

Castany, Laia; López-Bazo, Enrique; Moreno, Rosina

**Conference Paper**

## DIFFERENCES IN TOTAL FACTOR PRODUCTIVITY ACROSS FIRM SIZE - A DISTRIBUTIONAL ANALYSIS

45th Congress of the European Regional Science Association: "Land Use and Water Management in a Sustainable Network Society", 23-27 August 2005, Amsterdam, The Netherlands

**Provided in Cooperation with:**

European Regional Science Association (ERSA)

*Suggested Citation:* Castany, Laia; López-Bazo, Enrique; Moreno, Rosina (2005) : DIFFERENCES IN TOTAL FACTOR PRODUCTIVITY ACROSS FIRM SIZE - A DISTRIBUTIONAL ANALYSIS, 45th Congress of the European Regional Science Association: "Land Use and Water Management in a Sustainable Network Society", 23-27 August 2005, Amsterdam, The Netherlands, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/117450>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# **DIFFERENCES IN TOTAL FACTOR PRODUCTIVITY ACROSS FIRM SIZE. A DISTRIBUTIONAL ANALYSIS.**

**Laia Castany, Enrique López-Bazo, Rosina Moreno**

**45th Congress of the European Regional Science Association  
23-27 August 2005, Vrije Universiteit Amsterdam**

AQR Research Group  
Universitat de Barcelona and Parc Científic de Barcelona  
Avda Diagonal 690, 08034 Barcelona  
Tel: 93 4021825 93 4037041 934021012 FAX: +34 93 4021821  
Email: lcastany@ub.edu; elopez@ub.edu; rmoreno@ub.edu

**Abstract:** Most studies on productivity at firm level have found notable heterogeneity between firms, especially between large and small firms. Such differences might be caused either by differences in the distribution of the factors determining the level of Total Factor Productivity (TFP) across firms' size, and by differences in the return to such factors. To assess to what extent the observed differences in TFP between large and small Spanish manufacturing firms are caused by the reasons mentioned above we propose a methodology that, built on the traditional Oaxaca-Blinder decomposition, focuses the attention on the entire distribution of productivity. The TFP index used in our paper guarantees comparison of the level of productivity across firms in a given year and over time, and has been computed using the information in the Encuesta sobre Estrategias Empresariales (ESEE). Results confirm that the distribution of TFP in the large firms dominates that for the small firms and that the TFP differences between small and large firms in 1994 are equally explained by differences in firm characteristics and in their returns, while in 1998, 80% of them are explained only by differences in their characteristics. The joint effect of differences in returns to R&D, human capital and industries is actually significant, suggesting possible interactions between them. In addition, the evidence suggests that small firms with the lowest TFP levels would get the most benefit if they had returns from their characteristics as high as in large firms. Important policy issues are derived in connection with the possibility of increasing the aggregate productivity of the Spanish economy considering that the average firm size in Spain is smaller than in other European countries.

**JEL codes: D24, J24, L25**

**Key words:** Total Factor Productivity; skilled labour; R&D; firm size; Oaxaca decomposition.

## 1. Introduction

The evolution of productivity has been one of the issues of major concern among economists, especially in the last years, when the productivity growth has slowed down in many advanced economies. The Spanish economy has also suffered a deceleration process since the mid nineties, as aggregate studies show and even some of them attribute such phenomenon to the behavior of manufacturing sectors<sup>1</sup>. A common recommendation indicates that Spain should increase its competitiveness through efficiency instead of reducing prices to guarantee sustained growth and this requires both a more intense usage of technologies and human capital and a higher investment in these two productivity determinants. On the other hand, an average firm size and a number of large firms that are smaller than in other economies characterize the Spanish manufacturing sector. This picture of the predominance of SMEs in our economy is quite disappointing as larger firms are usually associated to higher productivity levels<sup>2</sup>. A strand of microeconomic literature that analyses the heterogeneity of productivity behavior 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<sup>3</sup>.

This paper contributes to the empirical evidence that there are some factors such as the innovative activity and the use of skilled labor that foster productivity and that firm size plays an important role in explaining differences in productivity between firms. Then firm size conditions the effect of R&D and employees' qualification on productivity, so that size is *indirectly* affecting productivity. We use the Oaxaca decomposition, a methodology widely used in labor economics, as a tool to analyze the impact of these factors, among others, in Total Factor Productivity by firm size. This methodology will permit us to assess the relative importance of firm characteristics and their returns in explaining productivity differences between small and large firms.

This heterogeneity by size is not constant at any point of the Total Factor Productivity (TFP) distribution and it follows a quite complex pattern that can only be analyzed by means of an analysis in the complete distribution. Thus, the traditional Oaxaca

---

<sup>1</sup> See Estrada and López-Salido (2001b)

<sup>2</sup> Ruano (2002) comments that young firms tend to be smaller and less efficient as they have not undergone the market selection mechanisms and they have also a higher failure rate.

<sup>3</sup> See Fariñas and Ruano (2004) for an analysis of firm dynamics; Delgado, Fariñas and Ruano (2002) for the exports effect; Huergo and Jaumandreu (2004b), Ornaghi (2003) and Mánez *et al.* (2003) among others for the innovative activity.

decomposition is modified and a counterfactual analysis is introduced to appreciate the heterogeneous pattern of behavior of TFP along the distribution.

The main results obtained in this paper are the following: (i) Firm size represents a source of heterogeneity because, even when conditioning for variables other than size, large firms are still significantly more productive than small ones, which is confirmed by parametric and non-parametric testing procedures. (ii) The TFP differences between small and large firms in 1994 are equally explained by firm characteristics and their returns, while in 1998, 80% of them are explained only by their characteristics. (iii) Human capital contributes to enhance firms technical progress and its returns for small and large firms are significantly different. The innovative effort does not seem to exert a significant effect on productivity. The industry in which firms operate explain a great part of the differences in productivity between small and large firms. The returns of these three factors do not make them statistically significant in our counterfactual analysis, while the joint effect of them is actually significant, suggesting possible interactions between them. (iv) The analysis of movements between the real and counterfactual distributions indicates that small firms with the lowest TFP levels would get the most benefit if they had returns from their characteristics as high as in large firms.

The rest of the paper is organized as follows: in Section 2, there is a review of the theoretical hypotheses underlying our model, the specification and empirical methodology used to perform our analysis. In Section 3, we describe the ESEE, the variables that shape our model and a first descriptive analysis. Section 4, contains the results and Section 5 the conclusions of the paper.

## **2. Factors determining firms productivity**

### *2.1. Theoretical framework*

The vast majority of firm-level based works 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 share some characteristics and they are all supposed to be affected by the economic forces in a similar way. Thus, such models may not explain the observed differences in productivity levels between small and large firms adequately, while the microeconomic approach permits a deeper analysis of the characteristics that

may explain such differences in productivity. Some models in industrial organization try to model these heterogeneity, for example, Lucas (1978) proposed a theory of the size distribution of business firms<sup>4</sup>.

The empirical literature on productivity at firm level agrees in considering size as a main source of heterogeneity of firms performance. Large firms are systematically found to be more productive than small ones, and, as Geroski (1998) argues, controlling for firm size in regressions can be even considered as a routine. Some common theoretical arguments to explain this regularity are the scale economies effect, the scope economies effect, the experience effect and organization effect<sup>5</sup>.

These phenomena does not only explain the way size affects productivity by itself, but they also involve factors other than size that are recognized to affect productivity as well. For example, a large firm may count on economies of scale when designing and implementing new technologies or a training strategy. The innovative activity of a firm or its human capital endowment have traditionally been considered two factors fostering their productivity.

On the one hand, the effort in 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 that increases its efficiency, but also because it rises its absorptive capacity, that is, it becomes more flexible and adaptable to benefit from spillovers than its rivals. Many empirical studies estimating the output elasticity of R&D capital at aggregate level have found a strong positive correlation between productivity growth and R&D investment. However, 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<sup>6</sup>, especially when including industry dummies.

Other studies have gone one step further by trying to relate the probability of a firm engaging in R&D activities with firm characteristics, such as size, finding a

---

<sup>4</sup> His model consists on a distribution of people by managerial talent which underlies the distribution of business firms by size. The individuals may become either employees (working for someone else and earning a salary) or managers (taking managerial decisions and obtaining its returns). One implication of this model is that people become more productive and their wages increase, they will prefer to work for someone else, and thus firm size will increase.

<sup>5</sup> Scale economies effect is the decrease of average costs by increasing the volume of output. Scope economies effect is the decrease of average costs of a product when the number of different products increases (for complementarity, interaction and indivisible resources). The experience effect is the fall in average costs as the volume of output accumulated over time increases. The organization effect takes place when smalls firms intend to reduce these effects by creating networks but then they face transaction costs (Carree and Thurik, 1998).

<sup>6</sup> See Mairesse and Sassenou (1991) for a survey, and Ormaghi (2003) for estimates of the returns to R&D for the Spanish manufacturing firms.

positive relation between them<sup>7</sup>. An application for Spain that uses innovative output instead of R&D effort obtains the same results and concludes that “innovation is strikingly related to size” (Huergo and Jaumandreu, 2004a). Klepper (1996) introduced a theoretical model that represents a new interpretation of the Product Life Cycle and that emphasizes the role of firm size in its appropriating the returns from innovation and engaging in R&D activities. The initial hypothesis is that innovating derives in a unitary cost reduction (in proportion to the volume of output). 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 promotes additional innovative effort (in both product and process R&D). One main implication of this model is that large firms have higher R&D returns and make a greater effort in R&D<sup>8</sup>.

On the other hand, the literature on 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 manner. Then, if education is translated into higher learning capacity to solve problems and to communicate, those workers with better education would be the most productive. The aggregate literature on the effects of human capital on productivity clearly recognizes its positive effects, although there is not consensus on whether it affects the long-run productivity level or the short-run productivity growth (rate effect). An interesting finding of this studies is that human capital has an effect on TFP by itself and also through enhancing the ability to develop and incorporate new technologies and capture knowledge spillovers at an international level, so that human capital would be a prerequisite for the incorporation of such technology<sup>9</sup>. The microeconomic literature, conscious of the effect of investing in human capital on productivity levels, has typically estimated mincerian equations<sup>10</sup>. Nevertheless a few studies have considered estimating such effect at firm level, and most of them approach the concept of human capital as training instead of education, as

---

<sup>7</sup> See Crépon *et al.* (1998) who obtain that “The probability of engaging in R&D for a firm increases with its size, while R&D capital intensity is strictly proportional to size” (p 15); Cohen and Klepper (1996) obtain similar results.

<sup>8</sup> The author argues that large firms have no innate advantage on innovation, but their capacity to capture the returns of R&D represents an incentive to dedicate their effort to it. Otherwise, if their knowledge was spread in the whole economy and they could not sell it (when selling their product) they could not capture its returns.

<sup>9</sup> See for instance Benhabib and Spiegel (1994) and Del Barrio-Castro, López-Bazo and Serrano-Domingo (2002).

<sup>10</sup> Acemoglu and Angrist (1999) estimate the private returns of education.

in our case. An exception would be the work by Black and Lynch (1996) who estimate a firm-level production function and include variables as average education, training and experience for the US economy. As far as we know, there are no studies analysing the role of firm size in determining either the way firms capture the returns to their employees' education, nor their decisions on investment in education. Even though, some studies have documented that workers in large firms earn higher wages, which, under the neoclassical hypothesis would mean that they are more productive. Thus, we are interested in analysing if the same effect happens at firm-level, that is, if returns to education are higher in large firms.

Finally, the industrial effect is recognized to capture a great deal of heterogeneity between firms, especially in relation with innovative activity, and then it is common practice to include industry dummies in the regression analysis. These control variables collect the effect that some sectors are more innovative than others by their own nature.

Given the evidence on the different patterns of productivity between small and large firms, and on the different impact that employees qualification and innovative effort may have on productivity according firm size, this paper analyses to what extent the productivity differentials between size class are due to both the levels in these firm characteristics and to the returns of such characteristics.

Geroski (1998) claims that size may have a direct effect on productivity, that is as a variable that *ceteris paribus* improves efficiency, or indirect, 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 analyzing separately the coefficients of small and large firms and evaluating to what extent they differ. Differences in the returns of firm characteristics (such as R&D and human capital) between small and large firms indicate that, even if two firms did a similar R&D effort or invested the same in human capital, the larger, for example, would make a more intense use of these two factors and would obtain higher returns from its investment. In this sense, Geroski argues that size is exerting an indirect effect on firm productivity, as it conditions the impact of other factors on productivity. Bearing this in mind, one of the main contributions of the present analysis is using different methodological approaches to assess the impact of different firm characteristics on productivity, and then, analyzing to what extent the heterogeneous pattern of productivity can be accounted for by either levels of these variables or by

their returns (supposing that the latter follow different patterns in small and large firms as suggested by Geroski).

## 2.2. Empirical Specification and Methodology

According with the theoretical arguments explained above, the specification of our analysis is a linear regression model where a TFP index is a function of the level of skilled labor and the innovative activity performed by every firm. We also control for industrial and temporal effects, as well as size and age, as the majority of the empirical literature agrees that they are narrowly related with firm behavior and characteristics. Hence, the empirical model can be expressed as follows:

$$TFP_{it} = \beta_0 + \beta_1 R \& D_{it-4} + \beta_2 H_{it-4} + \beta_3 SIZE_{it} + \beta_4 AGE_{it} + \beta_5 IND_{it} + \beta_6 YEAR_{it} + u_{it} \quad (1)$$

where  $TFP_i$  is the logarithm of the total factor productivity index in firm  $i$ ,  $R\&D$  is the innovative activity,  $H$  is the proportion of skilled labor,  $SIZE$  controls for the size of the firm<sup>11</sup>,  $AGE$  controls for the effect of the antiquity of firms in the industry<sup>12</sup>,  $IND$  are 19 industrial dummies,  $YEAR$  is a time dummy and  $u_i$  is an error term. The innovative activity and human capital variables enter the model with a lag of four periods for two reasons: first, because investing on R&D or employing a new qualified worker may not have an effect on productivity until some periods later. And second, because lagging this variables will minimize the effect of simultaneity problems caused by the likely endogenous character of these variables<sup>13</sup>.

We apply two different empirical methodologies to study the TFP differentials between small and large firms. On the one hand, we have performed the Oaxaca decomposition, which has widely been used to study the wage differentials between different types of workers, but in this paper it is applied to analyze productivity differentials at firm level. This methodology permits to analyze how much of the TFP differential between small and large firms can be explained by either, differences in their characteristics or differences in the returns of these characteristics.

---

<sup>11</sup> Firm size is not significant in the estimations for two separate subsamples and has not been included as control variable finally.

<sup>12</sup> We have contemplated the possibility of including  $AGE$  with a nonlinear relation in the model, to capture the effect by which young firms, being very productive, may enter the industry, while old firms accumulate a great deal of experience, which may also make them much more productive (see Huergo and Jaumandreu, 2004b). However, the linear specification was preferred in all cases.

<sup>13</sup> Hall (2000) claims that considerable lags may exist between R&D expenditures and the impact of innovation productivity. Hall and Mairesse (1995 p.278) and Sianesi and Van Reenen (2003 p.14) explain the likely endogeneity of the R&D and human capital stocks respectively in the production function.

On the other hand, we perform a counterfactual analysis, which will permit us to make hypothesis on the estimated TFP of firms under certain circumstances. Let us suppose that firms from a given size class have characteristics or returns from another size class and obtain a “virtual” estimated TFP, which can be compared with the “real” estimated TFP. With this methodology, one could, for example, compare TFP in small firms with TFP in small firms under the hypothesis that the returns to their characteristics were the same as in large firms.

A preliminary version of this methodology, which is very close to the traditional Oaxaca decomposition, evaluates such differences in the mean of the TFP distribution. Although it seems attractive to summarize the TFP differentials within a single number, for example the mean, Jenkins (1994) recognizes that these synthetic measures represent a big lost of information because they do not allow evaluating such differences along the whole TFP distribution and because the same statistical may be consistent with very different distributions<sup>14</sup>. This author suggested a modification of the traditional counterfactual analysis that permits studying TFP differentials in the complete distribution<sup>15</sup>.

The difference between the two methodologies is that the Oaxaca decomposition permits identifying the relative importance of firm characteristics and returns very easily, while the second one allows hypothesizing on different virtual contexts and permits, not only analyzing differences in the mean, but also in the whole distribution. We describe the two approaches next.

To perform the classical Oaxaca decomposition, we first estimate the regression model in (1) that relates TFP with a set of firm characteristics, which are, *a priori*, considered to determine firm productivity and a set of control variables. We estimate two separate regressions, for small and large firms, and obtain the returns of these characteristics in both size classes. The estimated regression models are (we omit the firm subscript to simplify the notation):

$$TFP_{Si} = X'_{Si} \beta_S + u_{Si}$$

$$TFP_{Li} = X'_{Li} \beta_L + u_{Li}$$

---

<sup>14</sup> By applying a similar argument to that used by Jenkins, if TFP differences attributable to firm size is 10%, one could extract dramatically different conclusions whether it occurred in the whole distribution or only in the first deciles (where small firms show very low TFP).

<sup>15</sup> An application for the case of wage differentials can be found in Del Río, Gradín and Cantó (2004).

where  $X_i$  is the vector of firm characteristics in specification (1),  $\beta$  are the returns of these characteristics,  $u_i$  is the error term and subscript  $L$  refers to large firms and  $S$  to small firms.

Secondly, we decompose the average TFP differential between small and large firms in two components: the first one attributes part of the TFP differential to the different characteristics of a representative small and large firm, or in other words, the observed productivity differential if characteristics of small and large firms had the same returns than large firms. The second component attributes the remaining part of the TFP differential to differences in returns between small and large firm under the hypothesis that they have the same characteristics<sup>16</sup>.

$$\mathbf{TFP}_L - \mathbf{TFP}_S = \underbrace{(\bar{X}_L' - \bar{X}_S') \hat{\beta}_L}_{\text{diff in charact}} + \underbrace{\bar{X}_S' (\hat{\beta}_L - \hat{\beta}_S)}_{\text{diff in returns}}$$

The Oaxaca decomposition evaluates the productivity differential in the mean of the distribution of characteristics, that is, comparing the mean characteristics of small and large firms.

Our second approach is based on a counterfactual analysis, the simpler version of which also analyses TFP differences in the mean of the distribution. It simply compares the actual level of estimated TFP for small firms with a counterfactual level obtained by applying the returns in large firms the characteristics of small firms.

$$\text{Real TFP: } \overline{\mathbf{TFP}}_S = \overline{X}'_S \hat{\beta}_S$$

$$\text{Counterfactual: } \overline{\mathbf{TFP}}_S^L = \overline{X}'_S \hat{\beta}_L$$

A t-test of equality of means is used to check for equality between the two means. But following the suggestion by Jenkins, we can rewrite these expressions to analyze the differences in the complete distribution:

$$\text{Real TFP: } \mathbf{TFP}_S = X'_S \hat{\beta}_S$$

$$\text{Counterfactual: } \mathbf{TFP}_S^L = X'_S \hat{\beta}_L$$

We obtain the estimated TFP for small firms,  $\mathbf{TFP}_S$ , as well as the hypothetical TFP of small firms if their characteristics were as efficient as in large firms,  $\mathbf{TFP}_S^L$ . We then

---

<sup>16</sup> The Oaxaca decomposition is based on the least squares estimation of specification (1). Hence, it has the property of “the mean of OLS estimators obtained from the regression equations” which guarantees that the estimated log of TFP evaluated in the mean characteristics is equal to the observed mean of the log of TFP. That is, this property guarantees that  $\overline{\mathbf{TFP}}$  equals  $\overline{\mathbf{TFP}}$ .

check for equality of both distributions by applying a battery of Kolmogorov-Smirnov tests. This procedure permits comparing the productivity distributions of different groups of firms and establish a ranking between them on the basis of the concept of first order stochastic dominance<sup>17</sup>. Let's suppose two independent random samples of size  $n$  and  $m$ . Let  $Z_1, \dots, Z_n$ , be a random sample corresponding to a group of firms from the cumulative distribution function  $F$ , and  $Z_{n+1}, \dots, Z_{n+m}$ , 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:  $F(z)-G(z) \leq 0 \quad \forall z \in \mathfrak{R}$ , with strict inequality for at least one  $z$ . The hypotheses we are testing are:

(i) Two sided test

$$H_0 : F(z) - G(z) = 0 \text{ all } z \in \mathfrak{R} \quad \text{vs} \quad H_1 : F(z) - G(z) \neq 0 \text{ some } z \in \mathfrak{R}$$

(ii) One-sided test

$$H_0 : F(z) - G(z) \leq 0 \text{ all } z \in \mathfrak{R} \quad \text{vs} \quad H_1 : F(z) - G(z) > 0 \text{ some } z \in \mathfrak{R}$$

In our case  $F$  and  $G$  represent the productivity distributions for small firms,  $F(\hat{TFP}_S)$ , and small firms evaluated under the returns of large firms,  $G(\hat{TFP}_S^L)$ , respectively. The hypotheses can be rewritten as:

(i) Two sided test

$$H_0 : \sup_{z \in \mathfrak{R}} |G(\hat{TFP}_S^L) - F(\hat{TFP}_S)| = 0 \quad \text{vs} \quad H_1 : \sup_{z \in \mathfrak{R}} |G(\hat{TFP}_S^L) - F(\hat{TFP}_S)| \neq 0$$

(ii) One-sided test

$$H_0 : \sup_{z \in \mathfrak{R}} \{G(\hat{TFP}_S^L) - F(\hat{TFP}_S)\} = 0 \quad \text{vs} \quad H_1 : \sup_{z \in \mathfrak{R}} \{G(\hat{TFP}_S^L) - F(\hat{TFP}_S)\} > 0$$

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, we conclude that  $F$  dominates  $G$ , and thus,  $G(\hat{TFP}_S^L)$  is on the right of  $F(\hat{TFP}_S)$ .

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

---

<sup>17</sup> This strategy has been recently applied in Delgado, Fariñas and Ruano (2002) to check for higher productivity among exporting firms.

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

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,  $\delta_N$  and  $\eta_N$ , are known under independence<sup>18</sup>.

### 3. Dataset

#### 3.1. Description of the Dataset and Variables

We use a sample of manufacturing Spanish firms from the *Encuesta de Estrategias Empresariales* (ESEE), carried out by the *Programa de Investigaciones Económicas* (PIE) of the *Fundación Empresa Pública* (FUNEP). This annual survey covers the period 1990-2001 and collects information on strategic decisions and behavior of firms. Every four years, firms answer a complete questionnaire and a reduced form of it (with those issues that are supposed to change yearly), the rest of the years, so that full information so far is available in 1990, 1994 and 1998. 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 CNAE-93, excluding division 23 (activities related to refinement of oil and fuel treatment). 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 six size strata, retaining about 5%, so that representativity for every industry and firm size class was guaranteed. The ESEE is designed to change as industry composition evolves. Newly created firms are selected using the original selection criteria. There are also exits in the survey, due to death and attrition, and these firms have been replaced by others in their industry and size group so as to maintain representativity. So, the ESEE is an unbalanced panel that attempts to capture the entry and exit of manufacturing firms over the sample period, which guarantees comparability between the TFP distributions in different time periods<sup>19</sup>.

The variables contained in this survey will permit us to build the output and input quantities and prices required to finally calculate a TFP index, as well as the

---

<sup>18</sup> Kolmogorov (1933) and Smirov (1939) showed that, under the assumption that all observations are independent, the limiting distributions of  $\delta_s$  and  $\eta_n$  under  $H_0$  are given by

$\lim_{s \rightarrow \infty} P(\delta_n > v) = -2 \sum_{k=1}^{\infty} (-1)^k \exp(-2k^2 v^2)$  and,  $\lim_{s \rightarrow \infty} P(\eta_n > v) = \exp(-2v^2)$ , respectively.

<sup>19</sup> See Fariñas and Jaumandreu (1999) for further details.

variables to account for its main determinants. Between 1990 and 2001, this survey has 33668 observations, for 3451 different firms. However it has not been possible to include them all in the analysis due to the following reasons: series of the stock of capital are not available for the last two years, some firms do not offer full information to build some of the variables to calculate the TFP index or their answers seem to be “nonsense” according with an economic criterion and so they have been removed (see the Annex for a full explanation of the cleaning procedure). After removing these observations, our sample consists on 10653 observations over 10 years. Data on skilled labor is only available every four years and so our analysis is restricted to 1990, 1994 and 1998, among which we lose the initial year because some variables enter the model with a lag of four periods. So our final dataset consists on an unbalanced panel of 523 firms for 1994 and 668 firms for 1998.

Total factor productivity is measured by the index suggested by Good, Nadiri and Sickles (1996), which is transitive, superlative (as it is based on a translog production function) and permits accounting for technological change. The analytical expression of this index is as follows:

$$\ln TFP_{it} = (\ln Y_{it} - \overline{\ln Y_t}) - \frac{1}{2}(S_{it} + \overline{S_t})(\ln X_{it} - \overline{\ln X_t}) \\ + \sum_{s=2}^t (\overline{\ln Y_s} - \overline{\ln Y_{s-1}}) - \sum_{s=2}^t \frac{1}{2}(S_s + \overline{S_{s-1}})(\overline{\ln X_s} - \overline{\ln X_{s-1}})$$

where  $Y$  is the quantity of output,  $X$  is a vector of the quantity of inputs (labor, capital and materials),  $S$  is a vector of the cost-based share of every input in the production function, as Hall (1990) suggested; subscripts  $i$  and  $t$  refer to firm and time period and the bar over the variables denotes their arithmetic mean. The productivity index for a given firm and year is expressed in relation to a hypothetical firm, calculated as the average of all firms in 1990. The logarithm of TFP equals zero for that firm in 1990, firms with lower productivity will show negative values and those with higher productivity, positive.

The output is measured by sales, the labor input is measured by number of hours, the capital input is measured by the stock of capital in equipment and constructions, and materials are measured by expenditure in intermediate inputs and services. As input shares are cost-based, we need information on the expenditures on each of these inputs: the cost of labor input is measured by the real wage bill, the cost of capital is measured by the user cost of capital and the cost of materials is calculated as

the quantity (see the Appendix for a complete explanation of the index main characteristics and building of its variables).

The regressors and control variables are defined as follows:

- The *innovative activity* is measured as the logarithm of the R&D expenditure expressed in constant pesetas of 1990 and weighted by the number of employees. Weighting R&D expenditure supplies a more reliable measure of innovative intensity as this measure is standardized by the relative presence of small and large firms in the economy.
- The *skilled labor* is measured by the ratio of qualified workers according to their education level. The category of qualified workers includes the employees with bachelor or higher degree level o studies.
- *Firm size* is measured by the number of employees.
- *Age* is the number of years since the constitution of the firm.
- *Industry* is a dummy variable for very industry as defined by the ESEE. The omitted category is “Other manufacturing industries”.
- *Year* is a dummy variable that takes value 1 in 1994 and 0 in 1998.

We consider that large firms are those with more than 200 employees, and SMEs are those with 10 to 200 employees. Although one could consider that firms between 200 and 250 employees are still SMEs we have established the mentioned separation to guarantee representativity by size strata, according with the characteristics of the ESEE. Our sample consists on an unbalanced panel data for years 1994, with 523 firms, 35% of which have more than 200 employees, and 1998, with 668 firms, 25% of which are large firms.

### 3.2. Descriptive Analysis

This section contains a descriptive analysis of the distribution of the TFP index and of the distribution of the other variables in the analysis. In doing so, we use synthetic measures and tools that allow comparisons of the complete distribution.

According with the macroeconomic literature<sup>20</sup>, Table 1 shows that TFP at firm level has increased during the period of analysis and, as quartiles show, this phenomenon takes place in the upper part of the distribution of small firms especially.

---

<sup>20</sup> See Estrada and Lopez-Salido (2001b)

As one might expect, this table shows that mean TFP is higher in large firms, which is confirmed by the test of equality of means that rejects that mean TFP is equal in the two subsamples<sup>21</sup>. Quartiles at 25%, 50% and 75% show that TFP is always higher in large firms, and moreover they permit appreciating that these differences reduce for the most productive firms and especially in 1998, when they are smallest. Thus, it confirms that differences in TFP across firm size are not homogeneous in the whole range of TFP levels. To shed some more light on this issue, we analyse the external shape of the TFP distribution for small and large firms by estimating the density function (using the non-parametric kernel method)<sup>22</sup>. Figures 1 and 2 show the empirical productivity distribution function that supplies additional information on the behaviour of TFP: by observing synthetic measures, one may think that TFP is always higher in large firms, but these figures show that in 1998 the most productive firms do not show this pattern. The test of stochastic dominance confirms the picture that these distributions are different and that TFP in large firms is higher (Test 1 in Table 7). Table 1 also shows that differences in TFP between large and small firms remain quite stable over time. In general terms, large firms seem to be more productive than small ones, which has widely been attributed to differences in R&D and human capital between firms.

The variables innovative activity and proportion of qualified workers are lagged four periods in our model, so this descriptive analysis corresponds to years 1990 and 1994. As Tables 2 and 3 show, our underlying hypotheses are confirmed by data: large firms invest more in R&D and make a deepest use of human capital, as is supported by the test of equality of means). Both small and large firms increase their R&D investment and human capital endowment during the period of analysis, but in the case of human capital, such increase is much more important in large than in small firms. It should be noticed that small firms in the three lower quartiles do not invest in R&D, while it only happens in the first quartile in the case of large firms. The differences in innovative activity between small and large firms are wider for the most innovative firms. Differences in skilled labour endowments between small and large firms double between 1990 and 1994 (and the highest increase takes place in the upper part of the distribution of this variable).

---

<sup>21</sup> Approximation suggested by Welch, as explained in Ruiz-Maya and Martín-Pliego (1995).

<sup>22</sup> We use a Gaussian kernel and a bandwidth estimated by applying the plug-in method proposed by Sheather and Jones (1991).

#### 4. Results

As the first step in our analysis, we estimate the empirical model in (1) for the whole sample of firms in 1994 and 1998. Results are summarized in the first column of Table 4. As expected, the coefficient for skilled labor is positive and significant, and large in magnitude. A 10% increase in firm's use of skilled workers causes a 28.7% increase in TFP of the average Spanish manufacturing firm. This is consistent with those studies that have revealed its contribution to enhance technical progress by the so-called level effect, although some other aggregate studies have cast doubt on the rate effect of human capital<sup>23</sup>. As for innovative activity, the elasticity of the R&D variable is positive but not significantly different from zero. This suggests that R&D investments are not among the list of factors that explain differences in productivity across firms in the Spanish manufacturing sectors<sup>24</sup>. Differences in TFP levels are strongly related to the industry in which firms operate. The set of coefficients for the industrial dummies are jointly significant and explain an important amount of the dispersion in firms' productivity.

The coefficients for the other control variables, the size of the firm and its age, are significant, though in the case of size it is only at 10%. Productivity level is increasing with age and, what is more interesting in our analysis, with size. This confirms that, still when conditioning to the other variables, large firms are more productive than the small ones, motivating our analysis for the two separate samples.

Results for the estimation of the samples of large and small firms are reported in the second and third column of Table 4. In both groups of firms, the innovative activity does not seem to exert a significant effect on TFP, while the conclusion derived from the whole sample regarding the role of skilled labor is confirmed for small and large firms. Moreover in the case of the separate sample, it can be clearly observed how the return associated to the use of highly qualified workers in small and large firm is quite different. An increase of 10 points in the proportion of skilled workers in a representative small firm is linked with an increase of 22% in its TFP level. But the same increase in a representative large firm causes a much stronger effect in its TFP level (a 40%). Therefore, the incentive to use highly qualified workers seems to be

---

<sup>23</sup> See Temple (2001)

<sup>24</sup> Mañez *et al.* (2003) obtain that the small firms that invest in R&D are more productive, while the same result is not true for large firms. We have checked the sensitivity of the estimate of the effect of "innovative activity" to the use of different measures of R&D that can be built from the information provided by the ESEE. Evidence favoring a positive effect of R&D has not been obtained in any case. These results will be provided upon request.

stronger in large firms, as the return they get almost double the one that is obtained by the small ones.

To shed some more light into the knowledge of the differences in the level of TFP between small and large firms, and to disentangle the contribution of the determinants of productivity to such differences, we use the estimated coefficients for both samples to compute the Oaxaca decomposition. Further, we compute the contribution of differences in characteristics and in returns between small and large firms for 1994 and 1998 to assess the effect of changes caused by the implementation of new technologies, that are likely to demand, for instance, a deeper use of human capital.

Results for the traditional Oaxaca decomposition are summarized in the first row of Table 5. In 1994, differences in productivity are equally explained by firm characteristics and by their returns. But surprisingly characteristics explain as much as 80% of the differences in productivity in 1998. Differences in TFP attributable to R&D seem to be due to returns of R&D basically, but we should take it cautiously as coefficients of this variable were not significant in the estimated models. In 1994, differences in TFP attributable to human capital are equally explained by firm characteristics and returns, but in 1998, returns remain quite stable while differences in human capital endowment have doubled. Differences in TFP attributable to the industry are the most important factor explaining TFP differences and they are basically due to firm characteristics, suffering a noticeable increase between 1994 and 1998.

The coefficients from the estimation of the two subsamples are also used to perform the counterfactual analysis. Firstly, we calculate the mean estimated TFP for small firms imposing the returns of large firms for all the variables. Secondly, we repeat the exercise imposing the returns of large firms for only one of the variables at a time, R&D, H or IND. The test of equality of means (Table 6) rejects the null hypothesis that the mean TFP distribution for small firms is equal to the mean of the counterfactual TFP distribution when we impose the coefficients of large firms for all the variables. Then, we can say that the real and the counterfactual TFP distribution for small firms have different means, meaning that differences in returns are significant at explaining TFP differences between firms. When we concentrate in one of the variables, this result holds for R&D and H in 1998, while in the other cases the means of the real and counterfactual distributions do not differ so much.

However, we have stressed in previous sections that comparison of synthetic measures of two or more distributions might only provide with a partial picture of

differences in their fundamental characteristics. Thus, following the suggestion of Jenkins (1994), we analyze differences in the external shape of the real and counterfactual distributions by comparing their density functions. Figures 3 and 4 show the estimated density functions in 1994 and in 1998. If TFP distribution for large firms (L) was on the right hand side of TFP distribution for small firms (S), but the counterfactual distribution (C) was between them, we could say that the distance between S and C is due to differences in returns between small and large firms, while the distance between L and C corresponds to differences in characteristics between firm size classes. This is a paradigmatic case and it corresponds to the results obtained from the analysis based on mean TFP, but actually, this explanation does not hold for any value of the TFP distribution, and so we should analyze the density functions more carefully. In 1994 and 1998, contrary to what the paradigmatic case indicates, the small firms with intermediate levels of TFP get lower productivity when they are evaluated under the returns of large firms. In 1998, the small firms with high levels of TFP get similar productivity when they're evaluated under the returns of large firms than when they're evaluated under their own returns (the real and counterfactual distributions are quite close), indicating that for these firms, the differences in TFP are mainly explained by differences in firm characteristics.

We now turn to the analysis of each potential factor explaining TFP differences between small and large firms separately. Figure 5 and 6 suggest that both in 1994 and 1998, the real and counterfactual TFP density functions are very close, indicating the small effect of differences in R&D returns, in accordance with the small magnitude of its effect in the regression analysis. The remaining TFP differences can be attributable to differences in R&D levels and to other factors. Either the returns of human capital do not seem to explain much of the differences in TFP between firm size classes, except for firms with high TFP levels, where this effect is quite notorious. The Oaxaca decomposition finds a notable effect of returns of human capital but the analysis for the whole distribution permits attributing most of this effect to firms with high productivity levels (see Figures 7 and 8). When isolating the industry effect, we should distinguish: firstly, firms with low levels of productivity, where the effect of returns of firms for belonging to a certain industry is unclear; secondly, firms with intermediate and high levels of productivity, where TFP of small firms evaluated under the returns of large firms is lower than the real TFP of small firms. One possible explanation for this effect is that many firms in this productivity levels may belong to industries with low

productivity and with large firm size. Thirdly, small firms with the highest productivity levels, where the effect of returns of firms for belonging to a certain industry explains a large proportion of TFP differences between small and large firms (see Figures 9 and 10).

The Kolmogorov-Smirnov test of stochastic dominance confirms the previous conclusions based on a visual inspection of the density functions. The test compares the real and counterfactual density functions, that is, TFP of small firms evaluated under their own returns and under those of large firms. First, we perform this test considering the joint effect of the whole set of factors likely affecting the level of TFP. The two-sided test rejects the null hypothesis that the real and counterfactual density functions are equal in favor of the alternative. Then, we perform the one sided test, where the null hypothesis is that the two distributions are equal, versus the alternative, where the difference between them is positive for the whole range of TFP levels. The alternative hypothesis is specified as the difference between the counterfactual and real density functions, and results clearly indicate that the null cannot be rejected. The test indicates that the counterfactual distribution stochastically dominates the real one and thus that the difference in returns has a statistically significant effect. This evidence points that difference in returns across firm size is responsible for, at least part, of the difference in TFP level in the whole range of values of productivity. We also test the stochastic dominance of the real TFP distribution for large firms over the counterfactual, as defined above, for the whole set of factors. The two-sided test rejects the null, while the one-sided test cannot, meaning that the TFP distribution of large firms under their own returns stochastically dominates the TFP of small firms evaluated under the returns of large ones, and thus that differences in characteristics have a significantly statistical effect, as we obtained with the Oaxaca decomposition (see Table 7).

Secondly, we perform the test for R&D, H and IND separately. The null hypothesis cannot be rejected in any of the three tests and so the counterfactual distribution does not seem to stochastically dominate the real one, which indicates that the returns of these variables separately are not statistically significant to explain differences in TFP between small and large firms. This is an interesting result, as it suggests that the factors under analysis might be interacting to originate differences in TFP levels in firms of different size.

The comparison of the external shape of the real and counterfactual distributions is informative of the changes in the shape of the distribution caused by differences in

returns, but tells nothing about which firms have played the major role in causing such changes. To get some information on this phenomenon, we estimate the bivariate density function for the real and counterfactual distributions. Figures 11 and 12 show the contour plots associated with the estimated bivariate density in 1994 and in 1998. Their lines indicate pairs of values in both distributions with the same probability, the most external lines correspond to low probability pairs of values (and the internal, high probability). When the mass of probability lies on the positive diagonal, it indicates that there is a high probability of reaching a similar TFP level when they are evaluated either under the returns of large or small firms. When these lines move upward and to the left, parallel to the diagonal, it indicates that there is a high probability of small firms to reach a higher TFP level when they are evaluated under the returns of large firms. One can observe that most movements within the distributions are due to small firms improving their TFP levels when their characteristics are evaluated at the returns of large firms. For levels of TFP around and above the average, the probability of increasing or decreasing the level of TFP is very scarce. Thus, we can conclude that small firms with the lowest TFP levels would be the firms that would benefit more of equalizing returns between firms of different size.

## **5. Conclusion**

This paper has explored the differences in TFP across firms size for the Spanish manufacturing sector. Our analysis has intended to assess to what extent can such differences be explained by differences in firms characteristics, such as R&D efforts or human capital, and by the returns to these characteristics.

We obtain that TFP differences between small and large firms in mid-nineties are equally explained by firm characteristics and their returns, while at the end of the decade, 80% of them are explained only by differences in their characteristics. The joint effect of the returns to R&D, human capital and industries is actually significant, suggesting possible interactions between them. We have also shown that small firms with the lowest TFP levels would get the most benefit if they had returns from their characteristics as high as in large firms.

If differences in TFP were only explained by firm characteristics, the policy recommendations would consist on encouraging the investment in R&D and human capital in small firms. While if differences in productivity are only explained by differences in returns, this would require a deeper analysis to understand the underlying

nature of phenomenon that makes these two factors more productive when they are employed in large firms. In this case, the policy implications would not consist on increasing investment in these factors, but maybe on a better usage of them in small firms. Actually, both differences in characteristics and differences in returns are causing TFP differences, so the two policy implications are required and a policy to increase returns in small firms with low TFP levels would be especially effective.

## References

- Acemoglu, D. and Angrist, J. (1999). "How large are the social returns to education? Evidence from Compulsory Schooling Laws", NBER Working Paper Series. Num 7444.
- Benhabib, J. and Spiegel, M.M. (1994). "The role of human capital in economic development: evidence from aggregate cross-country data", *Journal of Monetary Economics*, 34, pp 143-173.
- Black, S.E. and Lynch, L.M. (1996), "Human-capital investments and productivity", *American Economic Review*, Papers and Proceedings, May.
- Carree, M.A. and Thurik, A.R. (1998), "Small firms and economic growth in Europe", *American Economic Journal*, 26, iss 2, pp 137-146.
- Castiglionesi, F. and Ornaghi, C. (2004), "An empirical assessment of the determinants of TFP growth", presented at Encuentros de Economía Aplicada (2003)
- Cohen, W. M. and S. Klepper (1996). "Firm Size and the Nature of Innovation within Industries: The Case of Process and Product R&D", *Review of Economics and Statistics* 78, pp 232-243.
- Crépon, B., Duguet, E. and Mairesse, J. (1998). "Research, innovation, and productivity an econometric analysis at the firm level", NBER Working Paper, No. 6696, August.
- Del Barrio-Castro, T. López-Bazo, E. and Serrano-Domingo, G. (2002). "New evidence on international R&D spillovers, human capital and productivity in the OECD", *Economics Letters*, 77, pp 41-45.
- Del Río, Gradín and Cantó (2004) "El enfoque distributivo en el análisis de la discriminación salarial por razón de género", presented at Encuentros de Economía Aplicada (2004)
- Delgado, M.A, Fariñas, J.C. and Ruano, S. (2002), "Firm productivity and export markets: a non-parametric approach", *Journal of International Economics*, 57, PP- 397-422
- Estrada, A. and López – Salido, J. D. (2001), "Sectoral and aggregate technology growth", Banco de España - Servicio de Estudios, Documento de Trabajo n.º 0116
- Fariñas, J. C. and Jaumandreu, J. (1999), 'Diez años de Encuesta sobre Estrategias Empresariales (ESEE)', *Economía Industrial*, n.º 329, pp. 29-42.
- Fariñas, J.C. and Ruano, S. (2004), "The dynamics of productivity: a decomposition approach using distribution functions", *Small Business Economics*, 22, pp. 237-251
- Fingleton, B. y López-Bazo, E. (2003), "Explaining the Distribution of Manufacturing Productivity in the EU Regions", in *European regional growth*; Fingleton, B. (ed.), *Advances in Spatial Science*. Heidelberg and New York: Springer.
- Geroski (1998) "An Applied Econometrician's View of Large Company Performance", *Review of Industrial Organization*, Vol. 13, pp. 271-293.
- Good, D.H.; Nadiri, M.I. and Sickles, R.C. (1996), "Index Number and Factor Demand Approaches to the Estimation of Productivity". In M. H. Pesaran and P. Schmidt (eds.), *Handbook of Applied Econometrics*, Vol. II: Microeconomics, 14, Blackwell Publishers.
- Hall, R.E. (1990), "Invariance Properties of Solow's Productivity Residual", In Diamond (ed.), *Growth/Productivity/Employment*, Cambridge, MA: MIT Press, pp. 71-112.
- Hall, B.H. (2000), "Innovation and market value," in Barrell, Ray, Geoffrey Mason, and Mary O'Mahoney (eds.), *Productivity, Innovation and Economic Performance*, Cambridge: Cambridge University Press, 2000.
- Hall, B.H. and Mairesse, J. (1995), "Exploring the relationship between R&D and productivity in french manufacturing firms", *Journal of Econometrics*, Vol. 65:1, pp. 263-93.
- Huergo, E. and Jaumandreu, J. (2004a), "How does probability of innovation change with firm age?" *Small Business Economics*, 22, 193-207.
- Huergo, E. and Jaumandreu, J. (2004b), "Firms' age, process innovation and productivity growth", *Journal of Industrial Organization*, Vol. 22, pp. 541-559.
- Jenkins, S.P. (1994), "Earnings discrimination measurement. A distributional approach", *Journal of Econometrics*, 61, pp. 81-102
- Johnson, P. A. (2000), "A nonparametric analysis of income convergence across the US", *Economics Letters*, Vol 69, pp.219-223.
- Klepper, S., 1996, "Entry, Exit and Innovation Over the Product Life Cycle", *American Economic Review*, 86, pp 562-583.
- Kolmogorov, A.N., (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Ital. degli Attuari* 4, 83-91.
- López-Bazo, E. (2003), "Growth and Convergence Across Economies: The Experience of the European Regions", in *Regional economic growth, SMEs and the wider Europe*; Fingleton, B.; Eraydin, A. y Paci, R., (eds.), Aldershot, U.K. and Burlington, Vt.: Ashgate.
- Lucas, R.E. (1978), "On the size distribution of Business Firms", *Bell Journal of Economics*, 9:2, pp. 508-23.

- Mairesse, J. and Sassenou, M. (1991), "R&D and Productivity: A Survey of Econometric Studies at the Firm Level", NBER Working Paper, No. 3666.
- Máñez, J.A, Rincón, A., Rochina, M.E, and Sanchis, J.A. (2003), "Productividad e I+D: un análisis no paramétrico", *LINEEX Working Papers* 38/02.
- Martín-Marcos, A. y Suárez-Gálvez, C. (1997), "El Stock de Capital Para las Empresas de la Encuesta Sobre Estrategias Empresariales", Fundación SEPI, Programa de Investigaciones Económicas, Documento Interno nº 13.
- Martín-Pliego, F.J., De Juan, R., Merino, F., Rodríguez, D. and Suárez, C. (2001), "Las Empresas Industriales en 2001", Fundación SEPI, Programa de Investigaciones Económicas.
- Oaxaca, R. (1973), Sex Discrimination in Wages. In Ashenfelter, O. & Rees, A.: *Discrimination in Labor Markets*. Princeton University Press, 1973.
- OECD (1993), "Methods Used by OECD Countries to Measure Stocks of Fixed Capital", in *National Accounts: Sources and Methods* Nº2, OECD Statistics Directorate.
- Ornaghi, C. (2003), "Spillovers in Product and Process Innovation: Evidence From Manufacturing Firms", Universidad Carlos III de Madrid, Economics Series 13, Documento de Trabajo: 02-32.
- Ruano, S. (2002), "Dinámica del empleo y de la productividad. ¿Qué nos enseñan los datos microeconómicos?", *Economía Industrial*, Vol. 348, iss 6, pp. 95-110
- Ruiz-Maya, L. and Martín-Pliego, F.J. (1995), *Estadística II: Inferencia*, Colección Plan Nuevo, Ed. AC.
- Sheather, S. J. and Jones, M. C. (1991). "A reliable data-based bandwidth selection method for kernel density estimation", *Journal of the Royal Statistical Society -B* 53, 683-690.
- Sianesi, B. and Van Reenen, J.V. (2003), "The returns to education: macroeconomics", *Journal of Economic Surveys*, 17, iss 2.
- Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York
- Smirnov, N.V., (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bull. Math. Univ. Moscow* 2, 3-14.
- Temple, (2001), "Generalizations That Aren't? Evidence on Education and Growth", *European Economic Review*, Vol. 45, 4-6, pp. 905-918.

## Appendix

### *TFP Index*

We have used index numbers to measure Total Factor Productivity at firm level. There is a wide range of indexes derived from different production functions and with different properties, but basically, TFP indexes are calculated as the output of every firm minus a weighted sum of the inputs (labor, capital and materials). We have chosen the index proposed by Good, Nadiri and Sickles (1996), which is derived from a translog production function and which has the following desirable properties: transitivity (permits comparisons across individuals and time periods); it is superlative (as it is exact and derived from a flexible function, it provides a second order local approximation); it allows for technological change over time; input shares are characteristic of every firm; and it does not suppose perfect competition. The analytical expression is as follows:

$$\begin{aligned} \ln TFP_{it} &= (\ln Y_{it} - \overline{\ln Y_t}) - \frac{1}{2}(S_{it} + \overline{S_t})(\ln X_{it} - \overline{\ln X_t}) \\ &+ \sum_{s=2}^t (\overline{\ln Y_s} - \overline{\ln Y_{s-1}}) - \sum_{s=2}^t \frac{1}{2}(\overline{S_s} + \overline{S_{s-1}})(\overline{\ln X_s} - \overline{\ln X_{s-1}}) \end{aligned}$$

where  $Y$  and  $X$  are quantities of output and inputs respectively,  $S$  is the cost-based share of every input in the production function; subscripts  $i$  and  $t$  refer to firm and time period and the bar over variables means average. The upper part of the expression is the deviation of the firm output and inputs from those of a hypothetical firm, the reference point in year  $t$ . The lower part of the expression is the cumulative change in the output and inputs reference point between year  $t$  and the initial year. This second part introduces a productivity differential every year (as output, inputs and shares may change), and thus it permits the existence of technological change. Productivity index for a given firm and year is expressed in relation to the hypothetical firm in the base time period, 1990. As Hall (1990) suggested, weights are calculated as the share of every input in the total cost of inputs and this permits accounting for market power situations.

This index will provide a TFP value for each firm in each time period. It seems reasonable to consider the firm size when estimating the density function: two firms with the same productivity level but different size will have different impact on economy TFP. Even though, after comparing the density function of TFP and weighted TFP, not much difference is observed, then we have used the unweighted TFP.

### *Variables*

The variables on input and output quantities and prices are not directly drawn from the ESEE, so they have been approximated according to the following particularities. All the quantities are expressed in constant pesetas of 1990.

The *output* is measured as sales plus variation of stocks of sales and corrected with a firm-specific price index. The price increase of every firm is calculated as a weighted sum of the price increase in the main five markets where it operates. In 1990, the price index takes value 1 and in the following periods we add the price increase to the base time period index. As our dataset is an unbalanced panel, many firms do not appear in the sample in 1990, so we cannot build the whole price index series for these firms. To overcome this difficulty we calculate a hypothetical price increase for every year between 1990 and the first year the firm appears in the sample. This hypothetical price increase is computed as an average price increase by year and industry.

The *labor input* is measured as effective yearly hours of work (that is the normal yearly hours plus overtime yearly hours minus non-working yearly hours) multiplied by the number of workers at the end of the year (that includes part time, full time and eventual employees). When the firm answers that the number of eventual employees has changed considerably, it is computed as the average of eventual employees at the end of every quarter.

The *stock of capital* is computed following the perpetual inventory method and obtained from the series calculated by Martín-Marcos and Suárez-Gálvez (1997). According to the definition of physical capital by the OECD (1993) we consider both equipment and buildings<sup>25</sup>.

*Materials* are measured by the cost of intermediate inputs (which includes raw material purchases, energy and fuel costs) minus the variation of the stock of materials plus other services paid by the firm. This quantity is deflated using a firm-specific, materials price index computed as a weighted sum of the price indexes for raw materials, energy and other services, as suggested in Martín-Pliego *et al.* (2001). We overcome the difficulty related to unbalanced panels as explained above.

The *cost of labor input* is the real wage bill (deflated with the IPC).

---

<sup>25</sup> Some authors do not include buildings in the stock of capital as they are considered to have a smaller impact on production than workers or machines.

The *cost of capital* is computed as the *user cost of capital* (interest rates plus depreciation minus the variation of prices) multiplied by the stock of capital (explained above). We calculate separately cost of capital series for equipment and buildings. Interest rates are calculated as a weighted sum of the interest rates paid to credit entities and other organisms. In 1990 interest rates are not available in the ESEE, so we use those from 1991 (average by year and industry). Depreciation rates by industry are obtained from Martín-Marcos and Suárez-Gálvez (1997). Equipment price increase is drawn from INE. Constructions price index between 1990 and 1994 is obtained from Martín-Marcos and Suárez-Gálvez and since 1994, it increases following an implicit deflator of constructions obtained from INE.

The *cost of materials* is the quantity of materials used by the firm and it is computed as explained above.

### *Cleaning*

The ESEE is an unbalanced panel of 26786 observations between 1990 and 1999, due to entry and exit in the industry. Some firms don't respond to some or all the fields in the survey, and then it's impossible to calculate the output and input variables, which means that our sample is reduced to 12483 observations. Before computing the TFP index we need to clean the sample in a way that removes "nonsense" observations according with the following criteria. Ornaghi (2003), Castiglionesi and Ornaghi (2004) clean the survey with a similar criterion.

Firstly, we remove 102 observations with negative value added. Secondly, 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. And 1565 firms follow these criteria. Thirdly, we remove 116 observations where the share of labor input or materials is higher than 0.95 or lower than 0.05. Fourthly, we remove 47 more observations as they are isolated because of the previous cleaning steps. At the end, we obtain a sample of 10653 observations over 10 years, for 1945 different firms.

### *Density Function Estimation*

The main advantage of using synthetic measures to describe the behaviour of a variable is that they collect all the information in a single number, but they do not allow to analyse possible changes in the distribution function in detail and the same aggregate synthetic measure may be consistent with very different distributions of TFP. So, we analyse the behaviour in the complete distribution by estimating density functions (the external shape of the distribution) using non-parametric 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):

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

In this paper, we've used the Gaussian kernel for the  $K(\cdot)$  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 for different distributions, the bandwidth has been settled down to 0.06568, the arithmetic mean of bandwidths for the four different subsamples in our work: small and large firm in 1994 and 1998.

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.

Table 1: Descriptive analysis of the TFP index for the Spanish manufacturing firms

Ln PTF	1994		1998	
	small	large	Small	large
mean	-0,0013	0,0835	0,0616	0,1425
var	0,0507	0,0399	0,072	0,0529
test Eq mean♦	4,4308***		3,7746***	
<b>Percentiles</b>				
25%	-0,1425	-0,0474	-0,1094	-0,0092
50%	-0,0057	0,0888	0,0485	0,1376
75%	0,1303	0,1996	0,2329	0,261
n° observ	339	184	500	168

(♦)  $H_0: \overline{TFP}_L = \overline{TFP}_S$ ;  $H_1: \overline{TFP}_L \neq \overline{TFP}_S$ ;

\*\*\* denotes significant at 1%

Table 2: Descriptive analysis of R&D expenditures for the Spanish manufacturing firms

R&D (1000ptas/L)	1990		1994	
	small	large	Small	Large
mean	51,2071	126,4894	61,9224	141,2819
var	40502,6545	40313,6871	63674,8195	68302,4433
test Eq mean♦	4,0914***		3,4345***	
<b>Percentiles</b>				
25%	0	0	0	0
50%	0	45,9200	0	59,3410
75%	0	166,6340	0	160,9832
n° observ	339	184	500	168

(♦)  $H_0: \overline{R \& D}_L = \overline{R \& D}_S$ ;  $H_1: \overline{R \& D}_L \neq \overline{R \& D}_S$ ;

\*\*\* denotes significant at 1%

Table 3: Descriptive analysis of the skilled labour use for the Spanish manufacturing firms

H (skilled labor)	1990		1994	
	small	large	Small	Large
mean	0,0582	0,0782	0,0635	0,1108
var	0,0054	0,0055	0,0069	0,0116
test Eq mean♦	2,9655***		5,2035***	
<b>Percentiles</b>				
25%	0	0,0323	0	0,0409
50%	0,0357	0,0574	0,0400	0,0764
75%	0,0937	0,0864	0,0952	0,1367
n° observ	339	184	500	168

(♦)  $H_0: \overline{H}_L = \overline{H}_S$ ;  $H_1: \overline{H}_L \neq \overline{H}_S$ ;

\*\*\* denotes significant at 1%

Table 4: Regression results for small and large firms.

Dependent variable: lnTFP			
	Total sample	Small firms	Large Firms
R&D <sub>t-4</sub>	0,0004 (0,0012)	0,0013 (0,0016)	-0,0002 (0,0017)
H <sub>t-4</sub>	0,2871*** (0,0875)	0,2230** (0,1156)	0,4014*** (0,1303)
<b>control:</b>			
SIZE <sub>t</sub>	0,0096* (0,0057)		
AGE <sub>t</sub>	0,0013*** (0,0004)	0,0017*** (0,0005)	0,0009 (0,0006)
IND <sub>t</sub>	Yes	Yes	Yes
H <sub>0</sub> : IND <sub>t</sub> =0	9,2878***	6,1768***	4,4090***
n° observ	1191	839	352
R <sup>2</sup>	0,2000	0,172	0,226
Residual SS	56,2010	43,355	11,889

Standard deviation in parentheses

\*\*\* and \*\* denote significant at 1% and 5%.

Table 5: Oaxaca decomposition of the difference in TFP between large and small firms

	1994		1998	
	Caract $(\bar{X}_L - \bar{X}_s)\hat{\beta}_L$	Returns $\bar{X}_s(\hat{\beta}_L - \hat{\beta}_s)$	Caract $(\bar{X}_L - \bar{X}_s)\hat{\beta}_L$	Returns $\bar{X}_s(\hat{\beta}_L - \hat{\beta}_s)$
total	0,04408	0,0407	0,06491	0,01596
	52,00%	48,00%	80,30%	19,70%
R&D <sub>t-4</sub>	-0,00102	0,00928	-0,00109	0,00946
	1,20%	10,90%	1,40%	11,70%
H <sub>t-4</sub>	0,00806	0,01038	0,019	0,01133
	9,50%	12,20%	23,50%	14,00%
$\Sigma \text{IND}_t$	0,02497	-0,00523	0,03235	-0,0056
	29,40%	6,20%	40,00%	6,90%

Table 6: Test of equality of means of the real and the counterfactual distributions

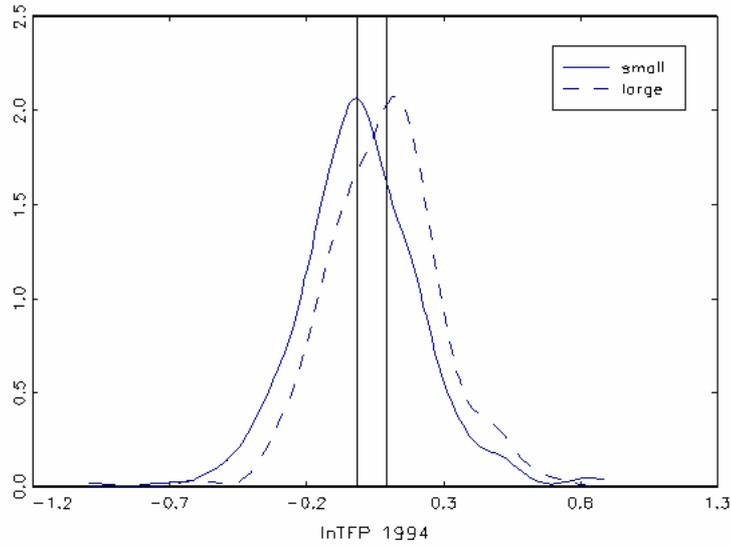
	1994		1998	
	t'	p-value	t'	p-value
Total	4,7142	0,0000	2,3261	0,0101
R&D <sub>t-4</sub>	1,1116	0,1334	1,4286	0,0767
H <sub>t-4</sub>	1,1913	0,1170	1,6322	0,0515
$\Sigma \text{IND}_t$	1,1116	0,1334	0,8214	0,2058

Table 7: Tests of stochastic dominance for the real and the counterfactual distributions

	1994					1998				
	KS statistic	crit val*	p- value*	crit val ( $\chi^2$ )	p-value ( $\chi^2$ )	KS statistic	Crit val*	p- value*	crit val ( $\chi^2$ )	p-value ( $\chi^2$ )
<b>Test 1</b>	$F(z) = F(TFP_S)$ and $G(z) = G(TFP_L)$									
2 sided	0,2386	0,1493	0			0,1908	0,1454	0,0002		
1 sided	0,0089	0,1392	1	0,0374	0,9815	0,0164	0,1356	1	0,1350	0,9347
<b>Test 2</b>	$F(z) = F(\hat{TFP}_S)$ and $G(z) = G(\hat{TFP}_S^L)$									
2 sided	0,2386	0,1249	0			0,1300	0,1029	0,0004		
1 sided	0,0089	0,1168	1	0,0374	0,9883	0,0320	0,0962	0,9600	1,0240	0,5993
<b>Test 3</b>	$F(z) = F(\hat{TFP}_S^L)$ and $G(z) = G(\hat{TFP}_L)$									
2 sided	0,2386	0,1493	0,0008			0,2443	0,1454	0		
1 sided	0,0089	0,1392	1	0,0374	0,9815	0,0000	0,1356	1	0,0374	1,0000

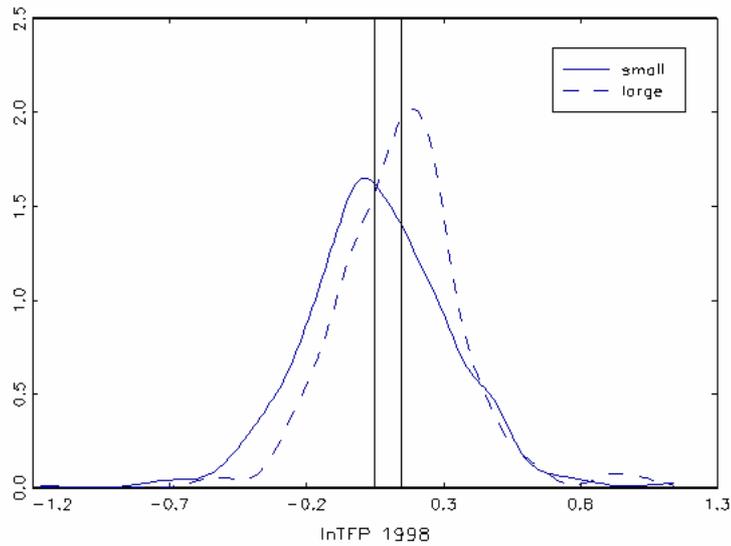
(\*) The Kolmogorov-Smirnov test, under independence of observations, follows a distribution as specified in section 2.2. It can be proved that the statistic of the one-sided KS test follows an asymptotic distribution of  $\chi^2$  with two degrees of freedom.

Figure 1: Estimated density function for small and large firms in 1994



Note: the lines represent the sample average for small (left) and large (right) firms.

Figure 2: Estimated density function for small and large firms in 1998



Note: the lines represent the sample average for small (left) and large (right) firms.

Figure 3: Estimated density function for the real and counterfactual distributions in 1994. Effect of all the factors.

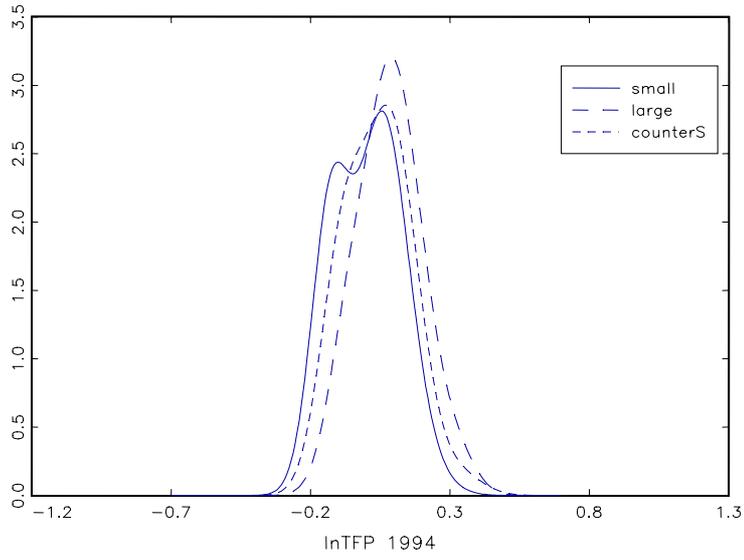


Figure 4: Estimated density function for the real and counterfactual distributions in 1998. Effect of all the factors.

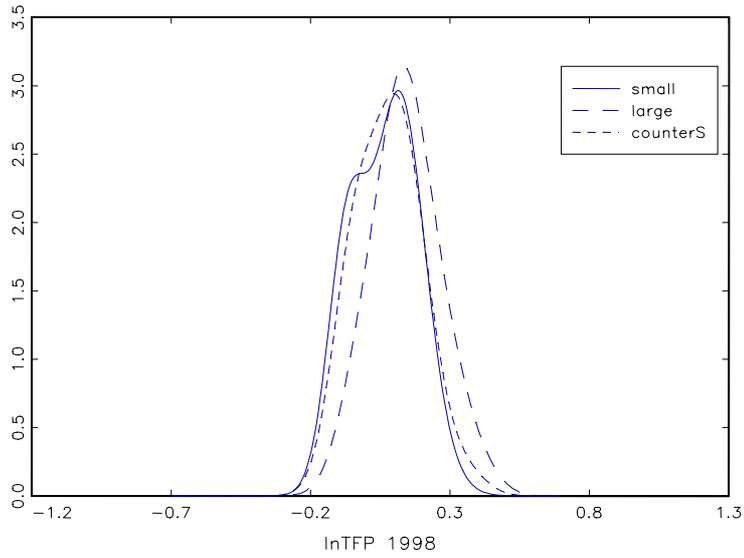


Figure 5: Estimated density function for the real and counterfactual distributions in 1994. Effect of R&D.

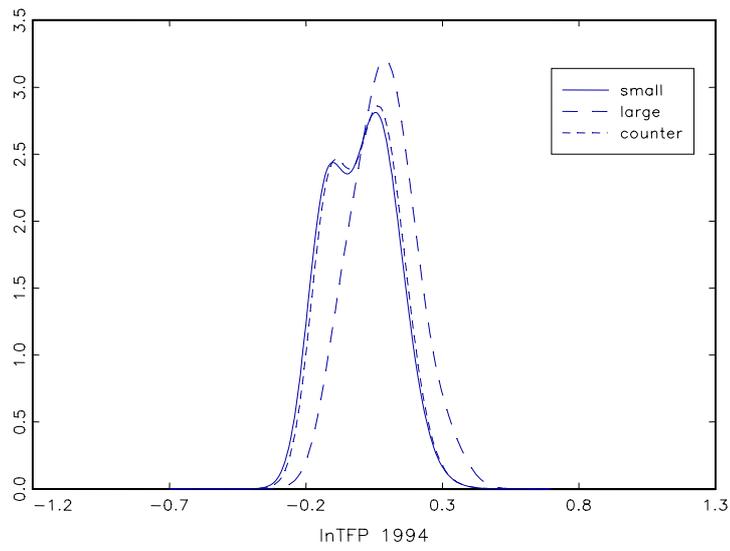


Figure 6: Estimated density function for the real and counterfactual distributions in 1998. Effect of R&D.

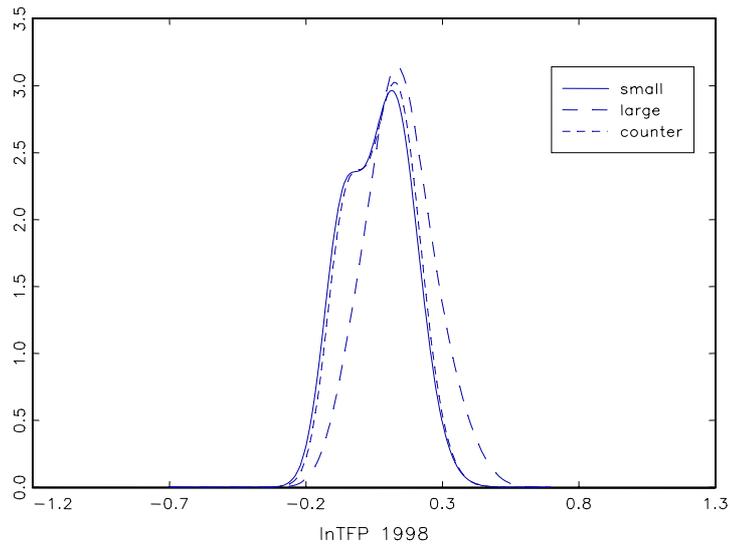


Figure 7: Estimated density function for the real and counterfactual distributions in 1994. Effect of skilled labour.

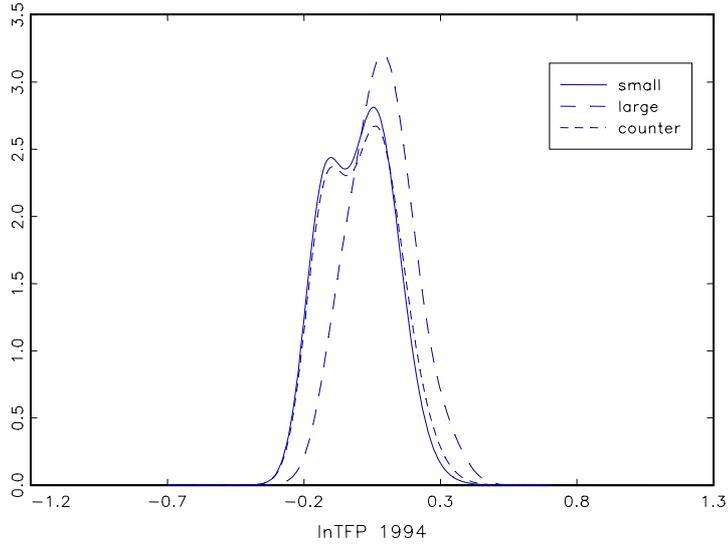


Figure 8: Estimated density function for the real and counterfactual distributions in 1998. Effect of skilled labour.

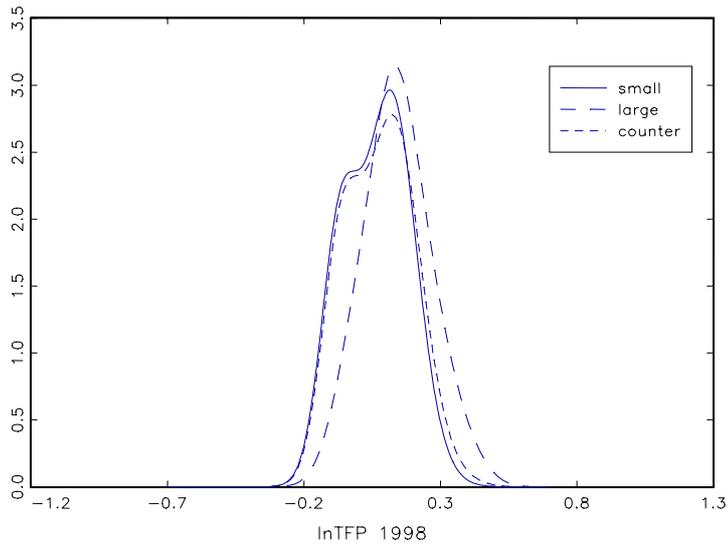


Figure 9: Estimated density function for the real and counterfactual distributions in 1994. Effect of industry.

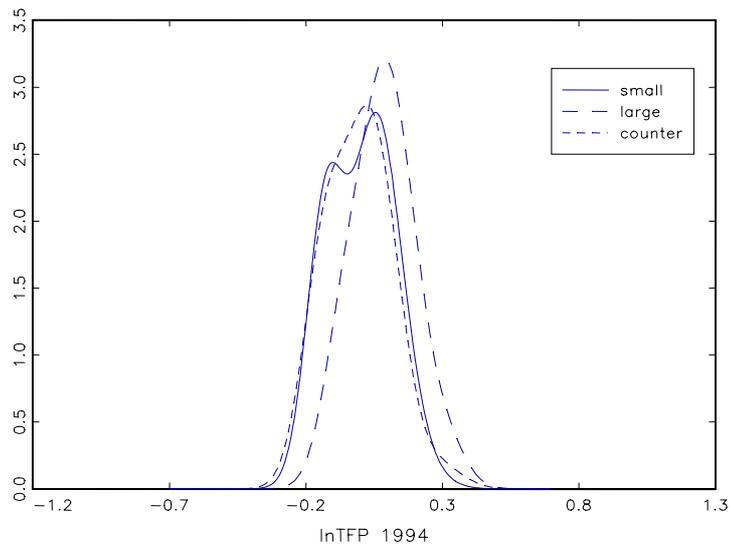


Figure 10: Estimated density function for the real and counterfactual distributions in 1998. Effect of industry.

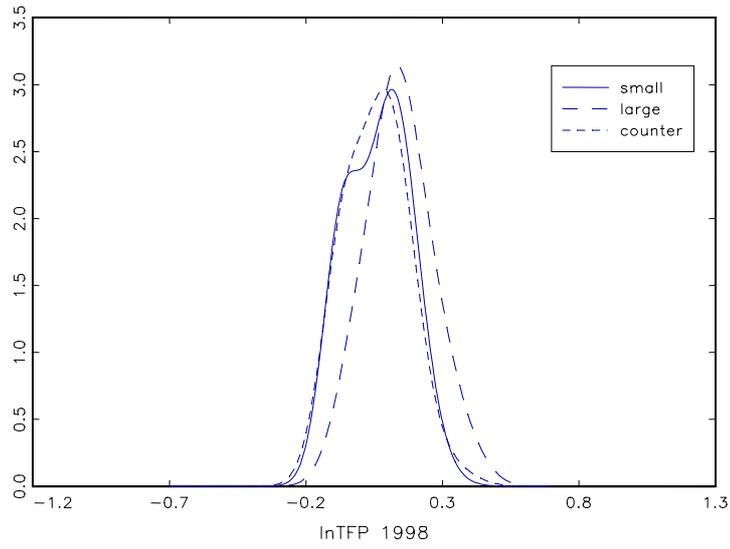


Figure 11: Movements between the real and the counterfactual distribution, 1994. Effect of all the factors.

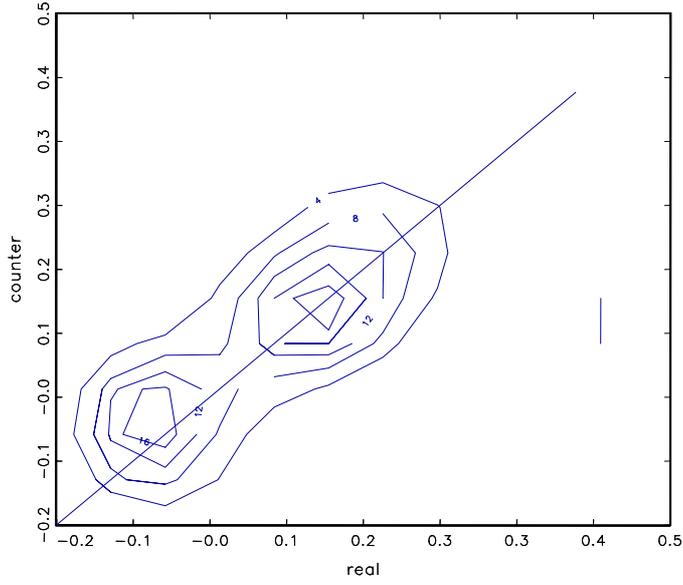


Figure 12: Movements between the real and the counterfactual distribution, 1998. Effect of all the factors.

