FACULTY OF LAW, ECONOMICS AND FINANCE

CREA

Discussion Paper

2009-19

Center for Research in Economic Analysis
University of Luxembourg

# Technology frontier, labor productivity and economic growth: Evidence from OECD countries

*available online : http://wwwfr.uni.lu/recherche/fdef/crea/publications2/discussion_papers/2009*


Théophile T. Azomahou, UNU-MERIT, Maastricht University

Bity Diene, University of Luxembourg

Mbaye Diene, Université Cheikh Anta Diop, Dakar


November, 2009

UNIVERSITÉ DU LUXEMBOURG


For editorial correspondence, please contact : elisa.ferreira@uni.lu

University of Luxembourg
Faculty of Law, Economics and Finance
162A, avenue de la Faïencerie
L-1511 Luxembourg

# Technology frontier, labor productivity and economic growth: Evidence from OECD countries*

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

[a] UNU-MERIT, Maastricht University, The Netherlands
[b] CREA, University of Luxembourg, Luxembourg
[c] CRES, Université Cheikh Anta Diop, Dakar, Sénégal

November 2009

**Abstract**

We use 29 OECD countries data spanning over 1960-2000 to study the growth strategy when countries are close to the technology frontier. Relying on a semi-parametric generalized additive model, we estimate labor productivity equations. We find that the number of agents enrolled in higher education is a determinant of growth. Moreover, when a country is sufficiently near the technology frontier thanks to an increasing R&D expenditure, it becomes optimal to invest in fundamental research, since after a short period of efficiency, business R&D can no longer ensure the transition toward the technology frontier, while higher education presents the opposite shape. These findings support the main assertion of Aghion and Cohen (2004) that countries which are near the technology frontier have to invest in higher education while those far away from the frontier make their technology level growing up by investing in primary and secondary schooling.

JEL Classification: I23, J24, O40
Keywords: Education, R&D, Labor Productivity, Economic Growth.

*We thank Raouf Boucekkine, Tapas Mishra and Bertrand Koebel for helpful comments. We only remain responsible for possible errors.

†Corresponding author. Théophile T. Azomahou. UNU-MERIT, Keizer Karelplein 19, 6211 TC Maastricht, The Netherlands. Tel. +31 43 3884440, Fax +31 43 3884499, E-mail: azomahou@merit.unu.edu

# 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, Romer, and Weil (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 (say, university research) generates a higher economic growth level. Therefore, expenditure devoted to higher education becomes very important for the welfare of an economy. 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 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 Pottelsberghue (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 of 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)[1] focus on the increasing importance of higher education when the technology in a country is near to the technology frontier.[2] They deal with the fact that the role of education in growth emphasizes on two mechanisms: 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. Some authors, viz, 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. But 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.

Our study complements Aghion and Cohen (2004) but departs from it on several respects. Specifically, we address the question: given the degree of development, what kind of education strategy policy a country should adopt. To this end, we investigate the interplay between labor productivity backwardness relative to USA, education, R&D expenditures and the relation between labor productivity growth labor productivity backwardness and education. An innovative aspect of this study concerns the econometric specification we used. Previous studies in the literature used parametric specifications. Here, we assumed a *semi-parametric generalized additive model.* To the best of our knowledge, our study is the first one which does adopt such specification to study growth empirics of labour productivity. This estimation strategy places less possible restrictions in the functional form to be estimated

[1]The study of Aghion and Cohen (2004) is based on the French case.

[2]The frontier is measured by the technology of USA

and then allows for nonlinearities in the relationship between the response variable and the explanatories. This specification also does address the issue regarding the well known "curse of dimensionality" in fully nonparametric regressions when several variables are included as regressors. Moreover, the parametric component in the specification is not predetermined as usually does in the literature. Here, the parametric component are based on specification tests.

To answer the above question of education strategy policy, we use 29 OECD countries data over the period 1960-2000 to estimate nonparametrically two models. In the first model, the dependent variable is labor productivity growth. In the second, the dependent variable is labor productivity backwardness. The main results that emerge are the following. We first find that the number of agents enrolled in higher education is a determinant of growth. This finding is partly consistent with Aghion and Cohen assertion. Indeed, the latter seems valid only if we have a high number of potential researchers, say the threshold. Secondly, there is a threshold above which business research can no longer ensure the transition toward the technology frontier. With this end in view, investment in fundamental research becomes one of the driving force of economies which are near the technology frontier.

The remaining of the study is organized as follows: sections 2 and 3 describes respectively data and variables used in this study and the econometric specification. Estimation results are discussed in section 4, and section 5 concludes the study.

## 2 Data and variables

Recall that we estimate two models. First, the dependent variable is GDP per worker growth rate (as a measure of labor productivity growth), and second, the explained variable is labor productivity backwardness (in logarithmic term). The explanatory variables in the first relation are labor productivity backwardness and human capital (school enrollment rate in primary, secondary and higher education). For the second relation, we use as explanaroties, human capital (school enrollment rates), R&D expenditure in percentage of GERD[3] financed by industry[4], R&D expenditure in percentage of GERD financed by government, and part of R&D expenditure funded from abroad.

Our empirical examination is carried for 29 OECD countries over the period 1960-

[3]Gross Domestic Expenditure on R&D as a percentage of GDP.

[4]This expenditure goes only to industries.

2000. The data is obtained from the Penn World Table 6.1, World Development and Eurostat (see Appendix A for details on variables definition, data sources and the list of countries). There are a lot of interactions between education labor productivity and economic growth. The human capital, ability of workers and their wages are quite related, agents with higher wages being simply those with higher qualification. For example Boucekkine et al. (2002) show that a longer life horizon increases the agent's incentive to study in a longer period since they anticipate higher wage level according to their qualification. Education increases individual labor productivity since acquired abilities and knowledge enable to produce more and newest services at better quality.

In studying labor productivity growth and labor productivity backwardness, we follow the bulk of the literature by not controlling for possible all 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.[5] 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 for labor productivity growth and labor productivity backwardness next year, say, but with the *shape* of the relationships. In this respect, determinants of labor productivity growth and labor productivity backwardness 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

[5]Is should be noticed that Aghion and Cohen (2004) used the Total Factor Productivity (TFP) which is the best measurement of labor productivity. Here, due to data limitations (since we include other variables in our specification), we end up with only 17 countries if we were to consider the TFP.

will not enable us to make precise policy prescriptions, we will be in a position to intervene convincingly in the long debate on the education strategy policy which allows to boost growth.

A final remark concerns the use of school enrollment rate as a proxy of human capital. Apart the fact that, it is a consensus choice in the empirical literature, at a first sight, estimations based on this may seem quite restrictive since school enrollment rates don't reflect the qualitative effect of education and the others aspects of human capital. However, even in school enrollment rates there are still many differences between respective countries; for example in 1996, the school enrollment rate in higher education was 52% in France against 18.2% in Turkey and 60.3% in Corea Republic. Also, according to Romer (2000), if R&D don't encourage a greater number of researchers to develop new ideas, they might be inefficient. Indeed, if we have a fixed number of researchers, an increase in R&D expenditures generates simply a higher wage level for them, and will have no impact on growth rate. So, to obtain an effect on growth, the increases in the wages of researchers might encourage many agents to adopt the research career. Also, with a general purpose technology, we need high skills, consequently an increase of the number of persons with higher formation favors growth. In this case, our choice concerning school enrollment rate for the three levels of education might be relevant. Descriptive statistics are presented in Table 1.

**Insert Table 1**

From this table we can notice big variations in the labor productivity backwardness, meaning many heterogeneities in this variable from one country to the other; we have very different levels of backwardness in the sample. For R&D expenditure, on average, percentage funded from abroad is higher than the others, and the percentage financed by government is the lowest one. Comparing the means of schooling, the secondary one is very high (85%), and only 36% for the higher level. This means that on average, there is not sufficient potential researchers, to ensure fundamental research which is very important for the renewal and the increase of the stock of knowledge in a country.

**Insert Figure 1**

Figure 1 displays the graphs of the density estimates of labor productivity and labor productivity backwardness. The densities show mainly unimodal distributions.

Moreover, it is clear that the two densities are different and then, both variables are not telling the same history.

## 3 Econometric specification

We consider estimating non parametric additive models to study labor productivity growth and labor productivity backwardness. Additive models are widely used in both theoretical economics and econometrics. Deaton and Muellbauer (1980) provides 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" which appears in nonparametric regressions when many explanatory variables are included. It also allows us to capture nonlinearities and heterogeneity in the effect of explanatory variables on the dependent one.[6] 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) and Ibragimov and Hasminskii, 1980).[7] Additive models also offer some simple testing procedures (see Appendix C). For example, the statistic "Gain" provides a 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 view of this, our econometric specification consists of a semi-parametric GAM specification for panel data the structure of which is given by

$$y_{it} = \alpha + \sum_{j=1}^{p} f_j\left(\mathbf{x}_{it}^{j}\right) + \mathbf{z}_{it}'\gamma + \mu_i + u_{it}, \qquad i = 1, \cdots, N, \quad t = 1, \cdots, T \tag{1}$$

where $y_{it}$ denotes the response variable representing the labor productivity growth in a first estimation and labor productivity backwardness in a second estimation, $\mathbf{x}_{it}$ for $j = 1, \cdots, p$ are continuous explanatory variables, $\mathbf{z}_{it}\gamma$ is the parametric component with $\mathbf{z}_{it}$ explanatories that do enter linearly in the specification, and where $\gamma$ represents the vector of parameters to be estimated, $\alpha$ denotes the regression intercept. The $f_j$ are unknown univariate functions to be estimated such that $\mathbb{E}\left[f_j\left(\mathbf{x}_{it}^{j}\right)\right] = 0$.

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

[7]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 $\ell$ an index of smoothness of $f$ and $p$ is the dimension of $f$.

The unobserved effect $\mu_i$ can be eliminated by differentiating or computing the within transformation. Lagging relation (1) by one period and subtracting yields

$$y_{it} - y_{i,t-1} = \sum_{j=1}^{p} f_j(\mathbf{x}_{it}^j) - \sum_{j=1}^{p} f_j(\mathbf{x}_{i,t-1}^j) + (\mathbf{z}_{it} - \mathbf{z}_{i,t-1})'\gamma + \eta_{it}, \tag{2}$$

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

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

which identifies the functions

$$\mathbb{E}\left[y_{it} - y_{i,t-1}|\mathbf{x}_{it}^j, \mathbf{x}_{i,t-1}^j\right] = \sum_{j=1}^{p} f_j(\mathbf{x}_{it}^j) - \sum_{j=1}^{p} f_j(\mathbf{x}_{i,t-1}^j), \tag{3}$$

with the norming condition $\mathbb{E}[f_j(\mathbf{x}_{it}^j, \mathbf{x}_{i,t-1}^j)] = 0$, 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 $\eta_{it}$ is such that $\mathbb{V}(\eta_{it}|\Delta\mathbf{x}_{it}, \Delta\mathbf{z}_{it}) = \sigma^2(\Delta\mathbf{x}_{it}, \Delta\mathbf{z}_{it})$. For a given $j$, let us denote $\hat{f}(\mathbf{x}_{it})$ and $\hat{f}(\mathbf{x}_{i,t-1})$ the estimates of $f(\mathbf{x}_{it})$ and $f(\mathbf{x}_{i,t-1})$ respectively. Then, a more precise estimator[8], say $\hat{\hat{f}}$, can be obtained as a weighted average of $\hat{f}(\mathbf{x}_{it})$ and $\hat{f}(\mathbf{x}_{i,t-1})$:

$$\hat{\hat{f}} = \frac{1}{2}\left[\hat{f}(\mathbf{x}_{it}) + \hat{f}(\mathbf{x}_{i,t-1})\right] \tag{4}$$

In practice, we base our estimation on a "backfitting algorithm" (see Appendix C for details on the computational methods).[9] 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 C for details). Finally, our confidence interval are constructed using the "wild bootstrap". As shown in Appendix D, the wild bootstrap has the advantage of being robust to heteroskedasticity and correlation between observations.

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

[9]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.

# 4 Estimation results

The first estimation explores the relationship between labor productivity growth, labor productivity backwardness and the three levels of education (school enrollment rate in primary, secondary and higher education). It is important to notice that here, if labor productivity backwardness is used as regressor in labor productivity growth equation, the reverse is not. As a result, here, we do not face a simultaneous regression specification issue, even thought it may be interesting to investigate such question as we point out later in the concluding section.

**Insert Table 2**

If we only look at the total gain, we can conclude that parametric model is not rejected, but as illustrated by the results in Table 2, it is apparent that the relation between labor productivity growth and school enrollment rate in higher education has nonlinear pattern. In fact, the gain statistic of higher education follows a $\chi^2$ with degrees of freedom equal to the individual degrees of freedom, is equal to $18.534 > \chi^2(10.006) = 18.315$ at the 5% level. These features are apparent in Figure 2.

**Insert Figure 2**

We can distinguish three phases in the evolution of labor productivity according to the evolution of higher education. In the first two phases, one observes a non significant increasing and decreasing shape in labor productivity when the number of persons enrolled in higher education increases. But the third phase reveals that labor productivity growth is significantly affected by school enrollment rate in higher education. This means that concrete results on growth will not appear unless we have a sufficient number of agents enrolled in higher education. These features strengthen Aghion and Cohen's analysis. However, in our context, the latter implies enough potential researchers in fundamental research area, in other words, enough students in higher education. In long run, a larger number of students might probably enlarge the number of researchers who propel the stock of knowledge and innovation.

Now, the point is, what role does higher education play in the transition of a country toward the technology frontier? As pointed out by Guellec and Pottelsberghue (2001), basic research performed mainly by universities enhances the stock of knowledge of the societies, and may open new opportunities to business research which in turn positively affect productivity. Therefore, it may be interesting to investigate the relationship between education, R&D and productivity. This is the aim of our second step.

In this part, the dependent variable is the labor productivity backwardness relative to USA. Indeed, it seems relevant to explore the way some variables influence countries technological path toward the technology frontier represented by the USA technology. For these variables, we use the enrollment rate for the three stages of education, R&D expenditure in percentage of GERD financed by industry, R&D expenditure in percentage of GERD financed by government and percentage of R&D expenditure financed from abroad. We still retain the GAM specification. Estimation results are represented in Table 3 and Figures 3 and 4.

**Insert Table 3**

The total gain statistic is equal to $44.569 > \chi 2(14.992) = 24.985$. As a result, the parametric model is rejected against the nonparametric one. For government expenditure, the positive or negative effect depends on the nature of the research, either the biggest part is guided to defense or civilian objectives. In fact, the defense-related part of public funding has a negative and significant effect on growth as concluded by Guellec and Pottelsberghue (2001). We notice that the relationship between labor productivity backwardness and the two explanatory variables, viz., higher education and percentage of R&D financed by industry is nonlinear, since $20.299 > \chi 2(4.99) = 11.061$ and $20.391 > \chi 2(5.998) = 12.588$ respectively. So nonlinearities in this model come mainly from these two variables. These features are illustrated by Figures 3 and 4.

**Insert Figure 3**

In Figure 3, labor productivity backwardness increases with higher education until reaching a threshold (computed as 30.6%) from which it starts decreasing. As a result, to ensure innovation and technological catch-up, a country should initially have an important potential of researchers. This view can be linked to the main hypothesis of Romer (1990), that human capital is essential element in production of new ideas, and a given increase in the stock of human capital generates an infinite speeding up of the growth rate. Therefore, in many endogenous growth model, human capital must reach some level to permit innovation.

**Insert Figure 4**

For R&D expenditure funded by industry (Figure 4), the shape of $\hat{f}$ is the opposite to what we obtain with higher education. Firstly, labor productivity backwardness decreases with the industry expenditure, implying that with the increase in these expenditures, countries are nearer the technology frontier, but there is a threshold (computed as 32.99%) from which labor productivity backwardness starts increasing, and finally displays a constant pattern. So it seems that when a country is sufficiently near the frontier, it is more efficient to invest in fundamental research, since taken alone, after a short period of efficiency, business R&D can no longer guaranty the transition toward the technology frontier. As an argument, we know that at this stage of development, imitation of new technologies from other countries is not adaptable, the concerned country must innovate to favor growth. However, through research done by universities, we have an increasing stock of knowledge in the society, and then it gives new perspectives to the research in industries.

# 5 Conclusion

In this paper, we justified the idea of Aghion and Cohen (2004) and find that things are more complicated than they seem, by including R&D expenditures, and explaining labor productivity and labor productivity backwardness with different tools. In fact, concrete results on growth will not appear unless we have a sufficient number of agents enrolled in higher education, meaning enough potential researchers in fundamental research area; a larger number of students might probably enlarge the number of researchers in long run.

Also, it seems that when a country is sufficiently near the technology frontier, it is more efficient to invest in universities (fundamental research), since after a short time, business R&D can no longer ensure the transition toward the technology frontier, while higher education presents the opposite shape. So, it seems important that relevant authorities have to connect universities and R&D, and make university laboratories and technology poles in synergy. They must promote wages differentiation to strengthen competition between universities in order to attract best researchers and students, and perform the education system quality. Direct subsidies on R&D must be accompanied by those on formations with research career target, and motivate actors in university.

An extension of this study would be to introduce a country-specific trend in the model. Another natural extension would be to investigate a VAR-type model to analyse the long-run and short-run patterns. Finally, structural nonparametric modelling (which incorporates potential endogeneity and simultaneity problems which may link labor productivity growth and labor productivity backwardness) may also deserve more attention. However, accounting for this in a nonparametric setting is by no means trivial.

## References

AGHION, P., AND E. COHEN (2004): *Éducation et Croissance*. Rapport du C A E, No 2, Paris, La Documentation française.

BASSANINI, A., AND S. SCARPETTA (2001): "Les Moteurs de la Croissance dans les Pays de l'OCDE: Analyse Empirique sur des Données de Panel," *Revue Économique de l'OCDE*, 33.

BOUCEKKINE, R., D. D. L. CROIX, AND O. LICANDRO (2002): "Vintage Human Capital, Demographic Trends, and Endogenous Growth," *Journal of Economic Theory*, 104, 340–375.

DEATON, A., AND J. MUELLBAUER (1980): *Economics and Consumer Behavior*. Cambridge University Press.

FAN, J. (1992): "Design-Adaptative Nonparametric Regression," *Journal of the American Statistical Association*, 87, 998–1004.

GUELLEC, D., AND B. V. POTTELSBERGHUE (2001): "Recherche, Développement et Croissance de la Productivité: Analyse des Données d'un Panel de 16 Pays de l'OCDE," *Revue Économique de l'OCDE*, 33.

HANUSHEK, E. A., AND D. KIM (1995): "Schooling, Labour Force Quality and Economic Growth," NBER Working Paper No. 5399.

HANUSHEK, E. A., AND D. KIMKO (2000): "Schooling, Labour Force Quality and the Growth of Nations," *American Economic Review*, 90, 1184–1208.

HASTIE, T. J., AND R. J. TIBSHIRANI (1990): "Generalized Additive Models," Chapman and Hall, London, New York.

HOROWITZ, J. L. (2001): "The Bootstrap," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. Leamer, chap. 52. North Holland, Elsevier Science, Amsterdam.

HÄRDLE, W. (1990): *Applied Nonparametric Regression*. Cambridge University Press, Cambridge, New York.

IBRAGIMOV, I., AND R. HASMINSKII (1980): "On Nonparametric Estimation of Regression," *Soviet Math. Dokl.*, 21, 810–4.

LINTON, O. (2000): "Efficient Estimation of Generalized Additive Nonparametric Regression Models," *Econometric Theory*, 16, 502–523.

LINTON, O., AND W. HÄRDLE (1996): "Estimating Additive Regression Models with Known Links," *Biometrika*, 83, 529–540.

MANKIW, G., D. ROMER, AND D. WEIL (1992): "A Contribution to the Empirics of Economic Growth," *Quartely Journal of Economics*, CVII, 407–437.

NELSON, R. R., AND E. S. PHELPS (1966): "Investment in Humans, Technological Diffusion, and Economic Growth," *American Economic Review*, 56, 69–75.

ROMER, P. M. (1990): "Endogenous Technological Change," *Journal of Public Economy*, 98, 71–102.

ROMER, P. M. (2000): "Should the Government Subsidize Supply or Demand in the Market of Scientists and Engineers?," NBER Working paper No 7723.

STONE, C. (1980): "Optimal Rates of Convergence for Nonparametric Estimators," *The Annals of Statistics*, 8, 1348–1360.

——— (1982): "Optimal Global Rates of Convergence for Nonparametric Regression," *The Annals of Statistics*, 8, 1040–1053.

——— (1985): "Additive Regression and Other Nonparametric Models," *The Annals of Statistics*, 13, 685–705.

——— (1986): "The Dimensionality Reduction Principle for Generalized Additive Models," *The Annals of Statistics*, 14, 592–606.

# Appendix

## Appendix A: List of countries

Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Italy, Japan, Korea, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Spain, Sweden, Switzerland, Turkey, United Kingdom, United State of America (USA).

## Appendix B: Variables and source

Table A. Variables and sources

| Variables | Definition | source |
|---|---|---|
| Labor productivity | Real GDP per worker, $Y_{it^*}$ | (a) |
| Labor productivity backwardness (relative to USA) | $\frac{Y_{it}-Y_{US_t^{**}}}{Y_{US_t^{**}}}$ | |
| Primary school enrollment | Rate | (b) |
| Secondary school enrollment | Rate | (b) |
| Higher education | Rate | (b) |
| R&D funded by industry | In percentage of GERD | (c) |
| R&D funded by government | In percentage of GERD*** | (c) |
| Part of R&D from abroad | | (c) |

* Country $i$ at period $t$; ** $(US_t)$ USA at period $t$.

*** Gross Domestic Expenditure on R&D (GERD) as a percentage of GDP.

(a) Penn World Table 6.1; (b) World Development Indicator. (c) Eurostat (OECD).

## Appendix C: Estimation procedure and specification test ("gain")

The GAM specification considered can be rewritten in compact form:

$$Y = \alpha + \sum_{j=1}^{p} f_j\left(X_j\right) + \mathbf{Z}'\gamma + \epsilon \tag{5}$$

The $f_j$ are unknown univariate functions to be estimated such that $\mathbb{E}\left[f_j\left(Xj\right)\right] = 0$. 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 $X_j$, $j = 1, ..., p$ by using the backfitting algorithm (see below). The resulting smooth is the first estimate of $f_j\left(X_j\right)$, $\hat{f}_j\left(X_j\right)$.

**Step 3**: Obtain the estimate of $\gamma$ by ordinary least squares

$$\hat{\gamma} = \mathbb{E}\left(Y - \hat{\alpha} - \sum_{j=1}^{p} \hat{f}_j\left(X_j\right)|\mathbf{Z}\right) \tag{6}$$

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 $\hat{\alpha} = \frac{1}{n}\sum_i^n Y_i,\ f_j\left(X_j\right) = f_j^0\left(X_j\right),\ j = 1, ..., p$

$(b)$ Cycle: $j = 1, ..., p, 1, ..., p...$

$$\hat{f}_j\left(X_j\right) = S_j\left(Y - \hat{\alpha} - \sum_{k \neq j} \hat{f}_j\left(X_k\right)|X_j\right) \tag{7}$$

where $S_j$ is the smoother, using $k$ nearest symmetric neighborhood for $f_j^0$, and $\hat{f}_j$ is the nonparametric estimtor of $f_j$.[10]

$(c)$ Continue $(b)$ until the individual functions don't 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 $Y - \hat{\alpha} - \sum\limits_{k \neq j} \hat{f}_j\left(X_k\right)$. In the case of linear estimator, we have $\mathbf{S}_j = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$, 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

$$J = \frac{\left(RSS_1 - RSS_2\right)/\left(df_2 - df_1\right)}{RSS_2/\left(n - df_2\right)} \sim F_{df_2 - df_1, n - df_2} \tag{8}$$

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" $J \times \left(df_2 - df_1\right)$ is approximated by $\chi^2\left(df_2 - df_1\right)$. 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.

[10] 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.

## Appendix D: 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

$$y = f(x) + \epsilon, \tag{9}$$

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:

$s = 1$

**Repeat**

**Step 1**: Generate the bootstrap errors $\epsilon^*$ using the two points distribution probability: $P(\epsilon^* = \hat{\epsilon}\lambda) = \delta$; $P(\epsilon^* = \hat{\epsilon}\mu) = 1-\delta$, with $\lambda = \left(1-\sqrt{5}\right)/2$, $\mu = \left(1+\sqrt{5}\right)/2$, $\delta = \left(5+\sqrt{5}\right)/10$.

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

**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.

Table 1: Descriptive Statistics

| Variables | #Obs. | Mean | Std.Dev. | Min. | Max. |
|---|---|---|---|---|---|
| Labor productivity growth rate | 958 | 0.036 | 0.47 | -0.01 | 1.02 |
| Labor productivity backwardness | 985 | -0.88 | 1.2 | -6.82 | 3.38 |
| Primary school enrollment rate | 603 | 0.11 | 0.01 | 0.08 | 0.14 |
| Secondary school enrollment rate | 615 | 0.85 | 0.23 | 0.11 | 1.48 |
| School enrollment rate in higher education | 280 | 0.36 | 0.19 | 0.04 | 0.98 |
| Government R&D expenditure | 196 | 0.4 | 0.14 | 0 | 0.7 |
| Industry R&D expenditure | 186 | 0.5 | 0.15 | 0 | 0.91 |
| R&D expenditure from abroad | 182 | 0.61 | 0.52 | 0 | 3.03 |

Table 2: GAM estimation for labor productivity growth

| Variables | Coef. | Std.Err | df. | Gain$^{(a)}$ |
|---|---|---|---|---|
| Labor productivity backwardness | -0.7 | 0.64 | 2 | 0.7 |
| Primary school enrollment rate | -0.12 | 0.12 | 5.99 | 5.636 |
| Secondary school enrollment rate | -0.12* | 0.05 | 20 | 18.44 |
| School enrollment rate in higher education | 0.36 | 0.11 | 10 | 18.53$^{(b)}$ |
| Intercept | 2.18** | 0.84 | 1 | - |

$^{(a)}$ The Gain statistic is computed as described in Appendix C.

$^{(b)}$ Significant Gain statistic.

$^{(**)}$ Significancy at the 1% level; $^{(*)}$ significancy at the 5% level.

Table 3: GAM estimation for labor productivity backwardness

| Variables | Coef. | Std.Err | df. | Gain[(a)] |
|---|---|---|---|---|
| Primary school enrollment rate | -0.08* | 0.04 | 2 | 0.64 |
| Secondary school enrollment rate | -0.02 | 0.013 | 2 | 1.8 |
| School enrollment rate in higher education | -0.02 | 0.02 | 4.99 | 20.29[(b)] |
| Government R&D expenditure | 0.03** | 0.01 | 3.99 | 1.24 |
| Industry R&D expenditure | 0.02 | 0.011 | 5.99 | 20.39[(b)] |
| R&D expenditure from abroad | -0.04 | 0.03 | 2 | 0.17 |
| Intercept | -0.18 | 0.16 | 1 | - |

[(a)] The Gain statistic is computed as described in Appendix C.

[(b)] Significant Gain statistic.

[(**)] Significancy at the 1% level; [(*)] significancy at the 5% level.

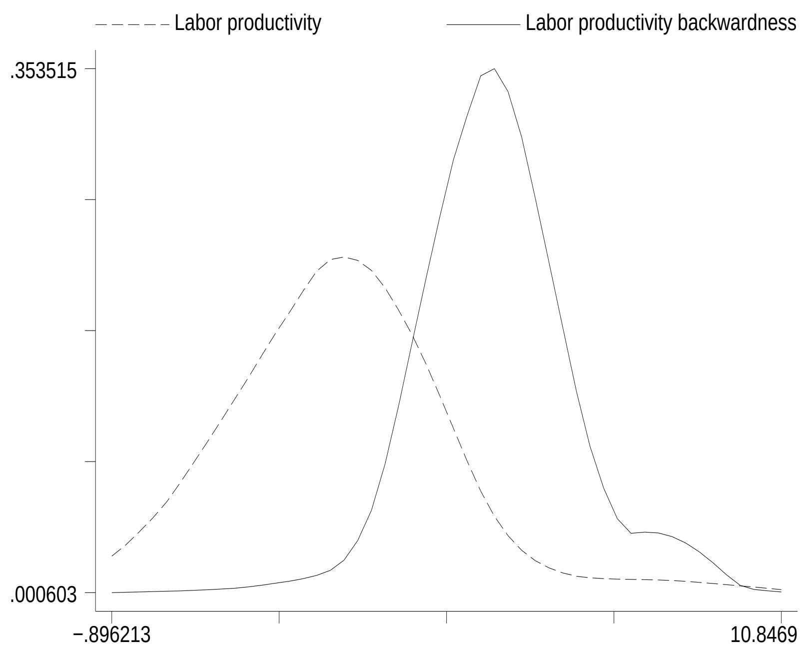


Figure 1: Distribution of labor productivity and labor productivity backwardness, kernel density estimate (Epanechnikov kernel).

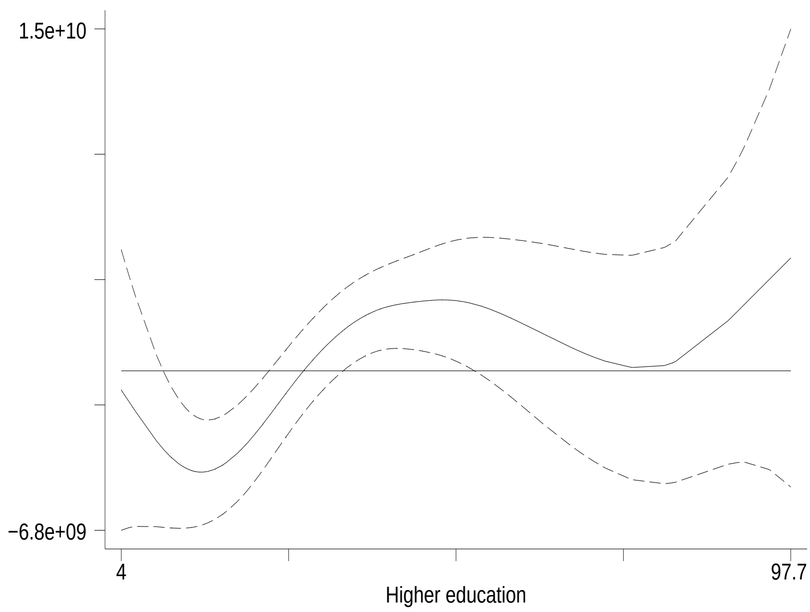


Figure 2: Nonparametric regression of the relation between labor productivity growth and school enrollment rate in higher education (from Table 2). The figure shows the estimated curve and the wild bootstrap confidence interval at 95%. The straight line is the zero line.

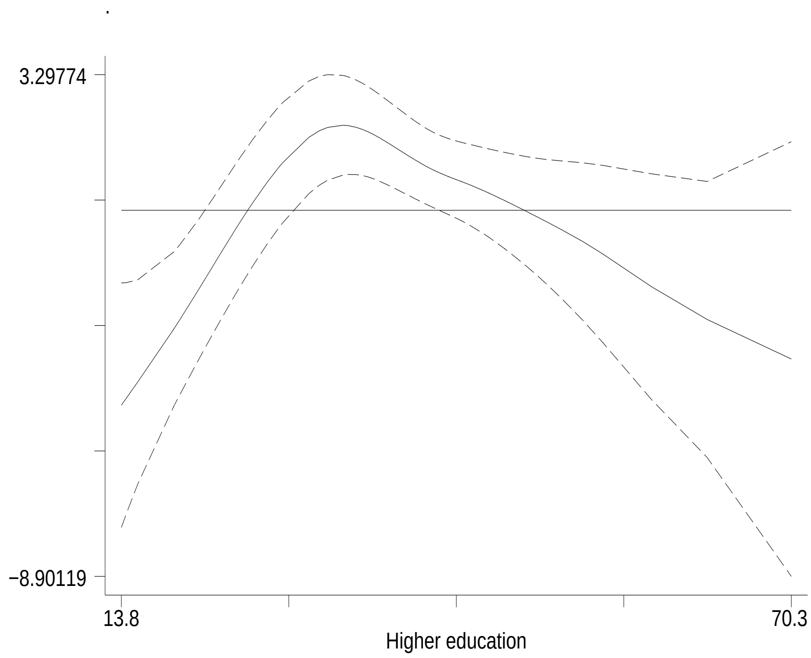


Figure 3: Nonparametric regression of the relation between labor productivity backwardness and school enrollment rate in higher education (from Table 3). The figure shows the estimated curve and the wild bootstrap confidence interval at 95% level. The straight line is the zero line.

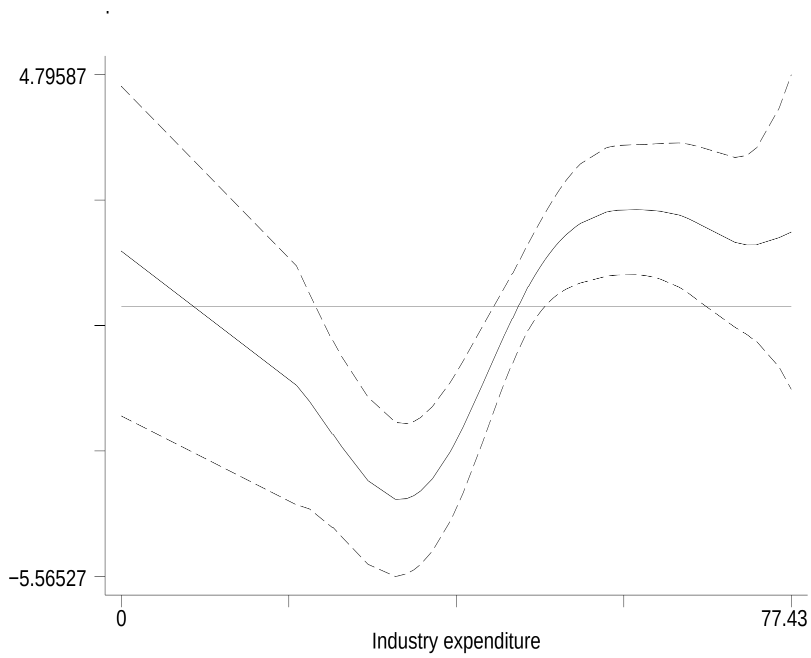


Figure 4: Nonparametric regression of the relation between labor productivity backwardness and the part of R&D expenditure in percent of GERD funded by industries (from Table 3). The figure shows the estimated curve and the wild bootstrap confidence interval at 95 % level. The straight line is the zero line.

# Center for Research in Economic Analysis

uni.lu

CREA
Discussion
Paper Series
2009

| Number | Date | Title | Author |
|---|---|---|---|
| 2009-01 | April 2009 | Currency Unions and International Assistance | Pierre M. Picard, Tim Worrall |
| 2009-02 | June 5, 2009 | On spatial equilibria in a social Interaction Model Financial Center | Pascal Mossay, Pierre M. Picard |
| 2009-03 | July 13, 2009 | Diaspora Externalities as a Cornerstone of the New Brain Drain Literature | Elisabetta Lodigiani |
| 2009-04 | July 28, 2009 | Migration and human capital in a endogenous fertility model | Luca Marchiori, Patrice Pieretti, Benteng Zou |
| 2009-05 | July 8, 2009 | Labor Market Pooling, Outsourcing and Labor Contracts | Pierre M. Picard, David E. Wildasin |
| 2009-06 | April 2009 | Does the Canadian economy suffer from Dutch Disease? | Michel Beine, Charles Bos, Serge Coulombe |
| 2009-07 | April 7, 2009 | Underinvestment in public goods: The influcence of state depended investment costs | Nikos Ebel, Benteng Zou |
| 2009-08 | April 2009 | International financial competition and bank risk-taking in emerging economies | Arnaud Bourgain, Patrice Pieretti, Skerdilajda Zanaj |
| 2009-09 | November 2009 | Remittances and Financial Openness | Michel Beine, Elisabetta Lodigiani, Robert Vermeulen |
| 2009-10 | September 2009 | A Harmonization of First and Second Natures | Pierre M. Picard, Dao-Zhi Zeng |
| 2009-11 | September 30, 2009 | Local social capital and geographical mobility | Quentin David, Alexandre Janiak, Etienne Wasmer |
| 2009-12 | December 7, 2009 | Fits and Misfits : Technological Matching and R & D Networks | R. Cowan, N. Jonard, B. Sanditov |
| 2009-13 | March 30, 2009 | On uncertainty when it a¤ects successive markets | Jean Gabszewicz, Ornella Tarola, Skerdilajda Zanaj |
| 2009-14 | January, 2009 | On tax competition, public goods provision and jurisdictions' size | Patrice Pieretti, Skerdilajda Zanaj |
| 2009-15 | February, 2009 | Diasporas | Michel Beine, Frédéric Docquiers, Çaglar Özden |
| 2009-16 | December 21, 2009 | The Determinants of Research Production by US Universities | Quentin David |
| 2009-17 | December 7, 2009 | Risk Premiums and Macroeconomic Dynamics in a Heterogeneous Agent Model | Ferre de Graeve, Maarten Dossche, Marina Emiris, Henri Sneessens, Raf Wouters |
| 2009-18 | October 12, 2009 | Competition Among the Big and the Small | Ken-Ichi Shimomura, Jacques-François Thisse |
| 2009-19 | November, 2009 | Technology frontier, labor productivity and economic growth: Evidence from OECD countries | Théophile T. Azomahou, Bity Diene, Mbaye Diene |