DEPARTMENT OF ECONOMICS
DISCUSSION PAPER SERIES

**THE TWO FACES OF R&D AND HUMAN CAPITAL:
EVIDENCE FROM WESTERN EUROPEAN REGIONS**

Johanna Vogel

Number 599
April 2012

Manor Road Building, Oxford OX1 3UQ

The Two Faces of R&D and Human Capital: Evidence from Western European Regions

Johanna Vogel^{*†}
University of Oxford

March 2012

Abstract

This paper investigates two channels through which research and development (R&D) and human capital may affect regional total factor productivity growth in the manufacturing sector, using panel data on 159 EU-15 regions from 1992 to 2005. Based on the endogenous growth model of Griffith, Redding and Van Reenen (2003), we allow R&D and human capital to influence productivity growth both directly, reflecting own innovation, and indirectly, reflecting imitation of frontier technology. Further, the model allows for conditional convergence to a long-run level of TFP relative to the frontier. We also develop an extension that captures geographically localised technology spillovers. Our preferred system-GMM estimates provide evidence of a positive and significant direct effect of human capital, and a positive and significant indirect effect of R&D on productivity growth. This may be interpreted as lending support to the recent focus of EU regional policy on raising educational attainment and R&D expenditures, although their channels of influence appear to differ. Our results also suggest that TFP convergence has taken place over our sample period and that geographic distance to the technology frontier matters.

Keywords: Total factor productivity, convergence, human capital, research and development, European regions

JEL Classification: O30, O47, I25, R11/12, C23

---

^{*}Correspondence: University College, Oxford OX1 4BH, UK; johanna.vogel@economics.ox.ac.uk.

[†]I am grateful to Steve Bond, Helen Simpson and participants of the Gorman Workshop at the University of Oxford, the 2009 ERSAs-Prepare Summer School in Volos, Greece, and the 56th North American Meetings of the Regional Science Association International in San Francisco for helpful comments and suggestions. All remaining errors are my own. I also thank Jonathan Stenning at Cambridge Econometrics for assistance with the European Regional Database and Jonny Johansson at Eurostat's Labour Force Survey Unit for access to the regional education data. Financial support by the ESRC (award no. PTA-031-2004-00246), the University College Old Members Trust Fund and the George Webb Medley Fund is gratefully acknowledged.

1 Introduction:

In Europe, the marked slowdown of productivity growth relative to the United States since the mid-1990s has placed productivity at the centre of the economic and political agenda. According to Van Ark, O'Mahony and Timmer (2008), labour productivity growth (measured as GDP per hour worked) in the U.S. increased from an average of 1.2% per annum during the period 1973-1995 to 2.3% during the period 1995-2006. For the EU-15 countries, by contrast, average labour productivity growth decreased from 2.4% per year to 1.5% between these two periods. Similar developments have been found for total factor productivity (TFP) growth.¹

In response to the competitive pressures arising from globalised manufacturing production, European policy-makers have come to regard innovation and human capital as essential for maintaining productivity growth and living standards over the long term. At the level of the European Union, the importance attached to these factors is illustrated by the Lisbon Strategy, the EU's growth strategy for the period from 2000 to 2010, and its successor for the following decade, Europe 2020. Both initiatives aim to raise productivity growth primarily by improving Europe's performance in the areas of research and development (R&D) and education. Creating an environment that is more conducive to innovation and equipping the workforce with the education necessary to carry it out is considered vital for dealing with the challenges presented by technological change, the catch-up of emerging economies and the current economic crisis. With the Lisbon Strategy, the EU set itself the ambitious goal of becoming "the most competitive and dynamic knowledge-based economy in the world".²

The European Commission has identified an "innovation gap" between Europe and its competitors, characterised by comparatively low European expenditure on R&D and a lower number of patent applications: in 2008 for example, spending on R&D in the EU-27 amounted to just 1.9% of GDP, while the figures for the U.S. and Japan were 2.8% and

---

¹This paper focuses on TFP as a measure of productivity, since its growth rate can be considered an underlying driver of both labour productivity growth and total output growth. More recently, Europe's productivity performance has improved slightly, but in 2008, average GDP per hour worked in the EU-27 was still 28% below that in the U.S. (European Commission, 2010a).

²Spring European Council Conclusion, Lisbon European Council, March 2000, available at http://consilium.europa.eu/ueDocs/cms_Data/docs/pressData/en/ec/00100-r1.en0.htm

3.4% respectively.³ Both the Lisbon and Europe 2020 strategies have therefore set a target for EU-wide R&D expenditures of 3% of GDP. Similarly, a smaller proportion of the EU-27 population aged 25 to 34, 30.9%, had completed tertiary education in 2008 than in the U.S. (40%) and Japan (54%). Hence, the Europe 2020 initiative aims to raise the share of the population aged 30 to 34 that has obtained a higher education degree to 40% by 2020.⁴

The focus of EU funding on R&D and education also benefits the European regions: with the introduction of the Lisbon and Europe 2020 strategies, a major share of the EU's regional policy budget has been orientated towards their objectives. For the period 2007 to 2013 for example, €228 billion, or 66% of the total regional policy budget, are allocated to the priority areas of the two strategies. Of this sum, €86 billion are directly earmarked for investments in knowledge and innovation.⁵

At a theoretical level, the emphasis of EU economic policy on innovation and human capital is related to endogenous growth theory, which emerged in the late 1980s. The models in this tradition emphasise these two factors as key drivers of technological progress and hence of long-run economic growth. A more recent strand of research, summarised in Aghion and Howitt (2006) for instance, extends the early endogenous growth models to a cross-country framework in which technological progress in all countries except the technological leader depends not only on their own innovative capacity, but also on the imitation (transfer) of technologies developed at the technological frontier. The distance of each country to the technological frontier determines the relative importance of innovation and imitation and consequently also the relative importance of factors that are conducive to each of the two activities respectively. Specifically, the policies and institutions that encourage productivity growth through imitation in countries further behind the frontier will be different from those required for growth through innovation in countries closer to the frontier.

For example, the Sapir Report (Sapir et al. 2004), commissioned by the EU to evaluate its policies in view of improving economic performance, takes this approach. It attributes

---

³Figures for 2007 - before the onset of the “great recession” - were very similar.

⁴These figures are taken from European Commission (2011) and the website of the Europe 2020 strategy at http://ec.europa.eu/europe2020/index_en.htm.

⁵See European Commission (2010b, 2010c).

Europe’s slow growth to a failure to fully adapt its economic institutions from those appropriate for catch-up growth based on factor accumulation and assimilation of U.S. technology in the decades following World War II to the requirements of the knowledge- and innovation-driven economic environment closer to the frontier that Europe finds itself in today. The report therefore calls for more investment in R&D and higher education, among other measures.

This paper contributes to the literature by applying this framework to the regions of the EU-15. So far, it has been applied empirically to manufacturing industries within and across countries, so our study explores a different level of disaggregation. We investigate the determinants of TFP growth, a standard empirical measure of technological progress, in the manufacturing sector across 159 NUTS-2 regions over the period 1992 to 2005. In particular, we focus on the dual roles of R&D and human capital as engines of productivity growth through both innovation and imitation. Our econometric analysis is based on Griffith et al.’s (2003) “two faces” extension to the early Schumpeterian framework of endogenous growth, which characterises TFP growth in transition to a long-run equilibrium level of TFP relative to the frontier. We extend this model to capture geographically localised technology spillovers.

We make use of a comprehensive dataset on the European regions from Cambridge Econometrics, from which we calculate an annual TFP index for each region. In addition, we use information from the EU Labour Force Survey to construct a continuous time series on regional educational attainment dating back to 1992. This allows us to cover a longer time period than previous empirical studies for the variables we consider. We also use panel data estimators to account for both unobserved region-specific fixed effects and the potential endogeneity of explanatory variables due to e.g. measurement error.

Our preferred econometric estimates suggest a positive and significant direct effect of human capital on productivity growth, but no significant indirect or imitation effect. On the other hand, we find that regional R&D activity may have played a more important role in facilitating the imitation or transfer of frontier technology. These results imply that improving both educational attainment and R&D performance - as advocated by the Lisbon

and Europe 2020 strategies - may raise productivity growth in the EU-15 regions, albeit through different channels. Further, our estimates are consistent with conditional TFP convergence relative to the frontier taking place over our sample period. This contrasts with some of the existing literature on the EU regions, which has found no evidence of TFP convergence. Finally, we find some support for the hypothesis that technology spillovers are to an extent geographically localised.

The remainder of the paper is organised as follows. Section 2 reviews the relevant theoretical and empirical literature. Section 3 outlines our theoretical framework, empirical specification and estimation methods. Section 4 describes the data and the construction of our variables. Section 5 presents and discusses our main results, and section 6 concludes.

2 Review of the Theoretical and Empirical Literature:

2.1 Endogenous Growth Theory:

Until the mid-1980s, the theory of economic growth was dominated by the neoclassical model of Solow (1956). This model, however, does not explain growth in the long term because its only source of growth is exogenously given technological progress. In the absence of technological progress, the assumption of diminishing returns to physical capital - the accumulation of which drives growth in the short run - implies that all growth must ultimately come to a halt.

The principal contribution of the “new” or “endogenous” growth theory is that it provides an explanation for long-run economic growth from within the model. Frankel (1962) and similarly Romer (1986), two early examples, achieve this by assuming that the level of technology or knowledge in the economy is a function of the aggregate stock of physical capital. While individual firms take the economy-wide level of technology as given and face diminishing returns to capital, the total of their investment decisions advances the state of technology available.⁶ This externality offsets diminishing returns to capital at the firm level and makes positive long-run growth possible.

---

⁶This is an example of Arrow’s (1962) “learning by doing”.

The seminal paper by Lucas (1988) avoids diminishing returns to capital by specifying what may be thought of as a broad class of capital to which there are constant returns, including human capital in the sense of skills. Human capital accumulation is assumed to be proportional to its existing stock, so that over time, it grows at an exponential rate. Hence, while there are diminishing returns to physical capital, human capital accumulation does not peter out. The property of aggregate constant returns thus again guarantees the existence of growth in the long run, and its key determinant is the growth rate of human capital.

In effect, the Frankel/Romer and Lucas models belong to the class of “AK” models, which counteract the zero long-run growth result of the neoclassical model by assuming constant returns to the accumulable factor capital. Meanwhile, technological progress remains an exogenously given constant or incompletely modelled.

By contrast, Romer (1990), Grossman and Helpman (1991), and Aghion and Howitt (1992) develop explicit theories of the determinants of technological progress, which drives the long-term equilibrium growth rates of all variables in these models. Technological progress is interpreted as the accumulation of ideas or knowledge through innovation, which is the outcome of purposeful economic activity by profit-maximising firms who invest in R&D to gain patents for their discoveries. A patent exploits the - at least partial - excludability of inventions and thereby provides the incentive for private firms to engage in research in the first place, because it allows the innovator to enjoy monopoly power and earn supernormal profits for a certain period of time. These elements represent an attempt to introduce greater realism into the analysis of economic growth by describing technological progress as a result of profit incentives, incorporating imperfectly competitive market structures, and investigating the economic processes underlying technological progress.

In Romer (1990), innovations made in the research sector lead to technological progress by increasing the available varieties of intermediate goods, while in Grossman and Helpman (1991) and Aghion and Howitt (1992), innovations improve the quality of existing intermediate goods. Because the latter concept incorporates the idea of “creative destruction” developed by Joseph Schumpeter (e.g. Schumpeter 1942), where a quality improvement

destroys the market for the old version of the good, this class of models is also referred to as Schumpeterian growth theory. In both branches of this literature, the growth rate of technology, i.e. of TFP, depends on the amount of resources devoted to the research sector, such as the level of human capital (Romer) or the number of workers (Aghion and Howitt) employed in research. These models therefore give both human capital and R&D activities a role in contributing to productivity growth. In contrast to the Solow or Lucas models, it is the level or stock of these inputs into innovation that drives growth, not their accumulation.

The driving force behind endogenous growth in Romer (1990) and the Schumpeterian models is an externality which arises as a by-product to the creation of new knowledge in research. A patent protects the use of an inventor’s idea in the production of a new good; but because knowledge is assumed to be nonrival and only partially excludable, the patent cannot prevent other inventors’ use of the idea in research, through e.g. patent documentation and the scientific literature.⁷ New knowledge thus spills over to all other researchers and adds to the total stock of knowledge or technology (in the form of product variety or product quality) available in the economy, on which all researchers can build for subsequent innovations. The creation of new ideas is therefore proportional to the existing aggregate stock of knowledge, and over time, the stock of knowledge grows exponentially. Since every innovation makes all researchers more productive, the accumulation of knowledge can offset diminishing returns to other factors of production. This allows for positive economic growth in the long-run equilibrium.

2.2 The “Extended” Schumpeterian Framework:

More recently, Griffith et al. (2003) have extended the Aghion and Howitt (1992) model in two ways. First, they introduce the idea that for countries lagging behind the current technological leader, technological progress may spring not only from their own innovations, but also from the transfer of technologies developed at the frontier, which occurs independently of own research activity (“autonomous” technology transfer) via knowledge

---

⁷See Romer (1990) p. S84-S85.

spillovers.⁸ This source of productivity growth for a follower country is modelled as a function of its distance to the technological frontier, measured in terms of TFP levels. The further behind the frontier an economy is, the greater is its potential for growth, so that there may be productivity catch-up or convergence relative to frontier TFP. An early exponent of this view is Gerschenkron (1962), who highlights the “advantage of backwardness”, which allows a less developed country to close the gap with advanced countries more quickly because by adopting the most modern knowledge available, it can make a relatively larger technological leap than those who went before.

Second, Griffith et al. (2003) allow the size of quality improvements in intermediate goods induced by innovation in a follower country’s research sector to be a function of its distance to the frontier. Thus, they introduce a role for own research activity in facilitating technology transfer. Intuitively, the ability of a lagging country to absorb and implement foreign technologies may depend on the level of education of its labour force and on the extent to which the country itself engages in R&D: the more educated the workforce is, or the more technological expertise it has from its own research efforts, the more able it is to implement new frontier technology, and the more a given distance to the frontier contributes to its TFP growth. This point is advanced in Abramovitz (1986) for instance, who argues that being technologically backward alone does not automatically lead to faster growth; rather, to exploit best-practice technology, a country must possess the “social capability” for growth, which encompasses among other factors the level and quality of education.

Similarly, in the simple but influential model of Nelson and Phelps (1966), education facilitates the adoption of new technologies, so that technological progress is an increasing function of the available level of human capital, interacted with the gap to the technology frontier. Regarding R&D, Cohen and Levinthal (1989) model a twofold effect of R&D investment, both generating inventions and improving firms’ “absorptive capacity” - that is, their ability to assimilate new knowledge available from elsewhere. Griffith et al. (2003) call the latter effect the “second face of R&D”.

Overall therefore, the rate of technological progress in a country or region behind the

---

⁸These spillovers are the cross-country analogue of the intertemporal spillovers that lead to growth over time within one country in Romer (1990) and the Schumpeterian models.

frontier depends on three factors in this framework: its own R&D activity and its stock of human capital, capturing the country’s own innovation as in the Schumpeterian models described above; its distance to the technological frontier, capturing its potential for catch-up growth through autonomous technology transfer; and an interaction term between the first two factors, capturing the idea that successful imitation of frontier technology also depends on the country’s own R&D activity and human capital. R&D and human capital thus affect productivity growth through two distinct channels: by stimulating innovation, which is a direct effect; and by providing the prerequisites necessary for successful imitation, which is an indirect effect via the country’s distance to the technological frontier.

2.3 Empirical Literature at the Country Level:

At the country level, a sizeable body of research is concerned with the effect of human capital and R&D on productivity growth in general, and with the extended Schumpeterian framework in particular. A common approach in this literature is to regress a measure of TFP growth computed using the growth accounting approach pioneered by Solow (1957) on the variables of interest, which is also the setup in this paper.

The available empirical evidence generally suggests that the direct effect of research activity, frequently measured by the ratio of R&D expenditures to output, on productivity growth is positive and substantial. An early example is the firm- and industry-level work of Griliches (e.g. Griliches 1980, Griliches and Lichtenberg 1984) for the United States. At the country level, the results of Coe and Helpman (1995) and Guellec and van Pottelsberghe de la Potterie (2004) support this conclusion for the OECD. Frantzen (2003) finds that domestic R&D Granger-causes TFP growth in a panel of manufacturing industries across OECD countries.

Griffith, Redding and Van Reenen (2004) (henceforth also GRVR) and Cameron, Proudman and Redding (2005) (henceforth also CPR) additionally investigate the “second face” of R&D - that is, its hypothesised indirect effect on TFP growth by facilitating technology transfer. While the country-level studies discussed above also examine the influence of knowledge spillovers from abroad on domestic productivity growth, they proxy for them

with the (import-weighted) R&D activity of a country's trading partners. GRVR and CPR on the other hand directly measure the follower's potential for absorbing knowledge spillovers as the difference in TFP levels between leader and follower.⁹ This term is then also interacted with a measure of R&D.

Using panel data on manufacturing industries across twelve OECD countries from 1974 to 1990, GRVR find a strong role for both direct and indirect effects of R&D. By contrast, CPR only find evidence of a positive and significant direct effect of R&D, while the interaction term representing the indirect effect is insignificant. Both GRVR and CPR also report significant evidence of autonomous technology transfer. CPR focus on a panel of UK manufacturing industries observed between 1971 and 1992. Khan (2006) considers French manufacturing industries and concludes that R&D affects TFP growth primarily through the direct innovation channel, similar to CPR. This difference in results between the cross-country analysis of GRVR and the single-country studies of CPR and Khan (2006) could reflect the additional variation that the country dimension adds by allowing for spillovers within manufacturing industries across countries.

Regarding the direct effect of human capital on growth, the empirical evidence is less uniform. Different approaches to measuring human capital, but also differences in estimation methods and datasets, have produced conflicting results.

On the one hand, models in the neoclassical tradition such as Mankiw, Romer and Weil (1992) (henceforth also MRW), but also the endogenous growth model of Lucas (1988), treat human capital like an input into the production function. Consequently, the long-run level (MRW), respectively the long-run growth rate (Lucas), of income per capita depend on the *accumulation* of human capital. Mankiw et al. (1992) proxy this variable with the secondary school enrolment rate and find that it has a positive and generally significant effect on the long-run income level in a large cross-country sample. However, the use of enrolment rates has been criticised.¹⁰ It is now more common to use educational attainment of the labour force - for example average years of schooling or the percentage of the labour

---

⁹This more general measure allows them to capture spillovers in a broad sense, including those that do not operate through R&D or trade.

¹⁰For example, Gemmell (1996) argues that enrolment rates conflate human capital accumulation and stock effects.

force that has attained a certain level of education - to measure the level or “stock” of human capital. The growth rate of this variable then serves as a measure of human capital accumulation. Benhabib and Spiegel (1994) and Pritchett (2001) take this route and find that the growth rate of average years of schooling has a negative but insignificant effect on income growth.¹¹ Early panel data studies also tend to obtain a negative coefficient on human capital accumulation, which is sometimes even significant.¹² Later contributions by Bassanini and Scarpetta (2002), De la Fuente and Doménech (2006) and Cohen and Soto (2007) suggest that these counterintuitive results may be related to quality issues with the human capital data used, particularly to measurement error.

On the other hand, in endogenous growth models à la Romer (1990) and Aghion and Howitt (1992), the stock or *level* of human capital - more precisely, the level of human capital engaged in research - determines TFP growth. Benhabib and Spiegel (1994) combine this approach with the Nelson-Phelps (1966) model and derive an empirical specification in which technological progress depends on both endogenous innovation through the level of human capital, as well as technology transfer from (distance to) the frontier, the rate of which also depends on the level of human capital. Their estimates indicate that for the richest third of countries in their sample, it is the direct innovation effect of human capital that matters, while for the poorest third, only the indirect effect via distance to the frontier is significant.

This provides some evidence of a dual impact of human capital on TFP growth as in the extended Schumpeterian framework. GRVR, who investigate this further at the industry level across countries, find a positive and significant coefficient on both the level of human capital and its interaction with distance to the frontier. In CPR’s study of UK manufacturing industries, neither effect is significant. Both GRVR and CPR use the percentage of the population aged 25 and above that has completed higher education to measure the level of human capital employed in research.

---

¹¹Both articles estimate a so-called “growth accounting regression”, i.e. a production function - here Cobb-Douglas augmented with human capital - in log-differences. Pritchett (2001) further finds a negative and significant effect of human capital accumulation on TFP growth.

¹²See, for instance, Islam (1995), Caselli, Esquivel and Lefort (1996), and Bond, Hoeffler and Temple (2001).

2.4 Empirical Literature at the Regional Level:

At the EU regional level, there is so far little empirical evidence on the dual roles that the extended Schumpeterian model envisages for R&D and human capital in driving long-run productivity growth. This may partly result from problems with the availability of data on these variables for a sufficient number of regions and time periods. In particular, studies using TFP as their outcome variable of interest have only recently started to emerge. Much of the early literature analyses the direct effects of R&D and human capital on per-capita output, i.e. on labour productivity, using specifications for short-run growth in transition to the steady state such as Mankiw et al. (1992).

Fagerberg, Verspagen and Caniëls (1997) is an early example of EU regional research that investigates growth through both endogenous innovation and imitation of frontier technology. They consider 64 regions during the 1980s and find that the initial level of GDP per capita, used to proxy for the potential for catch-up growth through technology transfer, and employment in business sector R&D, proxying for own innovation, are significant determinants of average per-capita GDP growth. Moreover, a decomposition of the difference in GDP growth rates between rich and poor regions in the sample indicates that poorer regions benefit more from catch-up, while R&D is more important for growth in richer regions.

One recent study that investigates TFP is Di Liberto and Usai (2010). These authors estimate MRW's transitional growth specification for 199 EU-15 regions using panel data methods and extract regional TFP levels from the estimated region-specific fixed effects. Comparing two sub-periods of their total sample period, 1985 to 2006, the distribution of regional TFP is analysed non-parametrically, in terms of its evolution over time, intra-distributional dynamics and geographic features. Di Liberto and Usai's (2010) main findings are that across all regions, no TFP convergence takes place between the two periods, and polarisation between clusters of low-TFP regions in Southern Europe and high-TFP regions in the Centre and North increases.

Turning to papers that look at both R&D and education, Bronzini and Piselli (2009) find evidence for a positive long-run cointegrating relationship between the level of TFP

and the stocks of R&D and human capital across the Italian regions from 1980 to 2001. This result is consistent with a positive effect of the rates of accumulation of R&D and human capital on TFP growth.

Sterlacchini (2008) assesses the impact of R&D and human capital on per-capita GDP growth using 197 EU-15 regions and cross-section OLS regressions for the period 1995 to 2002. Similar to Fagerberg et al. (1997), the initial level of GDP is included to capture technology catch-up. Sterlacchini's (2008) measure of R&D, the share of R&D expenditures in GVA, is only significant for those regions with an income level above 75% of the average across the EU-25. On the other hand, human capital - measured by the share of the population aged 25 to 64 with tertiary education - exerts a positive influence on growth for regions both above and below this threshold. One interpretation of these results is that R&D may be more conducive to growth in regions closer to the technological frontier, while human capital has a twofold effect similar to that described by the extended Schumpeterian model.

Two studies that are close to our empirical framework are Griffith, Redding and Simpson (2009) and Badinger and Tondl (2005). The former analyse TFP growth through catch-up to the frontier for a panel of British establishments in four-digit manufacturing industries from 1980 to 2000. Their empirical specification is similar to the model in Griffith et al. (2003, 2004) but focuses on the effect of technology transfer as captured by distance to the frontier in terms of TFP levels. TFP is measured using the same superlative index number approach that we employ. Within-groups estimates of Griffith et al.'s (2009) specification suggest that distance to the frontier - both national and regional industry frontier establishments are considered - has a positive and significant effect on establishment productivity growth, so that establishments further behind the frontier grow faster than those closer to the frontier. They conclude that catch-up effects to both types of frontier are economically important.

Badinger and Tondl (2005) use theoretical arguments similar to those underlying the extended Schumpeterian framework to motivate their study of per-capita output growth for 159 EU-15 regions from 1993 to 1999. Following Benhabib and Spiegel (1994), they

estimate a growth accounting regression based on a Cobb-Douglas production function in log-differences. The term for TFP growth on the right-hand side of this equation, $\Delta \ln A$, is replaced with terms for autonomous technology transfer as well as direct and indirect effects of R&D, human capital and trade. Each region’s potential for technology transfer is proxied with the gap between its own level of labour productivity and that of the best-performing region, similar to GRVR’s distance to the technology frontier. The direct or innovation effects of R&D and human capital are measured respectively by the ratio of regional patent applications to the European Patent Office over total regional employment and the share of the population with completed higher education. The indirect or imitation effects of R&D and human capital are represented by interaction terms with distance to the (labour productivity) frontier.

Applying OLS and a spatial econometric 2SLS estimator to their cross-section of regions, Badinger and Tondl (2005) find a significant and positive coefficient for human capital and its interaction with the labour productivity gap, as well as for their trade variable and its interaction with the gap. By contrast, this is not the case for the productivity gap term on its own, nor for R&D and its interaction. Therefore, the authors conclude that human capital and trade are important engines of EU regional growth, both directly and indirectly by facilitating imitation, while there is no evidence that autonomous technological catch-up to the frontier and R&D as measured by patents are important.

Our study differs from Badinger and Tondl (2005) in several respects. First, corresponding to our theoretical framework, we compute a regional TFP index from the rich dataset we have available and use it to construct our dependent variable. In addition, the data on human capital we compile allow us to investigate a longer and more recent time period than Badinger and Tondl (2005). Finally, we account for region-specific fixed effects in estimation by making use of the panel data framework.

These points also apply in comparison with much of the regional literature reviewed previously in this section. An additional difference of our approach is that we focus on the manufacturing sector, where technology transfer and domestic absorptive capacity may be particularly relevant.

3 Theoretical Framework and Empirical Specification:

In this section, we outline a simplified version of the extended Schumpeterian framework discussed in section 2.2, following Griffith, Redding and Van Reenen (2000).¹³ We derive our empirical specification and then discuss the estimation methods used.

First, assume that each region i produces output Y at time t according to a neoclassical production function of the form

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

where K and L are physical capital and labour input, and A represents total factor productivity. The production function F_{it} is characterised by constant returns to scale and diminishing marginal returns to capital and labour. TFP (A) may differ across regions and over time, and the region with the highest level of TFP at time t defines the technological frontier at t , with associated TFP level A_{Ft} .

Next, it is assumed that TFP in equation (1) depends on the stock of (R&D) knowledge capital D , as well as on other factors B , which include technology transfer from the frontier:

$$A_{it} = \Phi(B_{it}, D_{it}). \quad (2)$$

Taking logarithms and differentiating with respect to time yields

$$\frac{\dot{A}_{it}}{A_{it}} = \nu \frac{\dot{B}_{it}}{B_{it}} + \kappa \frac{\dot{D}_{it}}{D_{it}}, \quad (3)$$

where $\kappa = (\partial Y / \partial D) \cdot (D / Y)$ is the elasticity of output with respect to the knowledge stock. The second term on the right-hand side of this equation can be simplified further by assuming that the depreciation rate of the R&D knowledge stock D is negligible,¹⁴ so that equation (3) can be rewritten in terms of R&D intensity, i.e. the ratio of R&D expenditure

¹³The full theoretical model is given in Griffith et al. (2003), based on Aghion and Howitt (1992).

¹⁴Let D accumulate according to $\dot{D}_{it} = R_{i,t-1} - \varphi D_{i,t-1}$, where $R_{i,t-1}$ denotes R&D investment and φ is the rate of depreciation of D .

to output:

$$\frac{\dot{A}_{it}}{A_{it}} = \nu \frac{\dot{B}_{it}}{B_{it}} + \rho \left(\frac{R_i}{Y_i} \right)_{t-1}, \quad (4)$$

where $\rho = \partial Y / \partial D$ is the rate of return to R&D. Equation (4) expressed in discrete time is

$$\Delta \ln A_{it} = \nu \Delta \ln B_{it} + \rho \left(\frac{R_i}{Y_i} \right)_{t-1}. \quad (5)$$

$\Delta \ln B_{it}$ comprises the effect of technology transfer from the frontier on non-frontier productivity growth. The greater is region i 's distance to the frontier in terms of TFP levels, $\ln \left(\frac{A_F}{A_i} \right)$, the greater is its potential for TFP growth through autonomous technology transfer. The model therefore allows for convergence in relative TFP levels. We thus rewrite equation (5) as follows:

$$\Delta \ln A_{it} = \delta \ln \left(\frac{A_F}{A_i} \right)_{t-1} + \rho \left(\frac{R_i}{Y_i} \right)_{t-1} + u_{it}, \quad (6)$$

where δ is the rate of technology transfer or technology convergence, and u_{it} is a mean-zero error term.

Equation (6) contains two of the three key elements of the extended Schumpeterian framework developed in Griffith et al. (2003): first, the direct (innovation) effect of R&D on productivity growth (ρ), as in Aghion and Howitt (1992), which operates in all regions, i and F ; and second, growth through autonomous technology transfer (δ) for regions behind the frontier. The final element is the indirect (imitation) effect of R&D, which raises TFP growth in regions behind the frontier by improving their ‘‘absorptive capacity’’. This effect may be introduced by allowing the rate of technology transfer to depend on R&D activity, so that $\delta = \delta_1 + \delta_2 \left(\frac{R_i}{Y_i} \right)_{t-1}$. Equation (6) then becomes

$$\Delta \ln A_{it} = \delta_1 \ln \left(\frac{A_F}{A_i} \right)_{t-1} + \rho \left(\frac{R_i}{Y_i} \right)_{t-1} + \delta_2 \left(\frac{R_i}{Y_i} \right)_{t-1} \cdot \ln \left(\frac{A_F}{A_i} \right)_{t-1} + u_{it}. \quad (7)$$

In sections 2.1 and 2.2, we reviewed the theoretical rationale for modelling the effects of human capital on TFP growth analogously to those of R&D. Postulating that the knowledge capital stock D in (2) also has a human capital component allows for an extension of

equation (7) as below:

$$\begin{aligned} \Delta \ln A_{it} = & \delta_1 \ln \left(\frac{A_F}{A_i} \right)_{t-1} + \rho_1 \left(\frac{R_i}{Y_i} \right)_{t-1} + \rho_2 H_{i,t-1} + \\ & + \delta_2 \left(\frac{R_i}{Y_i} \right)_{t-1} \cdot \ln \left(\frac{A_F}{A_i} \right)_{t-1} + \delta_3 H_{i,t-1} \cdot \ln \left(\frac{A_F}{A_i} \right)_{t-1} + u_{it}, \end{aligned} \quad (8)$$

where $H_{i,t-1}$ represents the stock of human capital in region i at time $t - 1$. Note that equation (8) may be interpreted as an equilibrium correction model that describes adjustment to a long-run equilibrium distance $\ln \frac{A_F}{A_i}$ between TFP in frontier and non-frontier regions, with the speed of adjustment coefficient extended to depend on $\frac{R_i}{Y_i}$ and H_i .

For the frontier region F , innovation is the only source of TFP growth, so it is modelled as:

$$\Delta \ln A_{Ft} = \rho_1 \left(\frac{R_F}{Y_F} \right)_{t-1} + \rho_2 H_{F,t-1} + u_{Ft}. \quad (9)$$

In our empirical analysis, equations (8) and (9) for regions i and F are stacked together, which restricts ρ_1 and ρ_2 to be equal for non-frontier and frontier regions. We also investigate the sensitivity of our results to dropping the frontier regions from the sample.

The long-run equilibrium level of relative TFP implied by the model can be derived by first subtracting equation (8) from equation (9), which yields an expression for the change over time in the distance between frontier and non-frontier regions:

$$\begin{aligned} \Delta \ln \left(\frac{A_F}{A_i} \right)_t = & \rho_1 \left[\left(\frac{R_F}{Y_F} \right)_{t-1} - \left(\frac{R_i}{Y_i} \right)_{t-1} \right] + \rho_2 (H_{F,t-1} - H_{i,t-1}) - \\ & - \left[\delta_1 + \delta_2 \left(\frac{R_i}{Y_i} \right)_{t-1} + \delta_3 H_{i,t-1} \right] \cdot \ln \left(\frac{A_F}{A_i} \right)_{t-1} + (u_{Ft} - u_{it}). \end{aligned} \quad (10)$$

In the long-run or steady-state equilibrium, all right-hand side variables are constant over time, and $\Delta \ln \left(\frac{A_F}{A_i} \right) = 0$. That is, $\Delta \ln A_F = \Delta \ln A_i$, so that all regions grow at the constant equilibrium rate of TFP growth in the frontier. The frontier is the region where TFP grows fastest through innovation alone, i.e. the one with the highest levels of R&D and human capital as in equation (9); endogenous switchovers of technological leadership between regions are possible. The steady-state level of relative TFP between frontier and non-frontier regions is such that TFP growth from innovation and imitation combined in

non-frontier regions exactly equals growth from innovation alone in the frontier.

Setting $\Delta \ln \left(\frac{A_F}{A_i} \right) = 0$ on the left-hand side of equation (10) and dropping the error terms on the right-hand side, the long-run equilibrium level of relative TFP is obtained as

$$\ln \left(\frac{A_F}{A_i} \right)^* = \frac{\left(\rho_1 \frac{R_F}{Y_F} + \rho_2 H_F \right) - \left(\rho_1 \frac{R_i}{Y_i} + \rho_2 H_i \right)}{\delta_1 + \delta_2 \frac{R_i}{Y_i} + \delta_3 H_i} \quad (11)$$

The two terms in the numerator encompass those elements that promote innovation in regions at and behind the frontier, respectively. More domestic R&D, for example, reduces region i 's equilibrium distance to the frontier; more R&D in the frontier region raises this distance, since for TFP growth due to (now higher) innovation at the frontier to equal TFP growth due innovation and imitation combined in non-frontier regions, region i must be further behind the frontier and thus be able to grow faster through imitation. The terms in the denominator represent imitation (technology transfer) for regions behind the frontier. The higher the rate of technology transfer, both autonomous (δ_1) and via R&D and human capital (δ_2 and δ_3), the smaller region i 's distance to the frontier in equilibrium.

3.1 Empirical Specification and Estimation Methods:

We estimate a specification that stacks together equations (8) and (9), thus imposing common coefficients on the R&D and human capital terms in non-frontier and frontier regions. We use the panel data framework to allow for unobservable region-specific “fixed” effects that remain constant over time and may be correlated with other explanatory variables, such as economic institutions or physical geography. Our estimating equation is

$$\begin{aligned} \Delta \ln A_{it} = & \delta_1 \ln \left(\frac{A_F}{A_i} \right)_{t-1} + \rho_1 \left(\frac{R_i}{Y_i} \right)_{t-1} + \rho_2 H_{i,t-1} + \\ & + \delta_2 \left(\frac{R_i}{Y_i} \right)_{t-1} \cdot \ln \left(\frac{A_F}{A_i} \right)_{t-1} + \delta_3 H_{i,t-1} \cdot \ln \left(\frac{A_F}{A_i} \right)_{t-1} + u_{it} \end{aligned} \quad (12)$$

$$u_{it} = \mu_i + \eta_t + v_{it},$$

where μ_i are region-specific fixed effects and η_t are unobserved period-specific effects that are common to all regions but vary over time, such as macroeconomic shocks. We control

for the latter by including time dummies in estimation. v_{it} is a mean-zero disturbance.

3.1.1 Dynamic Panel Data Estimators:

Equation (12) can be equivalently represented as a dynamic model for $\ln A_{it}$, with all terms in $\ln A_{i,t-1}$ on the right-hand side. This lagged dependent variable is by construction correlated with the region-specific fixed effects μ_i , so that the standard pooled OLS (POLS) estimator suffers from endogeneity bias and is inconsistent. Similarly, the within-groups (WG) estimator has been shown to be inconsistent in the presence of a lagged dependent variable in panels with a small number of time periods (Nickell, 1981). Since the time dimension we have available ($T = 9.7$)¹⁵ is moderately short, this problem may be relevant for our application.

The POLS and WG estimates of the coefficient on the lagged dependent variable $\ln A_{i,t-1}$ are likely to be biased in opposite directions - upwards in the case of POLS and downwards in the case of WG (Bond, 2002). Therefore, these two estimators provide useful benchmarks against which to evaluate results obtained with other methods.

Equation (12) may, under certain conditions, be estimated consistently using the first-differenced and system-GMM estimators developed in Arellano and Bond (1991), Arellano and Bover (1995) and Blundell and Bond (1998). Both estimators remove the time-invariant fixed effect μ_i by first-differencing the equation. While first-differenced GMM (FD-GMM) is based on equation (12) in first differences only, the system-GMM estimator (S-GMM) uses a combination of this equation in first differences *and* in levels.

Both estimators address the endogeneity of some explanatory variables by employing lagged levels dated $t - 2$ and earlier as instrumental variables in the first-differenced equations (FD-GMM) as well as lagged first differences as instrumental variables in the levels equations (S-GMM). On the one hand, this allows us to deal with the first-differenced lagged dependent variable, which is by construction correlated with the first-differenced error term Δv_{it} . On the other hand, some explanatory variables may be contemporaneously correlated with v_{it} in equation (12). In particular, there is some concern that our

---

¹⁵This is the average number of time periods used in estimation in section 5, resulting from missing values in our measures of R&D and human capital (see section 4).

proxy for R&D activity measures this variable with error (see section 4.4), which could induce such correlation.

Consistency of the FD- and S-GMM estimators requires the instruments they employ to be valid and informative. An important condition for the validity of some of our instruments is that the error term v_{it} is serially uncorrelated. We investigate this using Arellano and Bond's (1991) test for serial correlation in the first-differenced regression residuals. As the number of available instruments is greater than the number of explanatory variables, the Sargan (1958)/Hansen (1982) test of overidentifying restrictions provides an additional tool for assessing the validity of the instruments.

Because some of the series we use in estimation exhibit a high degree of persistence over our sample period, the instruments for the first-differenced equations may only be weakly informative.¹⁶ This could lead to considerable finite-sample bias and imprecision in the FD-GMM estimates, where the bias in the coefficient estimate on the lagged dependent variable is likely to be downward, in the direction of WG. The S-GMM estimator counteracts the weak instruments problem in FD-GMM by introducing additional instruments for the levels equations. Overall therefore, S-GMM may be preferred.¹⁷

To limit the number of instruments used per equation in FD- and S-GMM, we restrict the lag length of the instruments used for the first-differenced equations, rather than using all available lags from $t - 2$ onwards. Since the number of time periods we have available is not very small, using all available lags as instruments may overfit the estimated equation and lead to small-sample bias.

3.1.2 Geographic Distance to the Frontier:

So far, our model assumes that knowledge spillovers from the frontier are global in the sense that they do not depend on the geographic distance between the frontier and the

---

¹⁶Estimating AR(1) models for each of our variables indicates that our measures of $\ln A_{it}$ and $\ln \left(\frac{A_{Ft}}{A_{it}} \right)$ in particular tend to be persistent: the estimated autoregressive coefficient is 0.98 when using POLS, 0.77 when using WG, and above 0.8 for both GMM estimators.

¹⁷Monte Carlo simulations conducted by Blundell and Bond (1998) for a univariate AR(1) model with persistent series indicate that the gains in terms of bias and precision from using the S-GMM estimator are substantial for T as large as 11. Simulations by Blundell, Bond and Windmeijer (2000) for the multivariate case suggest that this result extends to a model with additional explanatory variables, which is more relevant to equation (12).

regions receiving the spillovers. In equation (12), only technological distance to the frontier matters for a region's potential for technology transfer. However, while studies on the international diffusion of technology, such as Coe and Helpman (1995) and Eaton and Kortum (1999), have shown that knowledge spillovers across countries are substantial, there is also considerable evidence that they decline with geographic distance (e.g. Jaffe et al. 1993, Bottazzi and Peri 2003).¹⁸

Therefore, we consider an extension to equation (12) that allows a region's potential for growth through (autonomous) technology transfer to depend also on its geographic distance from the frontier. We capture this by introducing a geographically weighted version of technological distance to the frontier, $\left(w_{Fi} \cdot \ln \frac{A_F}{A_i}\right)_{t-1}$. The weights are defined as

$$w_{Fi} = \frac{d_{Fi}^{-2}}{\sum_i d_{Fi}^{-2}},$$

where d_{Fi} is the great-circle distance between the capitals of the frontier region F and region i .¹⁹ The weights w_{Fi} represent region i 's inverse squared distance from the frontier - where the frontier may vary over time - standardised by the total across regions.²⁰

A positive coefficient on $\left(w_{Fi} \cdot \ln \frac{A_F}{A_i}\right)_{t-1}$ would suggest that for regions at a given technological distance to the frontier, those that are geographically closer to it are able to exploit their potential for technology transfer more easily and catch up faster than more distant regions. This extension therefore allows us to investigate whether knowledge spillovers from the frontier are geographically localised. Econometrically, the additional term $\left(w_{Fi} \cdot \ln \frac{A_F}{A_i}\right)_{t-1}$ can be straightforwardly incorporated into our estimation strategy outlined in section 3.1.1.

---

¹⁸One explanation in the literature is the tacit nature of knowledge at the research frontier, which is new and complex and hence difficult to communicate or codify. Its transmission across regions and countries depends on personal contacts within the scientific community, which may in turn be facilitated by geographic proximity.

¹⁹The great-circle or geodesic distance between two points on earth is the shortest distance between these points measured along a path on the surface of the earth.

²⁰ w_{Fi} thus resembles a typical element of a row-standardised spatial weight matrix as commonly used in the spatial econometrics literature.

4 Data and Variables:

The data we use in this paper come from the 2007 edition of the Cambridge Econometrics (CE) European Regional Database and from Eurostat. From the CE database, we use the series on sector-level regional gross value added, gross fixed capital formation, employment, hours worked and employee compensation to calculate regional total factor productivity in the manufacturing sector. Data on regional patent applications, our measure of regional R&D activity, are taken from Eurostat’s REGIO database. Regional human capital data come from the European Union Labour Force Survey and were supplied by Eurostat directly. The latter are available for most regions only from 1992 onwards, so our analysis begins with that year.

The sample covers 159 regions at NUTS level 2 from eleven of the EU-15 countries over the period 1992 to 2005.²¹ For Denmark, education data are only available at the national level, so we use country-level data. We do not consider the French, Spanish and Portuguese small island territories in the Atlantic, the Caribbean and the Indian ocean because of their geographical remoteness from the European continent, nor Ceuta and Melilla, two small Spanish exclaves on the Moroccan coast. Further, we exclude eight East German regions, both Irish regions and seven UK regions, for which there are no data on human capital and/or patenting over the sample period. We also drop seven regions with very low economic activity in manufacturing as well as 21 regions from Luxembourg, Austria, Portugal, Finland and Sweden, for which there are issues with the human capital data (see sections 4.2 and 4.3). A list of included regions is provided in Appendix A.

The remainder of this section discusses measurement and construction of the individual variables. Further details as well as summary statistics are given in Appendix B. Appendix C illustrates the spatial distribution of the main variables across all regions in the sample.

---

²¹NUTS, the French acronym for Nomenclature of Territorial Units for Statistics, is the European Union’s regional classification system. Level 2 is recommended by Eurostat as the appropriate unit for analysing regional economic issues. The NUTS-2 regions in our sample are from Belgium (11 regions), Denmark (1), Finland (3), France (21), Germany (33), Greece (9), Italy (20), Netherlands (12), Portugal (3), Spain (16), and the United Kingdom (30). We do not consider the 12 countries that joined the EU in 2004 and 2007 because of limited data availability, especially in Eastern Europe during the early years of our sample.

4.1 Measuring Total Factor Productivity:

We construct an index of regional manufacturing TFP according to the superlative index number approach developed in e.g. Diewert (1976) and Caves, Christensen and Diewert (1982), which builds on the growth accounting approach pioneered by Solow (1957).²² In the latter framework, the starting point is an aggregate production function, the properties of which then define the resulting productivity index. In particular, consider the production function we used in equation (1), where technology A raises output in a Hicks-neutral fashion:

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

Totally differentiating this function with respect to time and dividing through by Y yields an expression for the growth rate of output as a function of the weighted growth rates of inputs and of technology:

$$\frac{\dot{Y}_{it}}{Y_{it}} = \frac{\partial Y}{\partial K} \frac{K_{it}}{Y_{it}} \frac{\dot{K}_{it}}{K_{it}} + \frac{\partial Y}{\partial L} \frac{L_{it}}{Y_{it}} \frac{\dot{L}_{it}}{L_{it}} + \frac{\dot{A}_{it}}{A_{it}}$$

Assuming that markets are competitive implies that the factor inputs are paid their marginal products, and thus the weights on the growth rates of inputs can be replaced by the factor shares in total income Y . TFP growth can therefore be expressed as below:

$$\frac{\dot{A}_{it}}{A_{it}} = \frac{\dot{Y}_{it}}{Y_{it}} - s_{it}^K \frac{\dot{K}_{it}}{K_{it}} - s_{it}^L \frac{\dot{L}_{it}}{L_{it}}, \quad (13)$$

where $s^K = \frac{rK}{Y}$ and $s^L = \frac{wL}{Y}$ are the shares of capital and labour in total income, and $r = \frac{\partial Y}{\partial K}$ and $w = \frac{\partial Y}{\partial L}$ are the real rental rate of capital and the real wage rate, respectively. Assuming in addition that the production function is characterised by constant returns to scale implies that $s^K + s^L = 1$, and since labour shares can be computed from wage data, equation (13) can be re-written as

$$\frac{\dot{A}_{it}}{A_{it}} = \frac{\dot{Y}_{it}}{Y_{it}} - (1 - s_{it}^L) \frac{\dot{K}_{it}}{K_{it}} - s_{it}^L \frac{\dot{L}_{it}}{L_{it}}. \quad (14)$$

---

²²In using the superlative index number approach, we follow Griffith et al. (2004), Cameron et al. (2005) and Griffith et al. (2009). For a methodological guide, see OECD (2001). Islam (1999) and Hulten (2001) provide historical surveys of TFP measurement.

Solow was thus able to derive a simple expression for TFP growth without imposing a particular functional form on the production function. $\frac{\dot{A}}{A}$ is an index number that can be calculated directly from indices of output and inputs, and it is also called the Solow residual. It captures the proportion of output growth that cannot be explained by the growth of inputs, and therefore it may contain not only technological change, but also other factors like cyclical effects, adjustment costs, economies of scale and measurement error. Accordingly, Abramovitz (1956) referred to the residual as a “measure of our ignorance”.

Empirical implementation of equation (14) requires the resolution of two remaining issues. First, we require a formulation in discrete rather than continuous time; and second, the framework laid out above is valid for within-region time-series TFP comparisons, but it does not allow the comparison of TFP levels *across* regions. Regarding the first point, (14) is the growth rate of a Divisia index, which can be approximated in discrete time by the Törnqvist index as follows:

$$\ln \left(\frac{A_{it}}{A_{i,t-1}} \right) \approx \ln \left(\frac{Y_{it}}{Y_{i,t-1}} \right) - (1 - \bar{s}_{it}^L) \ln \left(\frac{K_{it}}{K_{i,t-1}} \right) - \bar{s}_{it}^L \ln \left(\frac{L_{it}}{L_{i,t-1}} \right), \quad (15)$$

where $\bar{s}_{it}^L = \frac{1}{2} (s_{it}^L + s_{i,t-1}^L)$ is the average labour share in periods t and $t - 1$. Diewert (1976) shows that equation (15) holds exactly if the underlying production function takes the translog form, which is a very general functional form that approximates any twice-differentiable constant-returns-to-scale production function to the second order.²³ In Diewert’s terminology, the Törnqvist index is “superlative” in the sense that it is “exact for” (i.e. can be derived from) a “flexible” functional form (i.e. one that can provide a second-order approximation to any arbitrary twice-differentiable linear homogeneous (CRS) function).

We use equation (15) to compute the growth rate of TFP in region i , $\Delta \ln A_{it}$, on the left-hand side of our empirical specification (12). In our empirical analysis in section 5, this superlative index number measure of $\Delta \ln A_{it}$ is denoted as $\Delta \ln TFP_{it}$.

The superlative index number approach also offers a solution to the second problem encountered in empirical TFP measurement, namely the comparison of TFP levels across

---

²³The transcendental logarithmic, or translog, production function was developed by Christensen, Jorgenson and Lau (1973). The Cobb-Douglas production function, for example, is a special case of the translog function.

regions. Caves et al. (1982) extend the framework above to a multilateral (e.g. multi-region) setting and define the level of TFP in any region i (here also F) relative to the geometric mean across all regions, based on the assumption of translog production functions, as below:

$$\ln \left(\frac{A_{it}}{\bar{A}_t} \right) = \ln \left(\frac{Y_{it}}{\bar{Y}_t} \right) - (1 - \bar{s}_{it}^L) \ln \left(\frac{K_{it}}{\bar{K}_t} \right) - \bar{s}_{it}^L \ln \left(\frac{L_{it}}{\bar{L}_t} \right), \quad (16)$$

where $\bar{Y}_t$, $\bar{K}_t$ and $\bar{L}_t$ represent the geometric means of output, capital and labour across all i at time t , and $\bar{s}_{it}^L = \frac{1}{2} (s_{it}^L + \bar{s}_t^L)$ is the average of the labour share in region i and the geometric mean labour share. Caves et al. (1982) establish that if TFP levels for any two regions such as F and i are measured in this way, the superlative index number for their difference,

$$TFPGAP_{it} = \ln \left(\frac{A_{Ft}}{\bar{A}_t} \right) - \ln \left(\frac{A_{it}}{\bar{A}_t} \right), \quad (17)$$

has desirable properties.²⁴ $TFPGAP_{it}$ is our measure for the terms representing distance to the TFP frontier, $\ln \left(\frac{A_F}{A_i} \right)$, on the right-hand side of our empirical equation (12).

The superlative index number approach requires making the restrictive assumptions of competitive markets and constant returns to scale. An alternative approach to TFP measurement involves estimating the (aggregate) production function, either to obtain estimates of the factor elasticities with which TFP levels can be computed using data on capital and labour inputs, or to obtain TFP estimates directly by using panel data techniques (as in Di Liberto and Usai, 2010). This approach, however, usually imposes that the parameters to be estimated are common to all regions and/or constant over time, while the index number approach avoids this by using data on factor income shares. In addition, consistent estimation of production functions may be challenging, especially when there is correlation between factor inputs and the error term. A final advantage of the superlative index number approach arises when the true (aggregate) production function is not Cobb-Douglas, as the translog functional form is very general.

---

²⁴When comparing TFP levels between any two regions, the index is invariant to the base region chosen and transitive.

4.2 Calculating Regional Total Factor Productivity:

To measure regional output in the manufacturing sector, we use GVA at constant prices in 2000 euros from the CE database, which we adjust for differences in price levels across countries via national purchasing power parities (PPPs) defined relative to the EU average.²⁵ Inspection of the resulting output series reveals seven regions with very little economic activity in the manufacturing sector: the share of manufacturing output in total output is below 5% in Corsica, Sicily, the Algarve region in Portugal, and four Greek regions (Ionian Islands, Crete, and North and South Aegean). They are therefore excluded from the sample.

As our measure of capital input, we construct regional physical capital stocks from data on gross fixed capital formation (GFCF) for the manufacturing sector, expressed in PPP-adjusted 2000 euros, using the perpetual inventory method (PIM). This approach yields the following expression for the capital stock of region i at time t :

$$K_{it} = (1 - \delta)K_{i,t-1} + I_{it}, \quad (18)$$

where δ is the constant annual rate of depreciation and I_{it} is the flow of GFCF in period t . We assume that δ is equal to 0.06 for all regions and years.²⁶ To implement the PIM, an estimate of the initial-period capital stock is required. Since our capital stock series will be sensitive to this estimate, we use the full available time series for investment (from 1980 to 2005) to calculate the initial capital stock: assuming that capital accumulation prior to 1980 is also governed by equation (18), the capital stock in 1979 may be approximated as $K_{i,1979} = I_{i,1980}/(g_i + \delta)$, where g_i is the region-specific average annual growth rate of GFCF from 1980 to 2005. For our empirical analysis, which begins in 1992, any measurement error associated with the initial value of the capital stock should therefore be minimised.

We compute total annual hours worked for all persons in employment to measure regional labour input. To this end, we multiply employment in the manufacturing sector

---

²⁵The resulting artificial common European currency unit is the Purchasing Power Standard (PPS), where one PPS equals the average purchasing power of one euro across the European Union.

²⁶As long as depreciation rates remain relatively constant over time for each region, our estimation results should not be affected by our choice of δ , since it will be captured by the region-specific fixed effects.

with the average number of usual weekly hours worked per person in employment in that sector, both from the CE database, and the resulting figure with 52.

The regional share of labour in total income is calculated from data on labour income in the manufacturing sector from the CE database. We use their series on total annual employee compensation, which includes wages and salaries as well as employers' social contributions. The labour share s_{it}^L is obtained by dividing this series by regional manufacturing GVA. Assuming constant returns to scale, the capital share then equals one minus the labour share. We use these measures to construct $\bar{s}_{it}^L$ and $\tilde{s}_{it}^L$ in equations (15) and (16) above.

Since the labour share s_{it}^L thus computed is quite variable over time, we examined the robustness of our results to setting it equal to 0.67 for all regions and time periods, a value that is consistent with findings based on macroeconomic data for advanced countries (e.g. Gollin, 2002). This did not change our overall conclusions in this paper.

4.3 Regional Human Capital:

We follow the literature reviewed in section 2.3 and measure the stock of regional human capital H_{it} in equation (12) as the educational attainment of the working-age population. Specifically, we use the percentage of the population that has attained higher (tertiary) education, which we construct from information on the highest level of education or training successfully completed by those aged 15 and above from the European Union Labour Force Survey (EU LFS). These data are available from 1992 onwards, although not for all countries.²⁷

In terms of the Schumpeterian models described in section 2.1, higher educational attainment corresponds most closely to the human-capital dimension of the amount of resources devoted to research, one key determinant of technological progress, which is a highly skill-intensive activity. This measure is also employed in the three main studies that

---

²⁷Data over our full sample period 1992-2005 are available for the regions of Belgium, Denmark, France, Germany, Greece, Luxembourg, Spain, for two Portuguese regions and one UK region (Northern Ireland); from 1995-2005 for Austria and two, respectively six, regions of Finland and Sweden; from 1996-2005 for the Netherlands and the rest of the UK regions; 1998-2005 for 19 Italian regions; and 1999-2005 for Ireland and the remaining regions from Italy, Finland, Portugal and Sweden.

have investigated the extended Schumpeterian framework empirically, GRVR and CPR at the country-industry level and Badinger and Tondl (2005) at the level of European NUTS-2 regions.

The EU LFS provides information on educational attainment by several broad levels according to the International Standard Classification of Education (ISCED). For example, the version that applies for data since 1998, ISCED 1997, defines six such levels. Following guidelines by Eurostat, we aggregate these further into low, medium and high levels of education, corresponding respectively to lower secondary, upper secondary and tertiary education.

In 1998, the old ISCED classification dating from 1976 was replaced with ISCED 1997, introducing a structural break into the series. One major difference between the two conventions, illustrated in Appendix D, is the introduction of an additional education level in ISCED 1997. It captures educational programmes situated at the boundary between upper secondary and tertiary education from an international perspective such as certain kinds of short vocational courses. Under ISCED 1976, these were assigned to either upper secondary or tertiary education, so that both levels may be affected by the classification change.

Careful inspection of the time series for the periods 1992-1997 and 1998-2005 revealed that the break surrounding the year 1998 is small and mainly affects the medium attainment series, leaving the high attainment series intact for most countries. For 21 regions of Luxembourg, Austria, Portugal, Finland and Sweden, this is not the case, so they are dropped. For the remainder, we combine the two time periods, enabling us to analyse a comparatively long time series of regional educational attainment for the EU.

4.4 Regional R&D:

We use the ratio of regional patent applications to the European Patent Office (EPO) to employment in the manufacturing sector (in 10,000s), $(P/L)_{it}$, to proxy $(R/Y)_{it}$ in equation (12). The number of patent applications to the EPO is the indicator related to R&D activity with the most comprehensive coverage of regions and time periods in the

Eurostat REGIO database, and it has been employed by many region-level studies including Badinger and Tondl (2005). There are missing values for some Greek and Spanish regions but otherwise, the data are complete over the sample period.

Data on R&D expenditure, the measure used in the country-industry level studies by GRVR and CPR, are available consistently over our sample period only for the regions of France and Spain, and, with relatively few gaps, for Italy. For other countries, the series start only recently (e.g. in 2002 for the UK), have more gaps than observations (Germany), or exist only at the more aggregate NUTS-1 level (Belgium). Other indicators available in REGIO are R&D personnel, human resources in science and technology, or employment in high-tech sectors, but none of these presently cover the full sample period.

Patents represent an output of the innovation process. As a measure of R&D activity, they may therefore not be an ideal substitute for R&D expenditure or employees, which are an input into this process. However, the correlation between R&D expenditure and patents is frequently found to be strong, also at the EU regional level (Paci and Pigliaru, 2002). In addition, patent applications may be seen as an intermediate rather than a final output of the innovation process, since not all patented inventions find successful commercial implementation.²⁸ To allow for the possibility that patent applications measure R&D activity with error, we treat the variable as endogenous in estimation.

Patents currently take four years on average to be granted by the EPO. Therefore, Eurostat provides data on patent applications rather than patents granted, as the former are more closely linked to the date of invention. Patent applications to the EPO are counted by priority year, which is the year in which the application was first filed.²⁹ At the regional level, they are allocated according to the place of residence of the inventor. Applications with more than one inventor are divided equally between their places of residence to avoid double counting.

Patents are classified by technological field according to the International Patent Classification (IPC) rather than by economic activity in the REGIO database. We use the total

---

²⁸See Acs, Anselin and Varga (2002). This is reflected in the skewed distribution of patent value: a few patents are very valuable, but many are not.

²⁹For example, an application first filed at a national patent office can, within 12 months, be extended to the EPO. There, the application is assigned the date of the national application, the “priority date”.

number of patent applications across all sectors, although we would ideally like to capture patenting in manufacturing only. Since manufacturing tends to be the sector with the largest number of patents, this may in fact be less of a problem. Allowing for measurement error in the patents variable should also help to overcome it.

5 Estimation Results and Discussion:

In Table 1, we present the results of estimating equation (12) using pooled OLS, within groups, and the first-differenced and system-GMM estimators.³⁰ $\Delta \ln TFP_{it}$ denotes our superlative index number measure of the dependent variable $\Delta \ln A_{it}$, as described in section 4.1. Similarly, $TFPGAP_{i,t-1}$ denotes our measure of distance to the TFP frontier, $\ln \left(\frac{A_F}{A_i} \right)_{t-1}$, on the right-hand side. A full set of time dummies is included in all regressions to control for unobserved period-specific effects that are common to all regions.

At first glance, the pooled OLS estimates in column (i) do not appear very supportive of our model, as only the coefficient on the interaction between the TFP gap term and our measure of R&D activity, P/L , is significantly different from zero. This would suggest that technology transfer from the frontier depends more heavily on a region's own patenting activity rather than occurring autonomously or as a function of human capital. The positive sign of the coefficient estimate, $\hat{\delta}_2$, implies that for regions that lie a given distance behind the frontier, TFP growth is faster in those regions with higher domestic patenting activity.

In column (ii), on the other hand, the WG estimate of the coefficient on the linear TFP gap term, $\hat{\delta}_1$, is also highly significant and positive. Thus, the further a region lies behind the frontier, i.e. the larger is its TFP gap, the faster it should grow in transition to its long-run equilibrium TFP distance from the frontier. This finding is consistent with the idea that conditional convergence in relative TFP levels has taken place between frontier and non-frontier regions over our sample period. $\hat{\delta}_2$ remains positive and significant. As in column (i), there is no evidence of a significant indirect effect of human capital on productivity growth (δ_3), nor of significant direct effects of either patents or human capital.

---

³⁰We report results based on the sample including both frontier and non-frontier regions. Dropping the frontier regions has a negligible effect on the estimates; these are available on request.

The size of the WG estimate of δ_1 differs markedly from its OLS counterpart, which indicates that allowing for unobserved region-specific fixed effects may be important. In

Table 1: The Two Faces of R&D and Human Capital

Dependent variable:	(i)	(ii)	(iii)	(iv)
$\Delta \ln TFP_{it}$	POLS	WG	FD-GMM	S-GMM
$TFPGAP_{i,t-1}$	0.022 (0.014)	0.256*** (0.039)	0.175*** (0.056)	0.074* (0.044)
$(P/L)_{i,t-1}$	-0.029 (0.020)	-0.036 (0.032)	-0.017 (0.040)	-0.031 (0.035)
$H_{i,t-1}$	0.114 (0.079)	-0.296 (0.254)	-0.552 (0.382)	0.267 (0.270)
$(P/L \cdot TFPGAP)_{i,t-1}$	0.074*** (0.022)	0.082*** (0.031)	0.014 (0.038)	0.068* (0.041)
$(H \cdot TFPGAP)_{i,t-1}$	-0.017 (0.075)	0.072 (0.229)	0.489 (0.371)	-0.049 (0.317)
$\hat{\delta}_1 + \hat{\delta}_2 \cdot \overline{P/L} + \hat{\delta}_3 \cdot \overline{H}$	0.040*** (0.005)	0.291*** (0.027)	0.260*** (0.047)	0.085*** (0.023)
Test $\hat{\delta}_1 = \hat{\delta}_2 = \hat{\delta}_3 = 0$	28.93 (0.000)	43.90 (0.000)	34.70 (0.000)	23.38 (0.000)
AB-AR(1)	-1.41 (0.159)	-2.04 (0.042)	-4.03 (0.000)	-4.16 (0.000)
AB-AR(2)	-0.73 (0.468)	-3.26 (0.001)	-0.40 (0.689)	-0.52 (0.606)
Hansen J (p-value)			0.147	0.282
Dif-Hansen (p-value)				0.722
Time Dummies	Yes	Yes	Yes	Yes
Observations	1538	1538	1368	1538
Number of Regions	159	159	159	159
Number of Instruments			111	156

Notes: Standard errors, reported in parentheses, are robust to heteroskedasticity and serial correlation; for POLS and WG, they are Huber-White standard errors clustered on regions; GMM estimators are two-step estimators with standard errors corrected for small-sample bias as suggested by Windmeijer (2005); ***, **, and * indicate significance at 1%, 5% and 10% levels; AB-AR(1) and AB-AR(2) are Arellano and Bond's (1991) tests of first- and second-order residual serial correlation, asymptotically standard normal under the null of no serial correlation; Hansen J and Dif-Hansen are the p-values of the Hansen (1982) and Difference Hansen tests of m overidentifying restrictions, asymptotically $\chi^2(m)$ under the null that the overidentifying restrictions are valid; p-values in parentheses for all tests.

Instruments used for the first-differenced equations in columns (iii) and (iv) are $TFPGAP_{i,t-2}$, $(P/L \cdot TFPGAP)_{i,t-2}$, $(H \cdot TFPGAP)_{i,t-2}$, $(P/L)_{i,t-2}$ and all further lags, and $\Delta H_{i,t-1}$. Additional instruments used for the levels equations in column (iv) are $\Delta TFPGAP_{i,t-1}$, $\Delta(P/L \cdot TFPGAP)_{i,t-1}$, $\Delta(H \cdot TFPGAP)_{i,t-1}$, $\Delta(P/L)_{i,t-1}$ and $\Delta H_{i,t-1}$. $\Delta H_{i,t-1}$ is implemented as an "ivstyle" instrument in Roodman's (2009) `xtabond2` command for Stata.

this case, the OLS estimate of the coefficient on the lagged dependent variable in a dynamic model is likely to be biased upwards, while the WG estimate is likely to be biased downwards for moderately short T . In our model, the lagged dependent variable $\ln TFP_{i,t-1}$ appears in $TFPGAP_{i,t-1}$ as well as in its interactions with patents and human capital. Therefore, we also report the sum of the coefficients on these variables, evaluated at the sample means of $(P/L)_{it}$ and H_{it} , $\hat{\delta}_1 + \hat{\delta}_2 \cdot \overline{P/L} + \hat{\delta}_3 \cdot \overline{H}$ (see Table B.1 in the Appendix for the means). These sums are significantly different from zero at the 1% level in columns (i) and (ii). Their magnitude is consistent with the likely direction of bias in OLS and WG described above, supporting the view that these estimators may be inconsistent.³¹

Arellano and Bond’s (1991) tests of serial correlation, AB-AR(1) and AB-AR(2), find no evidence of serial correlation in the OLS residuals, but they detect significant first- and second-order serial correlation in the residuals of the WG specification.³² If the WG estimator is biased however, the diagnostic tests in column (ii) are unlikely to be reliable.

For the FD- and S-GMM estimates in columns (iii) and (iv), we use lagged levels dated $t-2$ of $TFPGAP_{it}$ and the interaction terms as instruments for the first-differenced equations, for reasons discussed in section 3.1.1.³³ Furthermore, we treat H_{it} as predetermined with respect to v_{it} .³⁴ Given the lag structure of our model, this makes possible a parsimonious choice of instruments: it allows us to treat $\Delta H_{i,t-1}$, which appears on the right-hand side of equation (12) in first differences, as exogenous. Hence, we include $\Delta H_{i,t-1}$ in the instrument set for the first-differenced equations.³⁵ On the other hand, we treat $(P/L)_{it}$ as endogenous with respect to v_{it} to allow for measurement error. This relatively weak assumption permits the use of lagged levels dated $t-2$ and earlier of $(P/L)_{it}$ as instruments in the first-differenced equations. As additional instruments for the equations in levels in

---

³¹Notice that $\ln TFP_{i,t-1}$ enters $TFPGAP_{i,t-1}$ with a negative sign, so a smaller OLS coefficient indicates upward bias and a larger WG coefficient indicates downward bias.

³²The AB-AR tests for WG are obtained from the least-squares dummy-variables variant of the estimator and are thus carried out on v_{it} , which we would hope to be serially uncorrelated.

³³All GMM estimators are implemented as two-step estimators with standard errors corrected for small-sample bias as suggested by Windmeijer (2005). The GMM estimates in this paper are computed using Roodman’s (2009) `xtabond2` command for Stata.

³⁴This assumption is not unreasonable since one would expect current shocks to TFP growth v_{it} to affect the proportion of the population with tertiary educational attainment only with a considerable lag. Results treating H_{it} as endogenous to allow for inward or outward migration in response to shocks are similar and available upon request.

³⁵We implement this by using the option “`ivstyle`” instead of “`gmmstyle`” for the H_{it} instruments in Roodman’s (2009) `xtabond2` command for Stata.

column (iv), we use first differences of all variables dated $t - 1$.

The Hansen and Difference Hansen tests in columns (iii) and (iv) do not reject the validity of our instrument sets for the first-differenced and the levels equations, respectively. In addition, the AB-AR tests provide significant evidence of first- but not of second-order autocorrelation in the residuals of the first-differenced equations. While negative first-order serial correlation is to be expected, the absence of significant second-order residual autocorrelation is consistent with serially uncorrelated v_{it} , a condition needed for instrument validity in FD- and S-GMM.

In column (iii), some of the FD-GMM results display the typical signs of weak instruments described in section 3.1.1. For example, the sum of the estimated coefficients on all terms involving the TFP gap, $\hat{\delta}_1 + \hat{\delta}_2 \cdot \overline{P/L} + \hat{\delta}_3 \cdot \overline{H}$, is close to the WG estimate in column (ii), suggesting downward bias in the coefficient on the lagged dependent variable $\ln TFP_{i,t-1}$. Moreover, the standard errors of some coefficients are noticeably larger. This points to potential gains from using the S-GMM estimator.

The S-GMM estimates in column (iv) improve in these respects compared to FD-GMM. The coefficient sum on the TFP gap terms is now substantially smaller than in columns (ii) and (iii), but it is still more than twice the size implied by POLS in column (i). Considering the likely direction of bias in the other estimators, the S-GMM results thus appear to be the most reliable in Table 1. Therefore, column (iv) is our preferred specification in this table, and we focus on the S-GMM estimator for the remainder of this section. Both the coefficients on the linear TFP gap term as well as on its interaction with patenting are positive and significant, albeit only marginally so.

Overall, the estimates in all columns of Table 1 agree that technology transfer, measured by distance to the TFP frontier, matters in some form for transitional productivity growth in regions behind the frontier. This is underlined by the fact that $\hat{\delta}_1 + \hat{\delta}_2 \cdot \overline{P/L} + \hat{\delta}_3 \cdot \overline{H}$ is always statistically significant at the 1% level. We also report joint significance tests which reject the null hypothesis $\hat{\delta}_1 = \hat{\delta}_2 = \hat{\delta}_3 = 0$ in each column. There is less agreement across columns as to whether the effect of distance to frontier is direct (autonomous technology transfer) or indirect (depending on own patenting activity), or both. Throughout Table 1,

there is no evidence that human capital has an indirect effect on productivity growth, nor that patenting and human capital have direct effects.

Collinearity between the linear terms in patents and human capital and their interactions with the TFP gap may be a reason why the coefficients on some of these terms cannot be separately identified in Table 1.³⁶ In column (iv), tests of the hypotheses that the two coefficients on $(P/L)_{i,t-1}$ and $(P/L \cdot TFP\text{GAP})_{i,t-1}$, as well as those on $H_{i,t-1}$ and $(H \cdot TFP\text{GAP})_{i,t-1}$, are jointly zero reject the null in both cases.³⁷ Therefore, we eliminate individually insignificant variables from our preferred specification in column (iv) one at a time, beginning with the least significant. The variables we drop are $(P/L)_{i,t-1}$ and $(H \cdot TFP\text{GAP})_{i,t-1}$, which are also jointly insignificant.³⁸ The resulting more parsimonious model is presented in Table 2.

In column (i) of Table 2, all coefficient estimates are positive and highly significant. Compared to column (iv) in the previous table, the point estimates are similar while their standard errors decline, sometimes considerably. For example, the magnitudes of $\hat{\delta}_1$ and the coefficient sum $\hat{\delta}_1 + \hat{\delta}_2 \cdot \overline{P/L}$ are virtually unchanged. The coefficients on human capital and the interaction between the TFP gap and patenting in particular are more precisely estimated, with both standard errors reduced by about two-thirds. From the Hansen and Difference Hansen tests, there is no evidence that our instruments may be invalid, and the AB-AR tests also point to an absence of serial correlation in v_{it} , as in Table 1.

The results in column (i) thus provide further evidence of productivity convergence relative to the frontier occurring for the European regions between 1992 and 2005. In addition, they indicate that regions with a greater proportion of highly educated workers are closer to the frontier in the long-run equilibrium and therefore experience a higher rate of transitional TFP growth. Similarly, for regions at a given distance behind the frontier, those with higher domestic patenting activity have grown faster.

In the context of the Griffith et al. (2003) model, these results may be interpreted to suggest that the direct, or innovation, effect of human capital and the indirect, or imitation, effect of R&D activity have been more important for productivity growth in our

---

³⁶See Table B.2 in the Appendix for the correlation matrix.

³⁷For the patents terms, the p-value of the test is 0.0265, and for the human capital terms, it is 0.007.

³⁸The p-value of the test in column (iv) is 0.6722.

sample of regions than the indirect effect of human capital and the direct effect of R&D. The recent emphasis of EU regional policy on raising both educational attainment and R&D performance therefore seems to be well placed, although their channels of influence on regional productivity growth appear to differ. This could be taken into account by policy-makers.

Table 2: Parsimonious S-GMM, Robustness and Geographic Proximity

Dependent variable: $\Delta \ln TFP_{it}$	(i) S-GMM	(ii) Alternative frontier region	(iii) Smaller instrument set	(iv) Distance-weighted TFP Gap
$TFPGAP_{i,t-1}$	0.074*** (0.025)	0.064*** (0.022)	0.066** (0.031)	0.053** (0.026)
$H_{i,t-1}$	0.216*** (0.083)	0.150* (0.090)	0.180** (0.084)	0.152** (0.071)
$(P/L \cdot TFPGAP)_{i,t-1}$	0.035*** (0.012)	0.037** (0.015)	0.035** (0.014)	0.037*** (0.012)
$(w_F \cdot TFPGAP)_{i,t-1}$				0.153* (0.091)
$\hat{\delta}_1 + \hat{\delta}_2 \cdot \overline{P/L}$	0.084*** (0.023)	0.074*** (0.018)	0.075*** (0.029)	0.063** (0.024)
AB-AR(1)	-4.16 (0.000)	-4.16 (0.000)	-4.18 (0.000)	-4.14 (0.000)
AB-AR(2)	-0.53 (0.593)	-0.53 (0.595)	-0.53 (0.597)	-0.53 (0.598)
Hansen J (p-value)	0.313	0.259	0.066	0.222
Dif-Hansen (p-value)	0.730	0.703	0.830	0.896
Time Dummies	Yes	Yes	Yes	Yes
Observations	1538	1538	1538	1538
No. of Regions	159	159	159	159
No. of Instruments	156	156	90	112

Notes: See notes to Table 1.

Columns (i) and (ii): Instruments used are as in column (iv) of Table 1.

Column (iii): Instruments used for the first-differenced equations are $TFPGAP_{i,t-2}$, $(P/L \cdot TFPGAP)_{i,t-2}$, $(H \cdot TFPGAP)_{i,t-2}$, $(P/L)_{i,t-2}$ and $\Delta H_{i,t-1}$; additional instruments used for the levels equations are $\Delta TFPGAP_{i,t-1}$, $\Delta(P/L \cdot TFPGAP)_{i,t-1}$, $\Delta(H \cdot TFPGAP)_{i,t-1}$ and $\Delta H_{i,t-1}$.

Column (iv): Instruments used for the first-differenced equations are as in column (iii) plus $(w_F \cdot TFPGAP)_{i,t-2}$; additional instruments used for the levels equations are as in column (iii) plus $\Delta(w_F \cdot TFPGAP)_{i,t-1}$.

$\Delta H_{i,t-1}$ is implemented as an “ivstyle” instrument in Roodman’s (2009) `xtabond2` command for Stata.

In comparison to the existing literature on TFP growth at the EU regional level, our finding of TFP convergence contrasts with Di Liberto and Usai (2010) and Badinger and Tondl (2005). While the former find some evidence of divergence, the coefficient on Badinger and Tondl's (2005) proxy for the TFP gap is insignificant. Apart from differences in sample periods and estimation approaches, one reason for this contrast to our results may lie with alternative approaches to measuring TFP. Regarding human capital and patenting, our results also differ from those in Badinger and Tondl (2005). Where we find that both variables affect TFP growth either directly or indirectly, in Badinger and Tondl (2005) both direct and indirect effects are significant for human capital, but neither is for patent applications.

In column (ii) of Table 2, we investigate the robustness of our results to using a different frontier region.³⁹ From 1998 onwards, the frontier in our sample is Southern and Eastern Ireland. For Ireland, there are known problems with value-added data resulting from its low corporation tax rate, currently at 12.5%. This may give multinational companies an incentive to register as much of their profits as possible in the country. Value-added figures that are inflated in this way could bias our TFP index upwards and create a false appearance of Irish technology leadership. Therefore, we replace Southern and Eastern Ireland with the region with the next-highest level of TFP in our dataset from 1998 to 2005, Outer London.⁴⁰ Our estimates remain fairly robust to this change, although the human capital coefficient is now only marginally significant.

In column (iii), we use a smaller instrument set on which we build in column (iv). First, we reduce the number of lags of $(P/L)_{it}$ used as instruments for the first-differenced equations. Following this, we exclude $\Delta(P/L)_{i,t-1}$ from the instrument set for the equations in levels, as the Difference Hansen test now rejects its validity. The coefficient estimates again remain similar to those in column (i).

Column (iv) presents the results of augmenting equation (12) with our distance-weighted

---

³⁹Arguably, what matters for estimation is not correctly identifying the true TFP frontier, but instead the correlation between true and measured distance from the TFP frontier. See Griffith et al. (2004) and Griffith et al. (2009).

⁴⁰As noted in section 4, the Irish regions drop out of the estimation sample because of lacking data on patenting. TFP leaders, on the other hand, are computed using all NUTS-2 regions from the EU-15 with available TFP data (198).

version of the TFP gap, $(w_F \cdot TFP_{GAP})_{i,t-1}$. Analogous to the other terms involving the TFP gap, we use the lagged levels $(w_F \cdot TFP_{GAP})_{i,t-2}$ as instruments for the first-differenced equations in column (iv), and the lagged first-differences $\Delta(w_F \cdot TFP_{GAP})_{i,t-1}$ as additional instruments for the equations in levels. The Hansen and Difference Hansen tests do not reject the validity of the augmented instrument sets, and there is no evidence of significant second-order serial correlation from the AB-AR tests.

The estimated coefficient on $(w_F \cdot TFP_{GAP})_{i,t-1}$ is positive and significant at the 10% level, providing some indication that regions at a given technological distance to the frontier grow faster if they are geographically closer to it. One interpretation of this finding is that technological knowledge spillovers are to a degree geographically localised in our sample of EU regions. Meanwhile, the coefficient on regions' own TFP gap, $\hat{\delta}_1$, remains robust to the inclusion of the distance-weighted TFP gap, although it declines slightly in magnitude. Thus, global technology spillovers from the frontier continue to play a role for regional productivity growth. The estimates of the remaining coefficients in column (iv) do not change substantially compared to column (iii).

As an alternative way to investigate localised spillovers, one could include a direct measure of TFP distance to geographically close regions, which would allow for technology transfer not only from the global frontier, but also from a local or country-level one. Griffith et al. (2009) take this route by allowing for productivity catch-up to both national and regional industry frontiers for British establishments. We leave this as an avenue for future research.

6 Conclusion:

This paper investigates the effects of R&D and human capital on total factor productivity growth in the manufacturing sector across 159 regions of the EU-15 from 1992 to 2005. We contribute to the literature by employing Griffith et al.'s (2003) "two faces" extension to the original framework of Schumpeterian growth, which models TFP growth in transition to a long-run equilibrium level of TFP relative to the frontier. Both R&D and human capital are allowed to have a dual effect on transitional growth, reflecting own innovation

and imitation of frontier technology. We then extend this model to capture geographically localised technology spillovers.

We use a rich dataset from Cambridge Econometrics to construct an annual index of TFP for each region, and we build a time series on regional educational attainment from EU Labour Force Survey data that begins in 1992. Thus, we are able to analyse TFP growth and its determinants at the EU regional level over a longer time period than existing work in the field. Finally, we use panel data estimation methods, which allows us to control for unobserved region-specific fixed effects.

Our preferred empirical results provide significant evidence of a positive direct effect of human capital and a positive indirect effect of R&D activity on TFP growth for the EU-15 regions. This suggests that raising both educational attainment levels and R&D activity is beneficial for regional productivity growth, although the effects seem to operate through different channels. Our results may be regarded as supportive of recent EU regional policy, which has emphasised education and R&D expenditures in light of the Lisbon and Europe 2020 strategies. Furthermore, our estimates are consistent with conditional convergence in TFP levels relative to the frontier over our sample period. Finally, we find some evidence that both global and geographically localised technology spillovers may be important determinants of TFP growth at the level of the EU regions.

An interesting extension of the results presented in this paper would be to also allow for direct and indirect effects of trade on TFP growth, in addition to the effects of patenting and human capital studied here. Trade integration features as a determinant of growth in several endogenous growth models (e.g. Rivera-Batiz and Romer 1991). It can be argued to have both direct and indirect effects on productivity growth, for instance through efficiency improvements via greater product market competition and larger market size, and by facilitating technology transfer through the reverse engineering of imported goods. Constructing a measure of regional trade remains an empirical challenge in this context.

References

- Abramovitz, M. (1956). Resource and Output Trends in the United States since 1870, *American Economic Review* **46**(2): 5–23.
- Abramovitz, M. (1986). Catching Up, Forging Ahead, and Falling Behind, *The Journal of Economic History* **46**(2): 385–406.
- Acs, Z. J., Anselin, L. and Varga, A. (2002). Patents and Innovation Counts as Measures of Regional Production of New Knowledge, *Research Policy* **31**(7): 1069–1085.
- Aghion, P. and Howitt, P. (1992). A Model of Growth through Creative Destruction, *Econometrica* **60**(2): 323–351.
- Aghion, P. and Howitt, P. (2006). Joseph Schumpeter Lecture: Appropriate Growth Policy - A Unifying Framework, *Journal of the European Economic Association* **4**(2-3): 269–314.
- Arellano, M. and Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, *The Review of Economic Studies* **58**(2): 277–297.
- Arellano, M. and Bover, O. (1995). Another Look at the Instrumental Variable Estimation of Error-Components Models, *Journal of Econometrics* **68**(1): 29–51.
- Arrow, K. J. (1962). The Economic Implications of Learning by Doing, *Review of Economic Studies* **29**(3): 155–173.
- Badinger, H. and Tondl, G. (2005). The Factors behind European Regional Growth: Trade, Human Capital and Innovation, *Review of Regional Research* **25**(1): 67–89.
- Bassanini, A. and Scarpetta, S. (2002). Does Human Capital Matter for Growth in OECD Countries? A Pooled Mean-Group Approach, *Economics Letters* **74**(3): 399–405.
- Benhabib, J. and Spiegel, M. M. (1994). The Role of Human Capital in Economic Development: Evidence from Aggregate Cross-Country Data, *Journal of Monetary Economics* **34**(2): 143–173.

- Blundell, R. and Bond, S. (1998). Initial Conditions and Moment Restrictions in Dynamic Panel Data Models, *Journal of Econometrics* **87**(1): 115–143.
- Blundell, R., Bond, S. and Windmeijer, F. (2000). Estimation in Dynamic Panel Data Models: Improving on the Performance of the Standard GMM Estimator, in B. H. Baltagi (ed.), *Advances in Econometrics, Volume 15: Nonstationary Panels, Panel Cointegration and Dynamic Panels*, Amsterdam: JAI Elsevier Science.
- Bond, S. R. (2002). Dynamic Panel Data Models: A Guide to Micro Data Methods and Practice, *Portuguese Economic Journal* **1**(2): 141–162.
- Bond, S. R., Hoeffler, A. and Temple, J. (2001). GMM Estimation of Empirical Growth Models, *CEPR Discussion Paper 3048*.
- Bottazzi, L. and Peri, G. (2003). Innovation and Spillovers in Regions: Evidence from European Patent Data, *European Economic Review* **47**(4): 687–710.
- Bronzini, R. and Piselli, P. (2009). Determinants of Long-Run Regional Productivity with Geographical Spillovers: The Role of R&D, Human Capital and Public Infrastructure, *Regional Science and Urban Economics* **39**(2): 187–199.
- Cameron, G., Proudman, J. and Redding, S. (2005). Technological Convergence, R&D, Trade and Productivity Growth, *European Economic Review* **49**(3): 775–807.
- Caselli, F., Esquivel, G. and Lefort, F. (1996). Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics, *Journal of Economic Growth* **1**(3): 363–389.
- Caves, D. W., Christensen, L. R. and Diewert, W. E. (1982). Multilateral Comparisons of Output, Input, and Productivity Using Superlative Index Numbers, *Economic Journal* **92**(365): 73–86.
- Christensen, L. R., Jorgenson, D. W. and Lau, L. J. (1973). Transcendental Logarithmic Production Frontiers, *The Review of Economics and Statistics* **55**(1): 28–45.
- Coe, D. T. and Helpman, E. (1995). International R&D Spillovers, *European Economic Review* **39**(5): 859–887.

- Cohen, D. and Soto, M. (2007). Growth and Human Capital: Good Data, Good Results, *Journal of Economic Growth* **12**(1): 51–76.
- Cohen, W. M. and Levinthal, D. A. (1989). Innovation and Learning: The Two Faces of R&D, *Economic Journal* **99**(397): 569–596.
- De la Fuente, A. and Doménech, R. (2006). Human Capital in Growth Regressions: How Much Difference Does Data Quality Make?, *Journal of the European Economic Association* **4**(1): 1–36.
- Di Liberto, A. and Usai, S. (2010). TFP Convergence across European Regions: A Comparative Spatial Dynamics Analysis, *CRENoS Working Paper 2010/30*.
- Diewert, W. E. (1976). Exact and Superlative Index Numbers, *Journal of Econometrics* **4**(2): 115–145.
- Eaton, J. and Kortum, S. (1999). International Technology Diffusion: Theory and Measurement, *International Economic Review* **40**(3): 537–570.
- European Commission (2010a). A Rationale for Action: Europe 2020 Flagship Initiative Innovation Union, Accompanying Document to the Communication from the Commission, *Commission Staff Working Document SEC(2010) 1161 final*.
- European Commission (2010b). Cohesion Policy: Strategic Report 2010 on the Implementation of the Programmes 2007-2013, *Communication from the Commission COM(2010) 110 final*.
- European Commission (2010c). Regional Policy Contributing to Smart Growth in Europe 2020, *Communication from the Commission COM(2010) 553 final*.
- European Commission (2011). *Innovation Union Competitiveness Report*, 2011 edn, Luxembourg: Publications Office of the European Union.
- European Commission-Eurostat (2008). *EU Labour Force Survey Database User Guide*.
- Eurostat (2008). *European Regional and Urban Statistics Reference Guide*, 2008 edn, Luxembourg: Office for Official Publications of the European Communities.

- Fagerberg, J., Verspagen, B. and Caniëls, M. (1997). Technology, Growth and Unemployment across European Regions, *Regional Studies* **31**(5): 457–466.
- Frankel, M. (1962). The Production Function in Allocation and Growth: A Synthesis, *American Economic Review* **52**(5): 996–1022.
- Frantzen, D. (2003). The Causality between R&D and Productivity in Manufacturing: An International Disaggregate Panel Data Study, *International Review of Applied Economics* **17**(2): 125–146.
- Gemmell, N. (1996). Evaluating the Impacts of Human Capital Stocks and Accumulation on Economic Growth: Some New Evidence, *Oxford Bulletin of Economics and Statistics* **58**(1): 9–28.
- Gerschenkron, A. (1962). *Economic Backwardness in Historical Perspective*, Cambridge, MA: Harvard University Press.
- Gollin, D. (2002). Getting Income Shares Right, *Journal of Political Economy* **110**(2): 458–474.
- Griffith, R., Redding, S. and Simpson, H. (2009). Technological Catch-Up and Geographic Proximity, *Journal of Regional Science* **49**(4): 689–720.
- Griffith, R., Redding, S. and Van Reenen, J. (2000). Mapping the Two Faces of R&D: Productivity Growth in a Panel of OECD Industries, *CEPR Discussion Paper 2457*.
- Griffith, R., Redding, S. and Van Reenen, J. (2003). R&D and Absorptive Capacity: Theory and Empirical Evidence, *Scandinavian Journal of Economics* **105**(1): 99–118.
- Griffith, R., Redding, S. and Van Reenen, J. (2004). Mapping the Two Faces of R&D: Productivity Growth in a Panel of OECD Industries, *The Review of Economics and Statistics* **86**(4): 883–895.
- Griliches, Z. (1980). Returns to R&D Expenditures in the Private Sector, in J. W. Kendrick and B. N. Vaccara (eds), *New Developments in Productivity Measurement*, NBER Studies in Income and Wealth, Chicago: University of Chicago Press.

- Griliches, Z. and Lichtenberg, F. (1984). R&D and Productivity Growth at the Industry Level: Is There Still a Relationship?, *in* Z. Griliches (ed.), *R&D, Patents, and Productivity*, Chicago: University of Chicago Press.
- Grossman, G. M. and Helpman, E. (1991). Quality Ladders in the Theory of Growth, *Review of Economic Studies* **58**(1): 43–61.
- Guellec, D. and van Pottelsberghe de la Potterie, B. (2004). From R&D to Productivity Growth: Do the Institutional Settings and the Source of Funds of R&D Matter?, *Oxford Bulletin of Economics and Statistics* **66**(3): 353–378.
- Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica* **50**(4): 1029–1054.
- Hulten, C. R. (2001). Total Factor Productivity: A Short Biography, *in* C. R. Hulten, E. R. Dean and M. J. Harper (eds), *New Developments in Productivity Analysis*, NBER Studies in Income and Wealth, Chicago: University of Chicago Press.
- Islam, N. (1995). Growth Empirics: A Panel Data Approach, *Quarterly Journal of Economics* **110**(4): 1127–1170.
- Islam, N. (1999). International Comparison of Total Factor Productivity: A Review, *Review of Income and Wealth* **45**(4): 493–518.
- Jaffe, A. B., Trajtenberg, M. and Henderson, R. (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations, *Quarterly Journal of Economics* **108**(4): 577–598.
- Khan, T. S. (2006). Productivity Growth, Technological Convergence, R&D, Trade, and Labor Markets: Evidence from the French Manufacturing Sector, *IMF Working Paper* 06/230.
- Lucas, R. E. (1988). On the Mechanics of Economic Development, *Journal of Monetary Economics* **22**(1): 3–42.

- Mankiw, N. G., Romer, D. and Weil, D. N. (1992). A Contribution to the Empirics of Economic Growth, *Quarterly Journal of Economics* **107**(2): 407–437.
- Nelson, R. R. and Phelps, E. S. (1966). Investment in Humans, Technological Diffusion, and Economic Growth, *American Economic Review of Economic Studies* **56**(1/2): 69–75.
- Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects, *Econometrica* **49**(6): 1417–1426.
- OECD (2001). *Measuring Productivity: Measurement of Aggregate and Industry-Level Productivity Growth, OECD Manual*.
- Paci, R. and Pigliaru, F. (2002). Technological Diffusion, Spatial Spillovers and Regional Convergence in Europe, in J. R. Cuadrado-Roura and M. Parellada (eds), *Regional Convergence in the European Union: Facts, Prospects and Policies*, Berlin: Springer.
- Pritchett, L. (2001). Where Has All the Education Gone?, *World Bank Economic Review* **15**(3): 367–391.
- Rivera-Batiz, L. A. and Romer, P. M. (1991). Economic Integration and Endogenous Growth, *Quarterly Journal of Economics* **106**(2): 531–555.
- Romer, P. M. (1986). Increasing Returns and Long-Run Growth, *Journal of Political Economy* **94**(5): 1002–1037.
- Romer, P. M. (1990). Endogenous Technological Change, *Journal of Political Economy* **98**(5): S71–S102.
- Roodman, D. (2009). How To Do xtabond2: An Introduction to Difference and System GMM in Stata, *Stata Journal* **9**(1): 86–136.
- Sapir, A., Aghion, P., Bertola, G., Hellwig, M., Pisani-Ferry, J., Vinals, J., Rosati, D., Buti, M. and Wallace, H. (2004). *An Agenda for a Growing Europe: The Sapir Report*, Oxford: Oxford University Press.

- Sargan, J. D. (1958). The Estimation of Economic Relationships using Instrumental Variables, *Econometrica* **26**(3): 393–415.
- Schumpeter, J. (1942). *Capitalism, Socialism and Democracy*, New York: Harper & Row.
- Solow, R. M. (1956). A Contribution to the Theory of Economic Growth, *Quarterly Journal of Economics* **70**(1): 65–94.
- Solow, R. M. (1957). Technical Change and the Aggregate Production Function, *Review of Economics and Statistics* **39**(3): 312–320.
- Sterlacchini, A. (2008). R&D, Higher Education and Regional Growth: Uneven Linkages among European Regions, *Research Policy* **37**(6-7): 1096–1107.
- UNESCO-OECD-Eurostat (2005). *UOE Data Collection on Education Systems Manual Volume 1: Concepts, Definitions and Classifications*, Montreal, Paris, Luxembourg.
- Van Ark, B., O'Mahony, M. and Timmer, M. P. (2008). The Productivity Gap between Europe and the United States: Trends and Causes, *Journal of Economic Perspectives* **22**(1): 25–44.
- Windmeijer, F. (2005). A Finite Sample Correction for the Variance of Linear Efficient Two-Step GMM Estimators, *Journal of Econometrics* **126**(1): 25–51.

Appendices

A List of Regions

Table A.1: List of NUTS-2 Regions in Sample by Country (159)

Country	Code	Region Name	Country	Code	Region Name
Belgium	BE10	Brussels	Germany	DE13	Freiburg
	BE21	Antwerp		DE14	Tübingen
	BE22	Limburg		DE21	Upper Bavaria
	BE23	East Flanders		DE22	Lower Bavaria
	BE24	Flemish Brabant		DE23	Upper Palatinate
	BE25	West Flanders		DE24	Upper Franconia
	BE31	Walloon Brabant		DE25	Middle Franconia
	BE32	Hainault		DE26	Lower Franconia
	BE33	Liège		DE27	Swabia
	BE24	Luxembourg (BE)		DE30	Berlin
	BE25	Namur		DE50	Bremen
Denmark	DK00	Denmark	DE60	Hamburg	
			DE71	Darmstadt	
			DE72	Gießen	
Finland	FI18	South Finland	DE73	Kassel	
	FI19	West Finland	DE80	Mecklenburg- Western Pomerania	
	FI1A	North Finland	DE91	Braunschweig	
France	FR10	Île de France	DE92	Hannover	
	FR21	Champagne-Ardenne	DE93	Lunenburg	
	FR22	Picardy	DE94	Weser-Ems	
	FR23	Upper Normandy	DEA1	Düsseldorf	
	FR24	Centre	DEA2	Cologne	
	FR25	Lower Normandy	DEA3	Münster	
	FR26	Burgundy	DEA4	Detmold	
	FR30	North-Pas de Calais	DEA5	Arnsberg	
	FR41	Lorraine	DEB1	Koblenz	
	FR42	Alsace	DEB2	Trier	
	FR43	Franche-Comté	DEB3	Rhine-Hesse-Palatinate	
	FR51	Loire Counties	DEC0	Saarland	
	FR52	Brittany	DEF0	Schleswig-Holstein	
	FR53	Poitou-Charentes	DEG0	Thuringia	
	FR61	Aquitaine	Greece	GR11	East Macedonia
	FR62	South Pyrénées		GR12	Central Macedonia
	FR63	Limousin		GR13	West Macedonia
	FR71	Rhône-Alpes		GR14	Thessaly
FR72	Auvergne	GR21		Epirus	
FR81	Languedoc-Roussillon	GR23		West Greece	
FR82	Provence-Alpes-Côte d’Azur	GR24		Central Greece	
Germany	DE11	Stuttgart		GR25	Peloponnese
	DE12	Karlsruhe	GR30	Attica	

Country	Code	Region Name	Country	Code	Region Name
Italy	ITC1	Piedmont	UK	ES43	Extremadura
	ITC2	Aosta Valley		ES51	Catalonia
	ITC3	Liguria		ES52	Community of Valencia
	ITC4	Lombardy		ES53	Balearic Islands
	ITD1	Bolzano-Bozen		ES61	Andalusia
	ITD2	Trento		ES62	Region of Murcia
	ITD3	Veneto			
	ITD4	Friuli-Venezia Giulia		UKC2	Northumberland and Tyne and Wear
	ITD5	Emilia-Romagna			
	ITE1	Tuscany		UKD2	Cheshire
	ITE2	Umbria		UKD3	Greater Manchester
	ITE3	Marche		UKD4	Lancashire
	ITE4	Latium		UKD5	Merseyside
	ITF1	Abruzzo		UKE1	East Riding and North Lincolnshire
	ITF2	Molise			
	ITF3	Campania		UKE2	North Yorkshire
	ITF4	Apulia		UKE3	South Yorkshire
	ITF5	Basilicata		UKE4	West Yorkshire
	ITF6	Calabria		UKF1	Derbyshire and Nottinghamshire
	ITG2	Sardinia		UKF2	Leicestershire, Rutland and Northamptonshire
Netherlands	NL11	Groningen		UKG1	Herefordshire, Worcestershire and Warwickshire
	NL12	Friesland			
	NL13	Drenthe		UKG2	Shropshire and Staffordshire
	NL21	Overijssel		UKG3	West Midlands
	NL22	Gelderland		UKH1	East Anglia
	NL23	Flevoland		UKH2	Bedfordshire and Hertfordshire
	NL31	Utrecht		UKH3	Essex
	NL32	North Holland		UKI1	Inner London
	NL33	South Holland		UKI2	Outer London
	NL34	Zeeland		UKJ1	Berkshire, Buckinghamshire and Oxfordshire
	NL41	North Brabant		UKJ2	Surrey, East and West Sussex
	NL42	Limburg		UKJ3	Hampshire and Isle of Wight
Portugal	PT16	Centre		UKJ4	Kent
	PT17	Lisbon		UKK1	Gloucestershire, Wiltshire and North Somerset
	PT18	Alentejo		UKK2	Dorset and Somerset
Spain	ES11	Galicia		UKK3	Cornwall and Isles of Scilly
	ES12	Principality of Asturias		UKK4	Devon
	ES13	Cantabria		UKL1	West Wales and The Valleys
	ES21	Basque Country		UKL2	East Wales
	ES22	Navarre		UKN0	Northern Ireland
	ES23	La Rioja			
	ES24	Aragón			
	ES30	Community of Madrid			
	ES41	Castile and León			
	ES42	Castile-La Mancha			

NUTS version: 2003.

B Database Description and Summary Statistics

The Cambridge Econometrics European Regional Database⁴¹ is one of the most commonly used datasets in EU-wide regional economic analysis. Its primary source is Eurostat’s freely accessible Regional Statistics Database (REGIO),⁴² which it complements with data from consultants and national and international statistical agencies to achieve more complete time series for all regions. The database contains annual series on a number of economic indicators, beginning in 1980 for the EU-15 and in 1990 for most Eastern European countries.

Within REGIO, Cambridge Econometrics take data from the regional branch accounts, where economic indicators are disaggregated into sectors according to the European Union’s industry classification system NACE (Revision 1.1). Therefore, most variables in the CE database are provided at breakdowns into five broad and 15 narrower sectors, where the broad sectors are: agriculture, hunting, forestry and fishing; energy and manufacturing; construction; market services; and non-market services.

B.1 Regional Output:

The primary output indicator collected at the regional level is gross value added, which is measured at basic prices. The original GVA data from Eurostat’s regional branch accounts are in current prices (euros). Cambridge Econometrics convert them into constant prices for the year 2000 by applying sectoral price deflators at the national level, obtained from the European Commission’s AMECO database, to the regional current-price data for each sector, and adding up across sectors to obtain the regional totals.⁴³ Given the sectoral composition of GVA in each region, individual region-specific price deflators can thus be derived implicitly. This requires the assumption that price changes over time within a given sector are the same across all regions of a country.

---

⁴¹<http://www.camecon.com>

⁴²http://epp.eurostat.ec.europa.eu/portal/page/portal/region_cities/regional_statistics/data/database

⁴³AMECO is the acronym for “annual macro-economic”. The database is available at http://ec.europa.eu/economy_finance/db_indicators/ameco/index_en.htm.

To adjust the data for price level differences across countries, we multiply CE's constant-price GVA series with the 2000 national PPS exchange rates from the CE database, which also come from AMECO. While this allows us to control for those differences in price levels across regions that are due to differences between countries, within-country variations in price levels remain unaccounted for. Since PPPs at the subnational level in Europe are not yet available, this is unfortunately unavoidable.

B.2 Regional Investment:

Like the data on GVA, those on regional gross fixed capital formation (GFCF) are given at constant prices in 2000 euros in the CE database. Cambridge Econometrics use the same procedure as for GVA to arrive at the constant-price GFCF series. That is, they apply the country-level sectoral GVA deflators for the year 2000 from AMECO to the regional current-price (euro) GFCF data from Eurostat and obtain regional price deflators. GVA deflators are used because investment price deflators are available from AMECO only by type of investment goods, such as dwellings or equipment, but not by sector. This is a shortcoming given that output and investment goods prices in a particular sector can be expected to vary in different ways over time. We convert the constant-price GFCF data from CE into PPS in the same way as for GVA, using national PPPs.

B.3 Regional Labour Input:

We use data on hours worked as our measure of labour input, which is recommended by OECD (2001). It is preferable to a simple headcount of persons in employment, which does not accurately reflect the actual productive services they provide if the average number of hours worked per person changes over time. For instance, this is likely to be the case over the business cycle, where hours worked are lower (higher) during periods of labour hoarding (labour scarcity), or if a shift towards more part-time work occurs.

The CE database provides data on hours usually worked, which is the typical (e.g. modal) value of hours actually worked in a job over a long reference period. CE take these data from the EU Labour Force Survey rather than the regional branch accounts.

Table B.1: Summary Statistics

Variable		Mean	Std. Dev.	Min.	Max.	Observations
$\Delta \ln TFP_{it}$	overall	0.012	0.058	-0.287	0.311	$N = 2066$
	between		0.019	-0.044	0.058	$n = 159$
	within		0.055	-0.211	0.353	$\bar{T} = 13$
$TFPGAP_{it}$	overall	0.915	0.328	0	2.482	$N = 2225$
	between		0.298	0.208	2.216	$n = 159$
	within		0.142	0.174	1.379	$\bar{T} = 14$
$(P/L)_{it}$	overall	0.277	0.238	0.001	2.151	$N = 1863$
	between		0.219	0.006	1.013	$n = 159$
	within		0.095	-0.244	1.415	$\bar{T} = 11.7$
H_{it}	overall	0.165	0.063	0.042	0.441	$N = 1898$
	between		0.063	0.054	0.385	$n = 159$
	within		0.022	0.076	0.259	$\bar{T} = 11.9$
$(P/L \cdot$ $TFPGAP)_{it}$	overall	0.233	0.213	0	1.542	$N = 1862$
	between		0.194	0.003	0.957	$n = 159$
	within		0.089	-0.135	1.126	$\bar{T} = 11.7$
$(H \cdot$ $TFPGAP)_{it}$	overall	0.148	0.069	0	0.575	$N = 1897$
	between		0.058	0.025	0.512	$n = 159$
	within		0.037	0.033	0.296	$\bar{T} = 11.9$

Table B.2: Correlation Matrix

	$\Delta \ln TFP_{it}$	$TFPGAP_{i,t-1}$	$(P/L)_{i,t-1}$	$H_{i,t-1}$	$(P/L \cdot TFPGAP)_{i,t-1}$	$(H \cdot TFPGAP)_{i,t-1}$
$\Delta \ln TFP_{it}$	1					
$TFPGAP_{i,t-1}$	0.1406***	1				
$(P/L)_{i,t-1}$	0.1415***	-0.1705***	1			
$H_{i,t-1}$	0.0997***	-0.3168***	0.4561***	1		
$(P/L \cdot TFPGAP)_{i,t-1}$	0.2085***	0.1929***	0.8852***	0.3046***	1	
$(H \cdot TFPGAP)_{i,t-1}$	0.2142***	0.6060***	0.2201***	0.5006***	0.4575***	1

Notes: *** indicates significance at the 1% level.

C Regional Variable Maps

Average annual manufacturing TFP growth (%), 1992-2005

Average manufacturing TFP gap, 1992-2005

Average no. of patent applications per 10,000s of manufacturing employees, 1992-2005

Average % of working-age population with completed higher education, 1992-2005

D Correspondence between ISCED 1976 and ISCED 1997

ISCED 1976	ISCED 1997
0 Education preceding the first level	0 Pre-primary level of education
1 Education at the first level	1 Primary level of education
2 Education at the second level, first stage	2 Lower secondary level of education (2A, 2B and 2C)
3 Education at the second level, second stage	3 Upper secondary level education (3A, 3B and 3C)
5 Education at the third level, first stage, of the type that leads to an award not equivalent to a first university degree	4 Post-secondary, non-tertiary education (4A, 4B and 4C)
6 Education at the third level, first stage, of the type that leads to a first university degree or equivalent	5 First stage of tertiary education, not directly leading to an advanced research qualification (5A, 5B and 5C)
7 Education at the third level, second stage, of the type that leads to a post-graduate university degree or equivalent	6 Second stage of tertiary education, leading to an advanced research qualification
9 Education not definable by level	

Source: UNESCO-OECD-Eurostat (2005)

The European Labour Force Survey (EU LFS) collects data on the indicator “highest level of education or training successfully completed” based on ISCED 1997 since 1998. The ISCED 1976 classification was used up to (and including) the year 1997.

To ensure comparability between the two ISCED regimes as far as possible, the guidelines to the EU LFS recommend grouping the education levels of both regimes into low, medium and high levels of educational attainment as follows:⁴⁴

Low: ISCED 1997 = ISCED 1976 = Levels 0-2 (at most lower secondary)

Medium: ISCED 1997 = Levels 3-4, ISCED 1976 = Level 3 (upper secondary)

High: ISCED 1997 = Levels 5-6, ISCED 1976 = Levels 5-7 (tertiary).

⁴⁴See European Commission-Eurostat (2008) pp. 27, 34, and 49; Eurostat (2008) p. 87; and the on-line EU Labour Force Survey Domain, Section 7 “Statistical Classifications used in the EU LFS”, “Education: ISCED 1997” at http://circa.europa.eu/irc/dsis/employment/info/data/eu_lfs/LFS_MAIN/Related_documents/ISCED_EN.htm.