation and human capital

CONCLUSIONS

In this Part II we studied the contribution of innovation and employees' qualification in explaining the differences in total factor productivity between small and large firms. Our hypothesis is that the TFP gap may not only be due to differences in the levels of innovation and human capital, but also to the returns to these factors. In this view, returns could play a central role in explaining the TFP differential.

In Chapter 3, we obtained that innovation and human capital have a significantly positive impact on productivity for manufacturing firms in the case of Spain. Similar results are obtained for the subsample of large firms. However, for small firms, these factors appear to have a smaller impact on productivity than for large firms or they are even non-significant. The results in Chapter 3 suggest that small and large firms seem to have different incentives to use these endowments, which could explain part of the TFP gap between small and large firms.

The contribution of differences in innovation and human capital and differences in returns to these characteristics to explain the average TFP gap is analysed using the Oaxaca-Blinder decomposition. We found that differences in the innovative activity between small and large firms seem to explain quite a small part of the gap. However, the differences in the human capital explain a larger part of the TFP differential, both as differences in the levels and differences in returns to these factors.

In Chapter 4, we transfer the idea of the Oaxaca-Blinder decomposition to the entire distribution by using the counterfactual distribution analysis. This analysis permits discovering that the contribution of differences in returns and differences in endowments are not homogeneous over the firms' TFP distribution. Another important result is that differences in returns to our variables of interest can explain only a small part of the TFP gap. Moreover, given the non-homogeneous behaviour, only the small firms with TFP levels above the average would improve their productivity when equalizing their returns to those of large firms.

A final comment is in order: the returns effect in the counterfactual analysis seems to be quite small in magnitude in relation to the results obtained in the Oaxaca-Blinder decomposition. However it should be noticed that the effects by the two methodologies are not directly comparable as the counterfactual analysis is based on the density functions of the TFP values, while the Oaxaca decomposition is based on TFP values.

PART III

DETERMINANTS OF FIRM-RELATED TRAINING IN SPAIN: THE ROLE OF FIRM SIZE AND SUBSIDIES

INTRODUCTION

The National Program for Spain in October 2005, framed in the Lisbon Strategy, highlights the necessity for Spain to increase and improve the quality of its human capital. The educational level of the Spanish labour force has considerably increased during the last decades. Concretely, the average years of education of the population in the private productive sector has increased from around 4 to 10 years (see for instance, López-Bazo and Moreno, 2007). Nowadays, almost 100% of the 16 year-old population has received formal education. Although the educational level of the Spanish labour force has considerably improved in the last three decades in relation to other advanced economies, this economy is still far from them. For example, the percentage of population with university studies over the population aged between 15 and 64 was 88% of the average EU-15 in 2004 (Gual et al., 2006).

However the qualification of the employees does not only depend on their schooling, but also on their life-long learning, which includes continuous and occupational training. Training is distinguished from formal school and post school qualifications (which are viewed as formal education) and is generally defined as courses designed to help individuals develop skills that might be of use in their job. The National Reform Program for Spain emphasizes the role of life-long learning as a key element for already occupied people to acquire knowledge and skills useful for their present and future employment, and for unoccupied people to reincorporate to the labour market. What is more, life-long learning also permits adapting workers' skills to the permanent evolution of job requirements and enhances the competitive position of workers and their employers. The main purpose of continuous training is to provide knowledge and adequate skills to occupied employees so that they could adapt to the changing requirements of firms at any moment. In this way, they become more competent and their professional performance is improved. This study is focused on continuous training provided by the firms to their employees.

Spain has a very low percentage of population aged 25-64 receiving continuous training: in 2003, this percentage was around 25%, while the average EU-25 is above 40% and Spain only performed better than Greece and Hungary. In 2004, 5.2% of the Spanish population received continuous training, while the average EU-15 is 10.7% and the average EU-25 is 9.9%.[91] According to the National Reform Program, a more intense effort regarding continuous training should be done, as it would help creating a more dynamic and competitive economy and it would contribute to workers' social integration.[92]

Since December 1992, organizations of workers and firms, as well as the Spanish administration have signed different agreements to impulse continuous training (*Acuerdos Nacionales sobre Formación Continua*, ANFC). Since the II ANFC in 1997, training policies are particularly concerned with certain collectives of workers that face more difficulties in keeping their employment and/or have more barriers to access training: this is the case of workers in small and medium firms (SMEs), disabled workers and employees above 46-years-old, women and unqualified workers.

The Tripartite Foundation for Employment Training (*Fundación Tripartita para la Formación en el Empleo*) is the national entity that supports and coordinates the execution of public policies aimed at improving continuous training in Spain. This entity is integrated by firms' organizations, trade unions and the Spanish government. These parties signed different agreements to encourage continuous training on December 1992, 1996 and 2000 (I, II and III National Agreements on Continuous Training, ANFC) and a reform on December 2003. These agreements established a system of subsidies to support and stimulate continuous training, which consists of

[91] National Reform Program for Spain (2005, pp 36, 68), from the Lisbon Strategy in March 2000.
[92] The National Reform Program has the objective of increasing the percentage of population that received training from 5.2% in 2004 to 10% in 2008 and 12.5% in 2010.

different training initiatives. In Part III, we focus on in-company training, which refers to those initiatives planned, organized and conducted by firms to improve their employees' skills. In 2001 and 2002, 64664 and 53324 firms obtained funds for providing subsidized training to their employees. Over 1.8 and 1.5 million workers participated in these actions and the public funds awarded to in-company training actions amounted to 364.63 million euros in 2001 and 378.69 in 2002.[93] The sources of these funds are European (European Social Fund, 24%) and domestic (firms and workers contributions to the Social Security System, 72%, and National Institute for Employment, INEM, 4%).[94]

In this Part III of the thesis, we study two different questions. Given that small firms are generally considered to have more difficulties in accessing training,[95] we intend to analyse the reasons for this more modest provision of training in small Spanish manufacturing firms. This first question is addressed in Chapter 5. The second question analysed here is whether the current training subsidies in our period of analysis, 2001 and 2002, had a positive impact on firms' provision of training. This second question is addressed in Chapter 6.

The two questions addressed here are analysed in the framework of a strand of literature that analyses the determinants of firm-related training. This approach basically estimates the impact of different firms' characteristics on their training provision decisions. Part III adds to the previous literature by paying special attention to the role of firm size and the subsidies for the case of Spain. A novelty of this study is considering the decision on the provision of training as a double-decision process and using the two-part models to estimate this question. Moreover, to the best of our knowledge, there are no empirical studies that perform a causal analysis on the impact of subsidies on continuous training for the Spanish manufacturing firms.

In Chapter 5 we address the question of why do small firms invest less in training. Departing from the idea that training is generally associated with certain firms' and employees' characteristics, we argue that large firms provide more training because they have certain characteristics that allow them to dedicate more efforts to training workers –such as having more qualified employees or less temporary workers– or that require

[93] Data drawn from the Tripartite Foundation web site: http://www.fundaciontripartita.org/
[94] Data for 2001 obtained from Otero et al. (2002).
[95] Fundación Tripartita para la Formación en el Empleo (2003), http://www.fundaciontripartita.org/

more training –such as technological activity or operating in international markets–. The hypothesis is that small firms are not associated with such characteristics or not with as much intensity as large firms, which could partially explain the differences in training provision between small and large firms.

On the basis of the different determinants of training suggested in the literature, we offer evidence that training is associated to certain firm characteristics and that among firms with such characteristics, large firms provide more training. As in the previous parts of this thesis, we use data drawn from the *Encuesta sobre Estrategias Empresariales* (ESEE). Next, we estimate the impact of these determinants in the firms' provision of training using a two-part model. These models consider the provision of training as a double decision process, where firms first decide whether to provide training or not and then, the quantity of it, if they do. Regarding the possible existence of sample selection biases, we consider the suitability of the *heckit* and the two-part model itself to model firms' decisions on training provision. After discussing the strengths and weaknesses of the two models, both theoretically and empirically, we finally select the two-part model. Given the previous evidence of important heterogeneity among firms, we estimate a two-part model with random effects. On the basis of the estimation for the small and large firms' subsamples, we use the Oaxaca-Blinder decomposition to analyse the differential in the provision of training by firm size –the differential in the probability of providing training and the differential in the quantity.

In Chapter 5, we confirm that small Spanish manufacturing firms face more restrictions in their access to training. The technological activity and the degree of competition of the markets where firms operate are the main reasons explaining the fact that small firms provide less training than their larger counterparts.

Using the same empirical framework, Chapter 6 intends to shed light on the impact of subsidies dedicated to increase the provision of training in Spanish manufacturing firms in 2001 and 2002, when subsidies were regulated by the III ANFC. The subsidies are defined as the percentage of hours of subsidized training over worked hours. Following the strategy in Chapter 5, we estimate a two-part model with random effects to assess the impact of subsidies on the decision of whether to provide training or not and its quantity.

A preliminary descriptive analysis shows that large firms receive more hours of subsidized training over worked hours. After controlling for a large set of training determinants as well as firm-specific effects, we do not find a clear positive effect of subsidies awarded to firms in 2001 and 2002 –neither to firms' probability to provide training, nor on their training expenditure. According with these results, we cannot ascertain that the economic policies dedicated to impulse training have had the expected positive result.