



LICOS Centre for Institutions and Economic Performance

Centre of Excellence



## LICOS Discussion Paper Series

Discussion Paper 321/2012

### **R&D and Non-linear Productivity Growth of Heterogeneous Firms**

*d'Artis Kancs and Boriss Siliverstovs*



#### **Katholieke Universiteit Leuven**

LICOS Centre for Institutions and Economic  
Performance  
Waaistraat 6 – mailbox 3511  
3000 Leuven

BELGIUM

TEL: +32-(0)16 32 65 98

FAX: +32-(0)16 32 65 99

<http://www.econ.kuleuven.be/licos>

# R&D and Non-linear Productivity Growth of Heterogeneous Firms<sup>§</sup>

D'ARTIS KANCS

*European Commission,  
DG JRC, LICOS and EERI  
Inca Garcilaso 3  
41092 Seville, Spain  
d'artis.kanacs@ec.europa.eu*

BORISS SILIVERSTOVS

*ETH Zurich,  
KOF Swiss Economic Institute  
Weinbergstrasse 35  
8092 Zurich, Switzerland  
boriss.silverstovs@kof.ethz.ch*

November 2, 2012

## Abstract

The present paper studies the relationship between R&D investment and firm productivity growth by explicitly accounting for non-linearities in the R&D-productivity relationship and inter-sectoral firm heterogeneity. In order to address these issues, we employ a two step estimation approach, and match two firm-level panel data sets for the OECD countries, which allows us to relax both the linearity and homogeneity assumptions of the canonical Griliches (1979) knowledge capital model. Our results suggest that: (i) R&D investment increases firm productivity with an average elasticity of 0.15; (ii) the impact of R&D investment on firm productivity is differential at different levels of R&D intensity – the productivity elasticity ranges from -0.02 for low levels of R&D intensity to 0.33 for high levels of R&D intensity; (iii) the relationship between R&D expenditures and productivity growth is non-linear, and only after a certain critical mass of R&D is reached, the productivity growth is significantly positive; (iv) there are important inter-sectoral differences with respect to R&D investment and firm productivity – high-tech sectors' firms not only invest more in R&D, but also achieve more in terms of productivity gains connected with research activities.

*Keywords: R&D investment, firm productivity, generalised propensity score.*

*JEL code: C14, C21, D24, F23, O32.*

---

<sup>§</sup>The authors acknowledge helpful comments from Andries Brandsma, Michele Cincera, Abraham Garcia, Fernando Hervás Soriano, Pietro Moncada-Paterno-Castello, Sandro Montresor, Enrico Santarelli, Peter Voigt, as well as participants at the 3rd European Conference on Corporate R&D and Innovation (CONCORD 2011) and the ECOMOD 2012 conference in Sevilla, Spain. We are grateful to Alexander Tübke for granting access to the EU Industrial R&D Investment Scoreboard data. The authors are solely responsible for the content of the paper. The views expressed are purely those of the authors and may not under any circumstances be regarded as stating an official position of the European Commission. Computations were performed in Ox 6.30 and PcGive 13.0 (Doornik, 2007; Doornik and Hendry, 2009).

# 1 Introduction

The present paper studies the relationship between R&D investment and firm productivity growth by explicitly accounting for non-linearities in the R&D-productivity relationship and inter-sectoral firm heterogeneity. Since the seminal work of Griliches (1958), the R&D-productivity question has increasingly become a topic of inquiry, and the following research on R&D investment and firm productivity has produced a sizeable amount of theoretical and empirical literature. Generally, both the theoretical models have assigned a substantial role to R&D as an important engine of productivity growth, and the empirical literature has confirmed that a significant share of variation in the observed productivity across firms can be explained by differences in R&D expenditures (Hall et al., 2010).

Whereas the general finding that firm investment in R&D is an important source of productivity growth is well established in the theoretical literature, in the empirical literature there is considerably less agreement on the magnitude of R&D contribution. Firm level studies have estimated the size of productivity elasticity associated with R&D investment ranging from 0.01 to 0.32, and the rate of return of R&D between 8.0 and 170.0 percent (see Mairesse and Sassenou, 1991; Griliches, 2000; Mairesse and Mohnen, 2001, for surveys). In addition, the often lacking robustness and statistical significance of the estimates challenges the conclusiveness of the empirical results (Mairesse and Sassenou, 1991; Czarnitzki et al., 2009; Luintel et al., 2010).<sup>1</sup>

The wide amplitude of the estimated R&D impact on firm productivity in light of the often lacking robustness and significance is, however, of little help to policy makers and R&D performers. Depending on whether a 1% increase in R&D investment boosts firm productivity by 0.01% or by 0.32% has very different implications for firm investment strategy. Similarly, depending on whether one Euro investment in R&D increases firm output by 0.08 or by 1.70 Euro has very different policy implications. In addition, both policy makers and innovators are more interested in specific issues, such as, how a particular amount of R&D investment affects the productivity of a particular (type of) firm at a particular level of technological sophistication.

In order to increase the precision of the R&D-productivity estimates while reducing the confidence interval, studies have attempted to control for inter-sectoral firm heterogeneity. Usually, firm-level studies find that R&D investment makes larger impact on firm productivity in high-tech sectors than in low-tech sectors. Griliches and Mairesse (1983) and Cuneo and Mairesse (1984) were among the first who controlled for inter-sectoral differences in R&D investment on firm productivity. Estimating firm-level production functions they found that the impact of R&D on firm productivity was significantly higher for science-based firms (elasticity 0.20) than for other sectors' firms (0.10). Verspagen (1995) studied inter-sectoral differences in R&D impact on productivity growth by employing a reduced-form production function estimator and sector-level data on value added, employment, capital expenditure and R&D investment for OECD countries, and found that R&D activities have a positive impact on firm output only in high-tech sectors, whereas in medium- and

---

<sup>1</sup>Surveying firm level studies on R&D impact Mairesse and Sassenou (1991) conclude that it is rather difficult to be sure whether differences between the econometric analyses concerning the relationship between R&D and economic performance of firms are real and a result, for example, of differences in the period, industries or countries considered, or simply the reflection of peculiarities of the individual studies.

low-tech sectors no significant effect was found. Harhoff (1998) used the direct production function approach of Hall and Mairesse (1995) to analyse the impact of R&D on labour productivity in manufacturing firms, by employing panel data for 443 German manufacturing firms over the 1977-1989 period, and found that the effect of R&D was considerably higher for high-tech firms than for other sectors' firms. Similarly, Kwon and Inui (2003) used the estimation strategy proposed by Hall and Mairesse (1995) to analyse the impact of R&D on labour productivity in manufacturing firms using a sample of 3,830 Japanese manufacturing firms over the 1995-1998 period, and found a significant impact of R&D on labour productivity. In addition, high-tech firms showed systematically higher and more significant coefficients than medium and low-tech firms. Tsai and Wang (2004) used a stratified sample of 156 large Taiwanese companies over the 1994 to 2000 period, and found that R&D investment had a positive and significant impact on the growth of firm productivity (elasticity 0.18), whereas the impact was considerably higher for high-tech firms (0.30) compared to firms in medium- and low-tech sectors (0.07). Employing the same Scoreboard data as in the present study, Ortega-Argiles et al. (2010) examined the top R&D investors in EU and concluded that the positive impact of R&D on firm productivity increases from low-tech through medium-high to high-tech sectors. Also Kumbhakar et al. (2010) employed the Scoreboard data and studied the impact of corporate R&D activities (measured by knowledge stocks) on firm performance (measured by labour productivity), and found that the overall elasticity ranged from 0.09 to 0.13, whereby the coefficient increased steadily from low-tech to medium-high and high-tech sectors (0.05 - 0.07 in low-tech sectors, and 0.16 - 0.18 in high-tech sectors).

More recent studies attempt to control also for non-linearities in productivity's response to R&D investment. Theoretical models have shown that, due to complementarities, economies of scale in the accumulation of knowledge and obsolescence of some of the previously acquired knowledge, the current and past investments in R&D do not have to increase firm productivity linearly (Furman et al., 2002; Doraszelski and Jaumandreu, 2013). According to Furman et al. (2002), the productivity of R&D investment may be sensitive to the level of technological sophistication (R&D investment in the past) in two opposite ways. On the one hand, due to the so-called "standing on shoulders" effect, prior R&D investment can increase the current productivity. On the other hand, due to the so-called "fishing out" effect, prior R&D investment may have discovered the ideas which are the easiest to find, making the discovery of new ideas and hence a further increase in productivity more difficult. Interactions between the two forces may result in non-linear R&D-productivity relationship.

Empirically, the absorption capacity and critical mass are found to be important causes of non-linearities in the R&D impact on firm productivity (Geroski, 1998; Gonzalez and Jaumandreu, 1998). Geroski (1998) reports that most of the analysed firms show no increasing returns to innovative activity until a certain threshold of R&D activity has been reached. Gonzalez and Jaumandreu (1998) analyse 2000 Spanish manufacturing companies for the 1990-1995 period and find that the R&D thresholds range across industries roughly between 0.2 and 0.5 of the median performing firm's R&D intensity. Bogliacino (2010) finds important non-linearities in the employment response to R&D investment. However, due to constraints of the employed approach, Bogliacino can capture non-linear effects only via a square term of R&D as an additional explanatory variable. Hence, he is not able to recover the entire underlying functional relationship between

R&D investment and firm productivity.

In the present study we follow both these recent lines of research and attempt to estimate the impact of R&D on firm productivity growth by explicitly accounting for non-linearities in the R&D-productivity relationship and inter-sectoral firm heterogeneity. We attempt to answer two questions: how R&D investment affects firm productivity at different levels of technological sophistication, and what the inter-sectoral productivity differences are with respect to productivity effects of R&D investment. These questions are highly relevant for both R&D performers and policy makers, but have not been answered in a satisfactory way yet.

In order to relax the linearity and homogeneity assumptions the standard knowledge capital framework of Griliches (1979), we employ a two-step estimation approach. In the first step, in order to retrieve firm productivity, we estimate firm-level production functions. We employ the structural production function estimator of Doraszelski and Jaumandreu (2013), which relaxes both the linearity and homogeneity assumptions of the canonical Griliches (1979) knowledge capital model. Instead of constructing a stock of knowledge capital under the above mentioned linearity assumptions, which is usually used to estimate the impact of R&D investment on firm productivity, Doraszelski and Jaumandreu (2013) propose to consider firm productivity as unobservable to the econometrician. In this way we can relax the linearity assumption of the R&D process, because we do not need to construct the stock of knowledge capital (under the linearity assumptions). In order to accommodate uncertainties of R&D process and firm heterogeneity, Doraszelski and Jaumandreu (2013) specify a controlled first-order Markov process, and assume that productivity shocks can accumulate over time. As a result, firms with the same R&D expenditures do not need to necessarily have the same productivity as in the standard knowledge capital approach of Griliches (1979).

In the second step, we employ the generalised propensity score (GPS) matching approach of Hirano and Imbens (2004) to estimate the impact of R&D on firm productivity.<sup>2</sup> In the context of the present study, two important advantages of the GPS estimator are the ability to capture potential non-linearities in the relationship between R&D expenditure and firm productivity, and heterogeneity across firms. Potential non-linearities can be captured, because the GPS is a non-parametric estimator. Hence, no specific functional form for the R&D-productivity relationship needs to be imposed a priori. Firm heterogeneity is accounted for by the fact that the GPS does not average the impact of R&D across the firms. Instead, the GPS matches pairs of similar firms, which differ solely by the treatment (R&D) level, and the estimated dose response functions yields the entire function of average and marginal treatment effects (firm productivity) over all possible treatment levels (R&D) of heterogeneous firms. This allows us to recover the full functional relationship between R&D investment and firm productivity, which is not possible in the standard knowledge capital approach.

In the empirical analysis we match two firm-level panel data sets: the EU industrial R&D investment Scoreboard data and the Orbis world wide company information from the Bureau van Dijk Electronic Publishing (BvDEP). The merged panel we employ in the empirical analyses contains 1129 companies from

---

<sup>2</sup>Other studies employing the propensity score matching approach in the context of R&D and firm productivity include e.g. Chang and Robin (2008); Czarnitzki et al. (2011); Sissoko (2011).

the OECD countries covering the 2006-2007 period. The panel structure of the employed micro data and the richness of the available variables (see Section 3) allow us to answer both questions: the impact of R&D investment on firm productivity at different levels of technological sophistication, and the firm productivity differences of R&D investment.

Our results suggest that: (i) R&D investment increases firm productivity with an average elasticity of 0.15; (ii) the impact of R&D investment on firm productivity varies with levels of R&D intensity – the productivity elasticity ranges from -0.02 for very low levels of R&D intensity to 0.33 for high levels of R&D intensity; (iii) the relationship between R&D expenditures and productivity growth is non-linear, and only after a certain critical mass of R&D is reached, the productivity growth becomes significantly positive; (iv) there are important inter-sectoral differences with respect to R&D investment and firm productivity – firms in high-tech sectors not only invest more in R&D, but also achieve more in terms of productivity gains connected with research activities. These results allow us to better interpret the wide distribution of estimates reported in previous studies, and to derive more specific policy conclusions.

The paper is structured as follows: the econometric strategy is explained in Section 2, the description of the data set used in the empirical analysis is given in Section 3, the results are presented in Section 4, and the final section concludes.

## 2 Econometric strategy

### 2.1 The traditional knowledge capital approach

Usually, the relationship between the R&D expenditures and firm productivity has been studied in the knowledge capital framework of Griliches (1979), where production function (usually a standard Cobb-Douglas function) is augmented by an input, which represents the efforts made by the firm to increase the knowledge capital. The knowledge capital is calculated from the accumulation of (depreciated) R&D/innovation expenditures over time. Estimating the impact of the knowledge capital yields a measure of the impact of innovation on multi-factor productivity (factor productivity, once the contribution of all the other factors is taken into account) (Kancs and Ciaian, 2011).

Despite its convenience in practical applications, the knowledge capital framework of Griliches (1979) has been found to have important drawbacks, such as linear R&D-productivity relationship and deterministic innovation process of homogeneous firms. First, both assumptions have been rejected in recent firm level empirical studies (Griliches, 2000). Second, neither is it possible to estimate a differential impact of R&D investment on firm productivity at different levels of technological sophistication if a linear R&D-productivity relationship is imposed, nor is it possible to analyse productivity differences of heterogeneous firms' productivity in a framework of homogeneous firms. Finally, the point estimates of studies based on the knowledge capital framework of Griliches (1979) do not allow to provide detailed answers to specific questions about R&D impact on firm productivity.

## 2.2 A two step approach

In the present study we employ a two step estimation approach, which allows us to relax both the linearity and homogeneity assumptions of the Griliches (1979) knowledge capital model. In the first step, we estimate firm-level production functions by employing the structural production function estimator of Doraszelski and Jaumandreu (2013). In the second step, we employ the generalised propensity score (GPS) matching approach of Hirano and Imbens (2004) to estimate the impact of R&D on firm productivity.

### 2.2.1 Firm productivity

In order to retrieve a measure of productivity for each firm, we follow the traditional literature (see Eberhardt and Helmers (2010), Van Beveren (2012) for surveys) and estimate firm level production functions. As usual, we assume that firm's production function takes a Cobb-Douglas form.<sup>3</sup> Output,  $y_{it}$ , of firm  $i$  in period  $t$  (in logarithmic form) can be expressed as:

$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \beta_m m_{it} + \omega_{it} + e_{it} \quad (1)$$

where  $\beta_0$  measures the mean efficiency level across firms and over time;  $k_{it}$ ,  $l_{it}$  and  $m_{it}$  are inputs of capital, labour and materials in natural logarithms, respectively. Following Olley and Pakes (1996) and Levinsohn and Petrin (2003), we assume that firms choose static (variable) inputs labour,  $l_{it}$ , and material,  $m_{it}$ , such as to maximise short-run profits. Capital input,  $k_{it}$ , is the only dynamic (fixed) input among the conventional factors of production and its amount in period  $t$  is determined by investment in period  $t - 1$ . Error term,  $e_{it}$ , is mean zero random shock, which is uncorrelated over time and across firms. The firm does not know the value of  $e_{it}$ , when it makes its decisions at time  $t$ . Productivity,  $\omega_{it}$ , is known to firm  $i$  in period  $t$ , but not to the econometrician, and therefore must be estimated.

The estimation of production function (1) is subject to the issue of endogeneity between firm productivity and inputs. As Marschak and Andrews (1944) have pointed out, inputs in the production function are not independently chosen, but determined by the characteristics of the firm, including its efficiency. Firms decide on the choice of inputs,  $l_{it}$  and  $m_{it}$ , based on the realised firm specific productivity shock,  $\omega_{it}$ , which they observe in period  $t$ . If production function (1) is estimated using Ordinary Least Squares (OLS) without separating the effect of inputs,  $l_{it}$  and  $m_{it}$ , on output,  $y_{it}$ , from the effect of productivity,  $\omega_{it}$ , then coefficients  $\beta_l$  and  $\beta_m$  are not identified. The resulting OLS estimates are biased, because the OLS estimator requires that all inputs in the production function are exogenous, i.e., determined independently from the firm's efficiency level.<sup>4</sup>

In order to address the issue of endogeneity, Olley and Pakes (1996) have proposed a structural production function estimation approach by adopting an explicit model with the firm's optimisation problem to derive a production function estimator. In the structural investment model of Olley and Pakes (1996) the endogeneity

---

<sup>3</sup>Note, however, that the estimation method is more general, and applies also to other functional forms, such as CES, provided some basic requirements are met. In particular, static inputs need to have positive cross-partial with productivity and the value of the firm has to be increasing in dynamic inputs (Ackerberg et al., 2007).

<sup>4</sup>The correlation between the level of inputs chosen and unobserved productivity shocks is often referred to as endogeneity of inputs or simultaneity bias (Marschak and Andrews, 1944).

problem is addressed by using information about the observed firm’s investment to proxy for unobserved productivity and by applying a control function estimator. The observed investment is a monotone function of unobserved productivity, which can be inverted to back out – and thus to control for – productivity. Building on Olley and Pakes (1996), Levinsohn and Petrin (2003) proposed to use the intermediate input demand instead of investment demand as a proxy for unobserved productivity.

Akerberg et al. (2006) have pointed out that the first step in the Olley and Pakes (1996) and Levinsohn and Petrin (2003) estimators fails to identify labour coefficient,  $\beta_l$ , except under very special assumptions. The reason is that labour demand is a function of the same state variable as investment and intermediate inputs, and therefore does not vary independently from the inverted investment or intermediate input functions used to proxy for unobserved productivity. In order to address the identification issue, Akerberg et al. (2006) have proposed to base the identification strategy by turning the “perfectly variable input” of labour into an “almost perfectly variable input” of labour. Akerberg et al. (2006) assume that firms chose labour input at time  $t - b$  ( $0 < b < 1$ ), after the capital stock was determined by investment at  $t - 1$ , but before the intermediate inputs are chosen at time  $t$ .

Doraszelski and Jaumandreu (2013) extend the dynamic firm investment model by allowing for uncertainties of the R&D process and heterogeneity of firms. They assume a controlled first-order Markov process with transition probabilities  $P(\omega_{it}|\omega_{it-1}, r_{it-1})$ , according to which productivity in period  $t$  is determined by productivity in period  $t - 1$ , and R&D expenditure in period  $t - 1$ :

$$\begin{aligned}\omega_{it} &= E[\omega_{it}|\omega_{it-1}, r_{it-1}] + \xi_{it} \\ &= g(\omega_{it-1}, r_{it-1}) + \xi_{it}.\end{aligned}\tag{2}$$

According to the controlled Markov law of motion (2),<sup>5</sup> the actual productivity,  $\omega_{it}$ , of firm  $i$  in period  $t$  can be decomposed into expected productivity,  $g(\omega_{it-1}, r_{it-1})$ , and unpredictable productivity innovation,  $\xi_{it}$ . The conditional expectation function  $g(\cdot)$  depends on past productivity,  $\omega_{it-1}$ , and past R&D expenditure,  $r_{it-1}$ . The inclusion of R&D expenditures,  $r_{it-1}$ , accounts for the fact that the firm may affect its future productivity by investing into R&D. Given that the firm knows its current productivity, and anticipates the effect of R&D on productivity in period  $t$ , when making the decision about investment in R&D in period  $t - 1$ , the firm knows the expected effect of R&D,  $g(\omega_{it-1}, r_{it-1})$ , made in period  $t - 1$  on productivity in period  $t$ . Random shock,  $\xi_{it}$ , is zero mean independent productivity innovation, which represents both uncertainties linked to past productivity,  $\omega_{it-1}$ , and uncertainties linked to past R&D,  $r_{it-1}$ . The controlled Markov law of motion (2) implies that stochastic shocks to productivity in period  $t$  will carry forward to firm’s productivity in future periods, i.e. firm-specific productivity shocks can accumulate over time.

Substituting the law of motion (2) into production function (1) yields an empirically estimable production function:

---

<sup>5</sup> $\omega_{it} = g(\omega_{it-1}, r_{it-1}) + \xi_{it} = \alpha_0 + \alpha_1\omega_{it-1} + \alpha_2(\omega_{it-1})^2 + \alpha_3(\omega_{it-1})^3 + \alpha_4r_{it-1} + \xi_{it}$ .



$$y_{it} = \beta_0 + \beta_t t + \beta_k k_{it} + \beta_l l_{it} + \beta_m m_{it} + (\omega(l_{it-1}, k_{it-1}, p_{it-1}, q_{it-1}, w_{it-1}), r_{it-1}) + \xi_{it} + e_{it} \quad (3)$$

where time trend,  $t$ , captures shifts in productivity that are not part of  $\omega_{it}$ , and  $w_{it}$ ,  $q_{it}$  and  $p_{it}$  are prices for labour, materials and output, respectively. Productivity,  $\omega_{it} = \beta_0^* - \beta_t t - \beta_k k_{it} + (1 - \beta_l - \beta_m) l_{it} + (1 - \beta_m)(w_{it} - p_{it}) + \beta_m(q_{it} - p_{it})$ , is derived by solving the firm's short-run profit maximisation problem (Levinsohn and Petrin, 2003).<sup>6</sup>

Finally, when estimating Equation (3), the endogeneity of static inputs needs to be addressed. Whereas capital input,  $k_{it}$ , is uncorrelated with  $\xi_{it}$  (because it is determined in period  $t - 1$ , and all lagged variables  $l_{it-1}, k_{it-1}, p_{it-1}, w_{it-1}, r_{it-1}$  are uncorrelated with  $\xi_{it}$ ), labour input,  $l_{it}$ , and material input,  $m_{it}$ , are correlated with  $\xi_{it}$  (because  $\xi_{it}$  is part of  $\omega_{it}$ , and  $l_{it}$  and  $m_{it}$  are functions of  $\omega_{it}$ ). The endogeneity of static inputs can be addressed by employing the instrumental variable estimator. The instrumenting for  $l_{it}$  and  $m_{it}$  requires that the instruments are not correlated with  $\xi_{it}$ . Given that  $l_{it-1}$  and  $m_{it-1}$  are uncorrelated with  $\xi_{it}$ , we use them as instruments for  $l_{it}$  and  $m_{it}$ .

## 2.2.2 R&D and firm productivity

In order to allow for non-linear effects of R&D investment on productivity of heterogeneous firms, in the second step we apply the GPS matching estimator developed by Hirano and Imbens (2004). Potential nonlinearities between R&D investment and firm productivity can be captured, because GPS is a non-parametric estimator. Firm heterogeneity is accounted for by the fact that the GPS does not average the impact of R&D across all firms. Instead, the GPS matches pairs of similar firms, which differ solely by the treatment (R&D) level. The estimated dose-response functions yield the entire function of the average and marginal treatment effects (firm productivity) over all possible treatment levels (R&D) of heterogeneous firms. This allows us to recover the full functional relationship between R&D investment and firm productivity, which is not possible in the standard estimation approach based on the knowledge capital model.

Usually, non-linear treatment effects are studied by employing the binary treatment propensity score (BPS) estimator proposed by Rosenbaum and Rubin (1983). In the present study we employ a generalised version of the BPS (GPS), because it has several important advantages with respect to our objective and the available data. First, it allows for a continuous treatment. As a result, the (very finite) sample we have in the present study can be used more efficiently. Second, the GPS estimator reduces bias caused by non-random treatment assignment, which is the case in the BPS. Third, the GPS methodology avoids that positive or negative trends would result in an overvaluation or undervaluation of the treatment effect. This is particularly important for our study, because economic trends are present at the same time as the treatment (R&D investment). Fourth, an important advantage of the GPS is that it allows to estimate the treatment effect also without a 'zero' control group, because there are no firms without R&D in our sample.

Following Hirano and Imbens (2004), we implement the GPS estimator in three steps. The first step

---

<sup>6</sup>  $\beta_0^* = -\beta_0 - (1 - \beta_m) \ln \beta_l - \beta_m \ln \beta_m$  is a constant.

is based on the assumption that the conditional distribution of treatment variable,  $r$ , is normal given the covariates:

$$r_{it}|X_{it} \sim N(X'_{it}\gamma; \sigma^2), \quad (4)$$

where  $X_{it}$  is a  $z \times 1$  vector of both contemporaneous and lagged values of discrete and continuous covariates. The parameters of the conditional distribution  $(\gamma, \sigma^2)$  are evaluated using a standard OLS regression. The estimated GPS is defined as follows:

$$\widehat{s}_{it} = \frac{1}{\sqrt{2\pi\widehat{\sigma}^2}} \exp \left[ -\frac{1}{2\widehat{\sigma}^2} (r_{it} - X'_{it}\widehat{\gamma}) \right]. \quad (5)$$

The propensity score in Equation (5) fulfils its purpose of measuring degree of similarity across heterogeneous firms if a so-called *balancing* property is satisfied, i.e. for those firms with assigned equal propensity scores (conditional on the firm-specific covariates) the associated treatment level is independent from firm characteristics.

Following equation (2), in the second step, the expected value of response variable,  $\omega_{it}$ , is modelled as a flexible parametric function of treatment (R&D investment) and the generalised propensity score,  $r_{it-1}$  and  $s_{it-1}$  ( $\omega_{it-1}$ ), respectively:

$$E[\omega_{it}|r_{it-1}, s_{it-1}] = \alpha_0 + \alpha_1 * r_{it-1} + \alpha_2 * r_{it-1}^2 + \alpha_3 * s_{it-1} + \alpha_4 * s_{it-1}^2 + \alpha_5 * (r_{it-1} * s_{it-1}), \quad (6)$$

where the latter is substituted with its estimates,  $\widehat{s}_{it-1}$ , from the first step. The flexibility of the functional form can be controlled for by varying the power of variables  $r_{it-1}$  and  $s_{it-1}$  and their cross-products.

The average expected response of target variable,  $\omega$ , for treatment dose,  $\tau$ , is estimated in the third step:

$$E[\widehat{\omega}_t(\tau)] = \frac{1}{N} \sum_{i=1}^N [\widehat{\alpha}_0 + \widehat{\alpha}_1 * \tau + \widehat{\alpha}_2 * \tau^2 + \widehat{\alpha}_3 * \widehat{s}(\tau, X_{it-1}) + \widehat{\alpha}_4 * \widehat{s}(\tau, X_{it-1})^2 + \widehat{\alpha}_5 * (\tau * \widehat{s}(\tau, X_{it-1}))], \quad (7)$$

where the coefficient estimates from the expected response (6) are used. The whole dose-response function is obtained by computing Equation (7) for each treatment level by using a grid of values in the corresponding range of treatment variable. We also provide a treatment effect and elasticity functions. The former function is a first derivative of  $E[\widehat{\omega}_t(\tau)]$  with respect to argument. The latter function is computed using a standard elasticity formula  $[\partial E[\widehat{\omega}_t(\tau)]/\partial \tau]/[E[\widehat{\omega}_t(\tau)]/\tau]$ . The latter function is of a particular interest for us, as it allows us to directly compare our results with those reported in the previous literature. Following Hirano and Imbens (2004), the confidence interval around the estimated dose-response function is obtained by using a bootstrap procedure.

## 3 Data and variable construction

### 3.1 Data sources

The principal data source is the EU Industrial R&D Investment Scoreboard data set. The Scoreboard is an annual data set compiled and provided by the European Commission. Firstly released in 2004, it comprises data on R&D investment, as well as other financial and economic variables (e.g. net sales, operating profits, employees) for the top 2,000 R&D global performers: 1,000 companies based in the EU and 1,000 based outside the EU.<sup>78</sup> In addition to economic and financial variables, the Scoreboard also identifies the industrial sector (of the parent company) as well as the geographical region of R&D investment (according to the location of company's headquarter). The Scoreboard data are reported in two ways. On the one hand, the Scoreboard data are presented as national aggregates broken down by NACE Rev.1.1 in the Eurostat dissemination database. On the other hand, given that the presentation of the aggregated statistics per economic activity and per country has no data for certain economic activities and certain countries, the full set of data is also presented as broken down by individual enterprise group.

The Scoreboard data set is compiled from companies' annual reports and accounts with reference date of 1 August of each year. For those companies, whose accounts are expected close to the cut-off date, preliminary information is used. In order to maximise the completeness and to avoid double counting, the consolidated group accounts of the ultimate parent company are used. Companies which are subsidiaries of another company are not considered separately. Where consolidated group accounts of the ultimate parent company are not available, subsidiaries are however included. In case of a demerger, the full history of the continuing entity is included, whereas the history of the demerged company goes only back as far as the date of the demerger to avoid double counting. In case of an acquisition or merger, the estimated figures for the year of acquisition are used along with the estimated comparative figures if available.

It is important to note that the Scoreboard data are different from the official R&D statistics provided by statistical offices. The Scoreboard data refers to all R&D financed by a particular company from its own funds, regardless of where the R&D activity is performed. Hence, because companies are identified with country of their registered head office which, in some cases, may be different from the operational or R&D headquarters. In contrast, the R&D statistics usually refers to all R&D activities performed by businesses within a particular sector and country, regardless of the location of the business's headquarters and regardless of the origin of the sources of finance. Second, the Scoreboard collects data from audited financial accounts and reports, whereas the R&D statistics are compiled on the basis of statistical surveys, in general covering the known R&D performer. Further differences concern sectoral classifications (R&D statistics follows the classification of economic activities in the European Community, NACE Rev.1.1, whereas the Scoreboard allocates companies in accordance to the sectoral classification as defined by the Financial Times Stock

---

<sup>7</sup>In the first edition (2004) the top companies were 500 EU and 500 non-EU; in the second edition (2005) were 750 for each area; since the third and up to the sixth (2009), they are 2,000 in total.

<sup>8</sup>Scoreboard data set may be criticised that it has a sample bias affecting the results, because it only represents the top R&D investors. However, this argument doesn't appear to be convincing since the 1,000 companies based in the EU and 1,000 based outside the EU altogether represent approximately 80% of business expenditure on R&D worldwide (Moncada-Paterno-Castello et al., 2010). While small R&D investors and non-R&D-performers are excluded from the sample, the objective of the present study is to focus on the impact of R&D investment on firm productivity, but not to examine the structure of the whole economy.

Exchange Index (ICB classification) and then converts them into NACE Rev.1.1. These differences need to be kept in mind when comparing the results reported in this paper to studies employing statistical R&D data.

For the purpose of TFP estimations, the EU industrial R&D investment Scoreboard data are augmented by the Orbis database, which contains worldwide company information and is commercialised by the Bureau van Dijk Electronic Publishing (BvDEP). Orbis reports annual accounts data on more than 100 million private and public companies world wide (50 million companies in Europe, 24 million companies in North America, 7 million companies in South and Central America, and 9 million companies in East and Central Asia) covering the 1996-2011 period. In order to enhance the comparison across countries, all firm accounts are transformed into a universal format. In addition to large multinationals, Orbis also covers a large fraction of new and small and medium sized companies (SMEs) across all industries (Kancs and Ciaian, 2011).

The Orbis database contains firm-level accounting data in a standardised financial format for 26 balance sheet items (e.g. fixed assets (intangible, tangible and other); current assets (stocks, debtors and other); cash and cash equivalent; total assets; shareholder's funds (capital and other); non-current liabilities (long-term debt and other); current liabilities (loans, creditors and other); total share funds and liabilities; working capital; net current assets; enterprise value; and number of employees), 25 income statement items (e.g. operating revenue/turnover; sales; cost of goods sold; gross profit; other operating expenses; operating profit/loss; financial revenue; financial expenses; financial profit/loss; profit/loss before tax; taxation; profit/loss after tax; extraordinary revenue; extraordinary expenses; extraordinary and other profit/loss; profit/loss for period; export turnover; material costs; costs of employees; depreciation; interest paid; cash flow; added value; earnings before interest and taxes; earnings before interest, taxes, depreciation and amortisation), and 26 financial ratios based on these variables. We use the European Central Bank period average exchange rates to convert all accounting data into EURO.

In addition to financial information, the Orbis database contains also other important information about the companies. First, there is information on the year of incorporation, which allows to calculate the age of the firm. Second, Orbis includes the national industry code and assigns companies a 3-digit NACE code – the European standard of industry classification – which we use to classify firms and construct industry dummy variables. In empirical analysis we use NACE Rev.1.1 codes on a 2-digit level to increase to a significant level the number of firms per industry. All firms in the Orbis database are uniquely identified by their VAT number, which allows us to match them with the Scoreboard data.

### 3.2 Dependent (response) variable

The dependent (response) variable is firm-specific TFP in 2007. In Section 2.2.2 it corresponds to variable  $\omega_{it}$ . In order to retrieve the unobserved firm-specific productivity, we estimate firm-level production functions as described in section 2.2.1. We apply the non-linear GMM estimator to estimate Equation (3). For robustness, we also estimate firm-level TFP by employing the non-linear least squares estimator. The estimation results are reported in Table 2.

According to Table 2, the estimated coefficients are reasonable and the returns to scale, as given by

$\hat{\beta}_l + \hat{\beta}_k + \hat{\beta}_m$ , are close to constant. Generally, our production function estimates are in line with those reported in the literature Olley and Pakes (1996), Levinsohn and Petrin (2003), Akerberg et al. (2006), Doraszelski and Jaumandreu (2013).

From the production function estimates we construct productivity series for each firm,<sup>9</sup> which in turn is used to construct the response variable – TFP in 2007 – for each firm. In order to eliminate the effects of productivity outliers, we censored the sample at the 0.5th and 99.5th percentiles. The qualitative results are similar to those without data censoring. After cleaning the response variable for outliers, and accounting for missing values in the variables and covariates, we are left with a total number of 1129 companies, which we employ in the empirical analysis.

### 3.3 Explanatory (treatment) variable

We define the explanatory (treatment) variable,  $r_{it}$ , as the share of R&D investment in the total capital expenditure. The constructed measure of R&D intensity includes all cash investment in R&D funded by the companies themselves, but excludes any R&D undertaken under contract for customers, such as governments or other companies, and the companies' share of any associated company or joint venture R&D investment. R&D expenditures are calculated based on the R&D accounting definition set out in the International Accounting Standard (IAS) 38 “Intangible assets”, which is based on the OECD “Frascati” manual. Research is defined as original and planned investigation undertaken with the prospect of gaining new scientific or technical knowledge and understanding. Expenditure on research is recognised as an expense when it incurred. Development is the application of research findings or other knowledge to a plan or design for the production of new or substantially improved materials, devices, products, processes, systems or services before the start of commercial production or use. Development costs are capitalised when they meet certain criteria and when it can be demonstrated that the asset will generate probable future economic benefits. Where part or all of R&D costs have been capitalised, the additions to the appropriate intangible assets are included to calculate the cash investment and any amortisation eliminated.

In order to account for intra-sectoral heterogeneity of firms with respect to R&D intensity, we regroup all firms into four sub-samples according to the level of technological sophistication. Following the OECD classification, firms are regrouped into four groups according to the 3-digit ICB classification: high-, medium-high-, medium-low-, and low-tech companies:

- **High-tech:** Technology hardware & equipment, Software & computer services, Pharmaceuticals & biotechnology, Health care equipment & services, and Leisure goods;
- **Medium-high-tech:** Industrial engineering, Electronic & electrical equipment, General industrials, Automobiles & parts, Personal goods, Other financials, Chemicals, Aerospace & defence, Travel & leisure, Support services, and Household goods & home construction;

---

<sup>9</sup>The firm-level productivity estimates,  $\hat{\omega}_{it}$ , are retrieved using:  $\hat{\omega}_{it} = -\hat{\beta}_l t + (1 - \hat{\beta}_l - \hat{\beta}_m) l_{it} - \hat{\beta}_k k_{it} + (1 - \hat{\beta}_m) (w_{it} - p_{it}) + \hat{\beta}_m (q_{it} - p_{it})$ , where  $\hat{\beta}_l$ ,  $\hat{\beta}_m$ , and  $\hat{\beta}_k$  denote production function estimates.

- **Medium-low-tech:** Food producers, Fixed line telecommunications, Beverages, General retailers, Alternative energy, Media, Oil equipment, services & distribution, and Tobacco;
- **Low-tech:** Gas, water & multi-utilities, Oil & gas producers, Nonlife insurance, Industrial metals & mining, Construction & materials, Food & drug retailers, Banks, Electricity, Industrial transportation, Mobile telecommunications, Forestry & paper, Mining, Life insurance.

The descriptive statistics of R&D activity for each group of companies is reported in Table 1. According to Table 1, the R&D activity of high-tech firms, measured both in absolute and relative terms, substantially exceeds that of medium-tech and low-tech companies.

Given that the unconditional distribution of the treatment variable is highly skewed, we take a logarithmic transformation in order to comply with the normality assumptions in the first step of the GPS regression. Following Hirano and Imbens (2004), we also take a logarithmic transformation of the estimated generalised propensity score in the second step.

### 3.4 Covariates

In Equation (5) the expected amount of treatment,  $r_{it}$ , that a firm receives in a given period  $t$  is evaluated given the covariates,  $X_{it}$ , i.e., the estimation of the impact of treatment is based on comparison of firms with similar propensity scores,  $\widehat{s}_{it}$ .<sup>10</sup> Adjusting for the propensity scores removes the biases associated with differences in covariates, which allows to estimate the marginal treatment effect for a specific treatment level on the outcome variable of firms that have received a specific treatment level with respect to firms that have received a different treatment level (counterfactual), whereas both groups of firms have similar characteristics.

In order to control for differences with respect to a specific treatment level (R&D productivity) on the outcome variable (productivity growth) of firms that have received a specific treatment level with respect to firms that have received a different treatment level (counterfactual), the expected amount of treatment,  $r_{it}$ , is evaluated given the covariates,  $X_{it}$ . The set of covariates are selected based on previous studies (Hall et al., 2010). Given the availability of variables in the merged Scoreboard and Orbis data set, we construct the following covariates:

- **Net sales:** In line with the accounting definition of sales, sales taxes and shares of sales of joint ventures & associates are excluded. For banks, sales are defined as the “Total (operating) income” plus any insurance income. For insurance companies, sales are defined as “Gross premiums written” plus any banking income.
- **Operating profit:** Profit (or loss) before taxation, plus net interest cost (or minus net interest income) and government grants, less gains (or plus losses) arising from the sale/disposal of businesses or fixed assets. Due to the fact that companies report both positive and negative operating profit, we cannot take a logarithmic transformation of this variable. In order to do so, we created the following two

---

<sup>10</sup>The adequacy of the estimated GPS is checked by assessing its balancing properties.

variables  $\log(\text{Operating profit})^{POS}$  and  $\log(\text{Operating profit})^{NEG}$ . The former variable is equal to the log of actual values whenever a firm reports positive profit and zero otherwise. The latter variable is equal to the log of absolute actual values multiplied by minus one whenever a firm reports negative profit and zero otherwise.

- **Capital expenditure:** The expenditure used by a company to acquire or upgrade physical assets such as equipment, property, industrial buildings. In company accounts capital expenditure is added to the asset account (i.e. capitalised), thus increasing the amount of assets. It is disclosed in accounts as additions to tangible fixed assets.
- **Number of employees:** The average number of employees or, if the annual average is not available, the number of employees at the end of the reference period.
- **Material costs:** The cost of material inputs of companies used for production of goods and services in the respective year.
- **Prices:** The relative producer prices for inputs and outputs calculated using the concept of total price performance (Jorgenson and Griliches, 1967), which measures the growth of input prices compared to the growth of output prices.
- **Market capitalisation:** The share price multiplied by the number of shares issued at a given date. Market capitalisation data have been extracted from both the Financial Times London Share Service and Reuters. These reflect the market capitalisation of each company at the close of trading on 4 August 2006. The gross market capitalisation amount is used to take into account those companies for which not all the equity is available on the market.
- **Industry sectors:** The industry sectors are based on the ICB classification. The level of disaggregation is generally the three-digit level of the ICB classification, which is then converted to NACE Rev.1.1.
- **Sector dummy variable:** Sectors are classified into high-tech, medium-high-tech, medium-low-tech, and low-tech.
- **Regional dummies:** “Asian Tigers”, “BRIC”, “EU”, “Japan”, “RoW”, “Switzerland”, and “USA”.
- **Company number:** This is a unique company identification number, which is assigned to each company and kept the same over the years.

## 4 Results

### 4.1 Non-linearities in the R&D-productivity relationship

The results of the first step GPS estimation procedure are reported in Table 3 for the pooled set of all companies, see Equation (4). From the table we note that variation in the R&D intensity is best captured by the number of employees, its squared value and operating profits. Also the industry-specific dummy

variables contribute substantially to the explanatory power of the first step of the GPS regression.<sup>11</sup> By means of this regression we are able to explain more than 70% in variation of treatment intensity variable, which is important in order to create a powerful GPS.

We verify whether the GPS is appropriately specified by testing the so-called balancing property. In order to do so, we subdivide each covariate into three groups of approximately similar sizes according to their treatment intensity. Then for each covariate in the respective group we test whether the mean is the same as the mean in the remaining treatment groups. The results of these tests are reported in Table 4 indicating that there is very strong heterogeneity among the covariates belonging to different groups. A well specified GPS should successfully account for these differences.

In order to check whether this is the case, we subdivided each group into blocs of approximately the same sizes according to the quintiles of the respective GPS. The dimensions of each group and bloc are reported in Table 5. Observe that now the total number of firms is less than previously reported. This is due to the fact that we imposed the so-called common support condition, which allows us to focus exclusively on the observations with similar GPS values but different treatment intensities. As argued in Becker et al. (2012), imposing the common-support condition substantially improves the balancing properties of the GPS and hence delivers more reliable results. The balancing properties of covariates adjusted for the GPS are reported in Table 6. Compared to the results for the unadjusted covariates, reported in Table 4, there is a substantial improvement with only five test statistics exceeding the nominal 5% significance level. The mean absolute value of all t-statistics reported in Table 6 drops to 1.062 from the corresponding value of 9.088 computed across all groups and covariates in Table 4. On basis of these encouraging results we conclude that the generalised propensity scores are appropriately defined allowing us to consistently estimate the dose-response relationship between the variables of interest.

Table 7 reports the second step GPS estimation results, see Equation (6), where the relationship between the response and treatment variables is specified, conditioning on the estimated GPS from the first step. According to Table 7, the treatment variable,  $\ln r_{it-1}$ , the generalised propensity score,  $\ln s_{it-1}$ , its squared value,  $(\ln s_{it-1})^2$ , and the cross-product term,  $\ln r_{it-1} * \ln s_{it-1}$ , have coefficient estimates that are significantly different from zero.

The results from the third step, summarised in Equation (7), can be illustrated at best graphically in Figure 1, where the average expected response of the TFP to each treatment dose (the so-called dose-response function) is shown. From the specified dose-response function it is possible to derive the corresponding treatment effect function as well as the elasticity function, shown in Figures 2 and 3, respectively.

Figure 3 reports the third step GPS estimation results expressed in terms of productivity elasticity with respect to R&D investment. The 90% confidence interval, which was computed using a bootstrap procedure based on 1000 draws, is marked by the dashed lines. According to Figure 3, the estimated elasticity of the average expected response of TFP in 2007 to R&D intensity in 2006 (GPS-adjusted) ranges between -0.02 and 0.33 (average 0.15). These results are in line with previous firm level studies, which have estimated the size of productivity elasticity associated with R&D investment ranging from 0.01 to 0.32 (see Mairesse and

---

<sup>11</sup>These results are not reported for the sake of saving space.



Sassenou, 1991; Griliches, 2000; Mairesse and Mohnen, 2001, for surveys).

At the aggregated level, the estimated elasticity of firm productivity is an increasing function in the R&D intensity, though with a decreasing rate. Figure 3 suggests that the higher is R&D investment, the larger is productivity growth per unit of R&D investment. The estimated concave relationship between R&D investment and firm productivity suggests that likely there is a maximum optimal level of R&D investment, after which productivity growth per unit of R&D investment would decrease again.

The impact of R&D investment on firm productivity is not significant (even slightly negative) at very low levels of R&D. These results are consistent with findings of Geroski (1998) and Gonzalez and Jaumandreu (1998), who find that a certain critical mass of R&D capacity is required, before significant productivity growth can be achieved from investment in R&D. Our results are also consistent with the hypothesis of absorptive capacity, which is found to be important particularly for firms with low levels of R&D. Firms must be capable of absorbing and using knowledge effectively, if they are to benefit from internal and external R&D investment (Cincera, 1997; Cohen and Levinthal, 1989; Griffith et al., 2004; Fabrizio, 2009).

## 4.2 Inter-sectoral firm heterogeneity

Given that our sample consists of very heterogeneous firms, for which R&D intensity may play rather different role in increasing firm productivity, the impact of equivalent R&D investment may be different between different types of firms. For example, firms in low-tech sectors may benefit from a “late-comer advantage”, while firms in high-tech sectors may be affected by diminishing returns to R&D, suggesting that the relationship between R&D and productivity growth might be stronger for firms in low-tech than in high-tech sectors (Marsili, 2001; Von Tunzelmann and Acha, 2005; Mairesse and Mohnen, 2005). On the other hand, due to the so-called “fishing out” effect, prior research may have discovered the ideas which are the easiest to find, making the discovery of new ideas more difficult (Furman et al., 2002).

In order to control for inter-sectoral firm heterogeneity in the impact of R&D investment and firm productivity growth, we regrouped all firms into four more homogeneous groups according to their level of technological sophistication. As above, we applied the GPS estimator to each of these groups, using the same empirical specification as in section 4.1. The estimation results are reported in Figures 4-5.

Figure 4 reports the third step GPS estimation results expressed in terms of productivity elasticity with respect to R&D investment of high-tech companies.<sup>12</sup> As before, the 90% confidence interval, which was computed using a bootstrap procedure based on 1000 draws, is marked by the dashed lines. According to Figure 4, the estimated elasticity of the average expected response of TFP in 2007 to R&D intensity in 2006 (GPS-adjusted) ranges between -0.04 and 0.54 (average 0.25) for high-tech companies. These results imply that high-tech sectors’ firms not only invest more in R&D, but also achieve more in terms of productivity gains connected with research activities. These results are in line with previous studies, which usually find that R&D investment makes larger impact on firm productivity in high-tech sectors than in low-tech sectors (Griliches and Mairesse, 1983; Cuneo and Mairesse, 1984; Verspagen, 1995; Harhoff, 1998; Kwon and Inui,

---

<sup>12</sup>The results for the first step and second step GPS estimations for low-, medium-low, medium-high and high-tech subsamples are qualitatively comparable to those reported for the full sample, and therefore are not reported separately. These results available from the authors upon request.

2003; Tsai and Wang, 2004; Ortega-Argiles et al., 2010; Kumbhakar et al., 2010).

For medium-low-tech and medium-high-tech companies we observe a less pronounced response of productivity growth to R&D investment. As expected, the estimated elasticities are between those for high-tech and low-tech companies, suggesting that for these companies R&D investment in technological innovations is less important for boosting firm productivity compared to high-tech sectors' firms, but more important compared to low-tech sectors' firms.

For companies in low-tech industries we estimate an average productivity elasticity with respect to R&D investment of 0.05 (Figure 5). At very low levels of R&D intensity the impact on productivity growth is slightly negative. The estimated productivity elasticity with respect to R&D investment increases continuously up to 0.12. Compared to firms in high-tech industries, the elasticity of firm productivity is around five times lower. These results confirm previous studies looking at inter-sectoral differences in the R&D impact on firm productivity.

The lower effect of R&D in low-tech industries can be explained by two factors. First, R&D is only part of innovation process, i.e. innovation does not stop at R&D. According to Potters (2009), the non-R&D innovation plays a particularly important role for firms in low-tech sectors, where design, logistics and organisation and other non-R&D innovations are at least as important for successful innovations as investment in R&D. Second, through knowledge spillovers, the R&D inputs in high-tech sectors may contribute importantly to the innovative power of low-tech sectors (Potters, 2009).

## 5 Conclusions and policy recommendations

The present paper studies the relationship between R&D investment and firm productivity growth by explicitly accounting for non-linearities in the R&D-productivity relationship and inter-sectoral firm heterogeneity. We attempt to answer two questions: how R&D investment affects firm productivity at different levels of technological sophistication, and what the inter-sectoral productivity differences are with respect to productivity effects of R&D investment? These questions are highly relevant for both R&D performers and policy makers, but have not been answered in a satisfactory way yet.

Given that such questions cannot be answered in the canonical knowledge capital framework of Griliches (1979), we employ a two step estimation approach, which allows us to accommodate both the potential non-linearities in the R&D-productivity relationship and to account for firm heterogeneity. In a first step, we estimate firm-level production functions by employing the structural production function estimator of Doraszelski and Jaumandreu (2013), which allows us to retrieve firm productivity. In a second step, we employ the generalised propensity score (GPS) matching approach of Hirano and Imbens (2004) to estimate the impact of R&D investment on firm productivity. By employing this two step estimation approach we are able to relax both the linearity and firm homogeneity assumptions of the Griliches (1979) knowledge capital model.

In the empirical analysis we match two firm-level panel data sets: the EU Industrial R&D Investment Scoreboard data and the Orbis world wide company information from the BvDEP. The merged panel we

employ in the empirical analyses contains 1129 companies from the OECD countries covering the 2006-2007 period.

Our results suggest that: (i) R&D investment increases firm productivity with an average elasticity of 0.15; (ii) the impact of R&D investment on firm productivity is differential for different levels of R&D intensity – the estimated productivity elasticity ranges from -0.02 for low levels of R&D intensity to 0.33 for high levels of R&D intensity; (iii) the relationship between R&D expenditures and productivity growth is non-linear, and only after a certain critical mass of R&D is reached, the productivity growth is significantly positive; (iv) there are important inter-sectoral differences with respect to R&D investment and firm productivity – firms in high-tech sectors not only invest more in R&D, but also achieve more in terms of productivity gains connected with research activities.

These results allow us to better interpret the wide distribution of estimates reported in previous studies, and to derive more specific policy conclusions. The identified relationship between R&D expenditures and firm productivity has particularly important implications for research and innovation policy. First, in presence of non-linear effects, policy makers may have several policy options to achieve the same objective, while different instruments may have very different opportunity costs. Reversely, the same policy instrument may have very different implications, if applied to heterogeneous firms with different levels of R&D intensity, e. g. firms with low level of R&D intensity vs. firms with high level of R&D intensity.

Second, in times of economic and financial crisis, an efficient use of public funds has become a more important issue than ever before. In order to undertake impact assessment of alternative policy instruments, one requires the productivity elasticity of R&D, which according to our results is not constant with respect to the level of technological sophistication – it ranges from -0.02 for low levels of R&D intensity to 0.33 for high levels of R&D intensity. Hence, taking the average value for all firms will generate wrong and/or non-efficient policy recommendations. Therefore, our results suggest that the entire functional relationship between the R&D investment and firm productivity has to be used, not only an average point estimate.

Third, in order to stimulate innovative activities, such as R&D investment, public policy measures should be expressly conceived according to the particular types of firms. For example, measures of policy support for high-tech sectors should be different from those addressing low-tech sectors. Given that according to our estimates higher productivity gains can be achieved in high-tech sectors, public policy should combine measures for stimulating R&D investment particularly in medium and high-tech sectors, while implementing incentive schemes to reinforce the absorption capacity in low-tech sectors. More generally, we advocate that the allocation of R&D efforts is as important issue as the increase of R&D expenditure.

Fourth, given that the relationship between R&D and productivity is stronger in the high-tech sectors, an alternative way to increase productivity could be an industrial policy based on incentives in favour of the expansion of high-tech sectors. In other words, reshaping the industrial structure, which is fixed in the short-term, should be targeted in the long-run, if knowledge-based economy is the long-term policy objective.

## References

- Akerberg, D., K. Caves, and G. Frazer (2006). Structural identification of production functions. Working paper, UCLA, Los Angeles.
- Akerberg, D., C. Lanier Benkard, S. Berry, and A. Pakes (2007). Econometric tools for analyzing market outcomes. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics* (1 ed.), Volume 6A, Chapter 63. Elsevier.
- Becker, S. O., P. H. Egger, and M. von Ehrlich (2012). Too much of a good thing? On the growth effects of the EU's regional policy. *European Economic Review* 56(4), 648–668.
- Bogliacino, F. (2010). Innovation and employment: A firm level analysis with European R&D Scoreboard data. IPTS Working Papers 2010-08, European Commission, DG Joint Research Centre.
- Chang, C.-L. and S. Robin (2008). Public policy, innovation and total factor productivity: An application to Taiwan's manufacturing industry. *Mathematics and Computers in Simulation* 79(3), 352–367.
- Cincera, M. (1997). Patents, R&D and technological spillovers at the firm level: Some evidence from econometric count models for patent data. *Journal of Applied Econometrics* 12(1), 265–280.
- Cohen, W. M. and D. A. Levinthal (1989). Innovation and learning: The two faces of R&D. *Economic Journal* 99(1), 569–596.
- Cuneo, P. and J. Mairesse (1984). Productivity and R&D at the firm level in French manufacturing. In *R&D, Patents, and Productivity*, NBER Chapters, pp. 375–392. National Bureau of Economic Research.
- Czarnitzki, D., P. Hanel, and J. M. Rosa (2011). Evaluating the impact of R&D tax credits on innovation: A microeconomic study on Canadian firms. *Research Policy* 40(2), 217–229.
- Czarnitzki, D., K. Kraft, and S. Thorwarth (2009). The knowledge production of 'R' and 'D'. *Economics Letters* 105(1), 141–143.
- Doornik, J. and D. Hendry (2009). *Empirical Econometric Modelling Using PcGive 13*. London: Timberlake Consultants Press and Oxford: www.doornik.com.
- Doornik, J. A. (2007). *Object-Oriented Matrix Programming Using Ox*. London: Timberlake Consultants Press and Oxford: www.doornik.com.
- Doraszelski, U. and J. Jaumandreu (2013). R&D and productivity: Estimating endogenous productivity. *Review of Economic Studies* 80, Forthcoming.
- Eberhardt, M. and C. Helmers (2010). Untested assumptions and data slicing: A critical review of firm-level production function estimators. Economics Series Working Papers 513, University of Oxford, Department of Economics.
- Fabrizio, K. R. (2009). Absorptive capacity and the search for innovation. *Research Policy* 38(2), 255–267.

- Furman, J. L., M. E. Porter, and S. Stern (2002). The determinants of national innovative capacity. *Research Policy* 31(6), 899–933.
- Geroski, P. A. (1998). An applied econometrician’s view of large company performance. *Review of Industrial Organization* 13(3), 271–294.
- Gonzalez, X. and J. Jaumandreu (1998). Threshold effects in product R&D decisions: Theoretical framework and empirical analysis. *Studies on the Spanish Economy* 45, FEDEA.
- Griffith, R., S. Redding, and J. V. Reenen (2004). R&D and absorptive capacity: From theory to data. *Scandinavian Journal of Economics* 105(1), 99–118.
- Griliches, Z. (1958). Research costs and social returns: Hybrid corn and related innovations. *Journal of Political Economy* 66(1), 419–431.
- Griliches, Z. (1979). Issues in assessing the contribution of research and development to productivity growth. *Bell Journal of Economics* 10(1), 92–116.
- Griliches, Z. (2000). *R&D, education and productivity: A retrospective*. Cambridge Massachusetts: Harvard University Press.
- Griliches, Z. and J. Mairesse (1983). Comparing productivity growth: An exploration of French and U.S. industrial and firm data. *European Economic Review* 21(1-2), 89–119.
- Hall, B. H. and J. Mairesse (1995). Exploring the relationship between R&D and productivity in French manufacturing firms. *Journal of Econometrics* 65(1), 263–293.
- Hall, B. H., J. Mairesse, and P. Mohnen (2010). Measuring the returns to R&D. In B. H. Hall and N. Rosenberg (Eds.), *Handbook of the Economics of Innovation*, Volume 2, Chapter 22, pp. 1033–1082. Elsevier.
- Harhoff, D. (1998). R&D and productivity in German manufacturing firms. *Economics of Innovation and New Technology* 6(1), 29–50.
- Hirano, K. and G. W. Imbens (2004). The propensity score with continuous treatments. In A. Gelman and X.-L. Meng (Eds.), *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, pp. 73–84. Chichester: Wiley.
- Jorgenson, D. W. and Z. Griliches (1967). The explanation of productivity change. *Review of Economic Studies* 34(3), 249–283.
- Kancs, D. and P. Ciaian (2011). Modelling the flow of knowledge and human capital: A framework of innovative capital. *International Journal of Public Policy* 7(1), 134–160.
- Kumbhakar, S. C., R. Ortega-Argiles, L. Potters, M. Vivarelli, and P. Voigt (2010). Corporate R&D and firm efficiency: Evidence from Europe’s top R&D investors. IPTS Working Papers 2010-11, European Commission, DG Joint Research Centre.

- Kwon, H. and T. Inui (2003). R&D and productivity growth in Japanese manufacturing firms. ERSI Discussion Paper Series 44, Economic and Social Research Institute.
- Levinsohn, J. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies* 70(2), 317–341.
- Luintel, K., M. Khan, and K. Theodoridis (2010). How robust is the R&D-productivity relationship? Evidence from OECD countries. Cardiff Economics Working Papers E2010/7, Cardiff University, Cardiff Business School.
- Mairesse, J. and P. Mohnen (2001). To be or not to be innovative: An exercise in measurement. NBER Working Papers 8644, National Bureau of Economic Research, Inc.
- Mairesse, J. and P. Mohnen (2005). The importance of R&D for innovation: A reassessment using French survey data. *Journal of Technology Transfer* 30(1), 183–197.
- Mairesse, J. and M. Sassenou (1991). R&D and productivity: A survey of econometric studies at the firm level. *Science, Technology Industry Review* 8, 9–43.
- Marschak, J. and W. H. Andrews (1944). Random simultaneous equations and the theory of production. *Econometrica* 12(3), 143–205.
- Marsili, O. (2001). *The Anatomy and Evolution of Industries*. Northampton, MA: Edward Elgar.
- Moncada-Paterno-Castello, P., C. Ciupagea, K. Smith, A. Tubke, and M. Tubbs (2010). Does Europe perform too little corporate R&D? A comparison of EU and non-EU corporate R&D performance. *Research Policy* 39(4), 523–536.
- Olley, G. S. and A. Pakes (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica* 64(6), 1263–1297.
- Ortega-Argiles, R., M. Piva, L. Potters, and M. Vivarelli (2010). Is corporate R&D investment in high-tech sectors more effective? *Contemporary Economic Policy* 28(3), 353–365.
- Potters, L. (2009). R&D in low-tech sectors. IPTS Working Papers 2009-08, European Commission, DG Joint Research Centre.
- Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1), 41–50.
- Sissoko, A. (2011). R&D subsidies and firm-level productivity: Evidence from France. Discussion Papers 2011/002, Catholic University of Leuven, IRES.
- Tsai, K.-H. and J.-C. Wang (2004). R&D productivity and the spillover effects of high-tech industry on the traditional manufacturing sector: The case of Taiwan. *The World Economy* 27(10), 1555–1570.

- Van Beveren, I. (2012). Total factor productivity estimation: A practical review. *Journal of Economic Surveys* 26(1), 98–128.
- Verspagen, B. (1995). R&D and productivity: A broad cross-section cross-country look. Open access publications from maastricht university, Maastricht University.
- Von Tunzelmann, N. and V. Acha (2005). Innovation in low-tech industries. In D. C. M. J. Fagerberg and R. R. Nelson (Eds.), *Oxford Handbook of Innovation*, pp. 407–432. New York: Oxford University Press.

Table 1: Descriptive statistics

	High-tech	Medium-high-tech	Medium-low-tech	Low-tech
R&D volume (2006), mln Euro	137780	115500	14511	14448
R&D intensity (2006) <sup>a</sup> , median value	3.60	0.77	0.25	0.10
Number of firms (Obs.)	462	458	79	130

Notes:

<sup>a</sup> R&D intensity is defined as a share of R&D expenditure in capital expenditure.

Table 2: Production function estimates

	All companies	High-tech	Medium-high-tech	Medium-low-tech	Low-tech
Materials	<b>0.691</b> (0.018)	<b>0.707</b> (0.025)	<b>0.693</b> (0.027)	<b>0.696</b> (0.026)	<b>0.651</b> (0.022)
Labour	<b>0.237</b> (0.010)	<b>0.210</b> (0.017)	<b>0.227</b> (0.016)	<b>0.232</b> (0.012)	<b>0.261</b> (0.013)
Capital	<b>0.120</b> (0.018)	<b>0.105</b> (0.023)	<b>0.102</b> (0.026)	<b>0.113</b> (0.027)	<b>0.153</b> (0.024)
Trend	<b>-0.016</b> (0.011)	<b>-0.015</b> (0.020)	<b>-0.015</b> (0.012)	<b>-0.017</b> (0.011)	<b>-0.019</b> (0.013)
Industry dummies	+	+	+	+	+
Regional dummies	+	+	+	+	+
Obs.	1129	462	458	79	130
R <sup>2</sup>	0.720	0.649	0.638	0.651	0.667

Notes:

Robust standard errors are reported in parentheses. The coefficient estimates that are significant at the 5% level are shown in the bold fold. The regressions contain regional and industry dummies.



Table 3: Regression results

	GPS: first-step regression, Equation (5) <sup>a</sup>			
	Coef.	S.E.	T-stat.	P-val.
Intercept	<b>15.298</b>	4.588	3.335	0.001
$\log(\text{Number of employees})_t$	<b>-1.235</b>	0.293	-4.222	0.000
$\log(\text{Number of employees})_t^2$	<b>0.041</b>	0.013	3.097	0.002
$\text{TFP}_{t-1}$	<b>-30.895</b>	13.856	-2.230	0.026
$\text{TFP}_{t-1}^2$	<b>24.590</b>	9.859	2.494	0.013
$\log(\text{Net Sales/Number of employees})_{t-1}$	-0.218	0.188	-1.160	0.246
$\log(\text{Net Sales/Number of employees})_{t-1}^2$	-0.026	0.046	-0.569	0.570
$\log(\text{Operating profit})_{t-1}^{POS}$	0.050	0.043	1.170	0.242
$\log(\text{Operating profit})_{t-1}^{NEG}$	-0.079	0.045	-1.737	0.083
$\log(\text{Operating profit})_t^{POS}$	<b>-0.115</b>	0.040	-2.903	0.004
$\log(\text{Operating profit})_t^{NEG}$	0.062	0.042	1.456	0.146
$\log(\text{Market capitalisation})_t$	0.233	0.336	0.693	0.489
$\log(\text{Market capitalisation})_t^2$	-0.013	0.022	-0.578	0.563
$\log(\text{Market capitalisation})_{t-1}$	-0.210	0.343	-0.613	0.540
$\log(\text{Market capitalisation})_{t-1}^2$	0.016	0.022	0.720	0.472
Industry dummies			+	
Regional dummies			+	
$R^2$	0.721			
Obs.	1129			

Notes:

<sup>a</sup> The dependent variable  $r_{it}$  in the first-step regression is the log of R&D intensity in 2006, defined as the share of R&D expenditure in capital expenditure in the same year, labelled as a period  $t$  in this table. The regression contains regional and industry dummies.

<sup>b</sup> The coefficient estimates that are significant at the 5% level are shown in the bold fold.

Table 4: Initial balancing properties of covariates

	Group 1	Group 2	Group 3
$\log(\text{Number of employees})_t$	<b>15.278</b>	<b>6.307</b>	<b>-21.260</b>
$\log(\text{Number of employees})_t^2$	<b>14.264</b>	<b>5.445</b>	<b>-21.249</b>
$\text{TFP}_{t-1}$	<b>16.387</b>	<b>5.315</b>	<b>-18.321</b>
$\text{TFP}_{t-1}^2$	<b>16.281</b>	<b>4.919</b>	<b>-18.881</b>
$\log(\text{Net Sales/Number of employees})_{t-1}$	<b>5.331</b>	<b>-3.236</b>	<b>-2.471</b>
$\log(\text{Net Sales/Number of employees})_{t-1}^2$	<b>-4.619</b>	0.961	<b>2.859</b>
$\log(\text{Operating profit})_{t-1}^{POS}$	<b>12.800</b>	<b>3.645</b>	<b>-15.607</b>
$\log(\text{Operating profit})_{t-1}^{NEG}$	<b>4.667</b>	<b>3.618</b>	<b>-6.826</b>
$\log(\text{Operating profit})_t^{POS}$	<b>15.326</b>	<b>3.402</b>	<b>-17.387</b>
$\log(\text{Operating profit})_t^{NEG}$	<b>5.791</b>	<b>3.068</b>	<b>-7.068</b>
$\log(\text{Market capitalisation})_t$	<b>11.551</b>	<b>1.981</b>	<b>-12.844</b>
$\log(\text{Market capitalisation})_t^2$	<b>10.786</b>	1.310	<b>-12.362</b>
$\log(\text{Market capitalisation})_{t-1}$	<b>10.741</b>	<b>2.167</b>	<b>-12.241</b>
$\log(\text{Market capitalisation})_{t-1}^2$	<b>10.037</b>	1.455	<b>-11.614</b>
Obs.	376	376	377

Notes:

Groups of approximately equal size were created using distribution of the continuous treatment variable, R&D intensity. Table entries are t-values of the test for the equal means between observations belonging to a particular group and those observations that do not belong to this group. The entries in the bold fold indicate significance at the 5% level. For the sake of brevity, the corresponding t-values for regional and industry dummy variables are not shown in the table.

Table 5: Cell size for testing the balancing property of GPS

	Group 1	Control 1	Group 2	Control 2	Group 3	Control 3
Bloc 1	56	431	72	208	45	529
Bloc 2	55	56	71	111	44	52
Bloc 3	55	45	72	96	45	21
Bloc 4	55	27	71	49	44	27
Bloc 5	56	22	72	36	45	6
Total	277	581	358	500	223	635

Notes:

The bloc size of each treatment group is held approximately the same. For each group it is determined by quintiles of the estimated GPS.

Table 6: GPS-adjusted balancing properties of covariates

	Group 1	Group 2	Group 3
$\log(\text{Number of employees})_t$	0.615	<b>2.699</b>	<b>-2.115</b>
$\log(\text{Number of employees})_t^2$	0.594	<b>2.617</b>	<b>-2.074</b>
$\text{TFP}_{2005}$	0.063	1.879	-1.278
$\text{TFP}_{2005}^2$	0.063	1.848	-1.269
$\log(\text{Net Sales/Number of employees})_{t-1}$	-1.882	-1.593	1.785
$\log(\text{Net Sales/Number of employees})_{t-1}^2$	<b>2.026</b>	1.053	-1.571
$\log(\text{Operating profit})_{t-1}^{POS}$	-0.902	1.829	-0.715
$\log(\text{Operating profit})_{t-1}^{NEG}$	-0.913	1.293	-0.271
$\log(\text{Operating profit})_t^{POS}$	0.540	1.283	-1.315
$\log(\text{Operating profit})_t^{NEG}$	-0.072	0.754	-0.770
$\log(\text{Market capitalisation})_t$	-0.604	0.704	-0.118
$\log(\text{Market capitalisation})_t^2$	-0.755	0.669	0.036
$\log(\text{Market capitalisation})_{t-1}$	-0.970	0.897	0.092
$\log(\text{Market capitalisation})_{t-1}^2$	-1.075	0.751	0.280
Obs.	277	358	223

Notes:

Table entries are t-values of the test for the equal means between observations belonging to a particular group and those observations that do not belong to this group, accounting for GPS. The entries in the bold fold indicate significance at the 5% level. For the sake of brevity, the corresponding t-values for regional and industry dummy variables are not shown in the table.

Table 7: Regression results

	GPS: second-step regression, Equation (6) <sup>b</sup>			
	Coef.	S.E.	T-stat.	P-val.
Incpt	<b>0.8192</b>	0.0045	183.740	0.000
$\ln r_{it-1}$	<b>-0.0176</b>	0.0021	-8.384	0.000
$(\ln r_{it-1})^2$	0.0003	0.0008	0.388	0.698
$\ln s_{it-1}$	<b>0.0114</b>	0.0046	2.456	0.014
$(\ln s_{it-1})^2$	<b>0.0015</b>	0.0007	2.019	0.044
$\ln r_{it-1} * \ln s_{it-1}$	<b>-0.0027</b>	0.0010	-2.728	0.007
$R^2$	0.127			
Obs.	858			

Notes:

<sup>a</sup> The dependent variable is the estimated level of TFP in 2007,  $\omega_{it}$ .

<sup>b</sup> The coefficient estimates that are significant at the 5% level are shown in the bold fold.

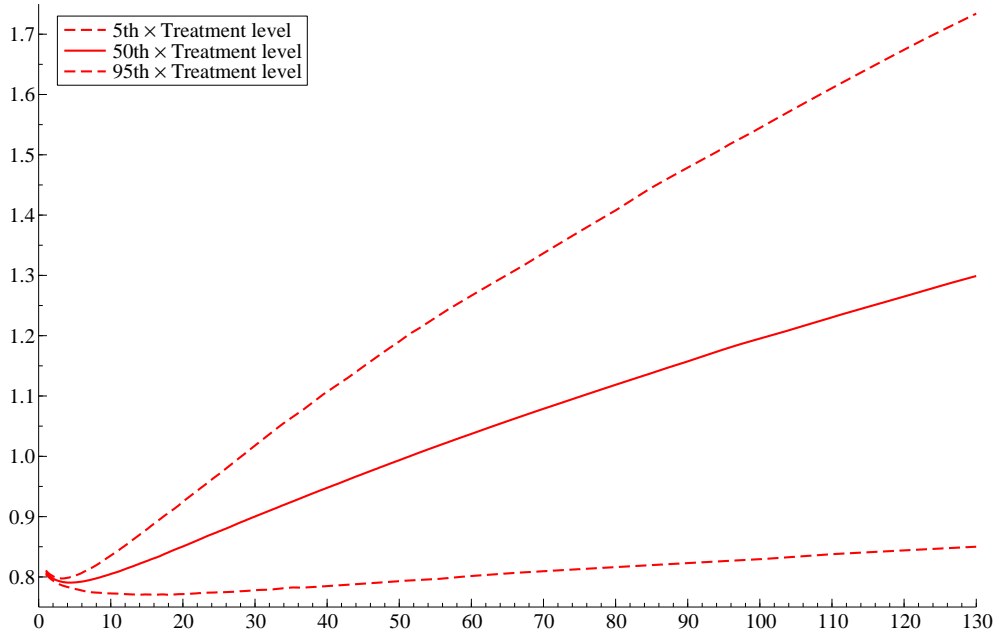


Figure 1: Dose-response function: **All companies**: Average expected response of TFP (2007) [Y-axis] to R&D intensity in 2006 [X-axis], GPS-adjusted. Dashed lines: bootstrapped 90 % confidence interval based on 1000 replications.

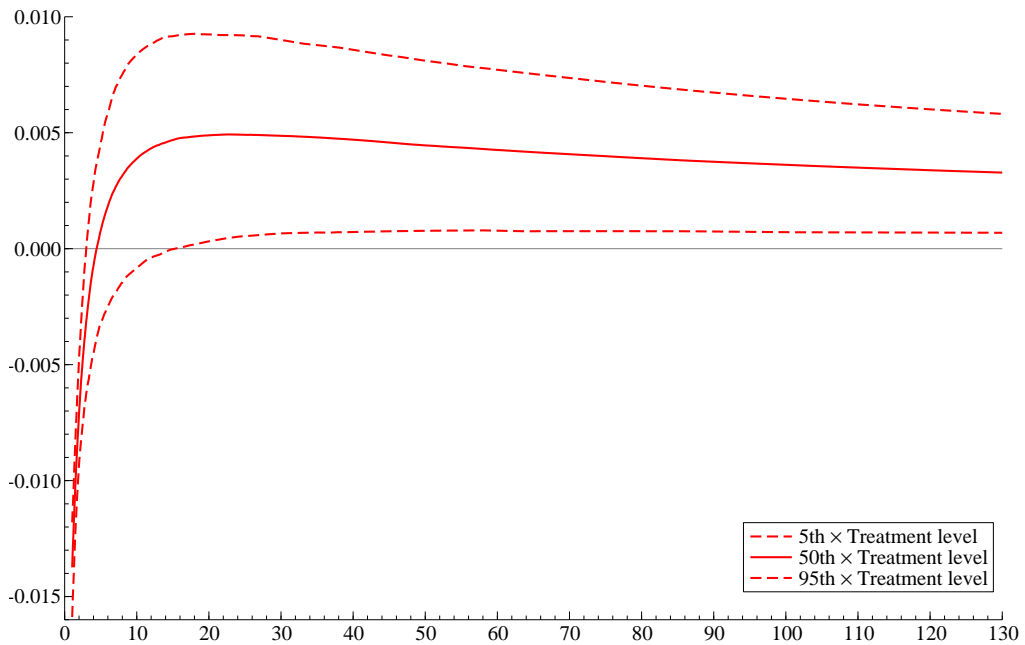


Figure 2: Treatment-effect function: **All companies**: Derivative of the average expected response of TFP (2007) [Y-axis] to R&D intensity in 2006 [X-axis], GPS-adjusted. Dashed lines: bootstrapped 90 % confidence interval based on 1000 replications.

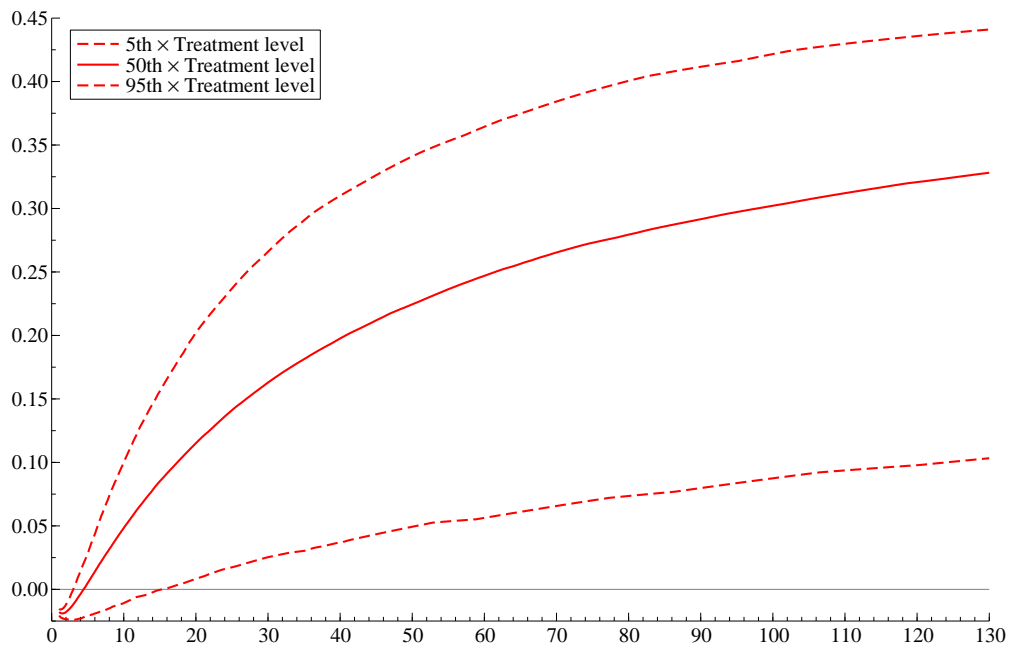


Figure 3: **All companies:** Elasticity of the average expected response of TFP (2007) [Y-axis] to R&D intensity in 2006 [X-axis], GPS-adjusted. Dashed lines: bootstrapped 90 % confidence interval based on 1000 replications.

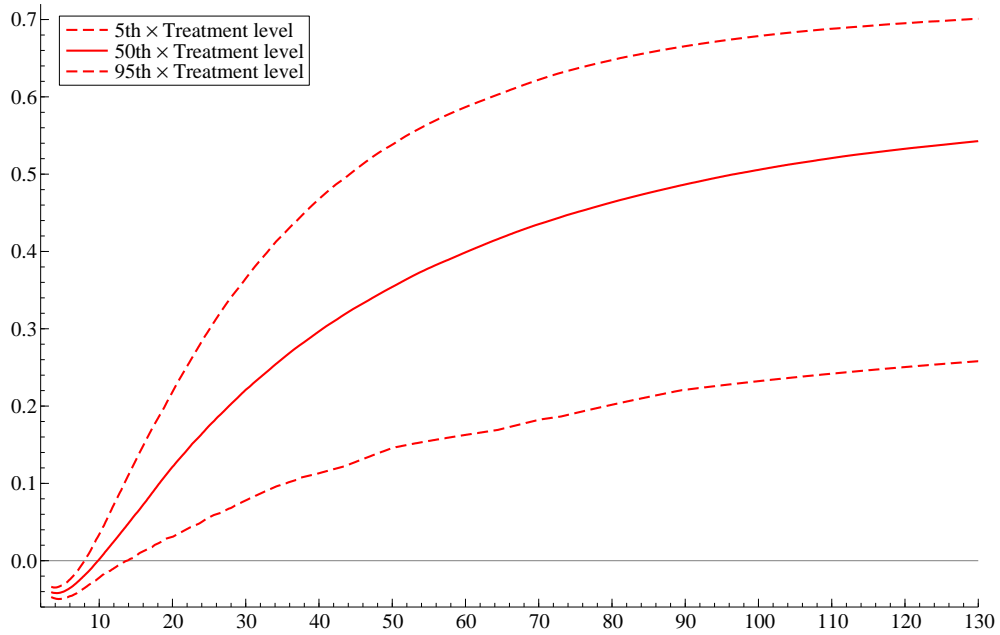


Figure 4: **High-tech companies:** Elasticity of the average expected response of TFP (2007) [Y-axis] to R&D intensity in 2006 [X-axis], GPS-adjusted. Dashed lines: bootstrapped 90 % confidence interval based on 1000 replications.

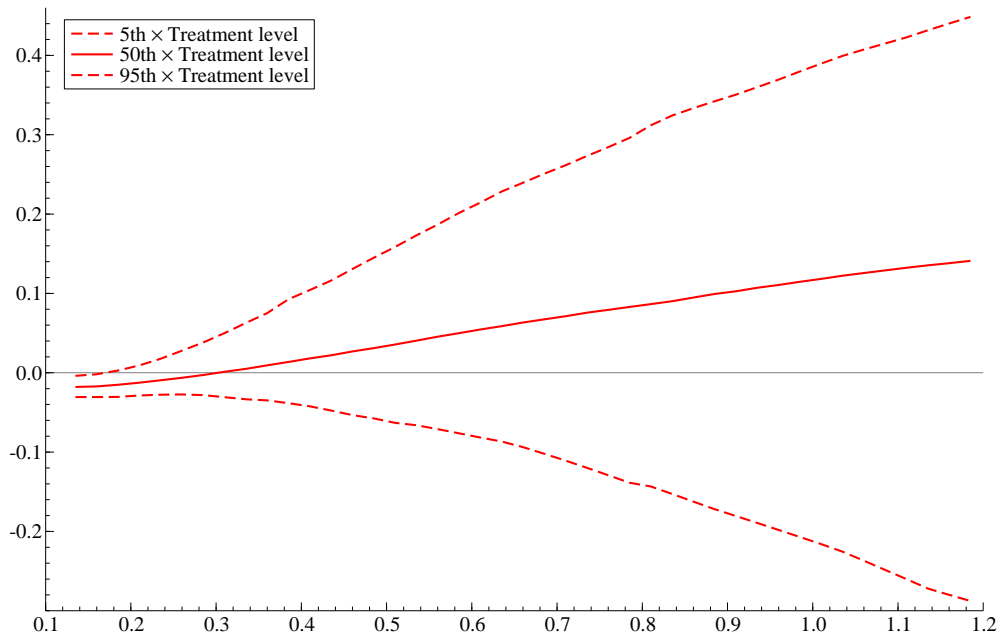


Figure 5: **Low-tech companies:** Elasticity of the average expected response of TFP (2007) [Y-axis] to R&D intensity in 2006 [X-axis], GPS-adjusted. Dashed lines: bootstrapped 90 % confidence interval based on 1000 replications.