Structural Change and Economic Dynamics 24 (2013) 45–75

Contents lists available at SciVerse ScienceDirect

Structural Change and Economic Dynamics

journal homepage: www.elsevier.com/locate/sced

Nonlinearities in productivity growth: A semi-parametric panel analysis

Théophile T. Azomahou[a,*], Bity Diene[b], Mbaye Diene[c]

[a] United Nations University (UNU-MERIT) and Maastricht University, Keizer Karelplein 19, 6211 TC Maastricht, The Netherlands
[b] University of Auvergne and CERDI, 65 boulevard François Mitterrand, Clermont-Ferrand Cedex 1, France
[c] University Cheikh-Anta-Diop and CRES, Rue 10 Prolongée, Cité Iba Ndiaye Diadji n° 1-2, Dakar, Senegal

ARTICLE INFO

Article history:
Received December 2009
Received in revised form January 2012
Accepted May 2012
Available online 1 June 2012

JEL classification:
C14
I23
O3
O4

Keywords:
R&D
TFP
Panel data
Nonparametric estimation
Reduced vs. structural form

ABSTRACT

We use country panel data spanning over 1998–2008 for both developed and developing countries to study the productivity growth when countries are close to the technology frontier. Relying on a semi-parametric generalized additive model, we estimate both reduced and structural forms for total factor productivity growth. We consider three measurements of frontier: the economy with the highest level of productivity growth, the world productivity growth and the productivity growth of the USA. We obtain a U-shape relation between productivity growth and the proximity to the world productivity growth. The relation between productivity growth and human capital displays an inverted U-shape form (res. U-shape) when the proximity to the highest productivity growth is used (res. the proximity to productivity growth of the USA). Total staff in R&D has an inverted W-shape effect on productivity growth. The share of R&D expenditure funded by government and from abroad impact positively the growth of productivity. However, the increase in government spending on R&D has a greater impact on productivity growth when the former is weak, and a smaller impact when R&D spending is already high. International trade has a positive effect on productivity growth. Specification tests show that our semi-parametric models provide a better approximation of the data compared to the parametric analogue, revealing a high degree of nonlinearity governing productivity growth.

1. Introduction

In endogenous growth theory (see e.g., Romer, 1990), human capital accumulation is one of the most important factors of growth. Assuming constant returns to technology, Mankiw et al. (1992) show that years of schooling increase the productivity. Nelson and Phelps (1966) have already asserted that the stock of human capital determines the ability to innovate or to catch up with developed countries. As pointed out by Hanushek and Kim (1995) and Hanushek and Kimko (2000), a high human capital accumulation and more fundamental research (university research) generates a higher economic growth. Therefore, expenditure devoted to higher education becomes a key factor for growth and development. This calls for the implementation of a government policy that would readily mop up the flow of financial means into the economic system so that quality higher education can be ascertained.

Most of the empirical studies have shown that human capital (usually measured empirically by years of education) and R&D have a significant positive effect on economic growth. Bassanini and Scarpetta (2001) have used a panel of 21 OECD countries for the period 1971–1998 to study the effect of human capital, R&D, demographic growth and investment on the real GDP per capita. Using "pooled

* Corresponding author. Tel.: +31 433 884 440; fax: +31 433 884 499.
E-mail addresses: azomahou@merit.unu.edu (T.T. Azomahou), bity.diene@udamail.fr (B. Diene), mdiene@cres-sn.org (M. Diene).

0954-349X/$ – see front matter
http://dx.doi.org/10.1016/j.strueco.2012.05.001

mean group estimator", they find that, whereas years of schooling, total R&D expenditure and industry R&D have a significant positive effect on GDP per capita growth rate, public R&D has a negative effect. The latter might be explained by the fact that the part of public R&D expenditure devoted to defence area is higher than those devoted to civilian area. Relying on 16 OECD countries over the period 1980–1998, Guellec and Pottelsberghe (2001) investigate the long term relationship between various types of R&D and multifactor productivity growth (hereafter MFP), within an error correction model and instrumental variables. They find that business R&D and foreign R&D have significant positive effect and only the defence-related part of public funding has a negative and significant effect on MFP. One main result is that the elasticity of public R&D is positively affected by the public research share done by universities.

Moreover, the endogenous growth theory suggested that, the difference in productivity growth rate between countries can be explained by differences in R&D and educational policy systems. In a recent theoretical and empirical study Aghion and Cohen (2004) focus on the increasing importance of higher education when the technology in a country is near to the technology frontier.[1] The authors emphasized two mechanisms through which education influences growth and development: the first one is that educated persons are more productive since they have a high human capital, and the second one concerns technological progress; a higher education level enables to adapt or to develop new technologies in a easier way. Aghion and Cohen (2004) state that countries which are near the technology frontier, have a kind of productivity gain achieved differently from those who are far away from the technology frontier. The authors assert that for countries located far away from the technology frontier, the productivity gain is obtained by the channel of adaptation and imitation of existing technologies. However for those who are near the frontier, innovation becomes the driving force of growth. Also, they develop a theoretical model where they find a critical threshold, below which to invest in primary and secondary education is more efficient and above which the country should invest in higher education.

Using data on 20 OECD countries, Aghion and Cohen (2004) studied the effect of years of schooling and countries labor productivity backwardness relative to USA on total productivity growth. They find that taking separately primary, secondary and higher education, the more a country is near the technology frontier, the more an additional year of schooling in primary or secondary level makes the marginal return to decrease. Their estimated threshold is 24% under the frontier and an additional year in higher education entails 8% effect on total factors productivity.

While this literature has emphasized the crucial role of productivity backwardness, different measurements of the same have been used in different contributions. Aghion and Cohen (2004) used labor productivity backwardness relative to the USA. Griffith et al. (2004) used a panel of industries from 12 OECD countries. The authors considered the economy with the highest level of TFP to study the role of technology transfer, absorptive capacity, human capital and R&D on productivity growth. Vandenbussche et al. (2006) firstly developed a theoretical model to answer the puzzle raised by Krueger and Lindahl (2001).[2] Then, the authors used 19 OECD countries to show empirical evidence of TFP growth by using the world productivity as frontier and also several measures of human capital among which the fraction of population with higher education. These studies and the others are based on parametric specifications of reduced-form productivity equations.

By emphasizing the crucial role of nonlinearity in the productivity process, our study contributes to this literature in several aspects. From a methodological point of view, an innovative aspect concerns our specification. Indeed, previous studies used parametric specifications. Alternative stories of nonlinearities have been investigated. For example, Griffith et al. (2004) include higher-order terms of R&D intensity in their regression. The authors find a negative effect which suggesting diminishing returns to R&D. However, this effect was insignificant. While parametric regressions enable to detect such nonlinearities (high-order terms such as squared, cubic, etc.), they still bear an inferential limitation: the parametric specification is always assumed as the true model. Here, we assumed a *semi-parametric generalized additive model*. To the best of our knowledge, our study is the first one which adopts such specification to study TFP growth. This framework places weak restrictions on the functional form to be estimated and then allows for nonlinearities of unknown form in the relationship between the TFP growth and the control variables.

The contribution of human capital, R&D and some other key factors like trade and FDI to TFP growth is now well established. In this context, there is a growing empirical (Griffith et al., 2003, 2004; Hu et al., 2005; Kneller, 2005; Kneller and Stevens, 2006; Madsen et al., 2010) and theoretical (Vandenbussche et al., 2006; Aghion et al., 2005, 2009) literature which depicts a clear and meaningful relation both at country and firm level. The empirical scrutiny has so far been restricted to the parametric domain. In parametric framework the functional relation between TFP growth and it determinants is assumed to be linear, relegating a complex feedback mechanism underlying the process, which can make the relation highly nonlinear and hence more complicated than it appears to be. While nonlinearity (due to the intricate way human capital and R&D act upon TFP growth and vice versa) can have substantial implications for long-term economic growth and policy, the empirical literature thus far have paid little attention to the importance of nonlinear relationship lacking to explain some puzzling results (Krueger and Lindahl, 2001).

In an attempt to delineate a clear relational structure between the two, we revisit the problem in a nonparametric setting, where the flexibility of the framework allows examining the true functional form of the human

[1] The frontier is measured by the technology of the USA.

[2] The puzzling finding of Krueger and Lindahl (2001) is that education is statistically significantly and positively correlated with growth only for economies with the lowest level of education.

capital–R&D–TFP relation. Indeed, while parametric specification of a TFP growth model has been extensively used in the empirical literature, its pitfalls against 'letting the data speak as it is' makes it less realistic. Unless we have strong reasons to believe that a linear or non-linear functional form of certain degrees could explain the human capital–R&D–TFP linkage, it is necessary to model them without pre-specified linear assumption. A few concerns may arise in this context: Can non-linear human capital and R&D structure cause TFP growth in developing and developed economies? What are the implications of nonlinear relations for long-term economic growth? This paper tries to answer these questions via nonparametric modeling using a panel data of developed and developing countries.

The second contribution of this study is the estimation of a structural form. The rationale behind is that proximity to productivity frontier may be endogenous. As a result, it can be framed as part of a semi-parametric triangular system (Newey et al., 1999). To our knowledge, the empirical literature on R&D and productivity was almost exclusively dominated by the parametric estimation of reduced forms. Scholars have successfully correct for endogeneity and measurement errors issues through the instrumental variables methods. For example, Griffith et al. (2004) rightly observe that the effect of R&D on TFP can be overestimated in the reduced form equation as firms invest heavily in R&D during periods where TFP is growing rapidly. Also the assumption of predetermination of R&D variable lagged may be violated due to feedback effects. The way in which measures of proximity are defined raises the issues of endogeneity and simultaneity. Better yet, endogenizing the proximity to frontier has an economic interest to the extent that the frontier represents a kind of reference to which other economies want to imitate or catch up. The estimation of a structural model is richer because it allows to estimate the determinants of the frontier. At the same time, this approach is more complicated because it involves the estimation of a system of equations. In the case of nonparametric methods, the estimate is even more delicate but still tractable. Last, while previous studies have considered one measure of proximity to the frontier, we consider three measures: the economy with the highest TFP growth, the world TFP growth and the US TFP growth.

The study delivers several key results: (i) While TFP growth is a decreasing mixing slope function of proximity to the highest TFP growth, it is increasing with proximity to the world and to TFP growth of the USA. (ii) The relation between TFP growth and human capital (measured by the percentage of graduate students in higher education) displays an inverted U-shape form (res. U-shape) when the proximity to the highest TFP growth is used (res. the proximity to TFP growth of the USA). (iii) Total staff in R&D has an inverted W-shape effect on TFP growth. (iv) The share of R&D expenditure funded by government and from abroad impact positively the growth of TFP. However, there is no evidence of R&D expenditure funded by the business sector on TFP growth. (v) International trade (measured by country openness) has a positive effect on TFP growth. However, we do not find evidence that FDI impacts productivity growth. (vi) The structural form estimation shows a U-shape relation between TFP growth and the proximity to the world TFP growth. This finding seems to reconcile the relations from two other proximity measures (decreasing for the proximity to the highest TFP growth and increasing for the proximity to the TFP growth of USA). (vii) Also, an increase in government spending in R&D has a greater impact on TFP growth when the latter is low, and a smaller impact when TFP growth is already high reflecting a S-shape nonlinearity. (viii) Last, specification tests show that our semi-parametric model provide a better approximation of the data compared to the parametric analogue, revealing a high degree of nonlinearity which governs the productivity growth process.

The remaining of the study is organized as follows. The next section reviews the literature. Section 3 describes the data and some key features derived from them. Section 4 presents the econometric specifications. In Section 5, we discuss the results. Section 6 offers a summary and some concluding remarks.

2. A quick overview of the literature

The endogenous growth theory emphasized many models that aim at clarifying the role of technology and human capital in explaining productivity growth across countries (Eaton and Kortum, 1999; Howitt, 2000; Xu, 2000; Keller, 2002a,b; Griffith et al., 2003, 2004; Hu et al., 2005; Kneller, 2005; Kneller and Stevens, 2006; Madsen et al., 2010). A leading country having the higher total factor productivity (TFP) is supposed to be the technological frontier and then, the diffusion of technology in a following country depends on its distance from the technology frontier: This is the concept of backwardness. The reduction of the distance depends on human capital, among others. But different natures of human capital lead to different ways for reducing backwardness. Indeed, Vandenbussche et al. (2006) and Aghion et al. (2005, 2009) show that though it is very crucial for growth, human capital does not affect innovation or imitation in the same way. As our paper is fully related to the strands of the literature dealing with backwardness and growth, we start first by discussing different types of backwardness. Then, we investigate the issue of its reduction by imitation, innovation and by human capital enhancement.

2.1. Higher growth and backwardness

Does backwardness imply higher growth? For some authors, the response is negative because of two facts: (i) the acquisition of the new technological knowledge developed elsewhere is expensive as countries have to master those technologies (Howitt, 2005), (ii) human resources, such as skilled workers need good training so as to adapt the new technologies (Hobday, 2003). But according to others, imitation and innovation can lead to a reduction of backwardness. It depends on the interaction between distance to the frontier, R&D intensity and educational attainment (Griffith et al., 2003, 2004; Kneller, 2005; Kneller and Stevens, 2006). There are many evidences in the ways to reduce the distance.

What about the countries which are far from the technological frontier? The distance can be reduced by R&D, which have been shown as having an innovative character in

OECD countries (Howitt, 2000; Griffith et al., 2003; Kneller, 2005). Research intensity reduces the distance to the frontier as pointed out by Griffith et al. (2004), Kneller (2005), Kneller and Stevens (2006) and Madsen (2007, 2008a). Zachariadis (2003, 2004) and Madsen (2008b) find evidence that R&D intensity has positive effects on TFP growth, and subsequently on the distance to the frontier, for OECD countries.

Countries far from the frontier can gain more benefits from investment in knowledge than those who are near to the frontier. The reason is that as they have adapted the foreign technology, they have additional economies of scale. Coe et al. (1997) have shown that TFP in developing countries is positively and significantly related to international R&D spillovers from developed countries. Savvides and Zachariadis (2005) also show that the growth of TFP for countries which are relatively near to the frontier may be significantly boosted by technological diffusion from the frontier countries. However countries which are far from the frontier have to implement alternative policies in order to reduce backwardness.

How is it possible to benefit from backwardness? Abromovitz (1986) shows that if a country has an absorptive capacity, it will be able to benefit from technological backwardness by exploiting the technology developed in the countries near to the frontier. The problem remains that all countries do not have the same ability to adopt new technologies. So more expenditures in R&D and more investment in higher education may increase their capacity to absorb foreign technology.

In the same line, Hobday (2003) outlines that the main element explaining the success of the New Industrialized Countries is their large investments in education, training and R&D, because they aim to adapt the technologies developed in the other advanced countries. But is there any link of interdependence between the countries near to the frontier and those far from it? The way to avoid that eventual interdependence relies on the enhancement of adequate local R&D investments so that knowledge developed in the frontier countries can be appropriately used in local conditions (Verspagen, 1991; Fagerberg, 1994; Aghion and Howitt, 2005; Howitt, 2005).

2.2. Productivity backwardness, human capital and growth

Productivity is crucial as it explains income differences between countries rather than physical or human capital accumulation do, as showed by: Klenow and Rodriguez-Clare (1997), Hall and Jones (1999), and Easterly and Levine (2001). Feyrer (2003) and Vandenbussche et al. (2006) have developed a model where the growth of productivity increases with the distance from the frontier and economies reduce their backwardness. Many explanations have been provided about the causes of productivity differences between countries. The main one is related to barriers (Parente and Prescott, 2004; Klenow and Rodriguez-Clare, 2004). Empirically, Cole et al. (2004) show that domestic and international competitive barriers have a great impact on productivity in Latin America. The domestic barriers are for example entry barriers and inefficient financial systems and subsidization of public corporate, low level of human capital level, etc. The degree of openness of economies is the main international barrier. Coe et al. (1997, 2009) show that countries can upgrade their productivity growth by importing larger varieties of intermediate goods and capital equipment, so as to import the foreign knowledge.

Griffith et al. (2000) recognize that R&D activity has the roles of stimulating innovation, and facilitating the imitation. But the statistical significance of the second role has not been rigorously investigated before them. Relying on a panel of industries in OECD countries, the authors show that R&D has a kind of “second face”: the industries that are far from the productivity frontier move rapidly to it when they provide more investments in R&D. They assume that innovation and technology transfer are sources of productivity growth for countries laying behind the technological frontier. Practically, the country with the highest level of total factor productivity (TFP) is defined as the frontier. The innovation is measured by the rate of TFP growth and they test whether it depends on a country’s distance from the frontier. They found that the potential for R&D to increase TFP growth through technology transfer is positively related to the distance from the frontier. Our paper is complementing Griffith et al. (2004) as we also examine to what extent R&D explains the growth rate of productivity.

Vandenbussche et al. (2006) have examined the contribution of human capital to technological improvements through innovation and imitation. The authors start with the puzzle of Krueger and Lindahl (2001) who show that economic growth is significantly and positively related to education only for countries which have low education levels. This means that in more developed countries the opposite relation is observed. Two reasons can be putted forward: Primarily for the latter countries, efforts to catch up the technological frontier are weak because they are already near to the frontier. Also their high level of education leads them to the adoption of new technologies (Nelson and Phelps, 1966). Secondly, instead of imitating only, those countries are able to make pure innovations. Nevertheless, different types of human capital are necessary for imitation and innovation. For example, unskilled labor is sufficient for imitation. The authors show empirically that the effects of each type of human capital on economic growth depend on the country’s distance to the technological frontier. They solve the puzzle by focusing on the distance to the technological frontier and on the composition of human capital. By using a panel data set covering 19 OECD countries between 1960 and 2000 they explain why previous studies failed to show positive relation between initial schooling level and economic growth in rich countries. Our paper contributes to the same strand of literature than Vandenbussche et al. (2006) though our methodology is different.

While some authors do not find direct robust relationships between educational attainment and growth (Benhabib and Spiegel, 1994), some others support the hypothesis of Nelson and Phelps (1966) which states that higher educated labor force increases the ability of countries behind the frontier, to absorb technology (Kneller and Stevens, 2006). But higher education is necessary for innovation in new technology. As a result, if a country is far from the frontier, it has to develop the skill of the labor suppliers by higher education, so as to catch up to the frontier. Nevertheless, the attainment of higher education for most of the labor suppliers is also a way to facilitate the assimilation of the foreign technology. This is shown by Nelson and Phelps (1966), Abromovitz (1986), Cohen and Levinthal (1989) and Engelbrecht (1997).

Educational attainment has two facets with respect to the production of knowledge – a direct effect and an indirect effect through enhancing the ability to absorb new technology (Kneller and Stevens, 2006). It is also commonly recognized that other levels of education, such as, primary and secondary, are sufficient for the imitation of the technology of countries near to the frontier.

3. Data and variables

Our data set consists of the most recent country panel spanning over 1998–2008 for both develop and developing countries. We combine three datasets: The Conference Board Total Economy Database (2010), the World Development Indicators (WDI, 2010) and the United Nations Educational, Scientific and Cultural Organisation database (UNESCO Statistical Yearbook, 2010).[3] The two dataset provides yearly data on economic, education, literacy, science and technology, culture and communication from 1960 to 2010. We restrict the time span of our sample to 1998–2008 by tacking the intersection of the two datasets. Based on the international classification, the proportion of countries covered by regions is: Europe & Central Asia (27.14%), East Asia & Pacific (16.66%), Latin America & Caribbean (18%), Middle East & North Africa (10%), North America (1.43%), South Asia (3.8%), Sub-Saharan Africa (22.8%). Also, 14.3% of countries are OECD members.

The variable of interest is the Total Factor Productivity (TFP) growth which is average output produced by a combination of multiple inputs, including labor and capital input, and with adjustments for changes in the quality of labor and changes in the composition of capital assets. To obtain the TFP measures, a growth accounting framework is used to compute the contribution of these inputs to aggregate Gross Domestic Product (GDP) growth. Based on the TFP and as stated earlier, three measurements of frontier have been computed: the distance to the economy with the highest TFP growth, the distance to world productivity growth and the distance to the TFP growth of the USA. The other variables considered in the analysis can be gathered into two categories: human capital indicators (percentage of graduate students in higher education, number of full-time equivalent workers engaged in R&D activities), R&D expenditures by origin of funding (share of R&D expenditure funded by government, share of R&D expenditure funded by firm and share of R&D expenditure funded from abroad).

For each of these human capital and R&D expenditure variables, we also use an interaction term with each measurements of frontier. As well documented by Griffith et al. (2004), the interaction between frontier and R&D variables is an indicator of the absorptive capacity, while the frontier indicates the technology transfer. We also include time dummies to reflect the effect macroeconomic or structural shock. Last, FDI and openness (trade) are also used as additional controls. The role of FDI and trade has been stressed in both cross sectional growth literature and studies on international R&D technology spillovers (Griffith et al., 2004). The theory puts forward several mechanisms through which trade impacts productivity growth (technology spillovers, firm size, increasing product market competition). Also, there are several ways to introduce international trade into an empirical model. Griffith et al. (2004) used the OECD bilateral trade database at country-industry level which provides information on import from trading partners. Then the authors construct measures of import penetration for each industry in each country. Here, as we are concerned with country level data and that our sample is not limited to OECD countries, we proceed with a more simple and intuitive approach by approximating trade with the openness. This is obtained by scaling the sum of import and exports of goods and services by GDP. All the variable used in this study and their definition and related data source are provided in Table 11.

In studying productivity growth, we follow the bulk of the literature by not controlling for all possible determinants. Of course, it is not our intention to deny the role of other factors. However, a number of points can be made in support of our choice. The first and the obvious one concerns data limitations. In this respect, it is important to note that using panel methods that sweep country effects away lets us control implicitly for any time invariant determinant. The second obvious point concerns comparability with existing studies. A more technical point concerns the curse of dimensionality in nonparametric studies: adding discrete regressors to a nonparametric specification does not alter the speed of convergence of the estimator, but adding continuous regressors does. More importantly, we are not concerned here with obtaining best predictions of productivity growth next year for example, but with the *shape* of the relations. In this respect, determinants of productivity growth which are not correlated with regressors become irrelevant. Moreover the impact of omitted determinants which are correlated with included regressors will be captured in the effect of those regressors. Depending on the question asked, this can be seen as a drawback or as an advantage. It is a drawback if we purport to determine the ceteris paribus impact of regressors – but what list of regressors would guarantee this? It is an advantage if we are interested in the global effects, including indirect effects linked with omitted variables. This is indeed the stance we take here. While the results of our study will not enable us to make precise policy prescriptions, we will be in a position to intervene convincingly in the long debate on the strategy policy which allows to boost growth.

Table 1 summaries descriptive statistics. The mean value of TFP growth is 0.839 with a big sample variation (3.793 as standard deviation), meaning many heterogeneity. The three measures of proximity give very different mean levels: 11.31 for the proximity to the highest TFP

[3] We are grateful to an anonymous referee to point us to the UNESCO database.

Table 1
Summary statistics.

Variable	Mean	Std. Dev.	Min.	Max.	# obs.
TFP growth	0.839	3.793	−27.546	24.502	1922
Proximity to the highest TFP growth	11.312	6.065	−12.149	46.441	1817
Proximity to the world TFP growth	−6.81e−10	3.704	−27.029	23.837	1817
Proximity to the TFP growth of USA	0.066	4.444	−21.146	46.441	1799
Higher educ.	54.012	11.487	12.232	80	746
Total staff R&D[a]	10.886	24.112	0.0005	196.535	721
R&D by gov. (share)	49.793	21.399	2.443	100	805
R&D by firm (share of business sector)	39.604	19.814	0.11	90.684	726
R&D from abroad (share)	9.747	11.654	0.06	73.675	692
FDI	5.702	34.627	−54.358	1095.278	3184
Openness	1.52e+07	4.88e+07	22,124.83	9.79e+08	5229
OECD countries	0.144	0.351			

[a] Scaled by 10,000.

growth, almost zero for the proximity to the world TFP growth (the latter is defined as the world productivity frontier calculated as the sample mean), and 0.066 for the proximity to the US TFP growth. Observe that the maximum of the highest TFP growth is the same as the US. In order to have a picture of the whole distribution of TFP growth as well as proximity measures, we compute the density estimates. Fig. 1 displays the corresponding graphs.

The densities show mainly unimodal distributions. Moreover, it is clear that the three distributions of proximity are not telling the same history. Indeed, starting from the proximity to the US TFP growth, we observe a shift to the right with respect to the two others. Also, the

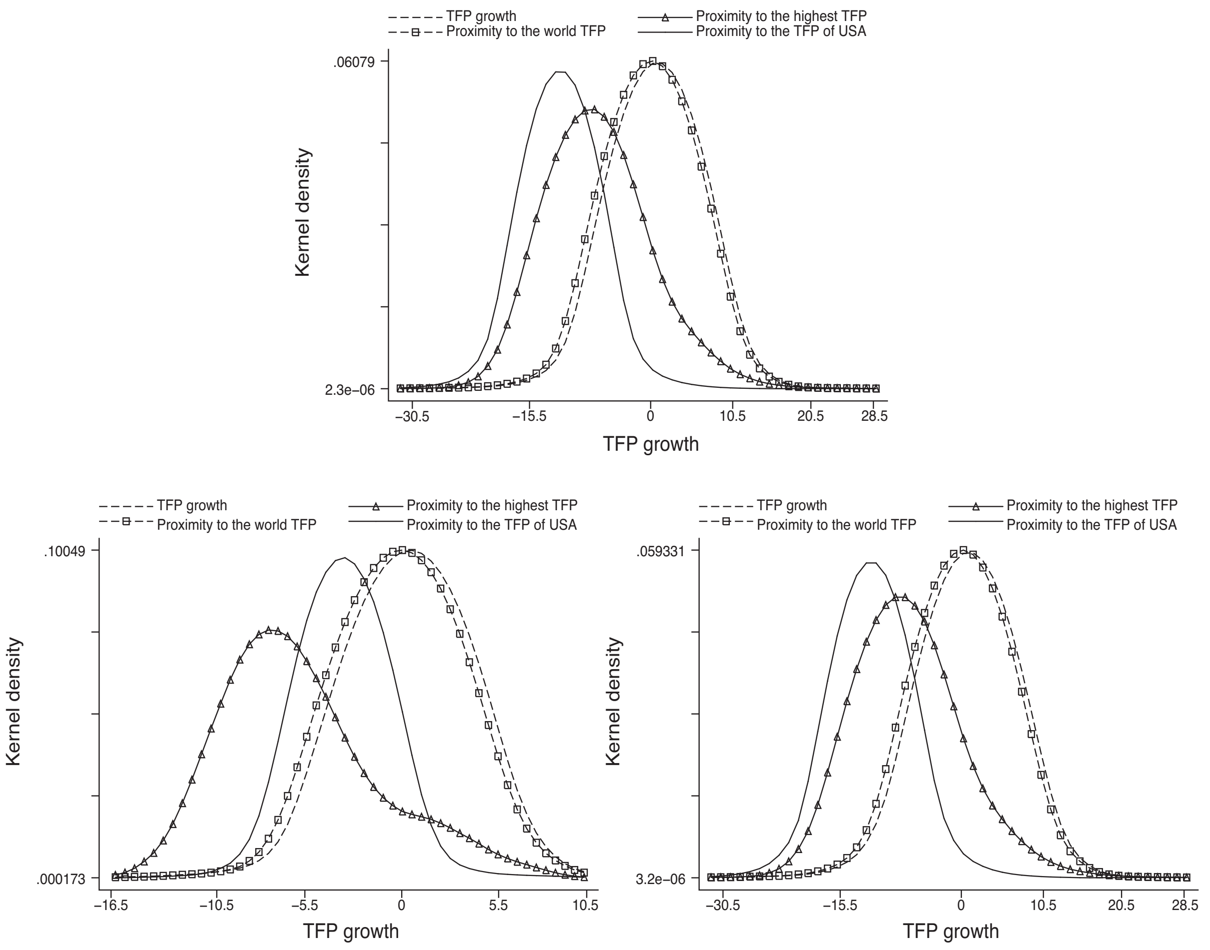

Fig. 1. Distribution of TFP growth and proximity measures for the whole sample, Kernel density estimate. [Top]: Full sample. [Bottom-left]: OECD countries. [Bottom-right]: non-OECD countries.

distribution of the proximity to the world is very close to TFP growth. This was already figured in Table 1 where the standard deviations are also close (3.7). This picture is quite different when we consider OECD countries only as shown in Fig. 1. Indeed, while the distributions of the proximity to the world and TFP growth are still very close, the proximity to the highest TFP growth shifts to the left compared to the proximity to the US TFP growth. The distributions for non-OECD countries show similar patterns to those of the full sample. We also observe a very high variability in R&D expenditure according to the source. R&D expenditure funded by government has the highest mean (49.79%) followed by R&D expenditure funded by the business sector (39.6%) and R&D expenditure funded from above (9.7%). In the case of OECD countries, these figures change to the following order, R&D expenditure funded by business sector (51.75%), R&D expenditure funded by government (38%), and from above (7.09%). For non-OECD countries, we have R&D expenditure funded by government (57.28%), R&D expenditure funded by the business sector (30.39%) and from above (11.76%).[4]

4. Conceptual framework and estimation strategies

In this section, we firstly set up a conceptual nonparametric framework for productivity growth. This framework is designed to emphasize the role of nonlinearities in that respect. Nonlinearity here means that very weak restrictions are putted in functional forms that link productivity to its determinant, viz. human capital, R&D, etc. Secondly, we elaborate on the estimation strategies that match with the model.

4.1. Conceptual framework

Let us denote countries by $i = 1, \ldots, N$, and time by $t = 1, \ldots, T$. The production Y of each country at time t needs to be inputed by labor L and physical capital K according to the neoclassical technology,

$$Y_{it} = A_{it} F_{it}(K_{it}, L_{it}) \tag{1}$$

where A_{it} is an index of technological progress, or *Total Factor Productivity* (TFP), F is assumed to be homogenous of degree 1 and exhibits diminishing returns with respect to production factors. TFP is allowed to vary across countries and time. We denote the frontier as $i = \mathcal{F}$ where $\mathcal{F} = \{\mathcal{F}_{\max}, \mathcal{F}_{\text{world}}, \mathcal{F}_{\text{USA}}\}$ meaning that we adopt three definitions of frontier: the country with the highest level of TFP growth, the world TFP (mean of the sample) and the TFP of USA. The remaining of this conceptual exposition, we will use $\mathcal{F}$ as long as there is no confusion. The empirical literature on R&D and productivity growth states that the rate of TFP growth is a function of R&D knowledge stock and other control variable, viz, human capital, trade, etc. (see e.g., Griffith et al., 2004; Vandenbussche et al., 2006). In a nonparametric setting, the TFP growth equation can be written as

$$\mathcal{A}_{it} = f(\ln R\&D_{it}, \mathbf{x}_{it-1}) + \mathbf{z}'_{it}\gamma + u_{it} \tag{2}$$

where $\mathcal{A}_{it} = \ln A_{it} - \ln A_{it-1}$, and $\mathbf{x}$ denote various continuous controls such as human capital (higher education), FDI, trade, and also two important determinants: the proximity to frontier or backwardness $\ln (A_i/A_{\mathcal{F}})_{t-1}$ and absorptive capacities (described by interactions terms between a measure of proximity and R&D or human capital variables); $\mathbf{z}$ components may be thought of as dummy variables. f is an unknown function to be estimated along with the vector of parameters γ and u_{it} is an error term, the distribution of which is specified later on. Eq. (2) has the advantage of being robust to parametric misspecification as the form of f is not set up ad hoc. Examples of fully parametric analogues of relation (2) are considered by Griffith et al. (2004) and Vandenbussche et al. (2006). However, at this stage, Eq. (2) is not easily tractable empirically due the issue "curse of dimensionality" which appears in nonparametric regressions when many explanatory variables are included. As a result, this slows the speed of convergence of the estimator of f. To overcome this problem, we make the assumption of additivity in f, so that relation (2) can be expressed as

$$\mathcal{A}_{it} = f_1(\ln R\&D_{it}) + f_2(\mathbf{x}_{it-1}) + \mathbf{z}'_{it}\gamma + u_{it} \tag{3}$$

Additive models are widely used in both theoretical economics and econometrics. Deaton and Muellbauer (1980) provide examples in which a separable structure is well designed for analysis and important for interpretability. From an econometric viewpoint, this specification has the advantage of avoiding the "curse of dimensionality". It also allows us to capture nonlinearities and heterogeneity in the effect of explanatory variables on the dependent one.[5] Moreover, the statistical properties (optimal rate of convergence and asymptotic distribution) of the estimator of the resulting regression function is well known (see e.g., Stone, 1980, 1982; Ibragimov and Hasminskii, 1980).[6] Additive models also offer some simple testing procedures. For example, we are able to test for nonlinearity against linearity for each regressor. As a result, our specification also provides a way to detect in a non ad hoc way the regressors which enter parametrically in the regression function. In what follows, we present the estimation strategies of the relation (3).

4.2. Reduced form estimation

Our econometric specification consists of a semi-parametric GAM specification for panel data. From relation

---

4 It should be noticed that others sources of R&D funding are those from higher education and from non profit institutions that we do not consider here due to lack of data availability.

5 See e.g., Hastie and Tibshirani (1990) and Stone (1985, 1986) for further details on GAM.

6 Consider the estimation of a regression function $f = \mathbb{E}(Y|X = x)$ based on a random sample $(Y_i, X_i)_{i=1}^n$ from this population. Stone (1980, 1982) and Ibragimov and Hasminskii (1980) showed that the optimal rate of estimating the regression function is $n^{-\ell/(2\ell+p)}$ with ℓ an index of smoothness of f and p is the dimension of f.

(3), let us stack $\ln R\&D_{it}$ and $\mathbf{x}_{it-1}$ as $\tilde{\mathbf{x}}_{it} = [\ln R\&D_{it}, \mathbf{x}_{it-1}]$. Then, we have

$$\mathcal{A}_{it} = \sum_{j=1}^{p} f_j(\tilde{\mathbf{x}}_{it}^j) + \mathbf{z}_{it}'\gamma + \mu_i + u_{it} \tag{4}$$

where the f_j are unknown univariate functions to be estimated such that $\mathbb{E}[f_j(\mathbf{x}_{it}^j)] = 0$. The unobserved effect μ_i can be eliminated by differentiating or computing the within transformation. Lagging relation (4) by one period and subtracting yields

$$\mathcal{A}_{it} - \mathcal{A}_{i,t-1} = \sum_{j=1}^{p} f_j(\tilde{\mathbf{x}}_{it}^j) - \sum_{j=1}^{p} f_j(\tilde{\mathbf{x}}_{i,t-1}^j) + (\mathbf{z}_{it} - \mathbf{z}_{i,t-1})'\gamma + \eta_{it}, \tag{5}$$

where $\eta_{it} = u_{it} - u_{i,t-1}$. We also assume that

$$\mathbb{E}(\eta_{it}|\tilde{\mathbf{x}}_{it}^j, \tilde{\mathbf{x}}_{i,t-1}^j) = 0, \quad i = 1, \ldots, N,\ t = 2, \ldots, T \tag{6}$$

which identifies the functions

$$\mathbb{E}[\mathcal{A}_{it} - \mathcal{A}_{i,t-1}|\tilde{\mathbf{x}}_{it}^j, \tilde{\mathbf{x}}_{i,t-1}^j] = \sum_{j=1}^{p} f_j(\tilde{\mathbf{x}}_{it}^j) - \sum_{j=1}^{p} f_j(\tilde{\mathbf{x}}_{i,t-1}^j), \tag{7}$$

with the norming condition $\mathbb{E}[f_j(\tilde{\mathbf{x}}_{it}^j, \tilde{\mathbf{x}}_{i,t-1}^j)] = 0$, since otherwise there will be free constants in each of the functions. It should be noticed that a special case under which first difference hypothesis is satisfied is strict exogeneity which drives the within estimator for parametric panel models. Furthermore, we assume that the error η_{it} is such that $\mathbb{V}(\eta_{it}|\Delta\tilde{\mathbf{x}}_{it}, \Delta\mathbf{z}_{it}) = \sigma^2(\Delta\tilde{\mathbf{x}}_{it}, \Delta\mathbf{z}_{it})$. For a given j, let us denote $\hat{f}(\tilde{\mathbf{x}}_{it})$ and $\hat{f}(\tilde{\mathbf{x}}_{i,t-1})$ the estimates of $f(\tilde{\mathbf{x}}_{it})$ and $f(\tilde{\mathbf{x}}_{i,t-1})$ respectively. Then, a more precise estimator,[7] say $\hat{\hat{f}}$, can be obtained as a weighted average of $\hat{f}(\tilde{\mathbf{x}}_{it})$ and $\hat{f}(\tilde{\mathbf{x}}_{i,t-1})$:

$$\hat{\hat{f}}(\tilde{\mathbf{x}}) = \frac{1}{2}[\hat{f}(\tilde{\mathbf{x}}_{it}) + \hat{f}(\tilde{\mathbf{x}}_{i,t-1})] \tag{8}$$

In practice, we base our estimation on a "backfitting algorithm" (see Appendix B for details on the computational methods).[8] We also test for the parametric analogue of the regression function against the non parametric one using the "gain" statistic. The "gain" is the difference in normalized deviance between the GAM and a model with a linear term for the corresponding regressor (see Appendix B for details). Finally, our confidence interval are constructed using the "wild bootstrap". As shown in Appendix C, the wild bootstrap has the advantage of being robust to heteroskedasticity and correlation between observations.

4.3. Structural form estimation

Relation (3) is a semi-parametric reduced form linking TFP growth to its determinants. However, there is clear endogeneity issue with the proximity variables. Indeed, the definition of a frontier relies on economic conditions. The frontier is a economic that perform better in some sense. This means that the frontier is also determined by economic performance, such as TFP growth. As a result, there might be a feedback effect from TFP to the frontier. This leads to a system of nonparametric equations.

In order to address this issue, we consider the triangular nonparametric simultaneous specification of Newey et al. (1999),

$$y = m(x, z_0) + \varepsilon \tag{9}$$

$$x = \pi(\mathbf{z}) + u, \qquad \mathbb{E}(\varepsilon|u, \mathbf{z}) = \mathbb{E}(\varepsilon|u) \neq 0, \quad \mathbb{E}(u|\mathbf{z}) = 0 \tag{10}$$

where y, x and z_0 denote respectively the dependent variable and the controls; $\mathbf{z}$ is a set of instruments that includes z_0. The system (9)–(10) is a generalization of the limited information simultaneous equations model to allow for structural nonparametric relation $m(x, z_0)$ between variables y, x and z_0, and a nonparametric reduced form $\pi(\mathbf{z})$. The conditional expectation of Eq. (9) yields the integral equation,

$$\mathbb{E}(y|\mathbf{z}) \equiv \pi(\mathbf{z}) = \mathbb{E}[m(x, z_0)|\mathbf{z}] = \int m(x, z_0)G(dx|\mathbf{z}) \tag{11}$$

where G denotes the conditional cumulative distribution function of x given $\mathbf{z}$. Thus, functions π and G are the nonparametric generalization of the reduced forms for y and x. Newey et al. (1999) discussed the identification of the system (9)–(10).[9] Starting from a preliminary estimation of the reduced forms $\hat{\pi}$ and $\hat{G}$:

$$\hat{\pi}(\mathbf{z}) = \int m(x, z_0)\hat{G}(dx|\mathbf{z}), \tag{12}$$

the authors developed an estimator for $\hat{m}$ that overcomes the well known ill-posed problem.[10]

In order to apply this methodology to analyze TFP growth, we specify a GAM for fixed effects panel data.[11] For Eq. (9), the GAM is

$$\mathcal{A}_{it} = \sum_{j=1}^{p} m_j(w_{it}^j) + \mathbf{z}_{0it}'\beta + \mu_i + \varepsilon_{it}, \tag{13}$$

where w_{it}^j is the jth component ($j = 1, \ldots, p$) of $\mathbf{w}_{it} \equiv (\tilde{\mathbf{x}}_{it}, z_{0it})$. For Eq. (10) we use a semi-parametric GAM specification the structure of which is given by

$$\tilde{x}_{it} = \sum_{k=1}^{q} \pi_j(z_{1it}^k) + \mathbf{z}_{2it}'\gamma + \lambda_i + u_{it}, \tag{14}$$

[7] This is particularly useful in case where the shape of the two estimates are closely related.

[8] Linton and Härdle (1996) propose an alternative estimation method based on the integration of a transformed pilot regression smoother. However, this estimator is not efficient and more recently, Linton (2000) suggested two-step procedures which are more efficient.

[9] Identification is needed as π and G are functionals of the distribution of observables $(y, x, \mathbf{z})$.

[10] The ill-posed inverse problem follows from non-continuity of $\hat{m}$. Indeed, lack of continuity of $\hat{\pi}$ and $\hat{G}$ can translate into large inaccuracies in $\hat{m}$.

[11] See e.g., Hastie and Tibshirani (1990) for further details on GAM.

where $\tilde{x}_{it}$ is the component of $\tilde{\mathbf{x}}_{it}$ corresponding to the proximity to frontier, that is $\ln(A_i/A_{\mathcal{F}})_{t-1}$, z^k_{1it}s is the kth component (k = 1, ..., q) of the set of continuous instruments $\mathbf{z}_1$ and $\mathbf{z}_{2it}$ corresponds to other instruments which do enter linearly in the specification. Following Vandenbussche et al. (2006), our instruments are proximity lagged three times ($\ln(A_i/A_{\mathcal{F}})_{t-3}$).[12] We also include the percentage of graduate students and its interaction with the measure of proximity under use, R&D expenditures by source and their interaction with the measure of proximity under use. The unobserved fixed effects μ_i and λ_i can be eliminated by first differences:

$$\mathcal{A}_{it} - \mathcal{A}_{i,t-1} = \sum_{j=1}^{p}[m_j(w^j_{it}) - m_j(w^j_{i,t-1})] + (\mathbf{z}_{0it} - \mathbf{z}_{0i,t-1})'\beta + \varepsilon_{it} - \varepsilon_{i,t-1} \quad (15)$$

$$\tilde{x}_{it} - \tilde{x}_{i,t-1} = \sum_{k=1}^{q}[\pi_j(z^k_{1it}) - \pi_j(z^k_{1i,t-1})] + (\mathbf{z}_{2it} - \mathbf{z}_{2i,t-1})'\gamma + u_{it} - u_{i,t-1} \quad (16)$$

Observe that the method of Newey et al. (1999) consists of estimating Eq. (15) by including an additional control variable which is the first difference residuals $\hat{u}_{it} - \hat{u}_{i,t-1}$ computed from Eq. (16). We perform estimation in two steps: (i) construct GAM semi-parametric first differences residuals $\hat{u}_{it} - \hat{u}_{i,t-1}$ of Eq. (16), (ii) estimate GAM semi-parametric model in Eq. (15) using the residuals $\hat{u}_{it} - \hat{u}_{i,t-1}$ from (i) as additional regressor. The estimation procedure still based on the ‘backfitting algorithm’ (Hastie and Tibshirani, 1990). Furthermore, as m_j is estimated twice, denoted as $\hat{m}^{(1)}_j$ and $\hat{m}^{(2)}_j$ for w^j_{it} and $w^j_{i,t-1}$ respectively, a simple and more precise estimator of m_j can be obtained by a weighted average: $\hat{m}_j = (\hat{m}^{(1)}_j + \hat{m}^{(2)}_j)/2$.

5. Estimation results

We present estimations results both for reduced and the structural forms. For the reduced form, we compute the semi-parametric GAM estimate and also the parametric estimations for comparison purpose. For the parametric estimate, we provide estimates based on three methods: OLS, the within estimator and the first difference estimator. As put forward above, the semi-parametric estimations are based on the fist difference assumption, hereafter FDA (6). Let us discuss very briefly the rationale of this assumption which central to the implementation of our semi-parametric estimations.

It is well known that strict exogeneity or *predeterminedness* is needed for the within estimator (Wooldridge, 2006). In our semi-parametric framework, strict exogeneity precludes any feedback from the current value of TFP growth on future values of time varying controls (for example proximity indicators, R&D, FDI, etc.), which is not a realistic assumption. It is also worthwhile to note that *predeterminedness* is neither necessary nor sufficient for (6). It is not sufficient, because under predeterminedness alone $\mathbb{E}(\eta_{it}|\tilde{\mathbf{x}}^j_{it}, \tilde{\mathbf{x}}^j_{i,t-1}) = -\mathbb{E}(u_{i,t-1}|\tilde{\mathbf{x}}^j_{it}, \tilde{\mathbf{x}}^j_{i,t-1})$, which will not be zero in general. Thus the extension of *predeterminedness* yielding (7) is

$$\mathbb{E}(u_{it}|\tilde{\mathbf{x}}^j_{i,t+1}, \tilde{\mathbf{x}}^j_{i,t}) = 0, \quad i = 1, \ldots, N,\ t = 1, \ldots, T-1.$$

with *predeterminedness* still holding for $t = T$. In our semi-parametric framework, this only precludes feedback from the current value of TFP growth on next year's value of time varying controls, but not on later values, which appears much less stringent than strict exogeneity. However, *predeterminedness* is not a necessary condition for the first difference assumption, since if u_{it} is a random walk, the first difference assumption is satisfied without further assumption on $\mathbb{E}(u_{it}|\tilde{\mathbf{x}}^j_{i,t}, \tilde{\mathbf{x}}^j_{i,t-1})$. This closes our discussion of assumption FDA, which we maintain in the sequel for estimations. Lastly, if there is enough variation in controls over the index i and between $t-1$ and t, then function f, m and π are identified up to a common constant. Thus, even in the nonpooled nonparametric model, the country-specific effects can be eliminated up to an additive constant.

5.1. Reduced form

As outlined earlier, the empirical literature considers reduced-form equations relating the TFP growth to education or human capital and the proximity to world frontier (Vandenbussche et al., 2006) and to R&D and proximity to the country with highest TFP growth (Griffith et al., 2004). In order to avoid conflation of concepts that will make it hard to track the main findings and in order to make our findings comparable to the previous ones, we follow the same road by estimating separately our models with three types of controls: higher education, total staff in R&D and R&D expenditure. For each case, we use three types of frontiers: the economy with the highest TFP growth, the world TFP growth and the TFP growth of the USA. We also include time dummies with 2008 as reference. Estimation results for reduced form are given in Tables 2, 4 and 6 for parametric estimations, and Tables 3, 5 and 7 and Figs. 2–15 for the semi-parametric estimations. As can be observed from Tables 3, 5 and 7, the ‘statistic of global gain’ (or nonlinearity χ_2) shows that our semi-parametric is not rejected against the parametric OLS analogue. As a result, our specification provides a better approximation of the data compared to the parametric model. The ‘statistic of individual gain’ allows us to test for the degree of nonlinearity of functions. We then plot figures only for the continuous controls variables for which the degree of nonlinearity is significant at the 5% level.

5.1.1. Higher education

We observe that the relation between TFP growth and proximity to the frontier depends on the type of frontier.

[12] The three times lagging follows from the fact relation (16) shifts the period to three as $\ln(A_i/A_{\mathcal{F}})_{t-1}$ already starts with one lag. Now, observe that this also allows to have a sufficiently distant variable to weaken as much as possible endogeneity while preserving sufficient observations to achieve convergence of estimates.

Table 2
Parametric estimates (reduced form): percentage of graduate students in higher education.

Variable	OLS		Within		First difference	
	Coef.	Std. Err.[a]	Coef.	Std. Err.[a]	Coef.	Std. Err.[a]
Proximity: $\mathcal{F}_{\max}$						
Prox. highest TFP	-0.388^{***}	0.125	-0.116	0.102	0.377^{***}	0.090
Higher educ.	0.011	0.031	0.076	0.059	0.094	0.076
Interaction prox. educ.	-0.0003	0.001	0.001	0.001	0.0005	0.0008
FDI	-0.006^{***}	0.002	-0.010	0.011	0.001	0.011
Openness	0.044	0.090	8.490^{***}	2.339	8.913^{**}	3.805
Year 1998	1.489^{**}	0.775	4.446^{***}	1.209	–	
Year 1999	0.337	0.727	3.515^{***}	0.893	1.215	1.142
Year 2000	3.299^{***}	0.572	4.759^{***}	1.020	1.669^{*}	0.911
Year 2001	0.994	0.711	3.179^{***}	0.945	1.166	0.921
Year 2002	5.142^{***}	0.766	4.140^{***}	1.182	-0.943	0.996
Year 2003	3.271^{***}	0.601	3.656^{***}	0.852	1.009	0.936
Year 2004	15.961^{***}	2.253	4.392^{*}	2.425	-10.548^{***}	2.043
Year 2005	2.207^{***}	0.448	2.688^{***}	0.631	1.745^{***}	0.668
Year 2006	3.201^{***}	0.550	2.679^{***}	0.618	1.039^{**}	0.512
Year 2007	1.321^{***}	0.483	2.040^{***}	0.406	2.355^{***}	0.418
OECD	-0.977^{***}	0.268	–		–	
Intercept	1.978	2.741	-144.292^{***}	38.446	–	
σ_μ				12.090		
σ_u				2.210		
ρ[b]				0.967		
# observations	458			458	356	
Proximity: $\mathcal{F}_{\text{World}}$						
Prox. World TFP	0.374	0.349	0.406	0.371	-0.588	0.388
Higher educ.	0.005	0.024	0.105^{***}	0.053	0.089	0.082
Interaction prox. educ.	0.0005	0.006	-0.007	0.006	0.004	0.006
FDI	-0.006^{***}	0.002	-0.011	0.012	0.002	0.011
Openness	0.043	0.091	8.450^{***}	2.397	6.681^{*}	3.795
Year 1998	-0.407	0.736	4.364^{***}	1.185	–	
Year 1999	0.323	0.729	3.573^{***}	0.931	-0.338	1.157
Year 2000	2.254^{***}	0.563	4.747^{***}	1.022	1.440	0.925
Year 2001	0.547	0.715	3.155^{***}	0.962	0.472	0.918
Year 2002	1.907^{***}	0.550	3.997^{***}	0.996	1.129	0.991
Year 2003	1.530^{***}	0.525	3.601^{***}	0.845	1.887^{*}	0.976
Year 2004	2.534^{***}	0.480	3.766^{***}	0.721	2.269^{***}	0.771
Year 2005	2.128^{***}	0.445	2.697^{***}	0.622	1.390^{**}	0.665
Year 2006	2.228^{***}	0.519	2.605^{***}	0.566	1.779^{***}	0.528
Year 2007	2.109^{***}	0.469	2.026^{***}	0.396	1.508^{***}	0.383
OECD	-0.977^{***}	0.271	–		–	
Intercept	-0.809	2.462	-145.379^{***}	39.507	–	
σ_μ				12.019		
σ_u				2.208		
ρ[b]				0.967		
# observations	458			458	356	
Proximity: $\mathcal{F}_{\text{USA}}$						
Prox. TFP USA	0.412	0.300	0.401^{***}	0.119	0.533^{***}	0.056
Higher educ.	0.012	0.022	0.032	0.072	0.041	0.062
Interaction prox. educ.	0.002	0.005	0.001	0.002	-0.001	0.001
FDI	-0.0002	0.003	-0.0005	0.011	0.006	0.006
Openness	-0.112	0.083	7.921^{***}	2.214	3.622	2.354
Year 1998	0.448	0.722	4.092^{***}	1.056	–	
Year 1999	-0.672	0.601	2.232^{***}	0.753	-0.940	0.624
Year 2000	0.381	0.544	3.587^{***}	0.666	-0.003	0.613
Year 2001	0.566	0.514	2.719^{***}	0.755	0.303	0.566
Year 2002	0.472	0.555	2.440^{***}	0.765	0.223	0.615
Year 2003	0.810	0.561	2.516^{***}	0.712	0.799	0.632
Year 2004	1.625^{***}	0.542	2.517^{***}	0.577	1.115^{**}	0.519
Year 2005	1.463^{***}	0.491	2.047^{***}	0.522	1.031^{**}	0.442
Year 2006	1.276^{***}	0.494	1.717^{***}	0.407	0.976^{***}	0.339
Year 2007	1.118^{**}	0.501	1.232^{***}	0.313	0.876^{***}	0.264
OECD	-1.282^{***}	0.234	–		–	
Intercept	2.321	2.359	-144.292^{***}	38.446	–	
σ_μ				11.483		
σ_u				1.640		
ρ[b]				0.980		
# observations	449			449	348	

[a] Robust standard errors (White, 1980).
[b] Fraction of variance due to μ_i.
* Significance level: 10%.
** Significance level: 5%.
*** Significance level: 1%.

Table 3
GAM semi-parametric estimates (reduced form): percentage of graduate students in higher education.

Variable	Dof.[a]	Lin. Coef.	Std. Err.	$p >$ gain[b]
Proximity: $\mathcal{F}_{\max}$				
Prox. highest TFP	5.013	−0.349	0.078	0.000
Higher educ.	5.003	−0.019	0.022	0.137
Interaction prox. educ.	5.005	0.002	0.001	0.000
FDI	5.001	−0.005	0.004	0.840
Openness	4.999	0.009	0.094	0.713
Year 1999	1	0.168	0.551	
Year 2000	1	3.136***	0.571	
Year 2001	1	0.996*	0.551	
Year 2002	1	4.574***	0.634	
Year 2003	1	3.458***	0.559	
Year 2004	1	10.036***	1.476	
Year 2005	1	2.196***	0.527	
Year 2006	1	3.349*	0.533	
Year 2007	1	0.932***	0.521	
Deviance		2563.63		
Dispersion		6.089		
Nonlinearity χ_2[c]	20.021			0.000
# observations		458		
Proximity: $\mathcal{F}_{\text{World}}$				
Prox. World TFP	5.004	0.549	0.190	0.000
Higher educ.	5.003	0.018	0.015	0.035
Interaction prox. educ.	5.002	−0.002	0.003	0.158
FDI	5.000	−0.005	0.004	0.786
Openness	4.999	0.022	0.091	0.663
Year 1999	1	−0.063	0.537	
Year 2000	1	1.556***	0.546	
Year 2001	1	0.415	0.537	
Year 2002	1	1.629***	0.521	
Year 2003	1	1.402***	0.524	
Year 2004	1	2.403***	0.511	
Year 2005	1	2.189***	0.513	
Year 2006	1	2.215***	0.510	
Year 2007	1	1.949***	0.504	
Deviance		2432.89		
Dispersion		5.778		
Nonlinearity χ_2[c]	20.008			0.000
# observations		458		
Proximity: $\mathcal{F}_{\text{USA}}$				
Prox. TFP USA	5.001	0.886	0.161	0.000
Higher educ.	5.004	−0.003	0.013	0.000
Interaction prox. educ.	5.002	−0.007	0.002	0.000
FDI	4.996	0.002	0.003	0.541
Year 1998	1	0.542	0.699	
Year 1999	1	−0.221	0.489	
Year 2000	1	0.459	0.511	
Year 2001	1	0.847*	0.485	
Year 2002	1	0.521	0.481	
Year 2003	1	0.765	0.478	
Year 2004	1	1.588***	0.467	
Year 2005	1	1.231***	0.464	
Year 2006	1	1.088**	0.465	
Year 2007	1	0.915**	0.458	
Deviance		1915.76		
Dispersion		4.649		
Nonlinearity χ_2[c]	20.002			0.000
# observations		449		

[a] Degree of freedom.
[b] Individual gain.
[c] Total gain.
* Significance level for parametric (linear) components: 10%.
** Significance level for parametric (linear) components: 5%.
*** Significance level for parametric (linear) components: 1%.

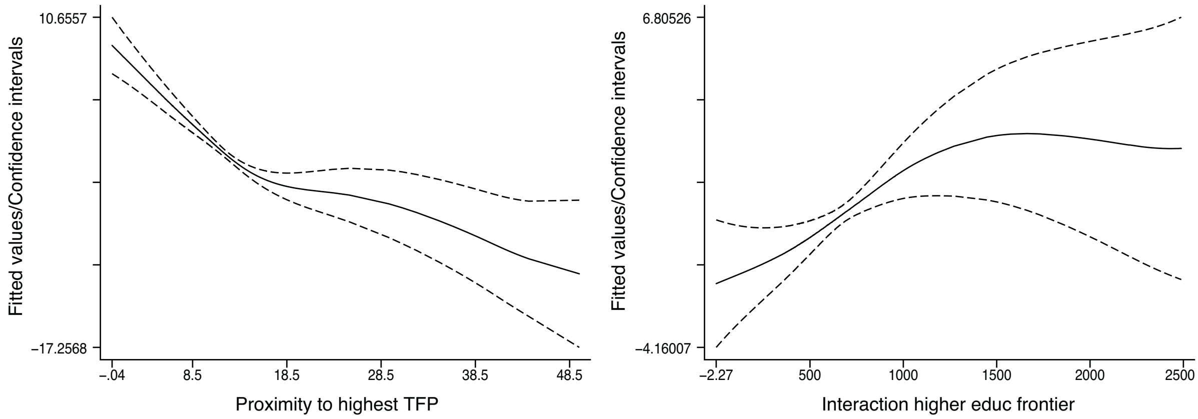

Fig. 2. Nonparametric estimation of the reduced form function $f(x)$. The solid curve represents the estimate $\hat{f}(x)$, the dashed curves correspond to the 95% pointwise bootstrap confidence intervals. [Left]: Relation between TFP growth and proximity to the highest TFP growth. [Right]: Relation between TFP growth and the interaction between the proximity to the highest TFP growth and the percentage of graduate students in higher education.

While it is decreasing with proximity to the highest TFP growth, the relation is increasing with both proximity to the world TFP growth and to the TFP growth of the USA. In the case of proximity to the highest TFP, the patterns of decrease vary, meaning that the speed of the decrease is much more pronounced for small values of proximity (see Fig. 2, left). The patterns of increase that we observe for the relation between TFP growth and proximity to the world TFP growth and to the TFP growth of the USA also differ. In the former case, there is a kind of exponential increase for values of proximity starting from −4.2% (Fig. 3, left). In the latter case, the relation is increasing at a decreasing rate (Fig. 4, top). There also a kind of opposite effect for the interaction variables. The relation between TFP growth and interaction between proximity to the highest TFP growth and the percentage of graduate students in higher education increases up to a certain level starting from which it became quite flat and non significant (Fig. 2, right). This finding is consistent with Vandenbussche et al. (2006) who used the world TFP growth as frontier. Its means that the proportion of graduate people in tertiary education is important for growth in economies closer to the frontier. In our case, this relation is not significant with respect to the world frontier. Though we find a decreasing relation when we use the TFP growth of the USA (Fig. 3, bottom-right) as frontier.

Interestingly is also the effect of higher education on TFP growth. Fig. 3 (right) shows (for the range of significancy) an inverted U-shape relation with a turning point around 56% of graduate students in higher education when the world frontier is used. In the case where the frontier is the USA, we obtain a U-shape relation (Fig. 4, bottom-left) with a turning point at about 36%. While Vandenbussche et al. (2006) and Griffith et al. (2004) find respectively a negative and a positive impact of human capital on TFP growth in a parametric context, we find here that the relation is highly nonlinear and much more complex and than it appears. Indeed, relying on the world frontier as Vandenbussche

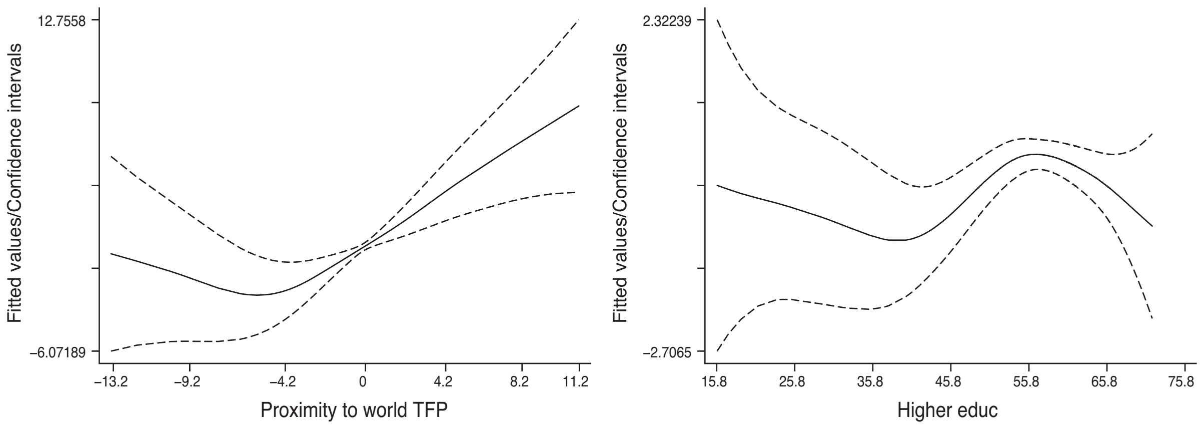

Fig. 3. Nonparametric estimation of the reduced form function $f(x)$. The solid curve represents the estimate $\hat{f}(x)$, the dashed curves correspond to the 95% pointwise bootstrap confidence intervals. [Left]: Relation between TFP growth and proximity to the world TFP growth. [Right]: Relation between TFP growth and the percentage of graduate students in higher education.

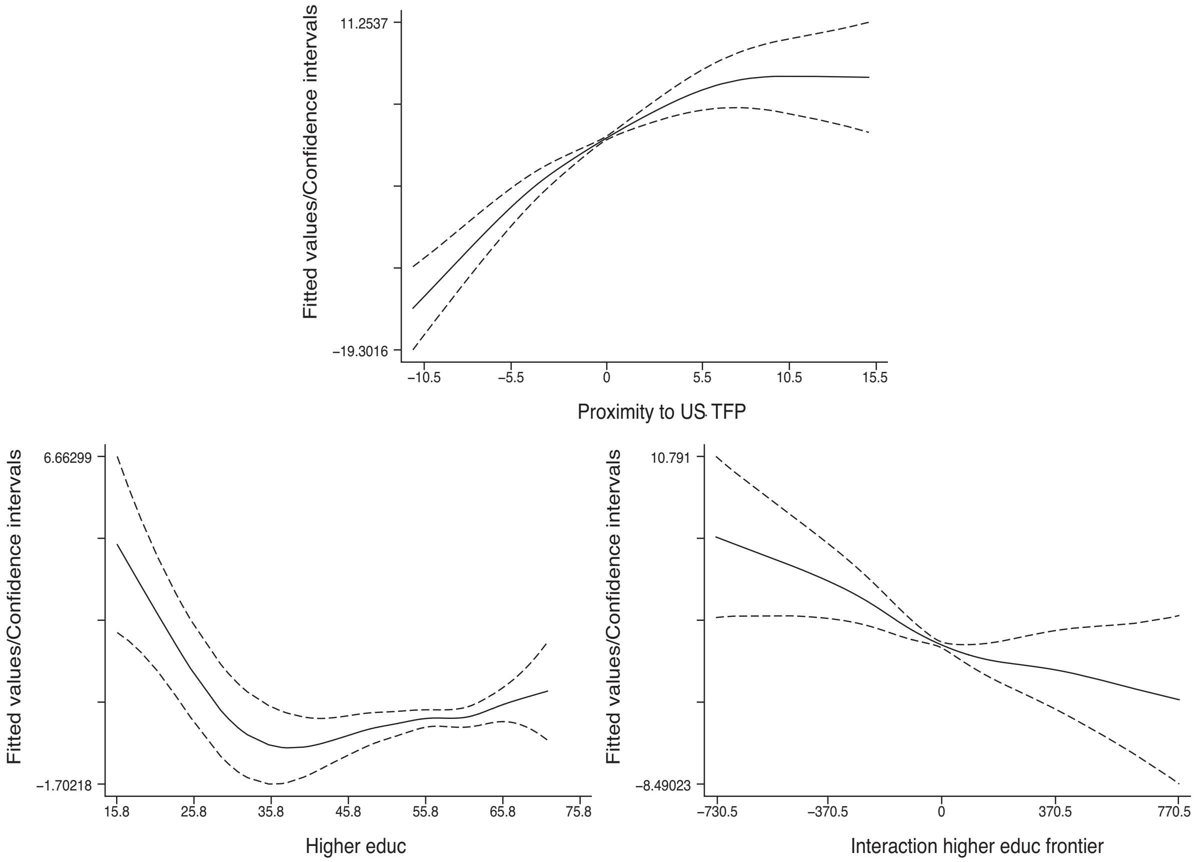

Fig. 4. Nonparametric estimation of the reduced form function *f*(*x*). The solid curve represents the estimate $\hat{f}(x)$, the dashed curves correspond to the 95% pointwise bootstrap confidence intervals. [Top]: Relation between TFP growth and proximity to the TFP growth of USA. [Bottom-left]: Relation between TFP growth and percentage of graduate students in higher education. [Bottom-right]: Relation between TFP growth and the interaction between the proximity to the TFP growth of USA and the percentage of graduate students in higher education.

et al. (2006) we obtain a positive relation up to the turning point of 56% after which the impact of higher education becomes negative. However, based on the TFP growth of the USA as frontier, the effect of higher education is strongly negative down to the point 36% after which the impact becomes positive. Nevertheless, the positive effect is still quite light. Griffith et al. (2004) also obtain a positive effect of human capital but their finding is based on a frontier defined as the highest TFP growth. Our results also highlight that the choice of the frontier does play a crucial role in the effect of human capital on TFP growth. Griffith et al. (2004) also emphasize the role of international trade and FDI in shaping TFP growth. In our semi-parametric estimations and whatever the proximity measure used, we do not find evidence of FDI and openness on TFP growth (Table 3). However, the within and first difference parametric estimations show evidence of positive effect of openness on TFP growth, except openness in the FD estimation when proximity to the TFP growth of the USA is used.

5.1.2. Total staff in R&D

The results when total staff in R&D is used are given in Tables 4 and 5, in Figs. 5–7. The relation between TFP growth and proximity to the frontier indicators displays the same global shape as the previous case. It is decreasing with respect to the highest TFP growth with different speeds (Fig. 5, left). In fact, we can distinguish two phases in the evolution of TFP growth with respect to proximity to highest TFP growth: The first phase is characterized by a sharp decline to the point 18.5 that we can call a break point (because the decline appears to have a break) from which the relationship continues to decrease but with a lower slope in the previous case (higher education, Fig. 2, left), and with identical slope in the current case. The other two proximity measures (world TFP growth and TFP growth of the USA) show equally similar curvature (Fig. 6) to the case of higher education when we consider the range of significancy. The effect of total staff in R&D on TFP growth is the same for both proximity to highest TFP (Fig. 5, right) and for TFP growth of the USA (Fig. 7, right). It increases to a certain point (about 8.5) and then decreases to some level (12.5) and increases again. It reflects an inverted W-shape relationship. However, we observe that the magnitude of the effect is much more pronounced when we use the TFP growth of the USA frontier.

Table 4
Parametric estimates (reduced form): total staff R&D.

Variable	OLS		Within		First difference	
	Coef.	Std. Err.[a]	Coef.	Std. Err.[a]	Coef.	Std. Err.[a]
Proximity: $\mathcal{F}_{\max}$						
Prox. highest TFP	−0.263***	0.041	−0.064	0.045	0.229***	0.048
Total staff R&D	0.017	0.176	−2.230*	1.333	−1.894	1.821
Interaction prox. staff	3.14e−08	3.83e−08	−4.18e−08*	2.22e−08	−3.64e−08	3.33e−08
FDI	−0.002	0.001	0.003	0.004	0.006***	0.001
Openness	−0.060	0.177	1.989	1.775	8.850***	3.396
Year 1998	−0.097	0.558	−0.216	0.534	–	
Year 1999	−1.344*	0.741	−0.643	0.741	0.890	0.690
Year 2000	1.023*	0.546	1.242*	0.639	1.920***	0.696
Year 2001	−0.961	0.611	−0.038	0.642	1.017	0.653
Year 2002	1.302*	0.754	0.797	0.656	−0.151	0.648
Year 2003	1.642***	0.548	1.660***	0.484	1.566***	0.574
Year 2004	9.213***	1.314	3.765**	1.470	−4.754***	1.338
Year 2005	0.353	0.429	0.786*	0.415	1.691***	0.489
Year 2006	1.110**	0.464	1.333***	0.385	1.436***	0.413
Year 2007	−0.324	0.481	0.674*	0.385	1.964***	0.417
OECD	−0.737*	0.380	–		–	
Intercept	4.674**	2.741	−8.081	23.792	–	
σ_μ				4.048		
σ_u				2.615		
ρ[b]				0.705		
# observations	556			556	441	
Proximity: $\mathcal{F}_{\mathrm{World}}$						
Prox. World TFP	0.469***	0.063	0.210***	0.059	−0.345***	0.068
Total staff R&D	0.059	0.157	−1.886*	1.125	−1.955	1.622
Interaction prox. staff	1.12e−08	1.17e−07	2.44e−08	8.73e−08	−3.45e−08	1.14e−08
FDI	−0.001	0.001	0.001	0.004	0.006***	0.001
Openness	−0.023	0.172	2.240	1.727	7.738**	3.370
Year 1998	0.027	0.510	−0.093	0.531	–	
Year 1999	0.083	0.688	−0.155	0.714	−0.113	0.673
Year 2000	1.739***	0.529	1.507**	0.631	1.671	0.744
Year 2001	0.006	0.590	0.201	0.598	0.623	0.673
Year 2002	0.658	0.697	0.673	0.561	0.976	0.682
Year 2003	1.765***	0.495	1.656***	0.461	2.056***	0.637
Year 2004	1.882***	0.371	1.717***	0.374	2.318***	0.566
Year 2005	1.694***	0.369	1.141***	0.391	1.422***	0.493
Year 2006	1.908***	0.412	1.513***	0.350	1.769***	0.436
Year 2007	1.561***	0.397	1.120***	0.354	1.468***	0.383
OECD	−0.622*	0.353	–		–	
Intercept	0.256	1.758	−16.641	22.170	–	
σ_μ				3.511		
σ_u				2.568		
ρ[b]				0.651		
# observations	556			556	441	
Proximity: $\mathcal{F}_{\mathrm{USA}}$						
Prox. TFP USA	0.535***	0.074	0.513***	0.025	0.503***	0.019
Total staff R&D	0.039	0.126	−1.694	1.230	0.262	0.809
Interaction prox. staff	−3.70e−08	1.39e−08	−4.26e−08	4.01e−08	−2.3e−08	6.33e−08
FDI	−0.0009	0.0007	0.005*	0.003	0.004***	0.001
Openness	−0.077	0.136	2.549	1.680	2.680	1.826
Year 1998	−0.078	0.529	−0.149	0.372	–	
Year 1999	−0.921	0.535	−0.949*	0.512	−0.686*	0.387
Year 2000	−0.132	0.554	−0.179	0.600	0.162	0.568
Year 2001	0.276	0.449	0.370	0.496	0.558	0.468
Year 2002	−0.381	0.575	−0.336	0.486	0.092	0.499
Year 2003	0.832	0.504	0.680	0.425	0.916**	0.444
Year 2004	1.192**	0.503	0.946**	0.415	1.374***	0.379
Year 2005	1.121***	0.415	0.642*	0.376	1.229***	0.327
Year 2006	0.777*	0.420	0.386	0.354	1.036***	0.283
Year 2007	0.733	0.454	0.301	0.324	1.005***	0.211
OECD	−0.989***	0.330	–		–	
Intercept	2.322	1.497	−23.027	12.881	–	
σ_μ				3.232		
σ_u				2.026		
ρ[b]				0.725		
# observations	556			556	441	

[a] Robust standard errors (White, 1980).
[b] Fraction of variance due to μ_i.
* Significance level: 10%.
** Significance level: 5%.
*** Significance level: 1%.

Table 5
GAM semi-parametric estimates (reduced form): total staff R&D.

Variable	Dof.[a]	Lin. Coef.	Std. Err.	p > gain[b]
Proximity: $\mathcal{F}_{\max}$				
Prox. highest TFP	5.005	−0.268	0.032	0.000
Total staff R&D	4.996	−0.026	0.144	0.041
Interaction prox. staff	5.004	−1.06e−09	4.12e−08	0.532
FDI	5.001	−0.002	0.001	0.720
Openness	4.996	−0.052	0.154	0.329
Year 1999	1	−1.093**	0.542	
Year 2000	1	1.522***	0.531	
Year 2001	1	−0.558	0.530	
Year 2002	1	1.358**	0.536	
Year 2003	1	2.220***	0.507	
Year 2004	1	9.907***	1.036	
Year 2005	1	0.757	0.514	
Year 2006	1	1.905***	0.494	
Year 2007	1	−0.320	0.522	
Deviance		4198.47		
Dispersion		8.089		
Nonlinearity χ_2[c]	20.002			0.000
# observations		556		
Proximity: $\mathcal{F}_{\text{World}}$				
Prox. World TFP	5.006	0.471	0.042	0.000
Total staff R&D	4.996	0.013	0.124	0.392
Interaction prox. staff	4.995	−6.90e−08	1.35e−07	0.167
FDI	5.001	−0.002	0.001	0.813
Openness	4.996	−0.050	0.147	0.539
Year 1999	1	−0.065	0.500	
Year 2000	1	1.719***	0.504	
Year 2001	1	0.097	0.491	
Year 2002	1	0.676	0.499	
Year 2003	1	1.841***	0.483	
Year 2004	1	2.050***	0.486	
Year 2005	1	1.879***	0.473	
Year 2006	1	2.159***	0.468	
Year 2007	1	1.723***	0.459	
Deviance		3811.84		
Dispersion		7.344		
Nonlinearity χ_2[c]	19.994			0.006
# observations		556		
Proximity: $\mathcal{F}_{\text{USA}}$				
Prox. TFP USA	5.002	0.531	0.035	0.000
Total staff R&D	4.996	−0.219	0.110	0.000
Interaction prox. staff	5.000	−6.04e−08	1.18e−07	0.062
FDI	5.001	−0.002	0.001	0.816
Openness	4.996	0.123	0.131	0.177
Year 1999	1	−0.751*	0.449	
Year 2000	1	0.064	0.458	
Year 2001	1	0.439	0.436	
Year 2002	1	−0.100	0.445	
Year 2003	1	0.779*	0.433	
Year 2004	1	1.131**	0.434	
Year 2005	1	0.860**	0.421	
Year 2006	1	0.628	0.418	
Year 2007	1	0.444	0.409	
Deviance		3010.11		
Dispersion		5.799		
Nonlinearity χ_2[c]	19.995			0.000
# observations		556		

[a] Degree of freedom.
[b] Individual gain.
[c] Total gain.
* Significance level for parametric (linear) components: 10%.
** Significance level for parametric (linear) components: 5%.
*** Significance level for parametric (linear) components: 1%.

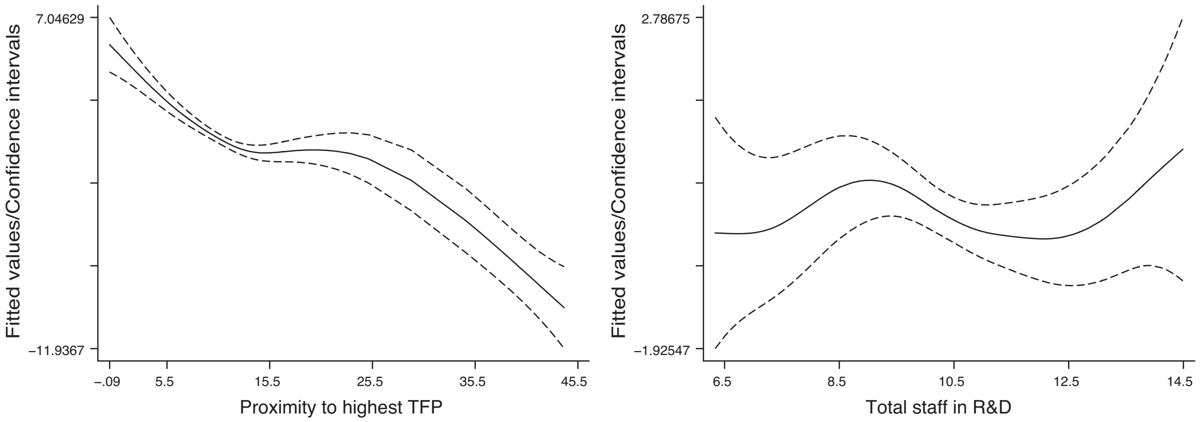

Fig. 5. Nonparametric estimation of the reduced form function $f(x)$. The solid curve represents the estimate $\hat{f}(x)$, the dashed curves correspond to the 95% pointwise bootstrap confidence intervals. [Left]: Relation between TFP growth and proximity to the highest TFP growth. [Right]: Relation between TFP growth and total staff in R&D.

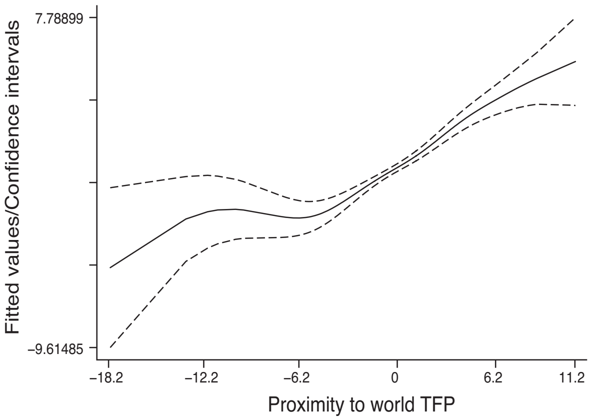

Fig. 6. Nonparametric estimation of the reduced form function $f(x)$. The solid curve represents the estimate $\hat{f}(x)$, the dashed curves correspond to the 95% pointwise bootstrap confidence intervals. Relation between TFP growth and proximity to the world TFP growth.

Here, while FDI and openness appear insignificant in the semi-parametric estimations (Table 5), they have significant and positive effect in the parametric within and fist difference estimations, except openness in the FD estimation when proximity to the TFP growth of the USA is used (Table 4).

5.1.3. R&D expenditure

The roles of R&D, also known as the two faces of R&D (stimulation of innovation and improvement of technology transfer or absorptive capacity), have been claimed and well documented in the literature. Griffith et al. (2004) provide empirical micro evidence using a panel of industries across 12 OECD countries. Our estimation also relates to this literature by including R&D expenditure by interacting R&D expenditure with the proximity measures. We distinguish three sources of R&D expenditure: Share of R&D expenditure funded by government, by business sector and from abroad. The estimations results are reported in Tables 6 and 7, and in Figs. 8–15. As before, the global shape of the relation between TFP growth and the three proximity measures are preserved (Fig. 8, left; Fig. 10, left and Fig. 12,

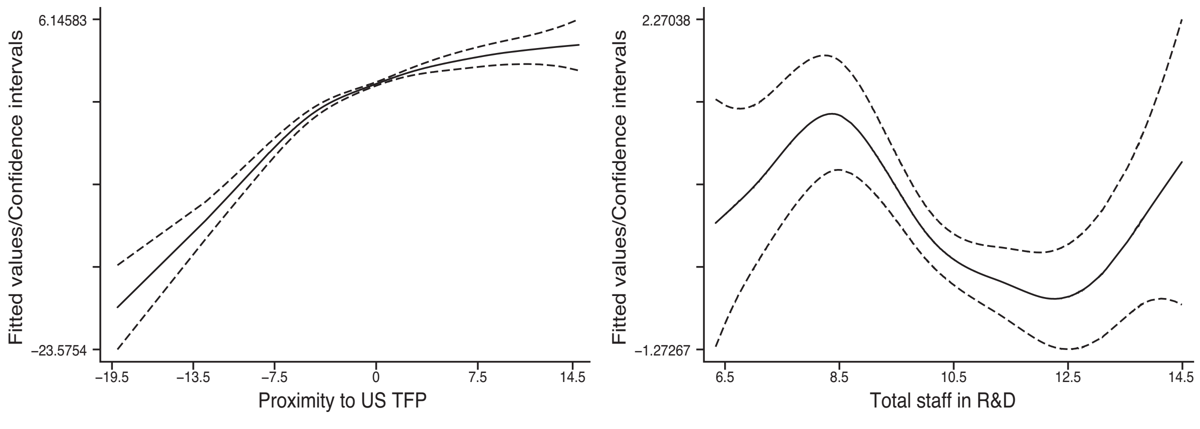

Fig. 7. Nonparametric estimation of the reduced form function $f(x)$. The solid curve represents the estimate $\hat{f}(x)$, the dashed curves correspond to the 95% pointwise bootstrap confidence intervals. [Left]: Relation between TFP growth and proximity to the TFP growth of USA. [Right]: Relation between TFP growth and total staff in R&D.

Table 6
Parametric estimates (reduced form): R&D expenditure.

Variable	OLS		Within		First difference	
	Coef.	Std. Err.[a]	Coef.	Std. Err.[a]	Coef.	Std. Err.[a]
Proximity: $\mathcal{F}_{\max}$						
Prox. highest TFP	−0.199	0.131	−0.067	0.143	0.224**	0.091
R&D gov.	0.052*	0.176	0.051	1.333	0.102	0.078
Interaction prox. R&D gov.	−0.0004	0.001	−0.00009	0.001	−0.0002	0.0009
R&D firm	0.050*	0.026	0.006	0.055	0.097	0.086
Interaction prox. R&D firm	−0.0004	0.001	−0.0002	0.001	−0.0004	0.0008
R&D abroad	0.057*	0.033	0.047	0.043	0.034	0.111
Interaction prox. R&D abroad	−0.0008	0.001	−0.00007	0.001	0.0003	0.001
FDI	−0.0009	0.001	−0.001	0.004	0.007	0.009
Openness	0.076	0.111	−0.015	1.409	10.894***	3.649
Year 1998	0.030	0.575	0.087	0.548	–	
Year 1999	−1.779**	0.802	−0.959	0.723	0.755	0.874
Year 2000	0.916*	0.547	1.222*	0.646	1.228**	0.671
Year 2001	−1.079*	0.658	−0.309	0.661	−0.874	0.665
Year 2002	1.386**	0.646	0.839	0.747	−0.318	0.663
Year 2003	1.208**	0.540	1.256**	0.505	1.409***	0.461
Year 2004	8.695***	1.436	4.043**	1.725	−5.553***	1.467
Year 2005	−0.160	0.431	0.551	0.545	5.449***	1.581
Year 2006	1.010*	0.532	1.264**	0.624	0.099	0.352
Year 2007	−0.732	0.492	0.523	0.592	0.518	0.366
OECD	−0.981***	0.346	–		–	
Intercept	−1.972	3.332	−0.956	21.318	–	
σ_μ				2.005		
σ_u				2.629		
ρ[b]				0.367		
# observations	520			520	399	
Proximity: $\mathcal{F}_{\mathrm{World}}$						
Prox. World TFP	0.770	0.546	1.007	0.856	0.828	0.515
R&D gov.	0.033**	0.016	0.039	0.033	0.092*	0.054
Interaction prox. R&D gov.	−0.005	0.005	−0.010	0.009	−0.011**	0.005
R&D firm	0.029*	0.015	0.004	0.040	0.092	0.066
Interaction prox. R&D firm	0.00002	0.006	−0.003	0.009	−0.013**	0.005
R&D abroad	0.034*	0.020	0.056	0.034	0.023	0.085
Interaction prox. R&D abroad	−0.003	0.009	−0.011	0.011	−0.011	0.009
FDI	−0.0006	0.001	−0.004	0.004	0.006	0.009
Openness	0.091	0.102	0.791	1.544	11.321***	3.568
Year 1998	0.216	0.503	0.239	0.542	–	
Year 1999	−0.320	0.721	−0.333	0.709	−0.170	0.812
Year 2000	1.023***	1.831	1.719**	0.659	1.761***	0.651
Year 2001	−0.023	0.626	0.006	0.622	−1.100*	0.639
Year 2002	1.069*	0.562	0.867	0.585	0.797	0.532
Year 2003	1.482***	0.477	1.251***	0.445	0.837**	0.409
Year 2004	1.863***	0.430	1.637***	0.463	−0.030	0.374
Year 2005	1.358***	0.392	1.036**	0.478	−1.073***	0.406
Year 2006	1.926***	0.503	1.515***	0.573	0.513	0.328
Year 2007	1.266***	0.403	1.115***	0.423	−0.265	0.304
OECD	−0.703**	0.324	–		–	
Intercept	−3.818*	2.232	−14.660	24.447	–	
σ_μ				2.145		
σ_u				2.523		
ρ[b]				0.419		
# observations	520			520	399	
Proximity: $\mathcal{F}_{\mathrm{USA}}$						
Prox. TFP USA	0.926*	0.490	0.993*	0.548	0.394*	0.225
R&D gov.	0.057***	0.014	0.039	0.029	0.038	0.034
Interaction prox. R&D gov.	−0.006	0.005	−0.006	0.005	0.0005	0.002
R&D firm	0.049***	0.013	−0.003	0.034	0.057	0.042
Interaction prox. R&D firm	−0.001	0.005	−0.002	0.005	0.001	0.002
R&D abroad	0.061***	0.017	0.006	0.033	−0.014	0.053
Interaction prox. R&D abroad	−0.006	0.007	−0.005	0.005	0.002	0.003
FDI	0.0002	0.0006	0.004	0.004	0.007	0.005
Openness	0.107	0.090	1.448	1.357	5.968***	2.059
Year 1998	0.091	0.569	0.064	0.395	–	
Year 1999	−1.017*	0.534	−0.911*	0.518	−0.628	0.499
Year 2000	−0.347	0.568	−0.475	0.666	0.757*	0.400
Year 2001	0.213	0.496	0.258	0.552	0.434	0.400
Year 2002	−0.601	0.552	−0.602	0.558	−0.192	0.312

Table 6 (*Continued*)

Variable	OLS		Within		First difference	
	Coef.	Std. Err.[a]	Coef.	Std. Err.[a]	Coef.	Std. Err.[a]
Year 2003	0.579	0.492	0.424	0.456	0.801***	0.285
Year 2004	0.919*	0.479	0.637	0.432	0.273	0.248
Year 2005	0.696*	0.405	0.487	0.439	−0.177	0.236
Year 2006	0.761*	0.434	0.298	0.484	−0.104	0.226
Year 2007	0.485	0.447	0.258	0.477	0.085	0.211
OECD	−1.274***	0.285	–		–	
Intercept	−4.991**	1.998	−23.970	22.117	–	
σ_μ				2.444		
σ_u				2.049		
ρ[b]				0.587		
# observations	520		520		399	

[a] Robust standard errors (White, 1980).
[b] Fraction of variance due to μ_i.
* Significance level: 10%.
** Significance level: 5%.
*** Significance level: 1%.

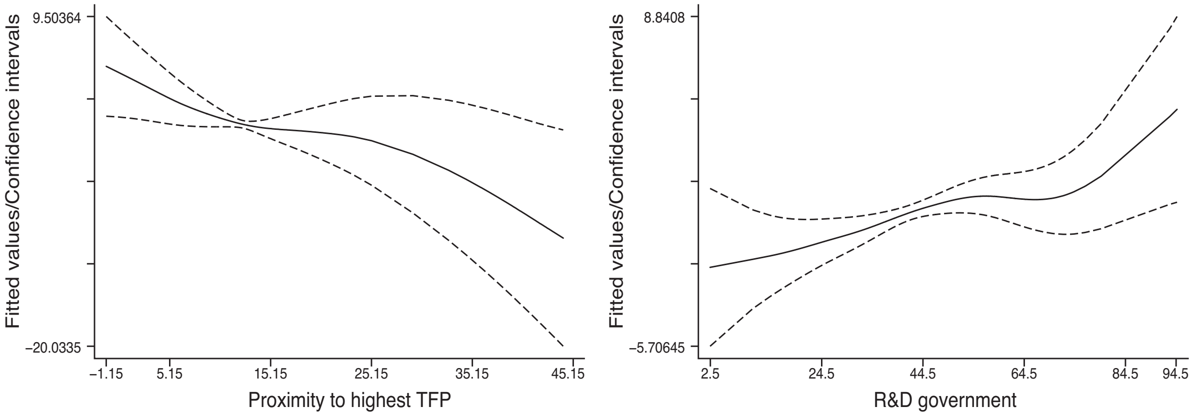

Fig. 8. Nonparametric estimation of the reduced form function $f(x)$. The solid curve represents the estimate $\hat{f}(x)$, the dashed curves correspond to the 95% pointwise bootstrap confidence intervals. [Left]: Relation between TFP growth and proximity to the highest TFP growth. [Right]: Relation between TFP growth and the share of R&D expenditure funded by government.

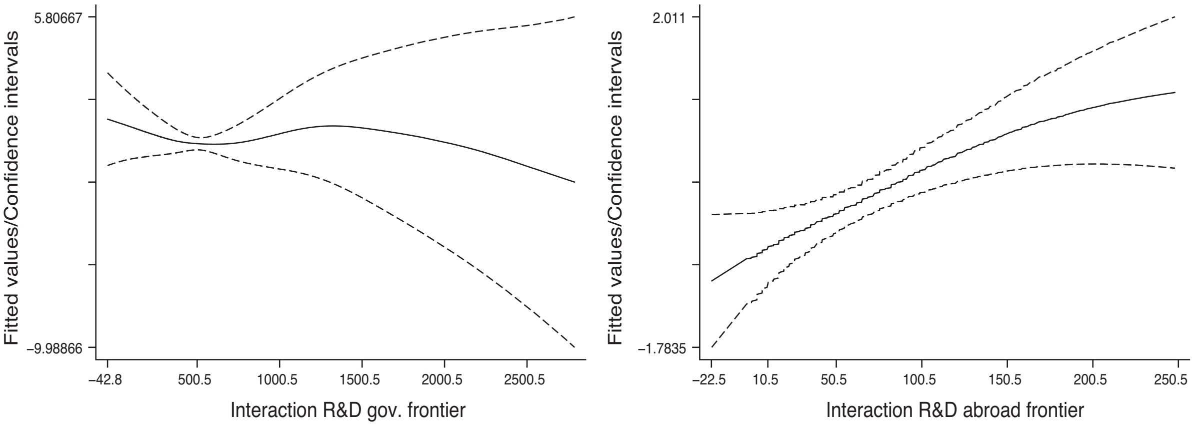

Fig. 9. Nonparametric estimation of the reduced form function $f(x)$. The solid curve represents the estimate $\hat{f}(x)$, the dashed curves correspond to the 95% pointwise bootstrap confidence intervals. [Left]: Relation between TFP growth and the interaction between proximity to highest TFP growth and share of R&D expenditure funded by government. [Right]: Relation between TFP growth and the interaction between proximity to the highest TFP growth and share of R&D expenditure funded from abroad.

Table 7
GAM semi-parametric estimates (reduced form): R&D expenditure.

Variable	Dof.[a]	Lin. Coef.	Std. Err.	p > gain[b]
Proximity: $\mathcal{F}_{\max}$				
Prox. highest TFP	5.006	−0.280	0.155	0.000
R&D gov.	5.004	0.052	0.027	0.013
Interaction prox. R&D gov.	5.003	−0.0002	0.001	0.041
R&D firm	4.995	0.050	0.026	0.068
Interaction prox. R&D firm	4.994	−0.0001	0.001	0.095
R&D abroad	5.005	−0.005	0.033	0.554
Interaction prox. R&D abroad	5.004	0.002	0.001	0.006
FDI	4.998	−0.001	0.002	0.708
Openness	4.997	0.020	0.099	0.531
Year 1999	1	−1.446***	0.560	
Year 2000	1	1.404***	0.529	
Year 2001	1	−0.737	0.527	
Year 2002	1	1.203**	0.536	
Year 2003	1	1.614***	0.497	
Year 2004	1	9.416***	1.056	
Year 2005	1	0.269	0.519	
Year 2006	1	1.677***	0.518	
Year 2007	1	−0.643	0.541	
Deviance		3548.59		
Dispersion		7.664		
Nonlinearity χ_2[c]	36.006			0.000
# observations		520		
Proximity: $\mathcal{F}_{\text{World}}$				
Prox. World TFP	4.990	0.761	0.426	0.000
R&D gov.	5.004	0.032	0.013	0.001
Interaction prox. R&D gov.	4.995	−0.004	0.004	0.004
R&D firm	4.995	0.037	0.013	0.234
Interaction prox. R&D firm	4.998	−0.0003	0.004	0.024
R&D abroad	5.005	0.049	0.019	0.152
Interaction prox. R&D abroad	5.006	−0.010	0.007	0.152
FDI	5.001	−0.001	0.002	0.545
Openness	4.997	0.083	0.091	0.618
Year 1999	1	−0.308	0.492	
Year 2000	1	1.685***	0.486	
Year 2001	1	−0.126	0.467	
Year 2002	1	0.933*	0.483	
Year 2003	1	1.355***	0.455	
Year 2004	1	1.710***	0.480	
Year 2005	1	1.300***	0.461	
Year 2006	1	1.907***	0.473	
Year 2007	1	1.270***	0.456	
Deviance		2990.02		
Dispersion		6.457		
Nonlinearity χ_2[c]	35.991			0.000
# observations		520		
Proximity: $\mathcal{F}_{\text{USA}}$				
Prox. TFP USA	4.995	0.928	0.335	0.000
R&D gov.	5.004	0.065	0.012	0.000
Interaction prox. R&D gov.	5.006	−0.005	0.003	0.009
R&D firm	4.995	0.040	0.012	0.000
Interaction prox. R&D firm	5.002	−0.002	0.003	0.010
R&D abroad	5.006	0.015	0.016	0.006
Interaction prox. R&D abroad	4.995	−0.006	0.005	0.000
FDI	5.001	0.0001	0.001	0.920
Openness	4.997	0.064	0.082	0.029
Year 1998	1	0.044	0.463	
Year 1999	1	−0.937**	0.448	
Year 2000	1	−0.178	0.452	
Year 2001	1	0.427	0.423	
Year 2002	1	−0.577	0.440	
Year 2003	1	0.375	0.417	
Year 2004	1	0.817*	0.437	
Year 2005	1	0.436	0.416	
Year 2006	1	0.578	0.431	
Year 2007	1	0.312	0.413	
Deviance		2454.68		
Dispersion		5.301		
Nonlinearity χ_2[c]	36.001			0.000
# observations		520		

[a] Degree of freedom.
[b] Individual gain.
[c] Total gain.
* Significance level for parametric (linear) components: 10%.
** Significance level for parametric (linear) components: 5%.
*** Significance level for parametric (linear) components: 1%.

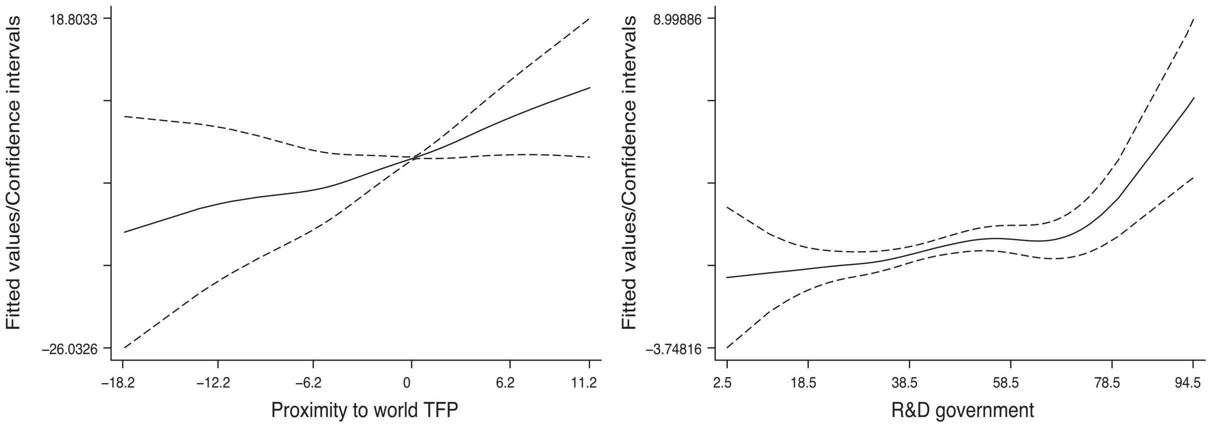

Fig. 10. Nonparametric estimation of the reduced form function $f(x)$. The solid curve represents the estimate $\hat{f}(x)$, the dashed curves correspond to the 95% pointwise bootstrap confidence intervals. [Left]: Relation between TFP growth and proximity to the world TFP growth. [Right]: Relation between TFP growth and the share of R&D expenditure funded by government.

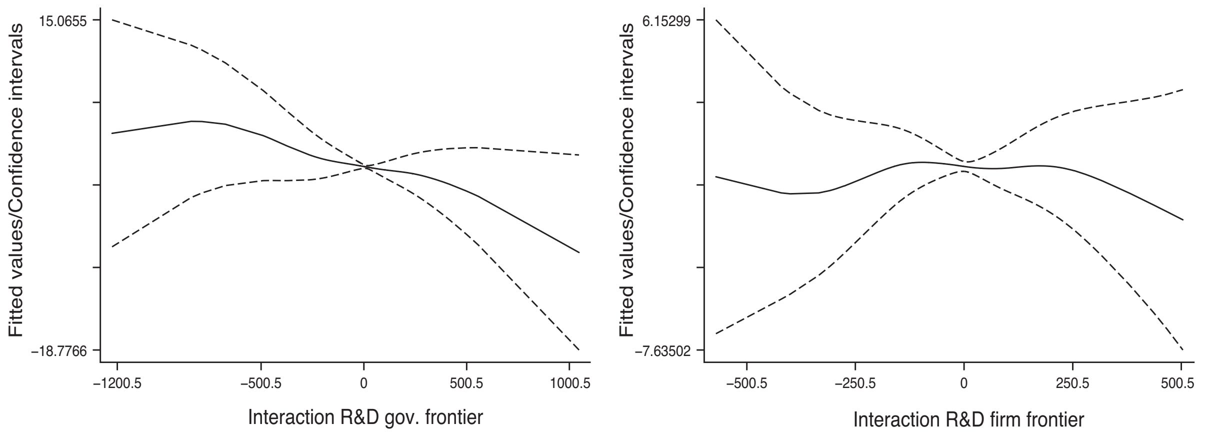

Fig. 11. Nonparametric estimation of the reduced form function $f(x)$. The solid curve represents the estimate $\hat{f}(x)$, the dashed curves correspond to the 95% pointwise bootstrap confidence intervals. [Left]: Relation between TFP growth and the interaction between proximity to World TFP growth and share of R&D expenditure funded by government. [Right]: Relation between TFP growth and the interaction between proximity to world TFP growth and the share of R&D expenditure funded by business sector.

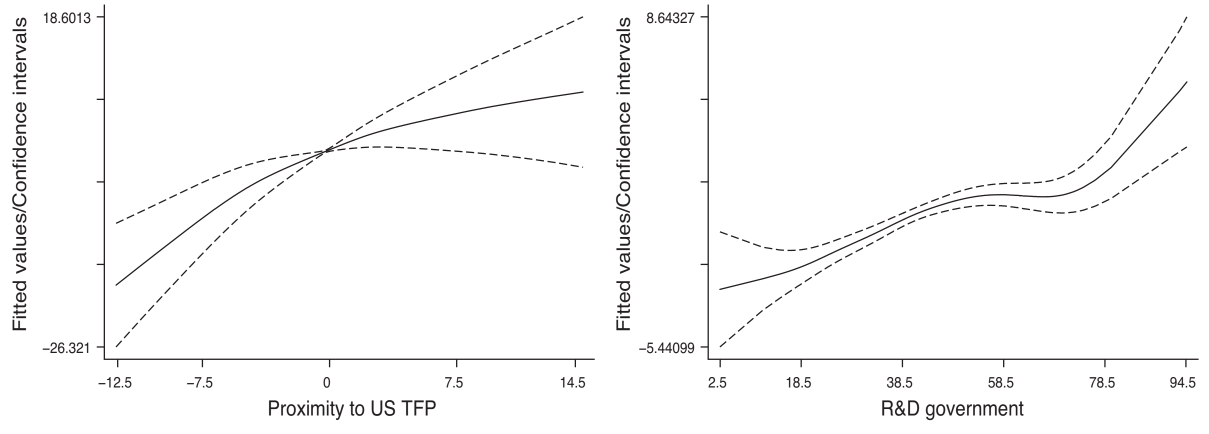

Fig. 12. Nonparametric estimation of the reduced form function $f(x)$. The solid curve represents the estimate $\hat{f}(x)$, the dashed curves correspond to the 95% pointwise bootstrap confidence intervals. [Left]: Relation between TFP growth and proximity to the TFP growth of USA. [Right]: Relation between TFP growth and the share of R&D expenditure funded by government.

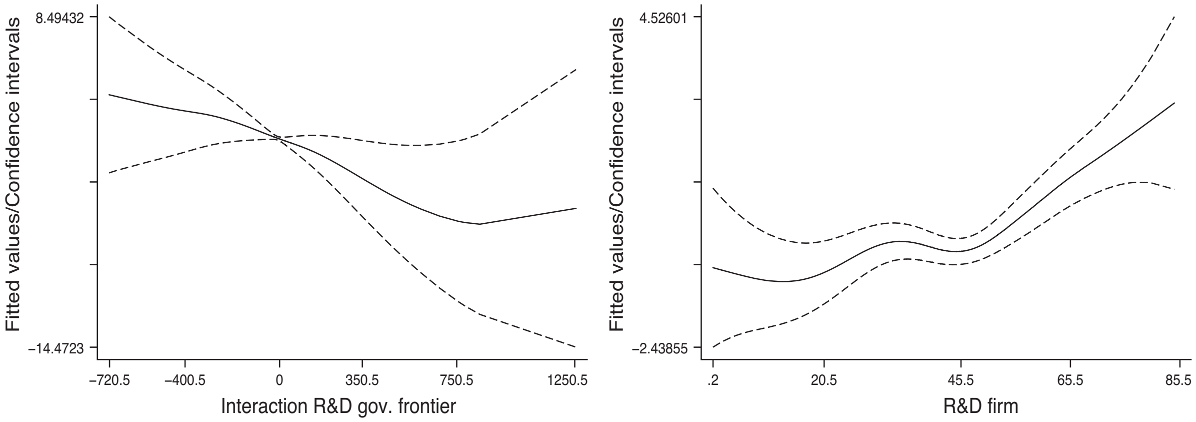

Fig. 13. Nonparametric estimation of the reduced form function $f(x)$. The solid curve represents the estimate $\hat{f}(x)$, the dashed curves correspond to the 95% pointwise bootstrap confidence intervals. [Left]: Relation between TFP growth and the interaction between proximity to the TFP growth of USA and the share of R&D expenditure funded by government. [Right]: Relation between TFP growth and the share of R&D expenditure funded by business sector.

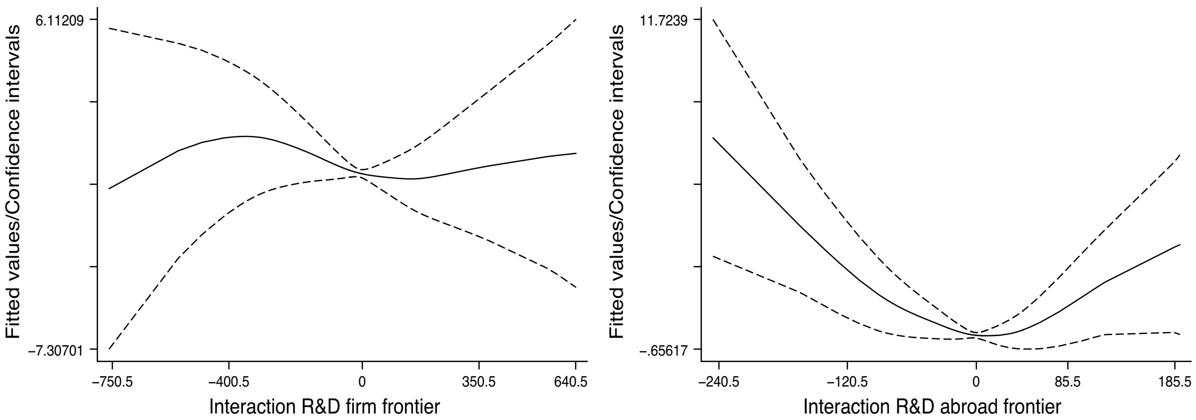

Fig. 14. Nonparametric estimation of the reduced form function $f(x)$. The solid curve represents the estimate $\hat{f}(x)$, the dashed curves correspond to the 95% pointwise bootstrap confidence intervals. [Left]: Relation between TFP growth and the interaction between proximity to the TFP growth of USA and the share of R&D expenditure funded by business sector. [Right]: Relation between TFP growth and the interaction between proximity to the TFP growth of USA and the share of R&D expenditure funded from abroad.

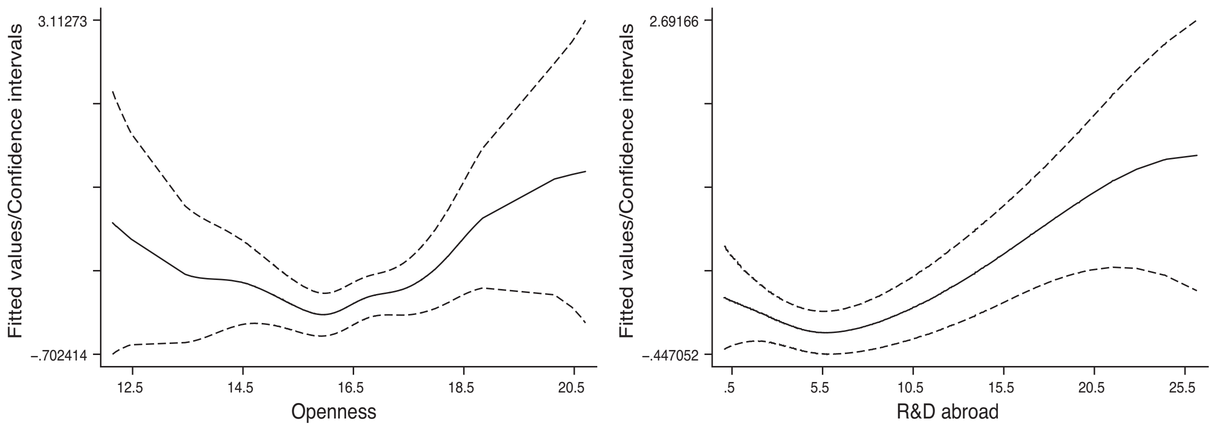

Fig. 15. Nonparametric estimation of the reduced form function $f(x)$. The solid curve represents the estimate $\hat{f}(x)$, the dashed curves correspond to the 95% pointwise bootstrap confidence intervals. [Left]: Relation between TFP growth and openness. [Right]: Relation between TFP growth and the share of R&D expenditure funded from abroad.

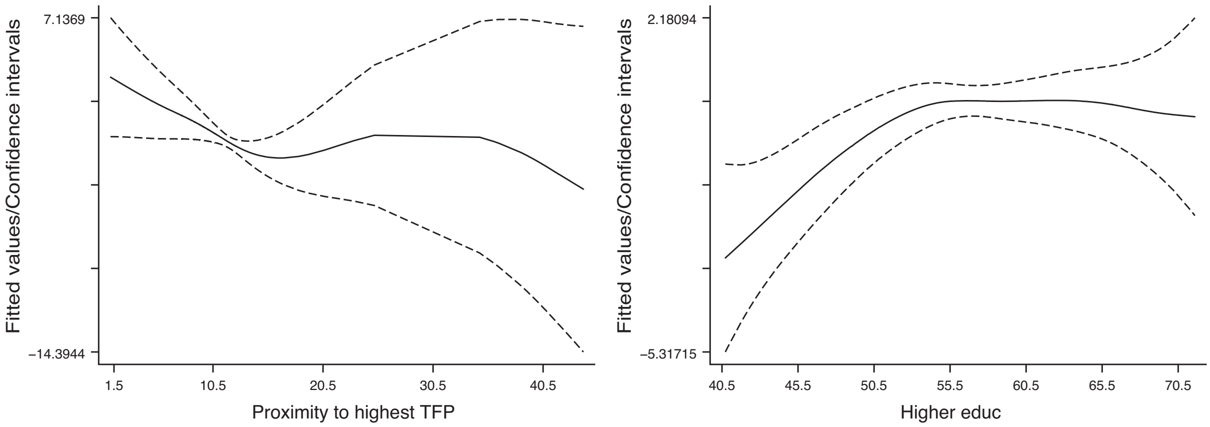

Fig. 16. Nonparametric estimation of the structural function $m(x)$. The solid curve represents the estimate $\hat{m}(x)$, the dashed curves correspond to the 95% pointwise bootstrap confidence intervals. [Left]: Relation between TFP growth and proximity to the highest TFP growth. [Right]: Relation between TFP growth and percentage of graduate students in higher education.

left). TFP growth is an increasing function of R&D expenditure funded by government and this relation is robust to the choice of proximity measures (Fig. 8, right; Fig. 10, right; Fig. 12, right). A similar patterns is observed for R&D expenditure funded by business sector when the frontier is the TFP growth of the USA (Fig. 13, right). These results are consistent with the findings of Griffith et al. (2004). However, we observe a U-shape relation for R&D expenditure funded from abroad when the frontier is the TFP growth of the USA (Fig. 15, right). The curve is decreasing down to 5.5% of the total of R&D expenditure coming from abroad. Contrarily the previous cases where openness does not have significant effect on TFP growth in the semi-parametric model, we observe now a U-shape relation (Fig. 15, left) with the world frontier. This means that the gain from increasing trade will impact positively TFP growth only starting from certain which is here approximately 16.5 point of openness. When we consider parametric estimations (Tables 6 and 7), the effect of openness highly positive and significant only in the FD estimates.

Estimates of the reduced form as described above enable us to identify a number of complex nonlinear relationships. Using the first difference estimator based on the more flexible assumption of extended *predeterminedness* or weak exogeneity of the explanatory variables, we were able to account for possible feedback as above mentioned. We used this estimator in both the parametric model as in the semi-parametric model. Remind that the parametric within estimator is based on the strong assumption of *predeterminedness* or strict exogeneity. In the next section, we propose estimation of the structural form.

5.2. Structural form

Relying on the triangular model of Newey et al. (1999), we estimated the system described in the structural relations (16) and 15. An important step in the implementation lies in the choice of instruments. In addition to the instruments suggested by Vandenbussche et al. (2006), we have included here the human capital (percentage of graduate students in higher education) and its interaction with proximity to the frontier under use, R&D expenditure (government, business sector and abroad) and their interactions with the proximity to the frontier under use. The explanatory power of our instruments are assessed by the exceptional degree of significance of nonlinearity of individual functions from the first step estimates (see Table 12, individual 'gain' statistics,) and also by the very low level of deviances and dispersions. The estimation results of the structural models are reported in Tables 8–10 and in Figs. 16–20. Below, we emphasize salient features in relation to the findings in the reduced form estimates.

5.2.1. Higher education

The results of the model with the percentage of graduate students in higher education are given in Table 8 and Figs. 16–18. Compared with estimates from the reduced form equation, there is now a slight difference in the second portion of the curve of the relationship between TFP

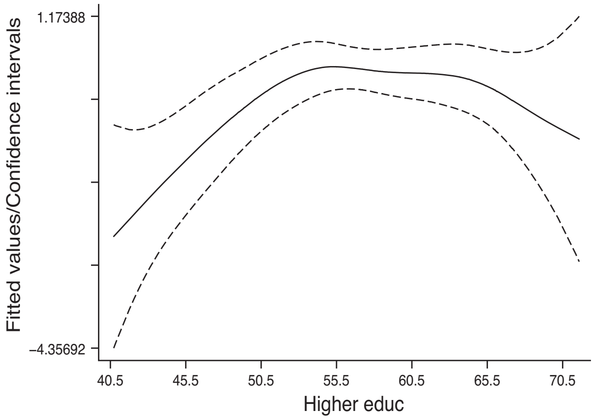

Fig. 17. Nonparametric estimation of the structural function $m(x)$. The solid curve represents the estimate $\hat{m}(x)$, the dashed curves correspond to the 95% pointwise bootstrap confidence intervals. Relation between TFP growth and the percentage of graduate students in higher education when proximity to the world TFP growth is used.

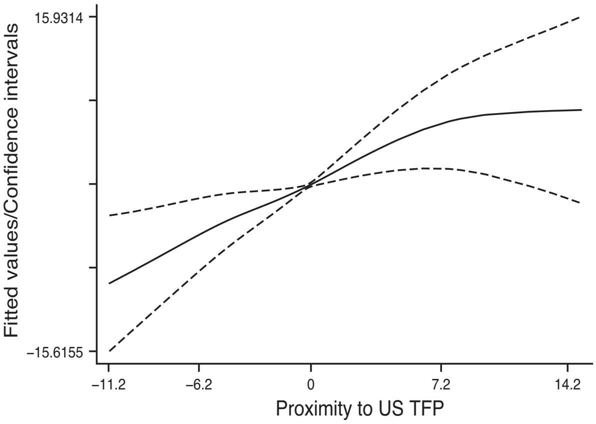

Fig. 18. Nonparametric estimation of the structural function $m(x)$. The solid curve represents the estimate $\hat{m}(x)$, the dashed curves correspond to the 95% pointwise bootstrap confidence intervals. Relation between TFP growth and proximity to the TFP growth of USA.

growth and proximity to the highest TFP growth (Fig. 16, left). The second part of the curve tends now to grow, but given the wide band confidence interval, this part of the curve is not significant. Two other developments appear. The first is that in estimating the reduced form, a significant nonlinearity does not show up in the relationship between TFP growth and percentage of graduate students in higher education. The structural form gives us a significant inverted U-shape relation (Fig. 16, right). The second is the relationship between TFP growth and the percentage of graduate students in higher education when using the proximity to the world TFP growth. The structural form confirms the inverse U-curve obtained in the reduced form analysis (Fig. 17), but here the structural specification detects clearly the range of significancy of the percentage of graduate students in higher education upon which the estimation is run.

5.2.2. *Total staff in R&D*

The results of this estimation are given in Table 9 and Fig. 19. First, it confirms the shape of the relationship between TFP growth and the proximity to the highest TFP growth (Fig. 19, top), and also with respect to the proximity to the TFP growth of the USA (Fig. 19, bottom-right). By cons, we get now a clear U-shape relation between TFP growth and the proximity to the world TFP growth (Fig. 19, bottom-left). Remind that this relation is increasing in the case of the reduced form estimation. This finding seems to

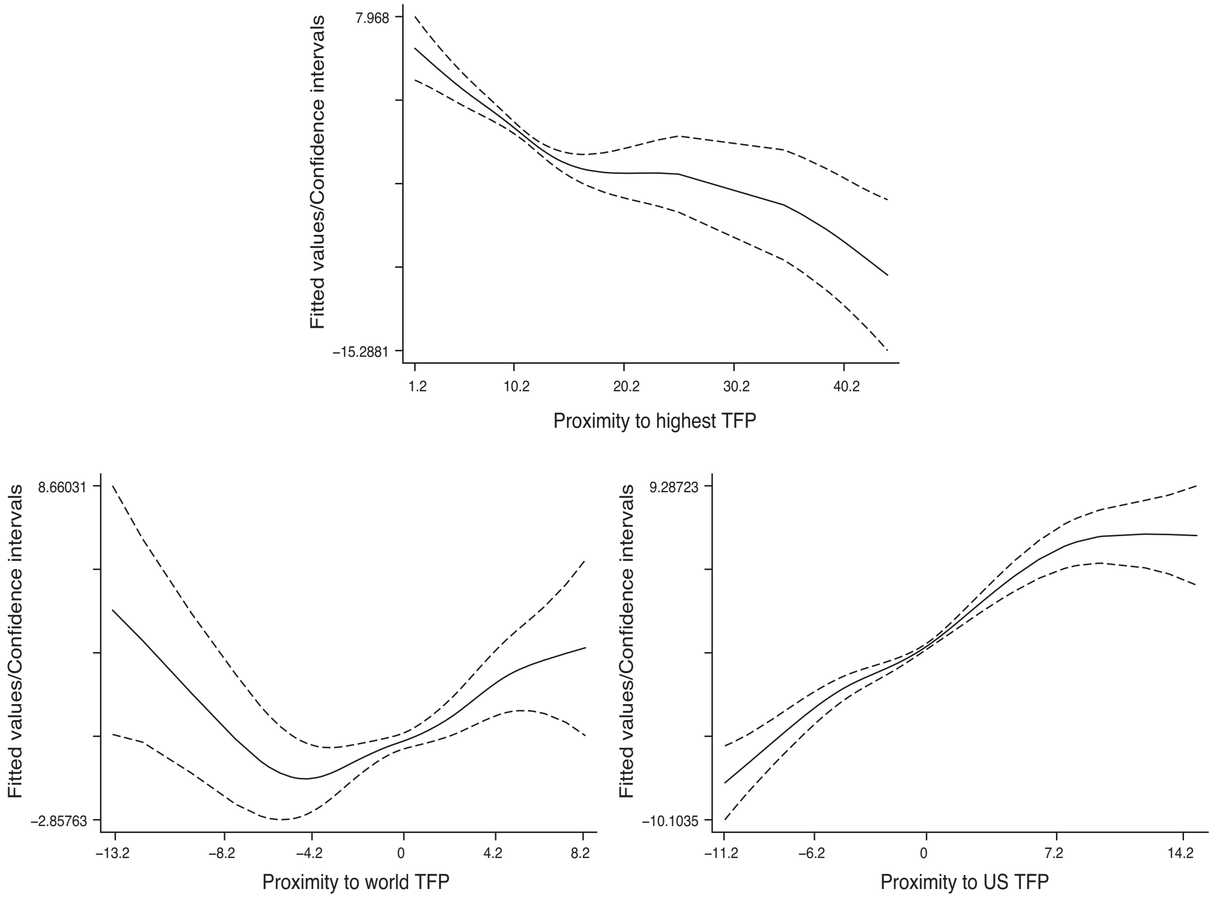

Fig. 19. Nonparametric estimation of the structural function $m(x)$. The solid curve represents the estimate $\hat{m}(x)$, the dashed curves correspond to the 95% pointwise bootstrap confidence intervals. [Top]: Relation between TFP growth and proximity to the highest TFP growth. [Bottom-left]: Relation between TFP growth and proximity to the world TFP growth. [Bottom-right]: Relation between TFP growth and proximity to the TFP growth of USA.

Table 8
GAM semi-parametric estimates (structural form): percentage of graduate students in higher education.

Variable	Dof.[a]	Lin. Coef.	Std. Err.	$p >$ gain[b]
Proximity: $\mathcal{F}_{\max}$				
Prox. highest TFP	4.996	−0.112	0.166	0.000
Higher educ.	5.003	0.079	0.042	0.012
Interaction prox. educ.	5.004	−0.002	0.002	0.863
FDI	5.000	−0.008	0.008	0.379
Openness	4.995	0.105	0.130	0.757
Residual step 1[d]	5.004	0.065	0.052	0.017
Year 2000	1	3.891^{***}	0.728	
Year 2001	1	1.361^{**}		
Year 2002	1	5.595^{***}	0.803	
Year 2003	1	4.242^{***}	0.683	
Year 2004	1	7.247^{***}	1.956	
Year 2005	1	2.463^{***}	0.636	
Year 2006	1	4.429^{***}	0.659	
Year 2007	1	1.339^{**}	0.640	
Deviance		1429.26		
Dispersion		5.627		
Nonlinearity χ_2[c]	24.002			0.000
# observations		296		
Proximity: $\mathcal{F}_{\mathrm{World}}$				
Prox. World TFP	4.994	−0.251	0.477	0.009
Higher educ.	5.003	0.042	0.028	0.017
Interaction prox. educ.	4.996	0.007	0.007	0.171
FDI	5.000	−0.006	0.009	0.382
Openness	4.996	0.099	0.130	0.679
Residual step 1[d]	4.995	0.211	0.112	0.200
Year 2000	1	2.173^{***}	0.722	
Year 2001	1	0.661	0.674	
Year 2002	1	2.128^{***}	0.673	
Year 2003	1	1.878^{***}	0.656	
Year 2004	1	2.651^{***}	0.680	
Year 2005	1	2.335^{***}	0.638	
Year 2006	1	3.098^{***}	0.644	
Year 2007	1	2.286^{***}	0.637	
Deviance		1441.69		
Dispersion		5.675		
Nonlinearity χ_2[c]	23.984			0.007
# observations		296		
Proximity: $\mathcal{F}_{\mathrm{USA}}$				
Prox. TFP USA	4.995	0.740	0.287	0.000
Higher educ.	5.003	0.038	0.021	0.148
Interaction prox. educ.	5.006	−0.001	0.004	0.006
FDI	5.000	0.002	0.007	0.544
Openness	4.995	0.142	0.100	0.732
Residual step 1[d]	5.005	−0.100	0.062	0.000
Year 2000	1	0.554	0.579	
Year 2001	1	1.056^{**}	0.521	
Year 2002	1	0.797	0.536	
Year 2003	1	0.848^{*}	0.516	
Year 2004	1	1.231^{**}	0.535	
Year 2005	1	0.978^{**}	0.497	
Year 2006	1	1.148^{**}	0.515	
Year 2007	1	0.914^{*}	0.502	
Deviance		867.725		
Dispersion		3.416		
Nonlinearity χ_2[c]	24.004			0.000
# observations		296		

[a] Degree of freedom.
[b] Individual gain.
[c] Total gain.
[d] Residual from first step estimation: $\hat{m}(\hat{u}_{it} - \hat{u}_{i,t-1})$.
* Significance level for parametric (linear) components: 10%.
** Significance level for parametric (linear) components: 5%.
*** Significance level for parametric (linear) components: 1%.

Table 9
GAM semi-parametric estimates (structural form): total staff R&D.

Variable	Dof.[a]	Lin. Coef.	Std. Err.	$p >$ gain[b]
Proximity: $\mathcal{F}_{\max}$				
Prox. highest TFP	4.996	−0.309	0.080	0.000
Total staff R&D	4.994	0.147	0.201	0.350
Interaction prox. staff	5.004	−7.56e−08	4.82e−08	0.759
FDI	5.000	−0.005	0.009	0.545
Openness	4.995	−0.056	0.182	0.726
Residual step 1[d]	5.004	0.068	0.051	0.164
Year 2000	1	3.737***	0.736	
Year 2001	1	1.399***	0.678	
Year 2002	1	5.897***	0.821	
Year 2003	1	4.307***	0.699	
Year 2004	1	10.325***	1.979	
Year 2005	1	2.501***	0.649	
Year 2006	1	4.378***	0.673	
Year 2007	1	1.330**	0.650	
Deviance		1482.79		
Dispersion		5.837		
Nonlinearity χ_2[c]	23.991			0.000
# observations		296		
Proximity: $\mathcal{F}_{\text{World}}$				
Prox. World TFP	4.994	0.224	0.113	0.000
Total staff R&D	4.994	−0.055	0.174	0.742
Interaction prox. staff	5.000	−2.29e−07	2.74e−07	0.691
FDI	5.000	−0.006	0.009	0.607
Openness	4.996	0.079	0.185	0.638
Residual step 1[d]	4.995	0.161	0.099	0.004
Year 2000	1	2.082***	0.716	
Year 2001	1	0.628	0.680	
Year 2002	1	2.220***	0.684	
Year 2003	1	1.867***	0.666	
Year 2004	1	2.579***	0.689	
Year 2005	1	2.211***	0.653	
Year 2006	1	2.947***	0.658	
Year 2007	1	2.215***	0.648	
Deviance		1488.96		
Dispersion		5.861		
Nonlinearity χ_2[c]	23.979			0.000
# observations		296		
Proximity: $\mathcal{F}_{\text{USA}}$				
Prox. TFP USA	4.995	0.635	0.063	0.000
Total staff R&D	4.993	−0.446	0.130	0.167
Interaction prox. staff	4.997	−5.84e−08	1.96e−07	0.427
FDI	5.000	−0.0003	0.006	0.565
Openness	4.995	0.340	0.137	0.212
Residual step 1[d]	5.005	−0.106	0.051	0.000
Year 2000	1	0.517	0.569	
Year 2001	1	0.899*	0.514	
Year 2002	1	0.675	0.533	
Year 2003	1	0.562	0.511	
Year 2004	1	1.008*	0.529	
Year 2005	1	0.798	0.496	
Year 2006	1	1.089**	0.511	
Year 2007	1	0.808	0.499	
Deviance		855.788		
Dispersion		3.369		
Nonlinearity χ_2[c]	23.985			0.000
# observations		296		

[a] Degree of freedom.
[b] Individual gain.
[c] Total gain.
[d] Residual from first step estimation: $\hat{m}(\hat{u}_{it} - \hat{u}_{i,t-1})$.
* Significance level for parametric (linear) components: 10%.
** Significance level for parametric (linear) components: 5%.
*** Significance level for parametric (linear) components: 1%.

Table 10
GAM semi-parametric estimates (structural form): R&D expenditure.

Variable	Dof.[a]	Lin. Coef.	Std. Err.	p > gain[b]
Proximity: $\mathcal{F}_{\max}$				
Prox. highest TFP	4.996	−0.368	0.947	0.000
R&D gov.	5.005	0.059	0.103	0.000
Interaction prox. R&D gov.	5.005	0.0008	0.009	0.147
R&D firm	5.006	0.076	0.102	0.008
Interaction prox. R&D firm	5.003	0.0007	0.009	0.599
R&D abroad	5.004	0.224	0.108	0.659
Interaction prox. R&D abroad	4.994	−0.005	0.008	0.577
FDI	5.000	−0.013	0.009	0.464
Openness	4.995	0.112	0.121	0.397
Residual step 1[d]	5.004	0.126	0.080	0.135
Year 2000	1	3.929***	0.732	
Year 2001	1	1.477**	0.671	
Year 2002	1	5.875***	0.822	
Year 2003	1	4.197***	0.692	
Year 2004	1	8.854***	1.994	
Year 2005	1	2.501***	0.641	
Year 2006	1	4.333***	0.667	
Year 2007	1	1.356**	0.645	
Deviance		1338.6		
Dispersion		5.720		
Nonlinearity χ_2[c]	40.012			0.000
# observations		296		
Proximity: $\mathcal{F}_{\text{World}}$				
Prox. World TFP	4.994	−7.154	1.709	0.010
R&D gov.	5.005	0.058	0.042	0.001
Interaction prox. R&D gov.	4.995	0.074	0.017	0.000
R&D firm	5.003	0.087	0.042	0.053
Interaction prox. R&D firm	4.995	0.072	0.017	0.625
R&D abroad	5.003	0.156	0.053	0.674
Interaction prox. R&D abroad	5.007	0.077	0.019	0.888
FDI	5.000	−0.009	0.009	0.570
Openness	4.996	0.151	0.120	0.559
Residual step 1[d]	4.995	0.433	0.135	0.000
Year 2000	1	2.522***	0.701	
Year 2001	1	0.927	0.661	
Year 2002	1	2.227***	0.659	
Year 2003	1	2.015***	0.645	
Year 2004	1	2.641***	0.669	
Year 2005	1	2.338***	0.631	
Year 2006	1	2.935***	0.635	
Year 2007	1	2.228***	0.626	
Deviance		1281.96		
Dispersion		5.478		
Nonlinearity χ_2[c]	39.993			0.000
# observations		296		
Proximity: $\mathcal{F}_{\text{USA}}$				
Prox. TFP USA	4.995	1.883	1.153	0.025
R&D gov.	5.005	0.090	0.029	0.015
Interaction prox. R&D gov.	5.002	−0.012	0.011	0.000
R&D firm	5.003	0.082	0.028	0.023
Interaction prox. R&D firm	5.006	−0.012	0.011	0.309
R&D abroad	5.004	0.080	0.036	0.472
Interaction prox. R&D abroad	5.006	−0.015	0.013	0.002
FDI	5.000	0.004	0.007	0.706
Openness	4.995	0.076	0.092	0.489
Residual step 1[d]	5.005	−0.104	0.075	0.000
Year 2000	1	0.524	0.577	
Year 2001	1	0.933*	0.514	
Year 2002	1	0.802	0.535	
Year 2003	1	0.722	0.509	
Year 2004	1	1.117**	0.531	
Year 2005	1	0.961*	0.492	
Year 2006	1	1.006**	0.512	
Year 2007	1	0.831*	0.497	
Deviance		773.981		
Dispersion		3.307		
Nonlinearity χ_2[c]	40.021			0.000
# observations		296		

[a] Degree of freedom.
[b] Individual gain.
[c] Total gain.
[d] Residual from first step estimation: $\hat{m}(\hat{u}_{it} - \hat{u}_{i,t-1})$.
* Significance level for parametric (linear) components: 10%.
** Significance level for parametric (linear) components: 5%.
*** Significance level for parametric (linear) components: 1%.

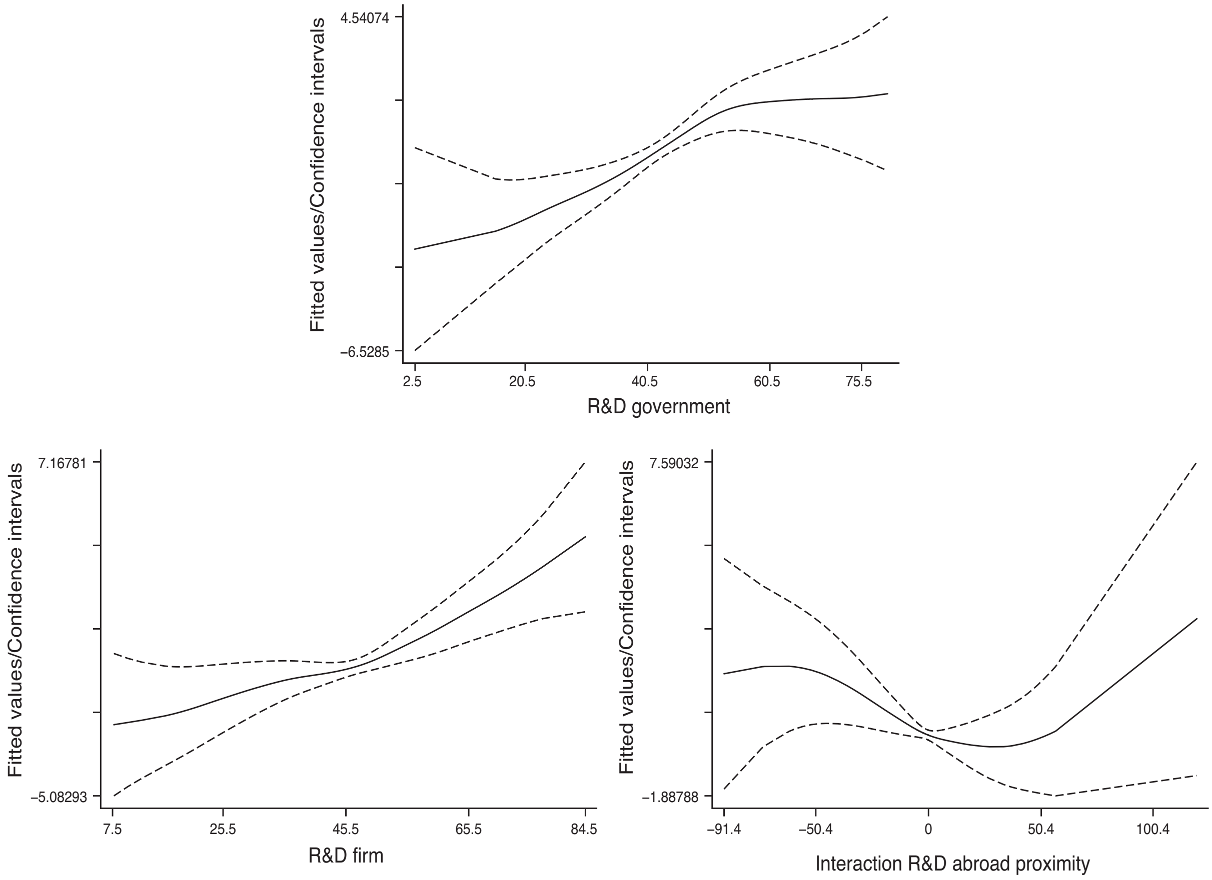

Fig. 20. Nonparametric estimation of the structural function $m(x)$. The solid curve represents the estimate $\hat{m}(x)$, the dashed curves correspond to the 95% pointwise bootstrap confidence intervals. [Top]: Relation between TFP growth and the share of R&D expenditure funded by government. [Bottom-left]: Relation between TFP growth and the share of R&D expenditure funded by business sector [Bottom-right]: Relation between TFP growth and the interaction between proximity to the TFP growth of USA and the share of R&D expenditure funded from abroad.

reconcile the relations from two other proximity measures (decreasing for the proximity to the highest TFP growth, and increasing for the proximity to the TFP growth of the USA).

5.2.3. R&D expenditure

The results of this estimation are given in Table 10 and Fig. 20. The most appealing finding the relationship between TFP growth and the share of R&D spending financed by government when the measure of proximity is regarded with respect to the TFP growth of the USA. In the estimates of the reduced form, we had obtained an overall increasing relationship. Here (Fig. 20, top), there is a nice S-shape nonlinearity or convex-concave curvature. Indeed, we see that TFP growth increases with respect to the share of government expenses in R&D but this increase is stronger for weak values of R&D expenditure. In other words, an increase in government spending in R&D has a greater impact on TFP growth when the former is weak, and a smaller impact when the expenditure is already high compared to the frontier defined by the USA.

6. Conclusion

The empirical study of the determinants of productivity has gone through many contributions. The literature establishes a consensus on the crucial roles of R&D not only as stimulus of innovation, but also as facilitator of imitation. The latter requires a transfer of technology from firms or holding country to imitators, hence the importance of technological gap between firms or countries, meaning the distance to the frontier. For the transfer is beneficial and become an engine of growth, one needs absorptive capacity. This highlights the importance of human capital, but also that of openness of a country as well as investment flows. The empirical study of these determinants was made until now in a parametric framework of reduced form models. Such a framework has the merit to clarify some mechanisms (technological externalities, diminishing returns to R&D, complementarity between the proximity to the frontier and the stock of human capital and the ability of the latter to be growth-enhancing) that aim to drive economic policy. The issues of endogeneity and measurement errors in these specifications have been also taken into account to correct for estimation bias. However, the parametric framework does not identify more complex mechanisms that may lead to nonlinearities in productivity growth.

The present study fills this gap by revisiting the effect of the determinants of productivity growth in a nonparametric paradigm. We considered not only reduced form estimation, but also structural forms thereby emphasizing the issues of endogeneity and simultaneity. We also studied the sensitivity of results to different choice of proximity to frontier. Our estimates show that there are strong nonlinearities in productivity growth, which also reflects important heterogeneity effects. Three key results different from those that have been found so far in the literature can be delivered: i) there is a U-shaped relationship between TFP growth and proximity to the frontier. Previous contributions based on parametric models find either a positive or a negative effect. ii) The growth rate of productivity increases with human capital up to a certain level and then decreases, describing an inverted U-shaped relationship. iii) The share of government spending on R&D has a positive effect on TFP growth. However, this effect is stronger for low values of governmental spending. The effect is lower when government spending is already high. This reflects two phases of the impact of government R&D spending on productivity growth. The first phase corresponds to a strong a positive effect whereas the second phase may be viewed as a decreasing efficiency of government spending (once a certain threshold) to boost productivity growth. In terms of economic policy, this suggests a diversification of sources of R&D funding starting from certain level of government spending. One can also think that from a certain level, government spending must be accompanied by other measures to make government spending more effective.

Some avenues for future research emerge from these results. The first concerns the development of a theoretical model that would provide an analytical framework for empirical results obtained from the structural model. A second track is to include analysis of panel data co-integration within our nonparametric framework with special attention to the issue of convergence. Both avenues are very promising.

Acknowledgements

This study has immensely benefited from the comments of two anonymous referees and the editor of the Journal. We thank Raouf Boucekkine, Tapas Mishra and Bertrand Koebel for helpful comments on an earlier draft of this paper. The usual disclaimer applies.

Appendix A. Additional tables

Table 11
Definition of variables.

Variable name	Definition	Nature	Source[a]
TFP growth	Total factor productivity growth (estimated as Tornqvist index)	Continuous	1
Proximity to highest TFP growth ($\mathcal{F}_{\max}$)	Distance to the economy with the highest TFP growth	Continuous	1
Proximity to the world TFP growth ($\mathcal{F}_{\text{World}}$)	Distance to the sample mean TFP growth	Continuous	1
Proximity to the TFP growth of USA ($\mathcal{F}_{\text{USA}}$)	Distance to the TFP growth of USA	Continuous	1
Higher educ.	Percentage of graduate students in higher education	Continuous	2
Total staff R&D	Number of people engaged in R&D activities, full-time equivalent	Continuous	2
R&D gov.	Share of R&D expenditure funded by government	Continuous	2
R&D firm	Share of R&D expenditure funded by business sector	Continuous	2
R&D abroad	Share of R&D expenditure funded from abroad	Continuous	2
FDI	Foreign Direct Investment (in % of GDP)	Continuous	3
Openness	(import of goods and services + export of goods and services)/GDP	Continuous	3
Interaction R&D frontier	Interaction between a source of R&D expenditure and a measure of frontier	Continuous	
Interaction higher educ frontier	Interaction between higher education and a measure of frontier	Continuous	
Interaction staff R&D frontier	Interaction between total staff R&D and a measure of frontier	Continuous	
Year	Dummy variable of year	Dummy (yes = 1)	
OECD	OECD country	Dummy (yes = 1)	

[a] 1: The Conference Board Total Economy Database (2010); 2: UNESCO Statistical Yearbook (2010); 3: WDI (2010).

Table 12
First step estimates of the structural model.

Variable	Dof.[a]	Lin. Coef.	Std. Err.	$p>$ gain[b]
Proximity: $\mathcal{F}_{\max}$				
Lag Prox. highest TFP	5.003	−0.0004	0.001	0.775
Higher educ.	4.998	−0.048	0.004	0.044
Interaction prox. educ.	5.003	0.003	0.0002	0.000
Total staff R&D	4.994	−0.078	0.013	0.000
Interaction prox. total staff R&D	5.006	5.19e−08	5.13e−09	0.000
R&D gov.	5.004	−0.067	0.003	0.000
Interaction prox. R&D gov.	4.989	0.008	0.0001	0.000
R&D firm	4.998	−0.078	0.003	0.000
Interaction prox. R&D firm	5.004	0.008	0.0001	0.000
R&D abroad	4.995	−0.044	0.006	0.404
Interaction prox. R&D abroad	4.996	0.006	0.0003	0.000
Deviance		16.581		
Dispersion		0.065		
Nonlinearity χ_2[c]	43.990			0.000
# observations		309		
Proximity: $\mathcal{F}_{\text{World}}$				
Lag Prox. World TFP	4.995	−0.0005	0.003	0.049
Higher educ.	4.999	0.0009	0.001	0.426
Interaction prox. educ.	4.994	0.006	0.0003	0.000
Total staff R&D	5.006	−0.002	0.006	0.000
Interaction prox. total staff R&D	5.000	3.21e−08	1.61e−08	0.698
R&D gov.	5.004	0.003	0.001	0.000
Interaction prox. R&D gov.	4.995	0.006	0.0002	0.000
R&D firm	4.998	0.003	0.001	0.000
Interaction prox. R&D firm	4.993	0.007	0.0002	0.000
R&D abroad	4.995	0.001	0.002	0.354
Interaction prox. R&D abroad	4.999	0.003	0.0007	0.033
Deviance		4.602		
Dispersion		0.018		
Nonlinearity χ_2[c]	43.978			0.000
# observations		309		
Proximity: $\mathcal{F}_{\text{USA}}$				
Lag Prox. TFP USA	5.005	−0.004	0.004	0.118
Higher educ.	4.998	0.004	0.002	0.294
Interaction prox. educ.	5.001	0.005	0.0004	0.000
Total staff R&D	5.005	0.012	0.009	0.000
Interaction prox. total staff R&D	4.997	1.18e−08	2.28e−08	0.825
R&D gov.	5.004	0.001	0.002	0.000
Interaction prox. R&D gov.	4.995	0.007	0.0002	0.000
R&D firm	4.998	0.001	0.002	0.000
Interaction prox. R&D firm	4.994	0.008	0.0002	0.000
R&D abroad	4.995	0.003	0.003	0.233
Interaction prox. R&D abroad	5.001	0.004	0.0007	0.067
Deviance		10.890		
Dispersion		0.043		
Nonlinearity χ_2[c]	43.993			0.000
# observations		309		

[a] Degree of freedom.
[b] Individual gain.
[c] Total gain.

Appendix B. Estimation procedure and specification test ("gain")

The GAM specification considered can be rewritten in compact form:

$$Y = \alpha + \sum_{j=1}^{p} f_j(X_j) + \mathbf{Z}'\boldsymbol{\gamma} + \epsilon \tag{B.1}$$

The f_j are unknown univariate functions to be estimated such that $\mathbb{E}[f_j(Xj)] = 0$. The estimation of this model might be implemented by the following steps.

Step 1: Center the data.

Step 2: Regress the residuals $\hat{\epsilon}$ on X_j, $j = 1, \ldots, p$ by using the backfitting algorithm (see below). The resulting smooth is the first estimate of $f_j(X_j)$, $\hat{f}_j(X_j)$.

Step 3: Obtain the estimate of γ by ordinary least squares

$$\hat{\gamma} = \mathbb{E}\left(Y - \hat{\alpha} - \sum_{j=1}^{p} \hat{f}_j(X_j)|\mathbf{Z}\right) \tag{B.2}$$

where as $\hat{\alpha} = \frac{1}{n}\sum_i^n Y_i$.

Step 4: Center the data again, and continue the process until convergence.

Backfitting algorithm

(a) Initialize $\hat{\alpha}=\frac{1}{n}\sum_i^n Y_i, f_j(X_j)=f_j^0(X_j)$, $j=1,\ldots,p$

(b) Cycle: $j=1,\ldots,p,1,\ldots,p\ldots$

$$\hat{f}_j(X_j)=S_j\left(Y-\hat{\alpha}-\sum_{k\neq j}\hat{f}_j(X_k)|X_j\right) \tag{B.3}$$

where S_j is the smoother, using k nearest symmetric neighborhood for f_j^0, and $\hat{f}_j$ is the nonparametric estimator of f_j.[13]

(c) Continue (b) until the individual functions do not change.

The degree of freedom df_j of the fit $\hat{f}_j$ might be approximated by the trace of $2\mathbf{S}_j-\mathbf{S}_j\mathbf{S}_j'$ where $\mathbf{S}_j$, is the smoothing matrix so that $\hat{\mathbf{f}}=\mathbf{S}_j\mathbf{w}$ ($\hat{\mathbf{f}}$ is the vector of $\hat{f}_j$ and w is the vector corresponding to $Y-\hat{\alpha}-\sum_{k\neq j}\hat{f}_j(X_k)$. In the case of linear estimator, we have $\mathbf{S}_j=\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, where $\mathbf{X}$ is the matrix of regressors $df_j=1$.

To compare two individual smooths $\hat{\mathbf{f}}_j^1=\mathbf{S}_{j,1}w$ and $\hat{\mathbf{f}}_j^2=\mathbf{S}_{j,2}\mathbf{w}$, we can use the approximative statistic

$$J=\frac{(RSS_1-RSS_2)/(df_2-df_1)}{RSS_2/(n-df_2)}\sim F_{df_2-df_1,n-df_2} \tag{B.4}$$

where RSS_1 and RSS_2 are respectively the deviance (or the residual sum of squares) of models corresponding to $\hat{f}_{j,1}$ and $\hat{f}_{j,2}$. The distribution of the statistic "gain" $J\times(df_2-df_1)$ is approximated by $\chi^2(df_2-df_1)$. Intuitively, the "gain" is the difference in normalized deviance between the GAM and a model with a linear term for the corresponding regressor. A large gain indicates a lot of nonlinearity, at least as regards statistical significance. The associated p-value is based on a chi-square approximation to the distribution of the gain if the true marginal relationship between that regressor the response variable was linear. Finally, it should be noticed that the df of the "gain" statistic may be fractional.

Appendix C. The wild bootstrap

Several *bootstrap* methods are available (see e.g., Horowitz, 2001). To construct the confidence bands for nonparametric estimators as well as the critical values of the nonparametric tests, we use the *wild bootstrap* as now described. Let us consider the univariate nonparametric regression model

$$y=f(x)+\epsilon, \tag{C.1}$$

where $f(x)$ represents a unknown function of x, whose nonparametric estimator is denoted $\hat{f}(x,h)$, h being the smoothing parameter. Let us denote by $\hat{\epsilon}=y-\hat{f}(x,h)$ the regression residuals. The different steps of the *wild bootstrap* algorithm are the following:

[13] Here, we use the local linear kernel estimator. This estimator is not adversely affected by the boundary of the data sample. Moreover, as proved by Fan (1992), it is the best linear smoother in the sense that it is the asymptotic minimax linear smoother when the unknown regression function is in the class of functions having bounded second derivative.

$s=1$
Repeat
Step 1: Generate the bootstrap errors ϵ^* using the two points distribution probability: $P(\epsilon^*=\hat{\epsilon}\lambda)=\delta$; $P(\epsilon^*=\hat{\epsilon}\mu)=1-\delta$, with $\lambda=(1-\sqrt{5})/2$, $\mu=(1+\sqrt{5})/2$, $\delta=(5+\sqrt{5})/10$.
Step 2: Generate new bootstrap samples $y^*=\hat{f}(x,h_b)+\epsilon^*$, where h_b is the bandwidth slightly greater than h. Then, $\hat{f}(x,h_b)$ is slightly over-smoothed compared to $\hat{f}(x,h)$. Compute $\hat{f}^*(x,h)$, that is the nonparametric estimator applied to the bootstrap sample $\{y^*;x\}$.
$s=s+1$
Until $s=B$ (number of bootstrap samples, here we set $B=1000$).

In order to compute the pointwise bootstrap confidence interval of level $(100-\alpha)$ for $\hat{f}(x,h)$, we define the lower and upper bounds as the $(\alpha/2)$th and $(100-\alpha/2)$ percentiles of the distribution of the bootstrap estimators $\hat{f}^*(x,h)$, respectively.

Remark 1. The *wild bootstrap* yields estimations which account for heteroskedasticity and correlation between observations. This can be easily observed from the resulting covariance structure. Indeed, let $\hat{u}_n$ denote a random variable, and u_n^* the associate bootstrap sample, where u_n^* has realization probabilities p and $1-p$ corresponding to $\beta\hat{u}_n$ and $\gamma\hat{u}_n$, respectively. Then, we can write, from the covariance decomposition,

$$cov(u_i^*,u_j^*)=\mathbb{E}[cov(u_i^*,u_j^*)\mid\hat{u}_i,\hat{u}_j] + cov[\mathbb{E}(u_i^*\mid\hat{u}_i,\hat{u}_j),\mathbb{E}(u_j^*\mid\hat{u}_i,\hat{u}_j)]. \tag{C.2}$$

Since $\mathbb{E}[cov(u_i^*,u_i^*)\mid\hat{u}_i,\hat{u}_j]=0$; and $\mathbb{E}(u_k^*\mid\hat{u}_i,\hat{u}_j)=\hat{u}_k$, $k=i$, j, we obtain $cov(u_i^*,u_j^*)=cov(\hat{u}_i,\hat{u}_j)$.

Remark 2. Another advantage of the bootstrap in constructing confidence intervals is that it avoids the computation of constants such as the bias of the estimator (see Härdle, 1990).

Remark 3. Other types of bootstrap confidence intervals can be used (for example, uniform confidence intervals) but their computation is not trivial.

References

Abromovitz, N., 1986. Catching up, forging ahead, and falling behind. Journal of Economic History 46, 386–406.
Aghion, P., Howitt, P., 2005. Growth with quality-improving innovations: an integrated framework. In: Aghion, P., Durlauf, S.N. (Eds.), Handbook of Economic Growth. Elsevier, Amsterdam, pp. 67–110.
Aghion, P., Howitt, P., Mayer-Foulkes, D., 2005. The effect of financial development on convergence: theory and evidence. Quarterly Journal of Economics 120, 173–222.
Aghion, P., Boustan, L., Hoxby, C., Vandenbussche, J., 2009. The causal impact of education on economic growth: evidence from U.S. Working Paper. Harvard University.
Aghion, P., Cohen, E., 2004. Education et Croissance. Rapport du C.A.E, No 2, Paris, La Documentation française.
Bassanini, A., Scarpetta, S., 2001. Les Moteurs de la Croissance dans les Pays de l'OCDE: Analyse Empirique sur des Données de Panel. Revue Economique de l'OCDE 33.
Benhabib, J., Spiegel, M.M., 1994. The role of human capital in economic development: evidence from aggregate cross-country data. Journal of Monetary Economics 34, 143–173.
Coe, D.T., Helpman, E., Hoffmaister, A.W., 1997. North-South R&D spillovers. Economic Journal 107, 134–149.
Coe, D.T., Helpman, E., Hoffmaister, A.W., 2009. International R&D spillovers and institution. European Economic Review 53, 723–741.
Cohen, W.M., Levinthal, D.A., 1989. Innovation and learning: the two faces of R&D. Economic Journal 99, 569–596.
Cole, H., Ohanian, L., Riascos, A., Schmitz Jr., J., 2004. Latin America in the rearview window. NBER Working Paper 11008.

Deaton, A., Muellbauer, J., 1980. Economics and Consumer Behavior. Cambridge University Press.
Easterly, W., Levine, R., 2001. It's not factor accumulation: stylized facts and growth models. World Bank Economic Review 15 (2), 177–219.
Eaton, J., Kortum, S., 1999. International technology diffusion: theory and measurement. International Economic Review 40, 537–570.
Engelbrecht, H.J., 1997. International R&D spillovers, human capital and productivity in OECD economies: an empirical investigation. European Economic Review 41, 1479–1488.
Fagerberg, J., 1994. Technology and international differences in growth rates. Journal of Economic Literature 32, 1147–1175.
Fan, J., 1992. Design-adaptative nonparametric regression. Journal of the American Statistical Association 87, 998–1004.
Feyrer, J., 2003. Convergence by parts. Mimeo. Department of Economics, Dartmouth College.
Griffith, R., Redding, S., Reenen, J.V., 2003. R&D and absorptive capacity: theory and empirical evidence. Scandinavian Journal of Economics 105, 99–118.
Griffith, R., Redding, S., Reenen, J.v., 2004. Mapping the two faces of R&D: productivity growth in a panel of OECD industries. Review of Economics and Statistics 86, 883–895.
Griffith, R., Redding, S.J., Van Reenen, J., 2000. Mapping the two faces of R&D: productivity growth in a panel of OECD industries. CEPR Discussion Papers 2457.
Guellec, D., Pottelsberghe, B.V., 2001. Recherche, Développement et Croissance de la Productivité: Analyse des Données d'un Panel de 16 Pays de l'OCDE. Revue Economique de l'OCDE 33.
Hall, R., Jones, C., 1999. Why do some countries produce so much more output per worker than others? Quarterly Journal of Economics 114 (1), 83–116.
Hanushek, E.A., Kim, D., 1995. Schooling, labour force quality and economic growth. NBER Working Paper No. 5399.
Hanushek, E.A., Kimko, D., 2000. Schooling, labour force quality and the growth of nations. American Economic Review 90, 1184–1208.
Härdle, W., 1990. Applied Nonparametric Regression. Cambridge University Press, Cambridge, New York.
Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models. Chapman and Hall, London, New York.
Hobday, M., 2003. Innovation in Asian industrialization: a Gerschenkronian perspective. Oxford Development Studies 31, 293–314.
Horowitz, J.L., 2001. The bootstrap. In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics. North Holland, Elsevier Science, Amsterdam (Chapter 52).
Howitt, P., 2000. Endogenous growth and cross-country income differences. American Economic Review 90, 829–846.
Howitt, P., 2005. Health human capital and economic growth: a Schumpeterian perspective. In: Lopez-Casasnovas, G., Rivera, B., Currais, L. (Eds.), Health and Economic Growth: Findings and Policy Implications. The MIT Press, Cambridge.
Hu, A.G.Z., Jefferson, G.H., Jinchang, Q., 2005. R&D and technology transfer: firm-level evidence from Chinese industry. Review of Economics and Statistics 7, 780–786.
Ibragimov, I., Hasminskii, R., 1980. On nonparametric estimation of regression. Soviet Math. Dokl. 21, 810–814.
Keller, W., 2002a. Geographic localization and international technology diffusion. American Economic Review 92, 120–142.
Keller, W., 2002b. Trade and the transmission of technology. Journal of Economic Growth 7, 5–24.
Klenow, P., Rodriguez-Clare, A., 1997. The neoclassical revival in growth economics: has it gone too far? In: Bernanke, B., Rotemberg, J. (Eds.), NBER Macroeconomics Annual. MIT Press, Cambridge/London, pp. 73–103.
Klenow, P., Rodriguez-Clare, A., 2004. Externalities and growth. NBER Working Paper No. 11009.
Kneller, R., 2005. Frontier technology, absorptive capacity and distance. Oxford Bulletin of Economics and Statistics 67, 1–23.
Kneller, R., Stevens, P.A., 2006. Frontier technology and absorptive capacity: evidence from OECD manufacturing industries. Oxford Bulletin of Economics and Statistics 68, 1–21.
Krueger, A., Lindahl, M., 2001. Education for growth: why and for whom? Journal of Economic Literature 39, 1101–1136.
Linton, O., 2000. Efficient estimation of generalized additive nonparametric regression models. Econometric Theory 16, 502–523.
Linton, O., Härdle, W., 1996. Estimating additive regression models with known links. Biometrika 83, 529–540.
Madsen, J.B., 2007. Technology spillover through trade and TFP convergence: 135 years of evidence for the OECD countries. Journal of International Economics 72, 464–480.
Madsen, J.B., 2008a. Economic growth, TFP convergence and the world export of ideas: a century of evidence. Scandinavian Journal of Economics 110, 145–167.
Madsen, J.B., 2008b. Semi-endogenous versus Schumpeterian growth models: testing the knowledge production function using international data. Journal of Economic Growth 13, 1–26.
Madsen, J.B., Saxena, S., Ang, J.B., 2010. The Indian growth miracle and endogenous growth. Journal of Development Economics 93, 37–48.
Mankiw, G., Romer, D., Weil, D., 1992. A contribution to the empirics of economic growth. Quarterly Journal of Economics CVII, 407–437.
Nelson, R.R., Phelps, E.S., 1966. Investment in humans, technological diffusion, and economic growth. American Economic Review 56, 69–75.
Newey, W.K., Powell, J.P., Vella, F., 1999. Nonparametric estimation of triangular simultaneous equations models. Econometrica 67, 565–603.
Parente, S., Prescott, E., 2004. A unified theory of the evolution of international income levels. In: Aghion, P., Durlauf, S. (Eds.), Handbook of Economic Growth. , Mimeo, draft paper.
Romer, P.M., 1990. Endogenous technological change. Journal of Public Economy 98, 71–102.
Savvides, A., Zachariadis, M., 2005. International technology diffusion and the growth of TFP in the manufacturing sector of developing economies. Review of Development Economics 9, 482–501.
Stone, C., 1980. Optimal rates of convergence for nonparametric estimators. The Annals of Statistics 8, 1348–1360.
Stone, C., 1982. Optimal global rates of convergence for nonparametric regression. The Annals of Statistics 8, 1040–1053.
Stone, C., 1985. Additive regression and other nonparametric models. The Annals of Statistics 13, 685–705.
Stone, C., 1986. The dimensionality reduction principle for generalized additive models. The Annals of Statistics 14, 592–606.
The Conference Board Total Economy Database, 2010, June. http://www.conference-board.org/economics/database.cfm
Vandenbussche, J., Aghion, P., Meghir, C., 2006. Growth, distance to frontier and composition of human capital. Journal of Economic Growth 11, 97–127.
Verspagen, B., 1991. A new empirical approach to catching up or falling behind. Structural Change and Economic Dynamics 2, 359–380.
White, H., 1980. A heteroscedasticity consistent covariance matrix estimator and a direct test for heteroscedasticity. Econometrica 48, 817–838.
Wooldridge, J.M., 2006. Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge, MA.
Xu, B., 2000. Multinational enterprises, technology diffusion, and host country productivity growth. Journal of Development Economics 62, 477–493.
Zachariadis, M., 2003. R&D, innovation, and technological progress: a test of the Schumpeterian framework without scale effects. Canadian Journal of Economics 36, 566–586.
Zachariadis, M., 2004. R&D-induced growth in the OECD? Review of Development Economics 8, 423–439.