R&D and Non-linear Productivity Growth[§]

d'Artis Kancs
European Commission
DG Joint Research Centre
Inca Garcilaso 3
41092 Seville, Spain
d'artis.kancs@ec.europa.eu

Boriss Siliverstovs
ETH Zurich
KOF Swiss Economic Institute
Leonhardstrasse 21
8092 Zurich, Switzerland
siliverstovs@kof.ethz.ch

March 4, 2015

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. We employ a two step estimation approach, and match two firm-level panel data sets for the OECD countries, which allows us to relax the linearity assumption 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 knowledge is accumulated, 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.

[§]The authors acknowledge helpful comments from Michele Cincera, Abraham Garcia, Fernando Hervás Soriano, Pietro Moncada-Paterno-Castello, Sandro Montresor, Enrico Santarelli, Peter Voigt, two anonymous referees as well as conference participants at the CONCORD and ECOMOD in Sevilla, RES in London, EEA-ESEM in Gothenburg. 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. Our main contribution to the sizeable existing literature on R&D-productivity relationship lies in providing evidence that and how the relationship between innovation and productivity growth is non-linear. On the one hand, our results are within the range of previous empirical evidence. On the other hand, they provide a more precise explanation for differences in previous econometric results.

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 (Griliches, 1973; Terleckyj, 1974), 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).

In the theoretical literature, there is a general consensus that R&D activities play a decisive role in fostering productivity growth. This relation has been first formalised by Griliches (1973) and Terleckyj (1974) and had widely been accepted since. Theoretical literature also recognises that, in addition to appropriable part, any innovative activity has also an information component that is almost completely non-appropriable and costless to acquire, an idea dating back to Marshall (1920); Nelson (1959); Arrow (1962). While being considered as one of the most obvious characteristic features of R&D (Leahy and Neary, 2007), the formalisation of this idea in a general equilibrium setup, though, came relatively recently, splitting research activities into an appropriable and non-appropriable knowledge, as for example in Goulder and Schneider (1999) in the context of climate studies or Diao *et al.* (1999) based on a theory of endogenous growth, based on the extension of product varieties (Romer, 1990; Grossman and Helpman, 1991; Aghion and Howitt, 1992; Jones, 1995). In the present study we focus specifically on the appropriable knowledge.

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 to R&D investment 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 these empirical results

(Mairesse and Sassenou, 1991; Czarnitzki *et al.*, 2009; Luintel *et al.*, 2010).[1]

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 for innovators. In addition, both policy makers and innovators are more interested in specific issues, such as, how a particular level of R&D investment would affect productivity in a particular sector at a particular level of technological intensity.

In order to improve 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 the impact of R&D 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 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 firms in other sectors. Kwon and Inui (2003) used the same estimation strategy to analyse the impact of R&D on labour productivity in manufacturing firms using a sample of 3,830 Japanese 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 impact 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 firm productivity growth (elasticity 0.18), whereas the impact was considerably higher for high-

[1] Surveying firm level studies on R&D impact Mairesse and Sassenou (1991) concluded 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 or a result, for example, of differences in the period, industries or countries considered, or simply the reflection of peculiarities of the individual studies.

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 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). They 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).

In order to control for non-linearities in productivity's response to R&D investment, more recent studies find 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 (Doraszelski and Jaumandreu, 2013). According to Furman *et al.* (2002), the productivity of R&D investment may be sensitive to the level of technological intensity (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 current productivity. On the other hand, due to the so-called "fishing out" effect, prior R&D investment may have discovered 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, a critical mass of existing knowledge is found to be an important cause 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 knowledge has been accumulated. 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 estimation approach, Bogliacino (2010) can capture non-linear effects only via a square term of R&D inserted 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 these lines of the recent 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. We attempt to answer two questions: how R&D investment affects firm productivity at different levels of technological intensity, and what are the inter-sectoral productivity differences with respect to productivity effects of R&D investment. These questions are highly relevant for both R&D performers and policy makers, though none of them has been answered in a satisfactory way in the literature yet.

In order to relax the linearity assumption of 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) consider firm productivity as unobservable to the econometrician. In this way, the linearity assumption of the R&D process can be relaxed, because we do not need to construct the stock of knowledge capital.

In the second step, we employ the generalised propensity score (GPS) approach of Hirano and Imbens (2004) to estimate the impact of R&D on firm productivity.[2] 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 eliminate selection bias arising due to a non-random assignment (choice) of treatment (R&D) intensity across firms by conditioning on observed firms' characteristics. Following Hirano and Imbens (2004), we estimate the relationship between R&D and productivity by means of OLS using a flexible parametric approach. In this way, potential non-linearities between the variables of interest can be captured. Hence, no specific functional form for the R&D-productivity relationship needs to be imposed a priori. Generalising the *unconfoundness* assumption underlying the binary treatment propensity score method, Hirano and Imbens (2004) introduce a *weak unconfoundness* assumption stating that after controlling for observed firm characteristics, firms' choice of treatment intensity can be considered as good as random, i.e. independent of the potential outcomes.

In particular, the most significant contribution of Rosenbaum and Rubin (1983) was to show that all the information about firm's decision to undertake treatment in case of binary outcomes, that allows drawing comparison among treated and untreated units, can be condensed in one scalar, called the *propensity score*, indicating a propensity or probability of a firm to receive a treatment given its characteristics. Hence, conditioning on a single variables (propensity score)

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

rather on the whole multitude of relevant firm's characteristics solves a problem of the "curse of dimensionality". This result of Rosenbaum and Rubin (1983) was extended to the case of continuous treatment in Hirano and Imbens (2004) by introducing the generalised propensity score, indicating the propensity of a firm to choose a given level of treatment conditional on its characteristics. Hence, the appealing feature of a propensity score based method is that it allows to identify the treatment effect in a parsimonious way. Having said that, it is also necessary to mention that the propensity score methodology is often referred to as a "data-hungry" approach, in that it requires that all important characteristics influencing the selection process to be accounted for; failure to do so, results in biased estimates of the treatment effect.

Based on the assumptions underlying the GPS method, it is possible to construct a dose-response function and its derivatives capturing average and marginal treatment effects as a function of different levels of treatment intensity. 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 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 intensity, and inter-sectoral productivity differences of R&D investment.[3]

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 the 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 knowledge is accumulated, 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

[3]The definition of high-tech, medium-high-tech, medium-low-tech and low-tech sectors follows the OECD classification of industrial sectors, according to their technological intensity.

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 R&D investment 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 an estimate 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 of its convenience in empirical implementation, 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. As noted by Griliches (2000), both assumptions have been rejected in recent firm level empirical studies. Conceptually, it is impossible to estimate a differential impact of R&D investment on firm productivity at different levels of technological intensity if a linear R&D-productivity relationship is imposed. In addition, 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 the linearity assumption 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) 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 approach in the 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.[4] Output, y_{it}, of firm i in period t (in logarithmic form) can be expressed as:[5]

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

where β_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, ω_{it}, is known to firm i in period t, but not observable to researchers, and therefore must be estimated.

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

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

According to the controlled Markov law of motion (2),[6] the actual productivity, ω_{it}, of firm i in period t is composed of expected productivity, $g(\omega_{it-1}, r_{it-1})$, and unpredictable productivity innovation, ξ_{it}. The conditional expectation function $g(\cdot)$ depends on past productivity, ω_{it-1}, and past R&D expenditure, r_{it-1}. Past R&D expenditures, r_{it-1}, account 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, $g(\omega_{it-1}, r_{it-1})$,

[4] 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-partials with productivity and the value of the firm has to be increasing in dynamic inputs (Ackerberg *et al.*, 2007).

[5] Given that we are closely following the methodology of Doraszelski and Jaumandreu (2013), we are adopting their notation as far as possible.

[6] $\omega_{it} = g(\omega_{it-1}, r_{it-1}) + \xi_{it} = \alpha_0 + \alpha_1 \omega_{it-1} + \alpha_2 (\omega_{it-1})^2 + \alpha_3 (\omega_{it-1})^3 + \alpha_4 r_{it-1} + \xi_{it}$.

made in period $t-1$ on productivity in period t. Random shock, ξ_{it}, is zero mean independent productivity innovation, which represents both uncertainties linked to past productivity, ω_{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.

As shown by Doraszelski and Jaumandreu (2013, p. 1346), the assumptions that labour and materials are static inputs in production function (1) imply the following labour demand function:

$$\begin{aligned} l\left(t, k_{it}, \omega_{it}, w_{it}, p_{Mit}, d_{it}\right) \quad = \quad & \frac{1}{1-\beta_l-\beta_m}\Bigg(\beta_0+\left(1-\beta_m\right)\ln\beta_l+\beta_m\ln\beta_m+\mu+\beta_t t+\beta_k k_{it}+\omega_{it} \\ & -\left(1-\beta_m\right)\left(w_{it}-p_{it}\right)-\beta_m\left(p_{Mit}-p_{it}\right)+\ln\left(1-\frac{1}{\eta\left(p_{it}, d_{it}\right)}\right)\Bigg) \end{aligned} \tag{3}$$

where $\mu=\ln E\left[\exp\left(e_{it}\right)\right]$, $\left(w_{it}-p_{it}\right)$ is the real wage, and $\left(p_{Mit}-p_{it}\right)$ is the real price of materials. As shown by Doraszelski and Jaumandreu (2013, p. 1346), solving labour demand function (3) for productivity, ω_{it}, one obtains the inverse labour demand function:

$$\begin{aligned} h_l\left(t, k_{it}, l_{it}, p_{it}, w_{it}, p_{Mit}, d_{it}\right) \quad = \quad & \lambda_l-\beta_t t-\beta_k k_{it}+\left(1-\beta_l-\beta_m\right) l_{it}+\left(1-\beta_m\right)\left(w_{it}-p_{it}\right) \\ & +\beta_m\left(p_{Mit}-p_{it}\right)-\ln\left(1-\frac{1}{\eta\left(p_{it}, d_{it}\right)}\right) \end{aligned} \tag{4}$$

where $\lambda_l=-\beta_0-\left(1-\beta_m\right)\ln\beta_l-\beta_m\ln\beta_m-\mu$ is a constant.

As shown by Doraszelski and Jaumandreu (2013, p. 1346), substituting firm's productivity (2) into production function (1) and using the inverse labour demand function (4) yields an empirically estimable production function:

$$y_{it}=\beta_0+\beta_t t+\beta_k k_{it}+\beta_l l_{it}+\beta_m m_{it}+\left(\omega\left(l_{it-1}, k_{it-1}, p_{it-1}, q_{it-1}, w_{it-1}\right), r_{it-1}\right)+\xi_{it}+e_{it} \tag{5}$$

where time trend, t, captures shifts in productivity that are not part of ω_{it}, and w_{it}, q_{it} and p_{it} are prices for labour, materials and output, respectively.

When estimating equation (5), the endogeneity of static inputs needs to be addressed. Whereas capital input, k_{it}, is uncorrelated with ξ_{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 ξ_{it}), labour input, l_{it}, and material input, m_{it}, are correlated with ξ_{it} (because ξ_{it} is part of ω_{it}, and l_{it} and m_{it} are functions of ω_{it}). The endogeneity of static inputs can be addressed by instrumenting for l_{it} and m_{it}, which requires that the instruments are not correlated with ξ_{it}. Our instrumenting strategy follows Doraszelski and Jaumandreu (2013): given that l_{it-1} and m_{it-1} are uncorrelated with ξ_{it}, we use them as instruments for l_{it} and m_{it}.

2.2.2 R&D and firm productivity

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 (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 also the case with the BPS. Third, 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. However, before describing the GPS methodology, it is worthwhile to outline the temporal framework of our analysis. The values of the response variable of interest (TFP) are recorded for the year 2007— a year before the outbreak of the global financial crisis, triggered by the collapse of Lehman Brothers in September 2008. in order to avoid a simultaneity bias, the values of our treatment variable (R&D intensity) are taken for the year 2006. Correspondingly, the GPS values are also defined for this year. The GPS is obtained by conditioning on the observed covariate values for the years 2006 as well as 2005.

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

$$r_i^{2006}|X_i^{2006} \sim N(X_i^{2006'^2}), \tag{6}$$

where $X_i^{2006} \equiv \left(Z_i^{2006'}; Z_i^{2005'}\right)'$ is a $z \times 1$ vector of both contemporaneous and lagged values of discrete and continuous covariates, Z_i. The parameters of the conditional distribution (γ, σ^2) are evaluated using a standard OLS regression. The estimated GPS is defined as follows:

$$\widehat{s}_i^{2006} = \frac{1}{\sqrt{2\pi\widehat{\sigma}^2}} \exp\left[-\frac{1}{2\widehat{\sigma}^2}(r_i^{2006} - X_i^{2006'}\widehat{\gamma})^2\right]. \tag{7}$$

The propensity score in equation (7) 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, ω_i^{2007}, is modelled as a flexible parametric function of treatment (R&D investment) and the generalised propensity score, r_i^{2006} and s_i^{2006}, respectively:

$$E[\omega_i^{2007}|r_i^{2006}, s_i^{2006}] = \alpha_0 + \alpha_1 * r_i^{2006} + \alpha_2 * \left[r_i^{2006}\right]^2 + \alpha_3 * s_i^{2006} + \alpha_4 * \left[s_i^{2006}\right]^2 + \alpha_5 * \left[r_i^{2006} * s_i^{2006}\right], \quad (8)$$

where the latter is substituted with its estimates, $\widehat{s}_i^{2006}$, from the first step. The flexibility of the functional form can be controlled for by varying the power of variables r_i^{2006} and s_i^{2006} and their cross-products.

The average expected response of target variable, ω, for treatment dose, τ, is estimated in the third step:

$$E[\widehat{\omega^{2007}(\tau)}] = \frac{1}{N}\sum_{i=1}^{N}\left[\widehat{\alpha}_0 + \widehat{\alpha}_1 * \tau + \widehat{\alpha}_2 * \tau^2 + \widehat{\alpha}_3 * \widehat{s}(\tau, X_i^{2006}) + \widehat{\alpha}_4 * \widehat{s}(\tau, X_i^{2006})^2 + \widehat{\alpha}_5 * (\tau * \widehat{s}(\tau, X_i^{2006}))\right], \quad (9)$$

where the coefficient estimates from the expected response (8) are used. The whole dose-response function is obtained by computing equation (9) 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 $\widehat{E[\omega_t(\tau)]}$ with respect to argument. The latter function is computed using a standard elasticity formula $[\partial \widehat{E[\omega_t(\tau)]}/\partial\tau]/[\widehat{E[\omega_t(\tau)]}/\tau]$. The latter function is of 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 in our study is the EU Industrial R&D Investment Scoreboard. The Scoreboard is an annual data set compiled and provided by the European Commission. The Scoreboard data set we use in the empirical analysis comprises data on R&D investment, as well as other financial and economic variables (e.g. net sales, operating profits, employees) for the top

1,500 R&D global performers: 750 companies based in the EU and 750 based outside the EU.[7][8] In addition to economic and financial variables, the Scoreboard identifies also 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 the 1st 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 classifi-

[7] In the first edition (2004) the top companies were 500 EU and 500 non-EU; in the eleventh edition (2014), they are 2,500 in total.

[8] 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 750 companies based in the EU and 750 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.

cation 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 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 accounting data 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 covering both large multinationals as well as small and medium sized companies (SMEs) across all industries world wide for the 1996-2011 period. In order to enhance the comparison across countries, all firm accounts are transformed into a universal format. We use the European Central Bank period average exchange rates to convert all accounting data into EURO.

The Orbis database contains firm-level accounting data in a standardised financial format for 26 balance sheet variables, 25 income statement variables, and 26 financial ratios based on these variables. In addition, it contains also important information about company characteristics. 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 ω_{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 (5). 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), Ackerberg *et al.* (2006), Doraszelski and Jaumandreu (2013).

From the production function estimates we construct productivity series for each firm,[9] which in turn are used to construct the response variable – TFP in 2007 – for each firm. After accounting for missing values in the main variables of interest 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 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 is 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 inter-sectoral heterogeneity of firms with respect to R&D intensity, we regroup all firms into four sub-samples according to the level of technological intensity. 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;

[9] The firm-level productivity estimates, $\hat{\omega}_{it}$, are retrieved using: $\hat{\omega}_{it} = -\hat{\beta}_t t + \left(1 - \hat{\beta}_l - \hat{\beta}_m\right) l_{it} - \hat{\beta}_k k_{it} + \left(1 - \hat{\beta}_m\right)(w_{it} - p_{it}) + \hat{\beta}_m (q_{it} - p_{it})$, 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.

A 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. Overall, the R&D intensity takes values in the range between the minimum value of 0.0026 and the maximum value of 126.38. We inserted the footnote with this information.

Given that the unconditional distribution of the treatment variable is highly skewed, we take a logarithmic transformation of explanatory variable in order comply with the normality assumptions in the first step of the GPS regression. As suggested by 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 (7) 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, $\widehat{s}_{it}$.[10] 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)

[10]The adequacy of the estimated GPS is checked by assessing its balancing properties.

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 variables log(Operating profit)^{POS} and log(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".

4 Results

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

The results of the first step GPS estimation procedure (equation 6) are reported in Table 3 for the pooled sample containing all companies. From the table we note that variation in the R&D intensity is best captured by the number of employees, its squared value, operating profits as well as lagged value of TFP and its squared value. Also the industry-specific dummy variables contribute substantially to the explanatory power of the first step of the GPS regression.[11] 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. The underlying assumption of normality of our treatment variable given the covariates was tested using the Kolmogorov-Smirnov test (Marzucchi *et al.*, 2015). The corresponding p-value is 0.55, indicating that we cannot reject the corresponding null hypothesis at the usual levels.

We verify whether the GPS is appropriately specified by testing the so-called balancing property. In doing so, we follow the procedure applied in Hirano and Imbens (2004, p. 80-82) by dividing data into three groups of approximately equal size according to the terciles of the treatment intensity distribution. This allows us to verify whether there are substantial differences in observables pertaining to different levels of treatment intensity, indicating a possibility of non-random selection into different levels of treatment which causes biases in causal inference of treatment effects. The results presented in Table 4 indicate that indeed there is a very strong heterogeneity among the covariates belonging to different groups. The entries in each column are t-values corresponding to the null hypothesis that the mean of a respective covariate in a group i is equal to the mean of the corresponding covariate in the remaining groups $j \neq i$. The corresponding null hypothesis is rejected in all but three cases at the 5% significance level. A well specified GPS should be able to successfully account for these differences.

Before verifying whether the balancing property holds for the GPS adjusted data, we followed Becker *et al.* (2012) and imposed a common-support condition on the observed ranges of GPS values for observations in each treated group i with those belonging to the control groups $j \neq i$. Observations falling outside of the overlapping interval of the GPS values were omitted from

[11] These results are not reported for the sake of saving space.

our further analysis. This allows us to focus on the observations that are sufficiently uniform in observable characteristics, concentrated in GPS, at each level of treatment intensity.

The balancing properties of covariates adjusted for the GPS are reported in Table 5. 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 5 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 6 reports the second step GPS estimation results, (equation 8), where the relationship between the response and treatment variables is specified, conditioning on the estimated GPS from the first step. The functional form represents a quadratic approximation also used in Hirano and Imbens (2004, p. 82). According to Table 6, the treatment variable, $\ln r_i^{2006}$, the generalised propensity score, $\ln s_i^{2006}$, its squared value, $\left[\ln s_i^{2006}\right]^2$, and the cross-product term, $\ln r_i^{2006} * \ln s_i^{2006}$, have coefficient estimates that are significantly different from zero.

The results from the third step, (summarised in equation 9), 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 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 also 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. Without sufficient existing knowledge, firms are not capable to absorb and use the new knowledge effectively, and hence are not able to benefit from internal and external R&D investment (Cincera, 1997; Cohen and Levinthal, 1989; Griffith *et al.*, 2004; Fabrizio, 2009). An even more interesting result seems to be the fact that, after becoming positive, the effect of R&D on productivity continues being positive but it's decreasing. These results jointly support both the "standing on the giants shoulders" and the "fishing-out effect" hypotheses discussed in Section 1. Previously accumulated knowledge "standing on giants shoulders" seems to be always beneficial for improving firm productivity. However, after a certain level (around 20%), prior research may have discovered the ideas which are the easiest to find and further improvements are more difficult (the marginal effect of R&D starts to decrease, as it can be seen in Figure 2).

4.2 Inter-sectoral firm heterogeneity

Given that our sample consists of very heterogeneous firms, for which R&D intensity may play rather differentiated role in increasing firm productivity, the impact of equivalent R&D investment may differ between different sectors. For example, given that persistent innovation is likely to be more important for productivity growth than temporary innovation (Raymond *et al.*, 2010), if persistent R&D performers reside more frequently in high-tech sectors than in low-tech sectors, then high-tech innovators would benefit more from R&D than low-tech innovators. Second, R&D may be only one part of innovation process particularly 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 (Potters, 2009). Third, through knowledge spillovers, R&D inputs in high-tech sectors may contribute importantly to the innovative power of low-tech sectors. On the other hand, firms in low-tech sectors may benefit from a "late-comer advantage", while firms in high-tech sectors may be more affected by diminishing returns to R&D due to the so-called "fishing out" effect, 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).

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 intensity. 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.[12] 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, 2003; Tsai and Wang, 2004; Ortega-Argiles *et al.*, 2010; Kumbhakar *et al.*, 2010).

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, though this estimate is not statistically significant from zero. Compared to firms in high-tech industries, the point estimate of 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.[13]

Our results seem to be in line with previous literature, which provides several reasons, why firms in low-tech industries may benefit less from own R&D investment than firms in high-tech industries. First, persistent innovation is found to be more important for productivity growth than temporary innovation and, according to Raymond *et al.* (2010), persistent R&D innovators reside more frequently high-tech sectors than in low-tech sectors. Second, innovation does not stop

[12]The results for the first step and second step GPS estimations for low-, medium–low, medium-high and high-tech sub-samples are qualitatively comparable to those reported for the full sample, and therefore are not reported separately. These results are available from the authors upon request.

[13]We compute the dose-response functions and their derivatives for medium-low-tech and medium-high-tech companies. It turns out that after imposing the common support assumption for medium-low-tech companies, there only 39 firms are left in this group, which is not sufficient, in our opinion, to obtain reliable estimates of the dose-response function. For medium-high-tech companies we found that the results, though these are based on a sufficiently large number of observations (442 firms), are qualitatively similar to those reported for the low-tech companies, i.e. we do not find a statistically significant evidence of existence of a close relationship between R&D and firm productivity for this group of companies. This is reflected in the low value of R^2 reported for the second-stage regression is 0.08 for medium-high-tech companies, which is much lower than the $R^2 = 0.23$ reported for the second-stage regression for high-tech companies. For the sake of brevity, we make the corresponding estimates available upon request.

at R&D investments, implying that R&D is only one part of innovation process. 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. Third, through knowledge spillovers, the innovative activities in high-tech sectors contribute importantly to the innovative power of low-tech sectors (Potters, 2009). The possibility of substitution between intramural R&D efforts and spillovers from high-tech sectors seem particularly plausible in light of our rather broad sectoral classification. For example, innovation in software & computer services (high-tech sector) could increase firm productivity also in many low-tech industries without their own intramural R&D efforts.

5 Conclusions

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 intensity, 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 in the literature 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 the potential non-linearities in the R&D-productivity relationship. 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) approach of Hirano and Imbens (2004) to estimate the impact of R&D investment on firm productivity. By employing a two step estimation approach we are able to relax the linearity assumption 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 knowledge is accumulated, 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.

An important caveat about our empirical analysis concerns the nature of the sample, which is made by the top innovators with regard to R&D investments. First, while other data sets, such as the OECD BERD data, can be considered fully representative of the OECD economies, in the EU Industrial R&D Investment Scoreboard used in the present study only the R&D "champions" are considered. This is a clear limitation of our analysis, the results of which cannot be straightforwardly extrapolated to e.g. SMEs. However – notwithstanding this source of sample selection – our analysis still provides interesting insights, and has even certain support from the empirical evidence on concentration of innovative activities. It is well documented in the previous literature that innovative activities are highly concentrated – only a small share of firms around the world innovate, the majority of firms in most regions around the world do not engage in any significant R&D activities, they imitate (Slivko and Theilen, 2014). Hence, by considering the top 1,500 innovators, which account for more than 80% of global R&D expenditure (top 2,500 account for more than 90% of global R&D expenditure), ensures also certain representativeness. Second, the data set used in the empirical analysis does not allow to account for knowledge spillovers. However, according to previous evidence, knowledge spillovers are an important form of externality associated with innovative activities. For example, according to Bloom *et al.* (2013), the positive effects deriving from technology spillovers counterbalance the negative business-stealing effects derived from increased productivity of competitors. In order to account for knowledge spillovers, other type of data, such as patent citations, would need to be used.

These results allow us to better interpret the wide distribution of estimates reported in previous studies, and brings more precision to the estimated effects of R&D, and the identified relationship between R&D expenditures and firm productivity allows to derive some 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 policy instruments may have very different opportunity costs. Reversely, the same policy instrument may have very different implications, if applied to different sectors 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, the productivity elasticity of R&D is required, which according to our results is not constant with respect to the level of technological intensity – it ranges from -0.02 for low levels of R&D intensity to 0.33 for high levels of R&D intensity. Hence, taking an average value for all firms would generate biased and/or inefficient policy recommendations. Our results provide the entire functional relationship between the R&D investment and firm productivity which can be readily 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 from policy perspective the allocation of R&D support is as important issue as measures to 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, but should be targeted in the long-run, if knowledge-based economy is a long-term policy objective.

References

Ackerberg, D., Caves, K. and Frazer, G. (2006). *Structural identification of production functions*. Working paper, UCLA, Los Angeles.

—, Lanier Benkard, C., Berry, S. and Pakes, A. (2007). Econometric tools for analyzing market outcomes. In J. Heckman and E. Leamer (eds.), *Handbook of Econometrics*, vol. 6A, *63*, 1st edn., Elsevier.

Aghion, P. and Howitt, P. (1992). A Model of Growth through Creative Destruction. *Econometrica*, **60** (2), 323–51.

Arrow, K. (1962). Economic Welfare and the Allocation of Resources for Invention. In *The Rate* *and Direction of Inventive Activity: Economic and Social Factors*, NBER Chapters, National Bureau of Economic Research, pp. 609–626.

Becker, S. O., Egger, P. H. and von Ehrlich, M. (2012). Too much of a good thing? On the growth effects of the EU's regional policy. *European Economic Review*, **56** (4), 648–668.

Bloom, N., Schankerman, M. and Reenen, J. V. (2013). Identifying Technology Spillovers and Product Market Rivalry. *Econometrica*, **81** (4), 1347–1393.

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 Robin, S. (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 Levinthal, D. A. (1989). Innovation and learning: The two faces of R&D. *Economic Journal*, **99** (1), 569–596.

Cuneo, P. and Mairesse, J. (1984). Productivity and R&D at the firm level in French manufacturing. In *R&D, Patents, and Productivity*, NBER Chapters, National Bureau of Economic Research, pp. 375–392.

Czarnitzki, D., Hanel, P. and Rosa, J. M. (2011). Evaluating the impact of R&D tax credits on innovation: A microeconometric study on Canadian firms. *Research Policy*, **40** (2), 217–229.

—, Kraft, K. and Thorwarth, S. (2009). The knowledge production of 'R' and 'D'. *Economics Letters*, **105** (1), 141–143.

Diao, X., Roe, T. and Yeldan, E. (1999). Strategic policies and growth: an applied model of R&D-driven endogenous growth. *Journal of Development Economics*, **60** (2), 343–380.

Doornik, J. and Hendry, D. (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 Jaumandreu, J. (2013). R&D and productivity: Estimating endogenous productivity. *Review of Economic Studies*, **80** (4), 1338–1383.

Eberhardt, M. and Helmers, C. (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., Porter, M. E. and Stern, S. (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 Jaumandreu, J. (1998). *Threshold effects in product R&D decisions: Theoretical framework and empirical analysis*. Studies on the Spanish Economy 45, FEDEA.

Goulder, L. H. and Schneider, S. H. (1999). Induced technological change and the attractiveness of CO2 abatement policies. *Resource and Energy Economics*, **21** (3-4), 211–253.

Griffith, R., Redding, S. and Reenen, J. V. (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.

— (1973). Research expenditures and growth accounting. In *Science and Technology in Economic Growth*, London: Palgrave Macmillan, pp. 59–95.

— (1979). Issues in assessing the contribution of research and development to productivity growth. *Bell Journal of Economics*, **10** (1), 92–116.

— (2000). *R&D, education and productivity: A retrospective*. Cambridge Massachusetts: Harvard University Press.

— and Mairesse, J. (1983). Comparing productivity growth: An exploration of French and U.S. industrial and firm data. *European Economic Review*, **21** (1-2), 89–119.

Grossman, G. and Helpman, E. (1991). *Innovation and Growth in the Global Economy*. Cambridge: MIT Press.

Hall, B. H. and Mairesse, J. (1995). Exploring the relationship between R&D and productivity in French manufacturing firms. *Journal of Econometrics*, **65** (1), 263–293.

—, — and Mohnen, P. (2010). Measuring the returns to R&D. In B. H. Hall and N. Rosenberg (eds.), *Handbook of the Economics of Innovation*, vol. 2, *22*, Elsevier, pp. 1033–1082.

Harhoff, D. (1998). R&D and productivity in German manufacturing firms. *Economics of Innovation and New Technology*, **6** (1), 29–50.

Hirano, K. and Imbens, G. W. (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*, Chichester: Wiley, pp. 73–84.

Jones, C. I. (1995). R&D-Based Models of Economic Growth. *Journal of Political Economy*, **103** (4), 759–84.

Jorgenson, D. W. and Griliches, Z. (1967). The explanation of productivity change. *Review of Economic Studies*, **34** (3), 249–283.

Kancs, D. and Ciaian, P. (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., Ortega-Argiles, R., Potters, L., Vivarelli, M. and Voigt, P. (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 Inui, T. (2003). *R&D and Productivity Growth in Japanese Manufacturing Firms*. ERSI Discussion Paper Series 44, Economic and Social Research Institute.

Leahy, D. and Neary, J. P. (2007). Absorptive capacity, R&D spillovers, and public policy. *International Journal of Industrial Organization*, **25** (5), 1089–1108.

Levinsohn, J. and Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies*, **70** (2), 317–341.

Luintel, K., Khan, M. and Theodoridis, K. (2010). *How Robust is the R&D-Productivity relationship? Evidence from OECD Countries*. Cardiff Economics Working Papers 2010/7, Cardiff University, Cardiff Business School.

Mairesse, J. and Mohnen, P. (2001). *To be or not to be innovative: An exercise in measurement.* NBER Working Papers 8644, National Bureau of Economic Research.

— and — (2005). The importance of R&D for innovation: A reassessment using French survey data. *Journal of Technology Transfer*, **30** (1), 183–197.

— and Sassenou, M. (1991). R&D and productivity: A survey of econometric studies at the firm level. *Science, Technology Industry Review*, **8**, 9–43.

Marshall, A. (1920). *Principles of economics.* London: Macmillan.

Marsili, O. (2001). *The Anatomy and Evolution of Industries.* Northampton, MA: Edward Elgar.

Marzucchi, A., Antonioli, D. and Montresor, S. (2015). Industry research co-operation within and across regional boundaries. What does innovation policy add? *Papers in Regional Science*, **Forthcoming**, dx.doi.org/10.1111/pirs.12079.

Moncada-Paterno-Castello, P., Ciupagea, C., Smith, K., Tubke, A. and Tubbs, M. (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.

Nelson, R. R. (1959). The Simple Economics of Basic Scientific Research. *Journal of Political Economy*, **67**, 297–298.

Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, **64** (6), 1263–1297.

Ortega-Argiles, R., Piva, M., Potters, L. and Vivarelli, M. (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.

Raymond, W., Mohnen, P., Palm, F. and van der Loeff, S. S. (2010). Persistence of Innovation in Dutch Manufacturing: Is It Spurious? *Review of Economics and Statistics*, **92** (3), 495–504.

Romer, P. M. (1990). Endogenous Technological Change. *Journal of Political Economy*, **98** (5), 71–102.

Rosenbaum, P. R. and Rubin, D. B. (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, Leuven.

Slivko, O. and Theilen, B. (2014). Innovation or imitation? The effect of spillovers and competitive pressure on firms R&D strategy choice. *Journal of Economics*, **112** (3), 253–282.

Terleckyj, N. (1974). *Effects of R&D on the productivity growth of industries: an exploratory study*. Washington: National Planning Association.

Tsai, K.-H. and Wang, J.-C. (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 Acha, V. (2005). Innovation in low-tech industries. In D. C. M. J. Fagerberg and R. R. Nelson (eds.), *Oxford Handbook of Innovation*, New York: Oxford University Press, pp. 407–432.

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), median value	3.60	0.77	0.25	0.10
Number of firms (Obs.)	462	458	79	130

Notes:
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	0.691***	0.707***	0.693***	0.696***	0.651***
	(0.018)	(0.025)	(0.027)	(0.026)	(0.022)
Labour	0.237***	0.210***	0.227***	0.232***	0.261***
	(0.010)	(0.017)	(0.016)	(0.012)	(0.013)
Capital	0.120***	0.105***	0.102***	0.113***	0.153***
	(0.018)	(0.023)	(0.026)	(0.027)	(0.024)
Trend	-0.016	-0.015	-0.015	-0.017	-0.019
	(0.011)	(0.020)	(0.012)	(0.011)	(0.013)
Industry dummies	+	+	+	+	+
Regional dummies	+	+	+	+	+
Obs.	1129	462	458	79	130
R^2	0.720	0.649	0.638	0.651	0.667

Notes:
Robust standard errors are reported in parentheses. The regressions contain regional and industry dummies. '***', '**' and '*' denote significance at the 1%, 5% and 10% level, respectively.

Table 3: Regression results

	GPS: first-step regression, Equation (7)[a]			
	Coef.	S.E.	T-stat.	P-val.
Intercept	15.298***[b]	4.588	3.335	0.001
$\log(\text{Number of employees})^{2006}$	-1.235***	0.293	-4.222	0.000
$\left[\log(\text{Number of employees})^{2006}\right]^2$	0.041***	0.013	3.097	0.002
TFP^{2005}	-30.895**	13.856	-2.230	0.026
$\left[\text{TFP}^{2005}\right]^2$	24.590**	9.859	2.494	0.013
$\log(\text{Net Sales/Number of employees})^{2005}$	-0.218	0.188	-1.160	0.246
$\left[\log(\text{Net Sales/Number of employees})^{2005}\right]^2$	-0.026	0.046	-0.569	0.570
$\log(\text{Operating profit})^{2006}_{POS}$	-0.115***	0.040	-2.903	0.004
$\log(\text{Operating profit})^{2006}_{NEG}$	0.062	0.042	1.456	0.146
$\log(\text{Operating profit})^{2005}_{POS}$	0.050	0.043	1.170	0.242
$\log(\text{Operating profit})^{2005}_{NEG}$	-0.079*	0.045	-1.737	0.083
$\log(\text{Market capitalisation})^{2006}$	0.233	0.336	0.693	0.489
$\left[\log(\text{Market capitalisation})^{2006}\right]^2$	-0.013	0.022	-0.578	0.563
$\log(\text{Market capitalisation})^{2005}$	-0.210	0.343	-0.613	0.540
$\left[\log(\text{Market capitalisation})^{2005}\right]^2$	0.016	0.022	0.720	0.472
Industry dummies		+		
Regional dummies		+		
R^2	0.721			
Obs.	1129			

Notes:

[a] The dependent variable r_i^{2006} 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. The regression contains regional and industry dummies.

[b] '***', '**' and '*' denote significance at the 1%, 5% and 10% level, respectively.

Table 4: Initial balancing properties of covariates

	Group 1	Group 2	Group 3
$\log(\text{Number of employees})^{2006}$	**15.278**	**6.307**	**-21.260**
$\left[\log(\text{Number of employees})^{2006}\right]^2$	**14.264**	**5.445**	**-21.249**
TFP^{2005}	**16.387**	**5.315**	**-18.321**
$\left[\text{TFP}^{2005}\right]^2$	**16.281**	**4.919**	**-18.881**
$\log(\text{Net Sales/Number of employees})^{2005}$	**5.331**	**-3.236**	**-2.471**
$\left[\log(\text{Net Sales/Number of employees})^{2005}\right]^2$	**-4.619**	0.961	**2.859**
$\log(\text{Operating profit})^{2006}_{POS}$	**15.326**	**3.402**	**-17.387**
$\log(\text{Operating profit})^{2006}_{NEG}$	**5.791**	**3.068**	**-7.068**
$\log(\text{Operating profit})^{2005}_{POS}$	**12.800**	**3.645**	**-15.607**
$\log(\text{Operating profit})^{2005}_{NEG}$	**4.667**	**3.618**	**-6.826**
$\log(\text{Market capitalisation})^{2006}$	**11.551**	**1.981**	**-12.844**
$\left[\log(\text{Market capitalisation})^{2006}\right]^2$	**10.786**	1.310	**-12.362**
$\log(\text{Market capitalisation})^{2005}$	**10.741**	**2.167**	**-12.241**
$\left[\log(\text{Market capitalisation})^{2005}\right]^2$	**10.037**	1.455	**-11.614**
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: GPS-adjusted balancing properties of covariates

	Group 1	Group 2	Group 3
$\log(\text{Number of employees})^{2006}$	0.615	**2.699**	**-2.115**
$\left[\log(\text{Number of employees})^{2006}\right]^2$	0.594	**2.617**	**-2.074**
TFP^{2005}	0.063	1.879	-1.278
$\left[\text{TFP}^{2005}\right]^2$	0.063	1.848	-1.269
$\log(\text{Net Sales/Number of employees})^{2005}$	-1.882	-1.593	1.785
$\left[\log(\text{Net Sales/Number of employees})^{2005}\right]^2$	**2.026**	1.053	-1.571
$\log(\text{Operating profit})^{2006}_{POS}$	0.540	1.283	-1.315
$\log(\text{Operating profit})^{2006}_{NEG}$	-0.072	0.754	-0.770
$\log(\text{Operating profit})^{2005}_{POS}$	-0.902	1.829	-0.715
$\log(\text{Operating profit})^{2005}_{NEG}$	-0.913	1.293	-0.271
$\log(\text{Market capitalisation})^{2006}$	-0.604	0.704	-0.118
$\left[\log(\text{Market capitalisation})^{2006}\right]^2$	-0.755	0.669	0.036
$\log(\text{Market capitalisation})^{2005}$	-0.970	0.897	0.092
$\left[\log(\text{Market capitalisation})^{2005}\right]^2$	-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 6: Regression results

	GPS: second-step regression, Equation (8)[a]			
	Coef.	S.E.	T-stat.	P-val.
Intercept	0.8192***[b]	0.0045	183.740	0.000
$\ln r_i^{2006}$	-0.0176***	0.0021	-8.384	0.000
$\left[\ln r_i^{2006}\right]^2$	0.0003	0.0008	0.388	0.698
$\ln s_i^{2006}$	0.0114**	0.0046	2.456	0.014
$\left[\ln s_i^{2006}\right]^2$	0.0015**	0.0007	2.019	0.044
$\ln r_i^{2006} * \ln s_i^{2006}$	-0.0027***	0.0010	-2.728	0.007
R^2	0.127			
Obs.	858			

Notes:
[a] The dependent variable is the estimated level of TFP in 2007, ω_{it}.
[b] '***', '**' and '*' denote significance at the 1%, 5% and 10% level, respectively.

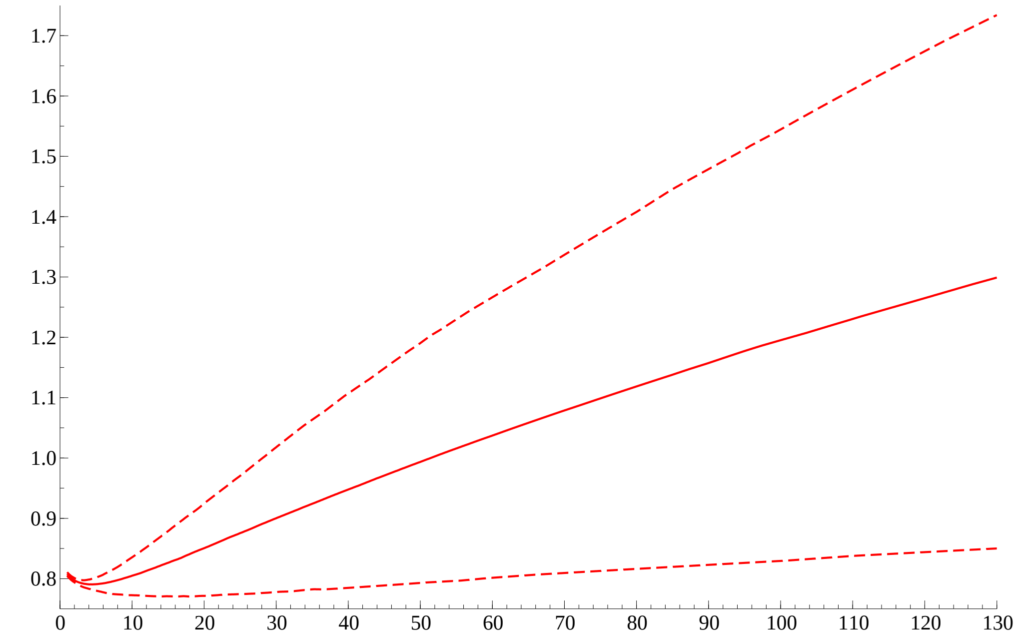

Figure 1: Dose-response function: **All companies**: Average expected response of TFP (2007) [Y-axis] to R&D intensity in 2006 [X-axis], GPS-adjusted. Dashed lines: bootstrapped 90 % confidence interval based on 1000 replications.

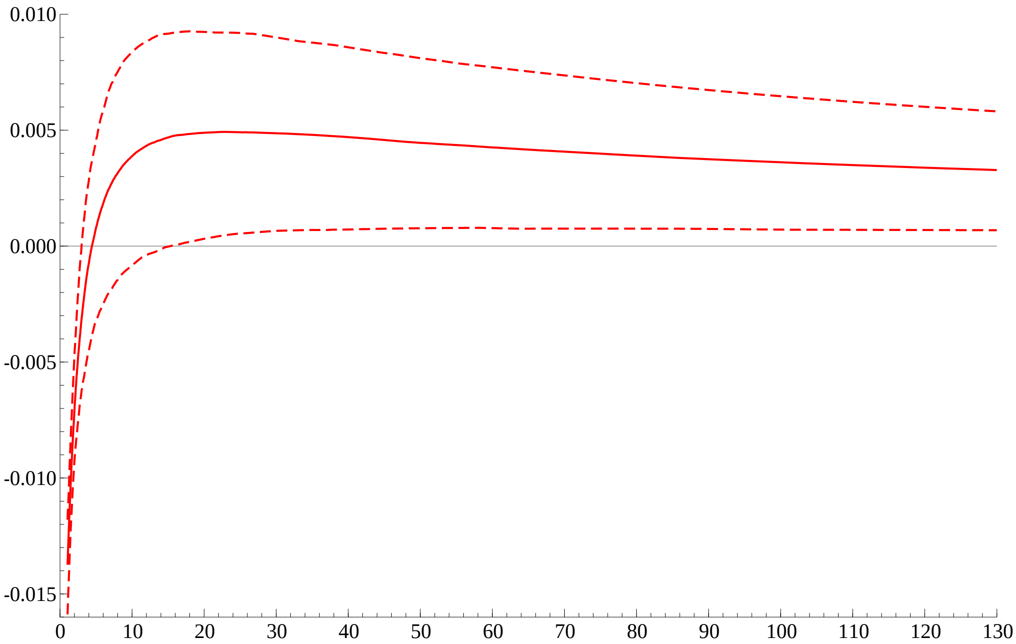

Figure 2: Treatment-effect function: **All companies**: Derivative of the average expected response of TFP (2007) [Y-axis] to R&D intensity in 2006 [X-axis], GPS-adjusted. Dashed lines: bootstrapped 90 % confidence interval based on 1000 replications.

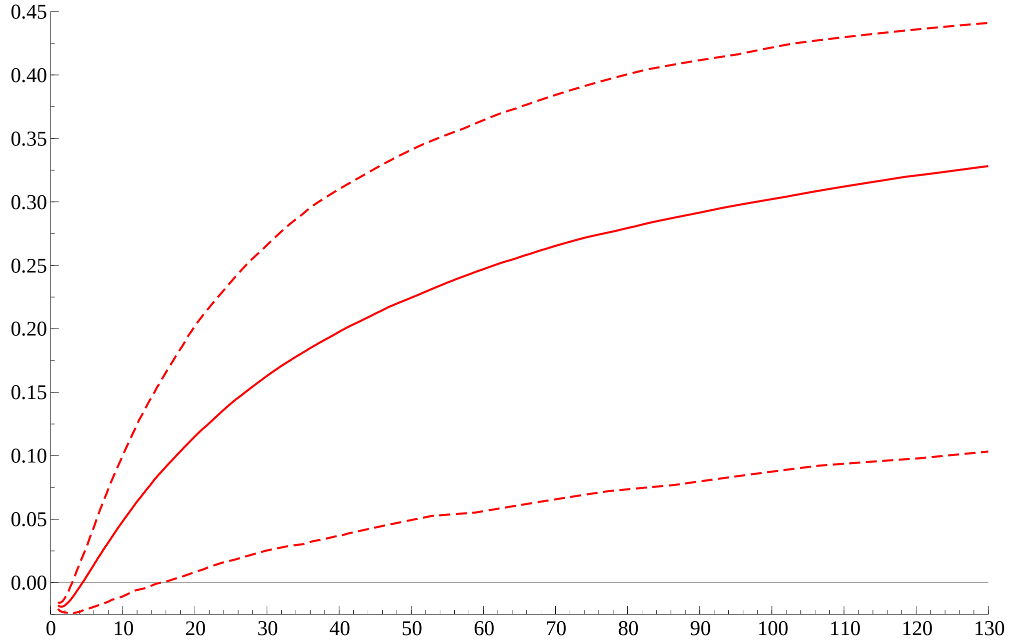

Figure 3: **All companies**: Elasticity of the average expected response of TFP (2007) [Y-axis] to R&D intensity in 2006 [X-axis], GPS-adjusted. Dashed lines: bootstrapped 90 % confidence interval based on 1000 replications.

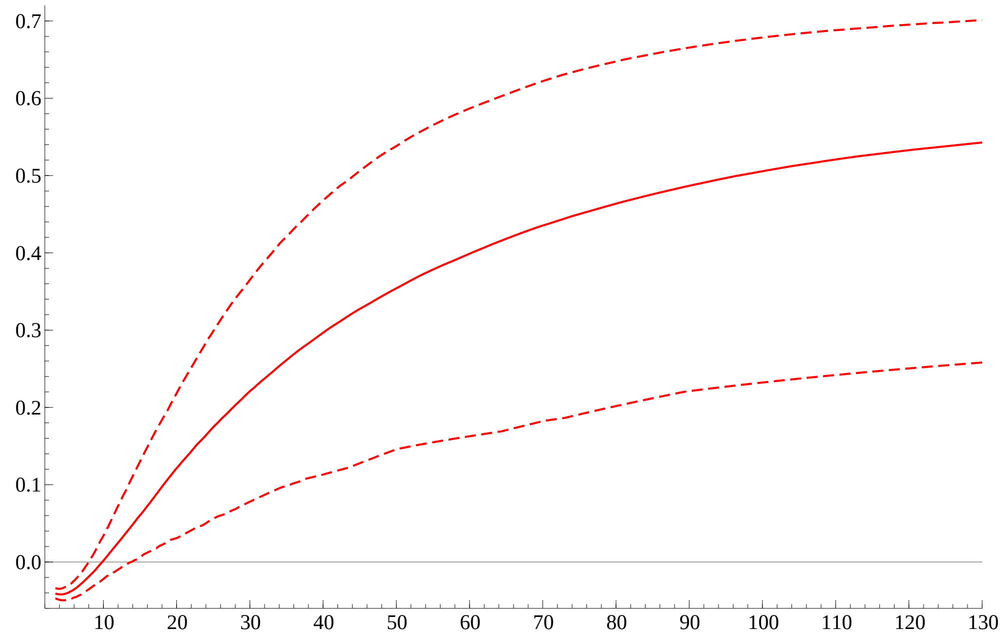

Figure 4: **High-tech companies**: Elasticity of the average expected response of TFP (2007) [Y-axis] to R&D intensity in 2006 [X-axis], GPS-adjusted. Dashed lines: bootstrapped 90 % confidence interval based on 1000 replications.

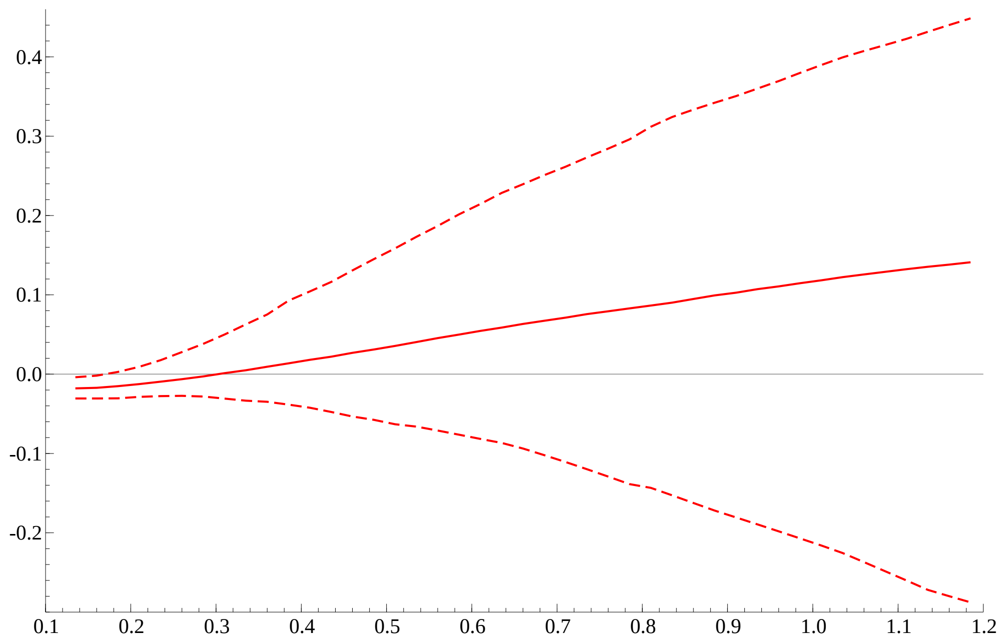

Figure 5: **Low-tech companies**: Elasticity of the average expected response of TFP (2007) [Y-axis] to R&D intensity in 2006 [X-axis], GPS-adjusted. Dashed lines: bootstrapped 90 % confidence interval based on 1000 replications.