Article

EUROPEAN URBAN AND REGIONAL STUDIES

European Urban and Regional Studies
2025, Vol. 32(4) 421–442

DOI: 10.1177/09697764251331563
journals.sagepub.com/home/eur
Sage

'Two faces' of human capital and research and development activities: Effects of related technologies and the regional technology portfolio on total factor productivity in Polish regions

Andrzej Cieślik
University of Warsaw, Poland

Tomasz Misiak
Rzeszów University of Technology, Poland

Tomasz Sierotowicz and Rafał Wisła
Jagiellonian University, Poland

Abstract

In this article, we study the role of technological relatedness/unrelatedness in explaining the relationship between research and development and human capital and total factor productivity growth at the regional level using system Generalized Method of Moments and statistical data for 16 Polish regions over the period 2003–2019. Using the research and development effects-based approach, we weight the number of patents relative to the employed in the research and development sector by measures that identify related technologies in the regional technology portfolio. Our results show positive and statistically significant effects of human capital and technological relatedness on both technological progress and technology diffusion. Moreover, technological unrelatedness exerts a negative effect on technological progress and does not lead to technology diffusion. The study also shows the importance of local technology resources.

Keywords

Human capital, knowledge diffusion, R&D activities, relatedness/unrelatedness

Introduction

A key determinant of economic growth is undoubtedly technological progress that is usually identified with the growth rate of total factor productivity

Corresponding author:
Tomasz Misiak, Department of Economics, Faculty of Management, Rzeszów University of Technology, al. Powstańców Warszawy 8, PL-35-959 Rzeszów, Poland.
Email: tmisiak@prz.edu.pl

(TFP). One possible approach to modelling TFP growth relates to the ‘two faces’ concept of its fundamental determinants that include both human capital and research and development (R&D) that may affect growth through two potential channels. The so-called ‘first face’ means that these variables can directly influence growth by generating and stimulating innovation. The ‘second face’ means that investment in human capital and R&D might also facilitate knowledge diffusion and improve absorptive capacity necessary to assimilate external knowledge. To the best of our knowledge, prior empirical research based on the ‘two faces’ concept did not study patterns of knowledge accumulation in regions, which is an important research gap and might be of great importance for R&D activities and the development of new technologies. Therefore, in this article, we propose to exploit the principle of technological relatedness in the empirical modelling of TFP growth at the regional level.

The principle of relatedness implies that countries or regions are more likely to develop new activities, such as new products, new industries or new technologies, if they already have related activities (Hidalgo et al., 2018). Thus, the technological relatedness and the technological portfolio of regions characterise the ‘regional knowledge base’. The technological relatedness structure can be proxied by patents and knowledge domains, which are reported as classes (subclasses) of patents. Therefore, the relatedness approach distinguishes technologies of the global knowledge space that are characterised by different degrees of relatedness to the regional knowledge base and uses patents and patent classes to define them. Some technologies may be unrelated to the regional knowledge base, while others are related to it to varying degrees.

The heterogeneity of the regional knowledge base and the relatedness between new and existing technologies may affect the type of innovations (e.g. radical versus incremental) that are introduced into the region and the technologies in which they can specialise. This becomes a key aspect in determining which new technologies will enter and expand the regional knowledge base. The heterogeneity of the regional knowledge base and the diversity of the related technologies can also have a significant effect on the knowledge-diffusion process. The more heterogeneous the regional knowledge base and the more related technologies are, the higher the probability of external knowledge absorption and the greater the knowledge diffusion. This can, therefore, also lead to a reduction in the gap to the technological frontier of the regions.

Therefore, in our empirical study, we use the concept of ‘two faces’ to both human capital and R&D activities. Considering the R&D sector from the effects side, we propose that the number of patents relative to those employed in the R&D sector should be weighted by measures that identify related technology against the regional technology portfolio. Inspired by the regional diversification literature towards new technologies that can leverage relevant regional capabilities, we use measures of relatedness/unrelatedness density that show how far a new technology is from the existing technology portfolio in the region (Balland et al., 2019; Boschma, 2017; Boschma et al., 2023).

By implementing a variable describing the R&D sector defined in this way, we contribute to better understanding the potential role of the impact of relatedness and unrelatedness between technologies on the growth rate of TFP. In particular, our approach to analysing the R&D potential of regional economies within TFP growth models can extend and complement the existing literature in four major ways. First, we modify the prior modelling approach of Griffith et al. (2004) by introducing a new variable that measures R&D sectoral effects, that is, the interaction between the number of patents in a region relative to the regional potential in this sector by weighting this variable by the density of technological relatedness and unrelatedness. We therefore introduce the principle of relatedness into TFP modelling based on the ‘two faces’ concept, which was an important research gap.

Second, we document the roles of related and unrelated technologies and the regional technology portfolio, not only through the creation of innovations but also through their impact on technology diffusion. Third, our model assumes a logistic function of technology diffusion determined by both human capital and the R&D sector. Fourth, we validate empirically the modified theoretical model by estimating panel regressions using a two-step system Generalized Method of Moments (GMM) for Polish Nomenclature of territorial units for statistics (NUTS2) regions over the period 2003–2019.

The remainder of the article is structured as follows. In the Literature review section, we provide a review of the related literature. In section 3, we describe the modelling framework and develop research hypotheses. In section 4, we explain how we measure TFP, human capital, technology relatedness and unrelatedness density and additional control variables and discuss our estimation methodology. In section 5, we report and interpret our estimation results. The last section summarises and concludes with policy recommendations and guidelines for future studies.

Literature review

A key role in stimulating economic growth and development is attributed to technological progress (Aghion and Howitt, 2008; Grossman and Helpman, 1991; Romer, 1994). Endogenous growth models emphasise the key role of human capital and R&D in driving productivity growth. In particular, Romer (1990) argues that innovations created in the R&D sector lead to technological progress by increasing the number of available varieties of intermediate goods. In the Schumpeterian models of Grossman and Helpman (1990, 1991) and Aghion and Howitt (1992), innovations improve the quality of existing intermediate goods. An important conclusion from the aforementioned models is that the growth rate of TFP depends on the amount of resources devoted to the R&D sector, and in particular, on the stock of human capital or the number of workers employed in research.

According to Griffith et al. (2003), improvements in intermediate goods induced by R&D investment allow closing the gap to the technological frontier. Following Cohen and Levinthal (1989), they argue that the R&D sector plays an important role in facilitating technology diffusion, which is known in the literature as the ‘second face of R&D’. They relate the dual effects of R&D investment to generating inventions (innovations) and improving the absorptive capacity of firms, (i.e. their ability to assimilate external knowledge). Subsequently, Griffith et al. (2004) attribute a similar role to human capital in the spirit of Nelson and Phelps (1966), where education facilitates the creation and adoption of new technologies.[1]

Empirical studies that use the concept of ‘two faces’ and TFP as the dependent variable are relatively scarce and usually investigate the effects of R&D and human capital on per capita income or productivity at the country level. An important problem is the selection of appropriate measures of human capital and R&D, as they are not directly observable.[2] While some standardised procedures were developed in calculating the level of human capital, such as that of Barro and Lee (2013), there is still no agreement how R&D efforts should be measured.

Some researchers use in their studies the shares of R&D expenditure in value added or GDP (Cameron et al., 2005; Capello and Lenzi, 2015; Griffith et al., 2004; Männasoo et al., 2018; Siller et al., 2021). The main advantage of this approach is that it is simple, and necessary data can be easily obtained from publicly available sources. However, results of these studies are mixed as the incurred R&D expenditures may not always lead to expected outcomes. Instead of expected innovations, increased R&D spendings may translate into higher operating costs and lower firm profits.

Therefore, some studies, especially at the regional level, advocate the use of an effects-based approach (Badinger and Tondl, 2005; Vogel, 2015). Patents are the direct effects of R&D activity that confirm the creation of a new technology as a result of investment in the R&D sector. In the case of regional analyses, the ratio of regional patent applications to total employment in the region is used rather than R&D expenditures. However, these studies also do not yield clear-cut results.

Difficulties in identifying the effects of R&D activity on regional productivity growth may arise from overlooking different patterns of regional knowledge accumulation. The relatedness principle can be used to account for potential knowledge-accumulation pathways. This principle was introduced by Hidalgo et al. (2007), who found that a country is more likely to export a new product if it already exports related products. In general, the principle of relatedness means that countries, regions or even cities are more likely to develop new activities, such as new industries, new products, new technologies or research areas, if they already have related activities (Hidalgo, 2021; Hidalgo et al., 2018).

In their seminal study, Hidalgo et al. (2007) introduced also the concept of product space. Using the product space, a measure of product relatedness called density can be calculated. This measure of relatedness determines how many related products a region already exports for each product that has not yet been produced. Neffke et al. (2011) identify related industries and created a product space at the regional level in Sweden. Their study shows that the number of related industries in a region influences the probability of adopting new industries. Boschma et al. (2015) find that the probability of entering a new technology increases with the number of related technologies. Farinha et al. (2019) demonstrate that new occupations related to the metropolitan area's current occupational portfolio are emerging, while existing occupations with low relatedness are disappearing in metropolitan areas in the United States. Similarly, new technologies are developed using knowledge drawn from existing technologies.

Balland et al. (2019) show that diversification into more complex technologies is attractive but, at the same time, more difficult to achieve for the European Union (EU) regions. Regions could overcome this diversification dilemma by developing new and complex technologies that would build on locally related capabilities. More recently, Boschma et al. (2023) studied the effects of relatedness and unrelatedness on breakthrough inventions. They used measures of relatedness and unrelatedness to show that the majority of breakthrough inventions were made between related technologies.

In contrast to the aforementioned research, Montresor and Quatraro (2017) argue that knowledge of key enabling technologies (KETs) could dilute the effect that regional branching attributes to technological relatedness, giving regions a greater scope for their technological diversification strategies. They also find that regions could benefit from the effects of KETs, even if they were followers in their development, through inter-regional spillover effects from proximity to KET leaders. Using KETs, as an alternative to Revealed Technological Advantages or relatedness in general, they show that the impact of this indicator on regional branching is significant and positive.

In principle, it is difficult to link R&D expenditure directly to the relatedness principle. However, R&D investments can be approached from the outcomes side in the form of patents (Badinger and Tondl, 2005; Vogel, 2015). There are many studies showing a strong positive correlation between R&D investments and patents (e.g. see Altuzarra, 2018; Artz et al., 2010; Roper and Hewitt-Dundas, 2015). Therefore, in our study, we approach R&D activity from the outcomes side, using patents as a proxy for R&D activity. This approach makes it possible to consider a path of knowledge accumulation in the region based on the principle of technological relatedness.

The principle of technological relatedness is flexible and versatile enough to consider its relevance to the 'two faces' concept of both R&D and human capital. As argued by Boschma et al. (2023), the stimulation of innovations is accompanied by novelty, and this is always associated with risk. In order to account for and mitigate this risk, successful innovations need to build on previous combinations. Balland et al. (2019) highlight that innovations are easier to introduce and less costly when diversifying into new technologies that can leverage relevant regional capabilities. Thus, we can expect innovations to be more likely developed in regions where they are related to local technology stocks (Kaplan and Vakili, 2015).

This argument links the relatedness principle to the generation of innovations and reflects its direct channel of influence (i.e. the 'first face') on technological progress. In fact, the relatedness principle can also influence the diffusion of knowledge, reflecting the 'second face' of R&D. Therefore, we can expect that the more heterogeneous the regional knowledge base is and the more technologies related in the region are, the greater will be the likelihood of external knowledge absorption and the degree of knowledge diffusion.

Modelling framework

As a theoretical background for our empirical study of the role of technological relatedness in explaining the relationship between R&D and human capital and TFP growth at the regional level, we use the modified framework originally developed by Griffith et al. (2004). Our modifications follow the reasoning explained in detail in the work by Benhabib and Spiegel (2005). In particular, following their theoretical framework, we introduce the rate of technology diffusion from the technology leader region *m* to the follower region *r* instead of the region's distance from the technological frontier in terms of TFP used in previous studies. This modification allows us to study simultaneously the implications of two types of processes often studied separately in the context of disaggregated technology diffusion models: confined exponential knowledge diffusion and the logistic model of technology diffusion (Banks, 1994).

Hence, we modify the Griffith et al. (2004) model in two following ways:

1. as confined exponential diffusion, which can be expressed as follows:

$$\frac{\Delta A_{r,t}}{A_{r,t}} = \alpha_0 + \alpha_1 H_{r,t-1} + \alpha_2 \left(H_{r,t-1}\right)\left(\frac{A_{m,t-1}}{A_{r,t-1}} - 1\right) + \beta_1 RD_{r,t-1} + \beta_2 (RD_{r,t-1})\left(\frac{A_{m,t-1}}{A_{r,t-1}} - 1\right) \quad (1)$$

where $A_{r,t}$ is the TFP of region *r* in year *t*, $H_{r,t-1}$ is the stock of human capital in region *r* in year *t-1*, $RD_{r,t-1}$ is the R&D activity in region *r* in year *t-1* and $\left(\frac{A_{m,t-1}}{A_{r,t-1}} - 1\right)$ denotes the rate of technology diffusion from the leader *m* region to the follower *r* region.

2. as the logistic technology diffusion processes, which can be expressed as follows:

$$\begin{aligned}
\frac{\Delta A_{r,t}}{A_{r,t}} &= \alpha_0 + \alpha_1 H_{r,t-1} + \alpha_2 \left(H_{r,t-1}\right)\left(1 - \frac{A_{r,t-1}}{A_{m,t-1}}\right) + \beta_1 RD_{r,t-1} + \beta_2 \left(RD_{r,t-1}\right)\left(1 - \frac{A_{r,t-1}}{A_{m,t-1}}\right) \\
&= \alpha_0 + \alpha_1 H_{r,t-1} + \alpha_2 \left(H_{r,t-1}\right)\left(\frac{A_{r,t-1}}{A_{m,t-1}}\right)\left(\frac{A_{m,t-1}}{A_{r,t-1}} - 1\right) + \beta_1 RD_{r,t-1} \\
&\quad + \beta_2 \left(RD_{r,t-1}\right)\left(\frac{A_{r,t-1}}{A_{m,t-1}}\right)\left(\frac{A_{m,t-1}}{A_{r,t-1}} - 1\right) = \alpha_0 + \left(\alpha_1 + \alpha_2\right) H_{r,t-1} - \alpha_2 \left(H_{r,t-1}\right)\left(\frac{A_{r,t-1}}{A_{m,t-1}}\right) \\
&\quad + (\beta_1 + \beta_2) RD_{r,t-1} - \beta_2 \left(RD_{r,t-1}\right)\left(\frac{A_{r,t-1}}{A_{m,t-1}}\right)
\end{aligned} \quad (2)$$

The key difference in dynamics between the logistic model of technology diffusion and the confined exponential model is the presence of the additional term, that is, $\left(\frac{A_{r,t-1}}{A_{m,t-1}}\right)$. As argued by Benhabib and Spiegel (2005), this term dampens the rate of diffusion as distance from the technological leader increases. This might result from the increasing difficulty of adopting more distant technologies. Moreover, as shown by Basu and Weil (1998), a frontier technology may not be immediately 'suitable' for a follower, especially if the differences in factor proportions between the leader and follower are too large. For this reason, we prefer to use the logistic model of technology diffusion.

Subsequently, we replace the total number of patents in relation to the R&D employment used in previous studies, such as Vogel (2015), by the total number of patents in relation to the personnel employed in the R&D sector weighted by densities of technological relatedness and unrelatedness. This allows us to incorporate the principle of technological relatedness into our empirical analysis. Hence, we modify equation (2) as follows:

$$\frac{\Delta A_{r,t}}{A_{r,t}} = \delta_0 + \delta_1 H_{r,\,t-1} - \delta_2 \left(H_{r,\,t-1}\right)\left(\frac{A_{r,\,t-1}}{A_{m,t-1}}\right) + \gamma_1 \left(\frac{Patents}{L_{RD}} * RelD\right)_{r,t-1} - \gamma_2 \left(\frac{Patents}{L_{RD}} * RelD\right)_{r,t-1} \left(\frac{A_{r,\,t-1}}{A_{m,t-1}}\right) + \gamma_3 \left(\frac{Patents}{L_{RD}} * UnrelD\right)_{r,t-1} - \gamma_4 \left(\frac{Patents}{L_{RD}} * UnrelD\right)_{r,t-1} \left(\frac{A_{r,\,t-1}}{A_{m,t-1}}\right) \quad (3)$$

where $\left(\frac{Patents}{L_{RD}} * RelD\right)_{r,t-1}$ is the total number of patents in the region relative to the R&D employment weighted by the technology relatedness density; $\left(\frac{Patents}{L_{RD}} * UnrelD\right)_{r,t-1}$ is the total number of patents in the region relative to the R&D employment weighted by the technology unrelatedness density.

Our approach is based on the assumption that it is not the total number of patents per se that determines technological progress but rather their structure due to both related technologies and the technology portfolio of the regions. A similar approach can also be applied with respect to variety. Following Frenken et al. (2007) and Castaldi et al. (2015), it can be argued that it is not variety per se but rather related variety (RV) that fosters innovation, as related knowledge domains can be more easily and efficiently combined.[3] On the other hand, unrelated variety (UV) might slow down inventions in regions because the cognitive distance between technologies could be too large and combining them would be too costly and too risky (Boschma et al., 2023).

In contrast, Saviotti and Frenken (2008) argue that UV could stimulate breakthrough inventions that can result from combinations of unrelated knowledge resources. Recombination of knowledge from technologically distant fields is believed to result in more novel and valuable innovations (Boschma et al., 2023; Dahlin and Behrens, 2005). Our approach goes beyond the aforementioned studies, whose primary goal was to measure relatedness/unrelatedness in terms of variety. Furthermore, the issue whether RVs and UVs are linked to the region's technology portfolio was not addressed in any of the previous studies, while it has been argued that new technologies tend to draw from local knowledge sources (Audretsch and Feldman, 2004; Boschma et al., 2023; Breschi and Lissoni, 2009; Grashof et al., 2019; Hervás-Oliver et al., 2018).

Therefore, in line with the methodology proposed by Boschma et al. (2023), our study uses densities of technological relatedness/unrelatedness. Relatedness density means that at least two different technologies are related within a single patent and show a link to the regional technology portfolio. Unrelatedness density means that only one technology related to the regional technology portfolio is assigned to a patent. The difference between these measures is that both are related to the regional technology portfolio, but in the case of relatedness density, there is technological branching (related to other technologies) that does not occur in the case of unrelatedness density.

The issues outlined earlier, including the growing phenomenon of the occurrence of technological relatedness, provide the basis for the development of the following research hypotheses that are the subjects of empirical tests in the subsequent part of the article:

Hypothesis 1 (H1). The interaction of technology relatedness density with the total number of patents in relation to the R&D employment positively affects the growth rate of TFP and leads to technology diffusion.

Patents reflect the creation of new knowledge (Davies and Maré, 2021; Rigby et al., 2022). New knowledge expands a region's knowledge base, which is determined by related technologies and the region's technology portfolio. Related technologies can lead to such recombination of knowledge in new ways, leading to innovation (Boschma et al., 2023). The creation of innovations is easier, and the costs of adaptation are lower when a diversification towards new technologies occurs based on relevant regional capabilities, that is, regional technology portfolio (Balland et al., 2019; Boschma et al., 2023). On the other hand, the existence of several related technologies implies a diversification of the regional knowledge base, which increases the likelihood of greater absorption of external knowledge and can determine knowledge-diffusion processes in the region.

> *Hypothesis 2 (H2). The interaction of technology unrelatedness density with the total number of patents in relation to the R&D employment negatively affects the growth rate of TFP and does not lead to technology diffusion.*

The lack of related technologies is not conducive to the broadening and heterogeneity of the regional knowledge base, which can result in over-specialisation. Over-specialisation means that there are few unrelated technologies in the region. Technological progress is achieved by advancing knowledge in a particular technology. The opposite of over-specialisation is a region where there are many different related technologies. Technological progress is achieved by combining knowledge from these different related technologies.

Although it might seem counter-intuitive, the principle of relatedness is not about over-specialisation. It is rather about understanding the unique pathways that lead to diversification (Hidalgo et al., 2018). The lack of related technologies, and thus the limited heterogeneity of the regional knowledge base, limits new ideas and unusual combinations. In turn, this limits the creation of innovation (Abbasiharofteh et al., 2023; Boschma et al., 2023; Mewes, 2019). Similarly, the reduction of related technologies might negatively affect the capacity to absorb external knowledge, which hinders knowledge diffusion.

Variables, data and estimation method

Estimating TFP

TFP is a measure of productivity that allows countries or regions to transform intermediate inputs and factors of production into final output. Importantly, TFP takes into account the assumption that countries and regions may differ with respect to their factor endowments and efficiency with which they use production factors. Thus, TFP has the advantage of taking both aspects into account (Beugelsdijk et al., 2018). However, an important limitation is that TFP is not directly observable to an econometrician, and its estimation requires the adoption of an appropriate empirical approach.

A significant number of prior studies in regional economics employed a production function regression approach to obtain TFP estimates (Dettori et al., 2012; Schatzer et al., 2019; Siller et al., 2021). An important aspect of this approach is to avoid burdening the estimates obtained with endogeneity errors, which undermines their robustness. Taking this into account, we estimate the TFP using the recent methodology proposed by Rovigatti and Mollisi (2018), which is a modified version of the well-known Wooldridge (2009) estimator.[4] Their approach allows controlling for endogeneity of capital and labour by generating a proxy for unobservable TFP with observable variables.

Wooldridge (2009) shows how to implement his approach within the system GMM framework.[5] Lagged variables (both free and state) are the key instruments in the Wooldridge approach. However, using lagged variables as instruments leads to a significant reduction in the number of observations that becomes problematic in panels with a small time dimension. Therefore, his approach was modified by Rovigatti and Mollisi (2018) who used the dynamic panel instruments à la Blundell and Bond (1998) to obtain improved efficiency and predictive power of estimations.[6] Hence, in this study, we follow the procedure proposed by Rovigatti and Mollisi (2018).

TFP shows the effects of factors that are not captured in the production function by the standard factors of production such as capital or labour. After taking natural logs of the generalised Cobb-Douglas production function, we obtain a linear function of the following form:

$$\ln VA_{r,t} = \beta_0 + \beta_k \ln K_{r,t} + \beta_L \ln L_{r,t} + \beta_M M_{r,t} + \omega_{r,t} + \xi_{r,t} \tag{4}$$

where $\ln VA_{r,t}$ denotes the natural log of the value added in the region *r (r=1,2, . . . N)* in year *t (t=1, 2, . . . T)*, $K_{r,t}$ denotes the physical capital (i.e. gross value of fixed assets), $L_{r,t}$ denotes the number of employees, $M_{r,t}$ denotes the use of intermediate inputs (i.e. raw materials, components and services) and energy consumption,[7] $\omega_{r,t}$ is unobservable productivity (i.e. $TFP_{r,t}$) and $\xi_{r,t}$ is a normally distributed idiosyncratic error term. Unobservable productivity $\omega_{r,t}$ is assumed to evolve following the first-order Markov chain process:

$$\omega_{r,t} = E\left(\omega_{r,t} \middle| \omega_{r,t-1}\right) + \varepsilon_{r,t} = f\left(\omega_{r,t-1}\right) + \varepsilon_{r,t}, \quad \mathrm{t} = 2,3,\ldots,\mathrm{T}$$

The estimation results of the production function are shown in Table 2 in the Appendix 1.

Furthermore, $\ln TFP_{r,t}$ can be obtained using the RM estimates as:

$$\ln TFP_{r,t} = \ln VA_{r,t} - \beta_k \ln K_{r,t} - \beta_L \ln L_{r,t} \tag{5}$$

Using equation (5), we first calculate the TFP values in the Polish regions and then the distance in relation to the regional technological leader. As a technological leader, we select Mazowieckie voivodeship because it is the region in Poland with the highest TFP value (see Figure 1 in the Appendix 1).

Measuring relatedness and unrelatedness

In order to identify the degree of relatedness/unrelatedness between technologies, we use the metadata of patent records (i.e. granted patents) obtained from the Patent Office of the Republic of Poland, which we classify accordingly at the level of NUTS2 regions. The database used is our empirical study covers the period 2003–2019 and includes a total of 32,643 patents. Each patent record is assigned at least one International Patent Classification (IPC) code. A single IPC code means a single technology. Using IPC at a subclass level, we identify a unique list of 622 codes. Thus, at the preliminary step of our analysis, we create 622×622 diagonal matrices with a number of pairs of each technologies *i* and *j*, in each year and region.

In the first step of our analysis, a probabilistic measure of co-occurrence between any of two technologies within each granted patent is calculated using formula (6) (for further details, see Antonietti and Montresor, 2021; Boschma et al., 2023; Kim et al., 2023).

$$\phi_{ijr,t} = \frac{m_{r,t} c_{ijr,t}}{s_{ir,t} s_{jr,t}} \tag{6}$$

where:

$\phi_{ijr,t}$ is the measure of the co-occurrence of two technologies *i* and *j* in the region *r* in year *t*;

$m_{r,t}$ denotes the total number of all patents in the region *r* in year *t*;

$c_{ijr,t}$ denotes the number of times the unique pair of subsequent technologies *i* and *j* occurred together in the same patent in the region *r* in year *t*;

$s_{ir,t}$ is the total number of times the technology *i* appeared in pairs with any other technologies in the region *r* in year *t*;

$s_{jr,t}$ is the total number of times the technology *j* appeared in pairs with any other technologies in the region *r* in year *t*.

Using calculated values of $\phi_{ijr,t}$ from formula (6), two separate 622×622 diagonal matrices can be obtained for each region and year. The first matrix describes unrelatedness of technologies. Each element in unrelatedness matrix $\Psi_{ijr,t}=1$, if $\phi_{ijr,t} \leqslant 1$, otherwise $\Psi_{ijr,t}=0$. Thus, this matrix indicates that technologies *i* and *j* co-occur by chance. The second matrix describes the relatedness of technologies. Each element in relatedness matrix $\eta_{ijr,t}=1$, if $\phi_{ijr,t} > 1$, otherwise $\eta_{ijr,t}=0$. Thus, this matrix indicates that technologies *i* and *j* co-occur purposely.

In the second step of our analysis, we use formula (7) to indicate how close each technology *i* is located to the technological base in each year and in each of 16 Polish regions. Thus, the calculated value indicates the revealed technological advance of each technology *i* in relation to the technological base of each region and each year of the research period (cf. Antonelli et al., 2017; Boschma et al., 2023).

$$RTA_{ir,t} = \frac{\dfrac{Patents_{ir,t}}{\sum_{i=1}^{622} Patents_{ir,t}}}{\dfrac{\sum_{r=1}^{16} Patents_{ir,t}}{\sum_{r=1}^{16}\sum_{i=1}^{622} Patents_{ir,t}}} \quad (7)$$

where:

$RTA_{ir,t}$ is the revealed technological advantage of each technology i in each region r and year t;

$Patents_{ir,t}$ represents the total number of patents with a unique ID number discover in each technology i in the region r in year t, regardless of other parameters describing patents (there are no records of patents with duplicates ID number within each technology i).

The calculated values $RTA_{ir,t}$ from equation (7) allow us to create regional matrices with 622 rows representing each technology i for each year where each element = 1 if $RTA_{ir,t} > 1$ and otherwise 0.

In the third step of our analysis, we calculate two separate matrices of density indicators describing unrelatedness and relatedness, respectively. The unrelatedness density indicator matrix is calculated using formula (8).

$$UnrelD_{ir,t} = \frac{\sum_{i,t \in r} \Psi_{ijr,t} RTA_{ir,t}}{\sum_{ir,t} \Psi_{ijr,t}} \quad (8)$$

where:

$UnrelD_{ir,t}$ is the unrelatedness density indicator of technology i in the region r in year t;

$\Psi_{ijr,\ t}$ is the unrelatedness technologies matrix obtained based on results in equation (6).

The unrelatedness density indicator $UnrelD_{ir,t}$ determines how the set of technologies unrelated to technology i is close to the set of technologies in each year y and region r described by $RTA_{ir,t}$ value. The calculated values obtained from formula (8) yield the matrix of unrelatedness density indicator of technologies, where rows represent technologies i and columns denote years t of the research period in each region r. This indicator takes a value in the range [0, 1], or from 0% to 100%. The interpretation of the unrelatedness density indicator is as follows. For example, $UnrelD_{ir,t} = 0.5$ implies that region r in year t has an RTA in 50 per cent of the technologies that are unrelated to technology i.

The relatedness density indicator is calculated using formula (9).

$$RelD_{ir,t} = \frac{\sum_{i,t \in r} \eta_{ijr,t} RTA_{ir,t}}{\sum_{ir,t} \eta_{ijr,t}} \quad (9)$$

where:

$RelD_{ir,t}$ is the relatedness density indicator of technology i in region r in year t;

$\eta_{ijr,\ t}$ is the relatedness technologies matrix obtained based on results in formula (6).

The relatedness density indicator $RelD_{ir,t}$ determines how the set of technologies related to technology i is close to the technologies in each region r and year t described by the $RTA_{ir,t}$ value. The calculated values obtained from formula (9) yield the matrix of relatedness density indicators of technologies, where rows represent technologies i and columns denote years t of the research period in each region r. The indicator takes a value in the range [0, 1], or from 0% to 100%. The interpretation of the relatedness density indicator is as follows: $RelD_{ir,t} = 0.4$ implies that region r in year t has an RTA in 40 per cent of the technologies that are related to technology i.[8]

Measuring human capital

The human capital indicator for Polish regions is calculated using information on the structure of the working-age population with different levels of education. This is the standard approach frequently employed in the labour economics literature to calculate the human capital index on the basis of the average number of years of schooling and the assumed rate of return on education as estimated by the Mincer equation (Caselli, 2005; Psacharopoulos, 1994).[9] The necessary data are obtained from the Eurostat regional database which contains information on the percentage of the working-age population (25–64) in European regions with a specific level of education according to the International Standard

Classification of Education (ISCED). The database includes the following levels of education:

- ISCED 2011 (level 0–2): Less than primary, primary, lower secondary.
- ISCED 2011 (level 3–4): Upper secondary, post-secondary non-tertiary.
- ISCED 2011 (level 5–8): Tertiary education (first and second stage).

Following Barro and Lee (2013) and according to ISCED 2011, we assume that on average, the level of education (0–2) corresponds to 9 years of education, level (3–4) to 12 years of education and level (5–8) to 16 years of education in Poland. Using these assumptions, the average years of schooling ($ays_{r,t}$) are calculated for each region by multiplying the shares by the years of education adopted at each level (Cieślik and Misiak, 2023). Then, in order to convert the average years of schooling into the human capital index, we use the Mincerian human capital function of the following form:

$$H_{r,t} = e^{\varphi(ays_{r,t})} \tag{10}$$

where: $\varphi(ays_{r,t})$ represents a piecewise linear function that is parameterized as follows (Beugelsdijk et al., 2018; Inklaar and Timmer, 2013):

$$\varphi(ays_{r,t}) = \begin{cases} 0.134 \cdot ays_{r,t} \;\; if \;\; ays_{r,t} \leq 4 \\ 0.134 \cdot 4 + 0.101\left(ays_{r,t} - 4\right) if \;\; 4 < ays_{r,t} \leq 8 \\ 0.134 \cdot 4 + 0.101 \cdot 4 + 0.068\left(ays_{r,t} - 8\right) if \;\; ays_{r,t} > 8 \end{cases}$$

In this way, we obtain a human capital index for each region and year, which we use later in our further analysis.

Control variables

We use several regional control variables in our regressions that include various measures of international openness and total variety. The use of control variables in the regressions stems from the regional economics literature, which shows that international openness may stimulate diffusion of technology from abroad. In addition, total variety may allow controlling for Jacobs externalities. Following Jacobs (1969), it can be argued that variety triggers new ideas, and this fosters the creation of innovation and TFP growth.

Total variety (hereafter variety) is measured according to the method proposed by Hartog et al. (2012). It is calculated as the sum of the entropy at the specific digit level and describes variety in the industrial structure of a specific region. The calculated values of this variable are higher in regions characterised by a highly diversified industrial structure.

$$Variety_{r,t} = \sum_{g=1}^{G} P_n \log_2(\frac{1}{P_n}) \tag{11}$$

where:

P_n is the share of employment in the S_n sector (where $n = 1, \ldots, N$) belonging to the same S_g sector.

International openness is taken into account in three possible ways. First, we include $open1_{r,t}$ defined as a ratio of the sum of exports and imports to GDP. Second, we use $open2_{r,t}$ defined as a ratio of only exports to GDP. Finally, we employ $open3_{r,t}$ defined as a ratio of only imports to GDP. Considering international openness in three ways allows us to identify which channel of potential international technology diffusion has the strongest effect on the rate of growth of TFP. In order to control for the possibility of knowledge transfers from abroad through channels other than international trade, such as foreign direct investment, we introduce two additional variables:

- The ratio of employment in companies with foreign capital to total employment in the region;

- The ratio of the number of firms with foreign capital participation to total number of firms in the region.

The majority of the data used to calculate the aforementioned variables come from the databases of *Statistics Poland* (i.e. the Polish Central Statistical Office). These include the number of employees, the employment shares, the stock of physical capital (gross value of fixed assets), the value added, intermediate consumption (data on raw materials, components and services used), energy consumption, regional price indices, R&D expenditures, GDP and other. As already mentioned, data on patent records are obtained from the Patent Office of the Republic of Poland. The data on the structure of the working age population with different levels of education come from the Eurostat's regional database. Finally, data on the values of exports and imports in the regions of Poland come from the resources of the National Revenue Administration. Nominal variables have been converted in the real terms using regional price indices and expressed in constant 2019 prices. The basic descriptive statistics and correlation matrix for these variables are summarised in the Appendix in Tables 3 and 4.

Estimation method

On the basis of the identified variables, using equations (2) and (3) and introducing additional regional control variables, we estimate the following equation:

$$\begin{aligned} grTFP_{r,t} &= const + \alpha_1 lnH_{r,t-1} - \alpha_2 lnH_{r,t-1} * \frac{TFP_{r,t-1}}{TFP_{m,t-1}} + \alpha_3 \left(\frac{Patents}{L_{RD}} * RelD \right)_{r,t-1} \\ &- \alpha_4 \left(\frac{Patents}{L_{RD}} * RelD \right)_{r,t-1} \left(\frac{TFP_{r,t-1}}{TFP_{m,t-1}} \right) + \alpha_5 \left(\frac{Patents}{L_{RD}} * UnrelD \right)_{r,t-1} \\ &- \alpha_6 \left(\frac{Patents}{L_{RD}} * UnrelD \right)_{r,t-1} \left(\frac{TFP_{r,t-1}}{TFP_{m,t-1}} \right) + \alpha_7 controlvariable_{r,t} + \varepsilon_{r,t} \end{aligned} \tag{12}$$

where:

$TFP_{r,t}$ is the total factor productivity in region r in year t;

$TFP_{m,t}$ is the TFP of regional technology leader m in year t;[10]

$\left(\frac{TFP_r}{TFP_m} \right)$ is the ratio of a region's TFP to the TFP of the regional technology leader. It is a measure that reflects the technology gap with respect to the leader and determines the diffusion of technology on a catch-up basis.[11]

$H_{r,t}$ is the index of human capital in region r in year t;

$\left(\frac{Patents}{L_{RD}} \right)_{r,t}$ is the number of patents in relation to the R&D employment in region r in year t;

$RelD_{r,t}$ is the technology relatedness density in region r in year t;

$UnrelD_{r,t}$ is the technology unrelatedness density in region r in year t;

control variables:

$variety_{r,t}$ is the total sectoral variety in region r in year t;

$open_{r,t}$ is international openness in region r in year t taken into account in three ways:

- $open1_{r,t}$ = as the ratio of exports + imports to GDP;
- $open2_{r,t}$ = as the ratio of exports to GDP;
- $open3_{r,t}$ = as the ratio of imports to GDP.

Share $empl_{r,t}$ is the ratio of employment in companies with foreign capital participation to total employment in region r in year t.

Share firms$_{r,t}$ is the ratio of the number of firms with foreign capital participation to the total number of firms in region r in year t.

$\varepsilon_{r,t}$ is the error term.

We estimate equation (12) assuming the logistic technology diffusion function and using a two-step system GMM estimator with Windmeijer (2005) correction. The potential choice of estimators used in this kind of studies is relatively wide. A frequently used estimation technique in older studies was ordinary least squares with fixed effects (hereafter OLS FE). However, it has been argued that the use of this estimator is problematic due to the measurement errors as it ignores variation between regions and a reduction in bias results in higher standard errors. Another approach that could be potentially employed is random effects (REs) when the assumptions of exogeneity and the lack of correlation between the independent variables and the individual effect are satisfied.

At the same time, however, it should be noted that these two aforementioned approaches are static. In a dynamic panel data model with individual effects together with the lagged values of the dependent variable, even assuming that the error term is independent and identically distributed, the FE and RE estimators tend to be inconsistent due to the well-known Nickell (1981) error. Therefore, OLS regressions may lead to inconsistent estimates due to the correlation between lagged values and the error term. Another problem is that in the case of equation (12), there might be a potentially strong problem of endogeneity.

Therefore, the choice of a proper estimation method is not random. First, equation (12) does not take into account the lagged value of the dependent variable which might suggest that from an econometric point of view, it is a static model. However, the dependent variable is the TFP growth rate, which means that the equation can be estimated using a dynamic panel data estimator frequently employed to estimate growth regressions in the macroeconomic literature.

Second, the use of the system GMM estimator allows us addressing two potentially important problems: the endogeneity problem and the fact that the series of some variables may be very persistent. Endogeneity may potentially result from the employed measure of R&D activity. The same might apply to the relationship between human capital and TFP, where the direction of causality is not obvious. Therefore, it can be argued that the use of system GMM allows mitigating the potential problem of endogeneity. In addition, the fact that it takes into account estimates of both levels and first differences solves the problem of poor identification even in the case of very persistent series (cf. Roodman, 2009).[12]

Estimation results

We report our estimation results in Table 1 in columns (1)–(11). In column (1), we show the results of the baseline estimation without the regional control variables. In column (2), we show the results with *variety*$_{r,t}$, as the only control variable, which reflects the aggregate sectoral variation in the Polish regions. Then, in columns (3)–(7), we control also for the potential effects of international knowledge diffusion. In addition, we separately test the effect of the density of technological relatedness by assuming that there is only technology relatedness (columns 8 and 9) and only unrelatedness (columns 10 and 11).

The baseline estimation results reported in column (1) show that the inclusion of technology relatedness/unrelatedness yields the expected effects.[13] In particular, the effect of human capital on TFP growth rate is positive and significant at the 1% level. The estimate of the indirect effect of human capital on technology diffusion is also significant at 1% with the expected negative sign. The interaction of the technology relatedness (unrelatedness) density with the total number of patents relative to the R&D employment shows that the direct effect on the TFP growth rate is positive (negative) and significant only at the 10% level.

In the baseline estimations, the findings on the indirect effect of our measure of R&D activity are mixed. Taking into account the technology relatedness density, we obtain the expected negative sign, but the estimated coefficient is not significant. On the other hand, taking into account the technology unrelatedness density, we obtain the result that is statistically significant at the 10% level but with a positive sign. This means that technology diffusion is not

Table 1. Dependent variable: $grTFP_{r,t}$.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
$grTFP_{r,t-1}$	1.1977***	1.3115***	1.3069***	1.2941***	1.3139***	1.3239***	1.2960***	1.2768***	1.3753***	1.1933***	1.3039***
	(0.0990)	(0.1073)	(0.1069)	(0.1039)	(0.1078)	(0.1225)	(0.1095)	(0.2375)	(0.2120)	(0.0896)	(0.0939)
$lnH_{r,t-1}$	0.3317***	0.3495***	0.3336***	0.3191***	0.3432***	0.3564***	0.3453***	0.2549*	0.2218	0.3404***	0.3567***
	(0.0721)	(0.0767)	(0.0730)	(0.0721)	(0.0745)	(0.0919)	(0.0702)	(0.1402)	(0.1468)	(0.0646)	(0.0678)
$lnH_{r,t-1} * \frac{TFP_{r,t-1}}{TFP_{m,t-1}}$	−0.0159**	−0.0163***	−0.0184***	−0.0147**	−0.0191***	−0.0167***	−0.0156***	−0.5993	−0.6798*	−0.0187***	−0.0215***
	(0.0056)	(0.0054)	(0.0062)	(0.0068)	(0.0065)	(0.0052)	(0.0053)	(0.3560)	(0.3482)	(0.0054)	(0.0061)
$\left(\frac{Patents}{L_{RD}} * RelD\right)_{r,t-1}$	0.0022*	0.0028**	0.028***	0.0029***	0.0027**	0.0031**	0.0036***	0.0056**	0.0062***	–	–
	(0.0011)	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0012)	(0.0012)	(0.0019)	(0.0017)		
$\left(\frac{Patents}{L_{RD}} * UnrelD\right)_{r,t-1}$	−0.0031*	−0.0038**	−0.0037**	−0.0036**	−0.0038**	−0.0039**	−0.0039**	–	–	−0.0023*	−0.0027**
	(0.0016)	(0.0014)	(0.0013)	(0.0014)	(0.0014)	(0.0013)	(0.0013)			(0.0013)	(0.0012)
$\left(\frac{Patents}{L_{RD}} * RelD\right)_{r,t-1} \left(\frac{TFP_{r,t-1}}{TFP_{m,t-1}}\right)$	−0.0044	−0.0068*	−0.0069*	−0.0082**	−0.0064*	−0.0066*	−0.0076**	−0.0184**	−0.0203**	–	–
	(0.0036)	(0.0037)	(0.0035)	(0.0038)	(0.0036)	(0.0033)	(0.0034)	(0.0068)	(0.0067)		
$\left(\frac{Patents}{L_{RD}} * UnrelD\right)_{r,t-1} \left(\frac{TFP_{r,t-1}}{TFP_{m,t-1}}\right)$	0.0063*	0.0085**	0.0083**	0.0080**	0.0085**	0.0088**	0.0083**	–	–	0.0045	0.0060*
	(0.0035)	(0.0034)	(0.0033)	(0.0033)	(0.0033)	(0.0031)	(0.0031)			(0.0033)	(0.0030)
control variable:	–	0.0271***	0.0273***	0.0276***	0.272***	0.0271***	0.0270***	–	0.0269***	–	0.0267***
$variety_{r,t}$		(0.0043)	(0.0044)	(0.0044)	(0.0043)	(0.0044)	(0.0043)		(0.0040)		(0.0036)
$openl_{r,t}$	–	–	0.0077	–	–	–	–	–	–	–	–
			(0.0050)								
$open2_{r,t}$	–	–	–	0.0234*	–	–	–	–	–	–	–
				(0.0125)							
$open3_{r,t}$	–	–	–	–	0.0083	–	–	–	–	–	–
					(0.0071)						
share $empl_{r,t}$	–	–	–	–	–	−0.0233	–	–	–	–	–
						(0.0586)					
share $firms_{r,t}$	–	–	–	–	–	–	−1.034	–	–	–	–
							(1.020)				
constant	−0.4130***	−0.5813***	−0.5659***	−0.5522***	−0.5750***	−0.5857***	−0.5712***	−0.1037	−0.1809	−0.4221***	−0.5849***
	(0.0911)	(0.1117)	(0.1073)	(0.1063)	(0.1090)	(0.1273)	(0.1026)	(0.2369)	(0.2429)	(0.0805)	(0.0036)
AR(1)	−3.03	−2.95	−2.94	−2.94	−2.94	−2.95	−2.93	−2.84	−2.90	−3.04	−2.97
	[0.002]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.004]	[0.004]	[0.002]	[0.003]
AR(2)	−0.15	0.07	0.09	0.07	0.09	0.06	0.08	−0.27	−0.07	−0.15	0.07
	[0.883]	[0.943]	[0.931]	[0.941]	[0.931]	[0.949]	[0.932]	[0.789]	[0.946]	[0.883]	[0.947]
Sargan test	1.19	0.51	0.50	0.50	0.51	0.51	0.45	1.91	1.01	1.12	0.47
	[0.276]	[0.474]	[0.478]	[0.479]	[0.477]	[0.477]	[0.504]	[0.167]	[0.314]	[0.290]	[0.495]
Hansen test	1.33	0.65	0.64	0.64	0.64	0.62	0.60	1.77	0.96	1.36	0.63
	[0.248]	[0.421]	[0.423]	[0.423]	[0.422]	[0.433]	[0.440]	[0.184]	[0.327]	[0.244]	[0.427]
F-test	103.96	74.40	56.33	52.83	61.79	65.79	57.56	10.20	21.57	66.07	71.82
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Number of observations	240	240	240	240	240	240	240	240	240	240	240
Number of instruments	9	10	11	11	11	11	11	7	8	7	8

Notes: Corrected standard errors (SEs) in parentheses; *p*-values for tests in square parentheses [*p*-value]. We perform the estimations using the xtabond2 command in Stata 17, which implements the system GMM with Windmeijer (2005) correction.
Source: Authors' estimations.
***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

observed, and the technology gap between the leader and the follower can be increasing over time.

The inclusion of the additional control variables improves the quality of our estimation results. In particular, $variety_{r,t}$ introduced into the estimated specifications summarised in columns (2)–(7) shows a positive and statistically significant effect on the growth rate of TFP at the 1% level. Moreover, already in the specification shown in column (2), the control variable $variety_{r,t}$ significantly improves the significance of the estimates of our measures of R&D activity. It turns out that direct effects with measures of technology relatedness (unrelatedness) density show the expected positive (negative) sign and are statistically significant at the 5% level. The technology-relatedness density also indirectly affects technology diffusion, and the sign of the estimated parameter is, as expected, negative and significant at the 10% level. In contrast, the technology unrelatedness density, that is significant at the 5% level, does not lead to technology diffusion.

Finally, international openness is included in columns (3)–(5) in three ways: $open1_{r,t}$ *to* $open3_{r,t}$, respectively. The estimation results show that only in the specification in column (4), where we include $open2_{r,t}$, the measure of international openness calculated as the ratio of exports to GDP positively affects the TFP growth rate and is significant at the 10% level. In addition, in this specification, the introduction of international openness affects both the value of the estimated parameter that determines the effect of indirect technology relatedness density on the diffusion process and its level of statistical significance (significant at the 5% level). This result is not surprising, since firms with foreign capital participation are responsible for over 50% of Polish exports. Foreign-owned companies often equip their export-oriented subsidiaries in Poland with technologically advanced production facilities that produce high-quality products for export. Thus, they can directly influence, through exports, the growth rate of TFP, and these advanced technologies and their R&D departments can indirectly contribute to technology diffusion.

In addition, in columns (6) and (7), we also directly control for the importance of foreign-owned firms in generating technological progress using two additional variables $shareempl_{r,t}$ and $sharefirms_{r,t}$. However, these variables turned out to be statistically not significant but affected the parameter estimates of the variables accounting for technological relatedness on both innovation creation and diffusion effects. Additional sensitivity controls using the results of the impact on the regional TFP growth rate of separately technology relatedness and unrelatedness density, reported in columns (8)–(11), show that our results remain robust. Hence, we can derive conclusions very similar to those based on the results summarised in columns (1)–(7).

Summing up, our results summarised in Table 1 do not allow rejecting any of our research hypotheses. All post-estimation tests performed for estimates in columns (1)–(11) show that the instruments used are valid and that the model is not overidentified (see tests in Table 1). In addition, the results summarised in columns (2)–(7) show the robustness of the estimated parameters and the introduction of additional regional variables with which we control for sectoral variety within regions and international openness improve the significance levels of some estimates. Thus, the estimation results remain robust for all specifications after controlling for sensitivity.

Conclusion

Based on the 'two faces' approach, we examined the role of human capital and the R&D sector in technological progress and convergence to the technological frontier in Poland at the level of NUTS2 regions. We used the analytical approach, which, compared to previous studies, allowed for a broader analysis and a more detailed description of the effects of R&D on the level of technological development of the regional economy. This innovative approach allowed us to show that technological relatedness had a positive effect on the creation of innovations and led to technology diffusion. In contrast, technological unrelatedness had a negative effect on innovation and did not lead to technology diffusion.

These new and original findings significantly contribute to the existing pool of knowledge and have several important policy implications regarding effective ways to accelerate the technological development of the economy at the regional level.

In particular, it can be argued that it is technological relatedness that drives technological development. Technology unrelatedness negatively affects progress (i.e. retards progress) and can lead to over-specialisation, which is generally not conducive to technology diffusion. Our results remained robust with respect to a number of sensitivity tests. In general, our results did not allow rejecting any of our research hypotheses. Furthermore, our results supported the relevance and reliability of the theoretical and empirical frameworks we used in the study of the effects of human capital and R&D on the growth rate of TFP.

Our findings clearly indicate the need to employ more general research frameworks in future studies of these phenomena than it has been done so far. The empirical results let us conclude that in light of the observed acceleration of the introduction of increasingly advanced technologies, both human capital and the R&D sector are two key elements stimulating regional development. Therefore, both the creation of human capital and the development of the R&D activities should be supported (but only when they lead to creation of complex, multi-diversified technology solutions), especially in lagging-behind regions.

The application of the principle of relatedness within the TFP growth model clearly shows that over-specialisation is ineffective. In particular, it does not stimulate technological progress. Unique pathways that lead to diversification and related technologies and take into account the regional technological potential determine both the creation of new innovations and fostering technology diffusion. This allows us also to argue that economic policies should focus not only on identifying promising technologies but also, above all, on identifying mechanisms to facilitate knowledge flows between industries and regions (see also Balland et al., 2019; Hidalgo et al., 2018).

In particular, these policies should focus on promoting the flows of people who can bring with them the knowledge that is lacking in the region. Public policies at different spatial scales (EU, national, regional) should play an important role in this regard. In particular, they should support the development of research infrastructure (universities), the education system (primary, secondary), patent legislation and regional policies. However, it is important to note that such policies should take into account local technology resources and the technological capabilities of the regions, which can be combined through research cooperation programmes (see Boschma et al., 2023).

Our findings on technology relatedness, despite important methodological differences compared to earlier studies, are generally consistent with the prior findings of Rigby et al. (2022). In particular, they find that regions that diversified into related technologies make the local economy more complex and show higher growth rates in terms of both GDP and employment. Our results suggest that technological relatedness is more important than unrelatedness for the technological development of the local economy. These results allow us to conclude that the greater the number of related technologies, the greater the development potential for technological growth. This finding is important not only in the limited context of the local economy but also in the wider context of developing countries attempting to reach higher levels of economic development.

The main limitation of our study is the use of data only for single-country regions. Therefore, this kind of research should be repeated for a larger set of regions including, for example, the regions of the entire EU or other groups of countries. An important challenge for future research is also the appropriate modification of measures of the technology relatedness/unrelatedness density in the direction of intra-sectoral and inter-sectoral relatedness and more comprehensive causality studies.

Authors' Note

The article is a result of research collaboration between the University of Warsaw, Rzeszow University of Technology, Jagiellonian University in Kraków and the Patent Office of the Republic of Poland based on cooperation agreements (2021, 2023).

Acknowledgements

The authors are grateful to the Patent Office of the Republic of Poland for providing a database of patent records (patents granted) that made it possible to determine the technology relatedness and unrelatedness densities in the regions.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Tomasz Misiak https://orcid.org/0000-0002-4296-0291

Notes

1. A similar approach was employed by Benhabib and Spiegel (2005); however, their study focuses only on the role of human capital.
2. It is worth noting that there is a peculiar intertwining between these two variables. In particular, human capital and knowledge it embodies are used in R&D activities. However, in empirical studies, both human capital and R&D activities should be treated separately since not all knowledge that human capital represents can be used in R&D activities.
3. See also related studies by Aarstadt et al. (2016) and Martynovich and Taalbi (2020).
4. The estimations were obtained using the *prodest* function in STATA statistical package.
5. According to his method, productivity is an unknown function of lagged productivity, which allows the application of the system GMM.
6. The use of dynamic panel instruments allows taking advantage of additional information embodied in lagged instruments without losing observations and estimation power.
7. $M_{r,t}$ in equation (4) only is used as a proxy for unobservable TFP to control for endogeneity and thus better estimate parameters for K and L.
8. At the level of the co-occurrence matrix, any two technologies are either related or unrelated (being unrelated and related are opposites). Once only technologies in which a region has an RTA are included and considered, they are no longer opposites (see Boschma et al. 2023).
9. The same procedure is employed in the calculation of the human capital index at the country level in Penn World Table.
10. We select the Mazowieckie Voivodeship as the regional technology leader because TFP values for this region are the highest, and the TFP gap with respect to the other regions is the largest (see Figure 1 in the Appendix 1). This does not mean, however, that we restrict our analysis to the diffusion of technology from the leading region only. Although we do not distinguish between the specific channels of knowledge diffusion in our study, we use TFP to check whether diffusion is actually taking place and whether the technology gap is narrowing.
11. This term appears in the model in interaction with both human capital and the variable defining R&D activity showing the effects of these variables on technology diffusion. A negative sign indicates diffusion, and a positive sign indicates a lack of diffusion and the problem of a widening technology gap between regions and the regional technology leader. In essence, this term reflects the 'second face' of both human capital and R&D activity.
12. We also tried to estimate equation (12) using panel spatial models. However, the NUTS2 regions analysed in our study are too large, and the spatial relationships are not statistically significant. This is supported by the results of Moran tests for spatial dependence, which are shown in Table 6 in the Appendix 1.
13. For the sake of comparison, in Table 5 in the Appendix 1, we report also the estimation results based on equation (12) that uses alternative (so far, most commonly used in research) measures of R&D activity. However, the majority of these alternative measures are not statistically significant.

References

Aarstad J, Kvitastein O and Jakobsen SE (2016) Related and unrelated variety as regional drivers of enterprise productivity and innovation: a multilevel study. *Research Policy* 45(2016): 844–856.

Abbasiharofteh M, Kogler DF and Lengyel B (2023) Atypical combinations of technologies in regional co-inventor networks. *Research Policy* 52(10): 104886.

Aghion P and Howitt P (1992) A model of growth through creative destruction. *Econometrica* 60(2): 323–351.

Aghion P and Howitt P (2008) *The economics of growth*. Cambridge, MA: MIT Press.

Altuzarra A (2018) R&D and patents: is it a two way street? *Economics of Innovation and New Technology* 28(2): 180–196.

Antonelli C, Crespi F, Mongeau Ospina CA and Scellato G (2017) Knowledge composition, Jacobs externalities and innovation performance in European regions. *Regional Studies* 51(11): 1708–1720.

Antonietti R and Montresor S (2021) Trajectories and Key Enabling Technologies (KETs) in Italian Regions. *Economic Geography* 97(2): 187–207.

Artz KW, Norman PM, Hatfield DE and Cardinal LB (2010) A longitudinal study of the impact of R&D, Patents, and product innovation on firm performance. *Journal of Product Innovation Management* 27(5): 725–740.

Audretsch DB and Feldman MP (2004) Knowledge spillovers and the geography of innovation. In Henderson V and Thisse JF (ed.) *Handbook of Regional and Urban Economics*, vol. 4. Amsterdam: Elsevier North Holland, pp., 2713–2739.

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.

Balland PA, Boschma R, Crespo J and Rigby DL (2019) Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification. *Regional Studies* 53(9): 1252–1268.

Banks RB (1994) *Growth and Diffusion Phenomena*. Berlin: Springer Verlag.

Barro RJ and Lee JW (2013) A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics* 104(C): 184–198.

Basu S and Weil DN (1998) Appropriate technology and growth. *Quarterly Journal of Economics* 113(4): 1025–1054.

Benhabib J and Spiegel MM (2005) Human capital and technology diffusion. In: Aghion P and Durlauf SN (eds) *Handbook of Economic Growth*, vol. 1A. Amsterdam: Elsevier, pp. 935–966.

Beugelsdijk S, Klasing MJ and Milionis P (2018) Regional economic development in Europe: the role of total factor productivity. *Regional Studies* 52(4): 461–476.

Blundell R and Bond S (1998) Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 24(1): 115–143.

Boschma R (2017) Relatedness as driver behind regional diversification: a research agenda. *Regional Studies* 51(3): 351–364.

Boschma R, Balland PA and Kogler DF (2015) Relatedness and technological change in cities: the rise and fall of technological knowledge in U.S. metropolitan areas from 1981 to 2010. *Industrial and Corporate Change* 24(1): 223–250.

Boschma R, Miguelez E, Moreno R and Ocampo-Corrales DB (2023) The role of relatedness and unrelatedness for the geography of technological breakthroughs in Europe. *Economic Geography* 99(2): 117–139.

Breschi S and Lissoni F (2009) Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *Journal of Economic Geography* 9(4): 439–468.

Cameron G, Proudman J and Redding S (2005) Technological convergence, R&D, trade and productivity growth. *European Economic Review* 49(3): 775–807.

Capello R and Lenzi C (2015) Knowledge, innovation and productivity gains across European regions. *Regional Studies* 49(11): 1788–1804.

Caselli F (2005) Accounting for cross country income differences. In Aghion P and Durlauf S (eds.), *Handbook of Economic Growth*, vol. 1A. Amsterdam: Elsevier, pp. 679–741.

Castaldi C, Frenken K and Los B (2015) Related variety, unrelated variety and technological breakthroughs: an analysis of US state-level patenting. *Regional Studies* 49(5): 767–781.

Cieślik A and Misiak T (2023) Related and unrelated variety and convergence to technological frontier: Empirical evidence for Polish regions. *European Planning Studies* 32(3): 648–667.

Cohen WM and Levinthal DA (1989) Innovation and learning: the two faces of R&D. *Economic Journal* 99(397): 569–596.

Dahlin KB and Behrens DM (2005) When is an invention really radical? Defining and measuring technological radicalness. *Research Policy* 34(5): 717–737.

Davies B and Maré DC (2021) Relatedness, complexity and local growth. *Regional Studies* 55(3): 479–494.

Dettori B, Marrocu E and Paci R (2012) Total factor productivity, intangible assets and spatial dependence in the European Regions. *Regional Studies* 46(10): 1401–1416.

Farinha T, Balland PA, Morrison A and Boschma R (2019) What drives the geography of jobs in the US? Unpacking relatedness. *Industry and Innovation* 26(9): 988–1022.

Frenken K, Van Oort FG and Verburg T (2007) Related variety, unrelated variety and regional economic growth. *Regional Studies* 41(5): 685–697.

Grashof N, Hesse K and Fornahl D (2019) Radical or not? The role of clusters in the emergence of radical innovations. *European Planning Studies* 27(10): 1904–1923.

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.

Grossman G and Helpman E (1990) Comparative advantage and long-run growth. *American Economic Review* 80(4): 796815.

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

Hartog M, Boschma R and Sotarauta M (2012) The impact of related variety on regional employment growth in Finland 1993–2006: high-tech versus medium/low-tech. *Industry and Innovation* 19(6): 459–476.

Hervás-Oliver JL, Albors-Garrigos J, Estelles-Miguel S and Boronat-Moll C (2018) Radical innovation in Marshallian industrial districts. *Regional Studies* 52(10): 1388–1397.

Hidalgo C, Klinger B, Barabassi A and Hausmann R (2007) The product space conditions the development of nations. *Science* 317(5837): 482–487.

Hidalgo CA (2021) Economic complexity theory and applications. *Nature Reviews Physics* 3(2): 92–113.

Hidalgo CA, Balland PA, Boschma R, Delgado M, Feldman M, Frenken K and Zhu S (2018) The principle of relatedness. In Morales AJ, Gershenson C, Brah D, Minai AA and Bar-Yam Y (eds), *Unifying Themes in Complex Systems IX: Proceedings of the Ninth International Conference on Complex Systems 9*. Cham: Springer International Publishing, pp. 451–457.

Inklaar R and Timmer MP (2013) *Capital, Labor and TFP in PWT8.0 (Mimeo)*. Groningen: University of Groningen.

Jacobs J (1969) *The Economy of Cities*. New York: Vintage Books.

Kaplan S and Vakili K (2015) The double-edge sword of recombination in breakthrough innovation. *Strategic Management Journal* 36(10): 1435–1457.

Kim SH, Jun B and Lee JD (2023) Technological relatedness: how do firms diversify their technology? *Scientometrics* 128(1): 1–3.

Männasoo K, Hein H and Ruubel R (2018) The contributions of human capital, R&D spending and Convergence to total factor productivity growth. *Regional Studies* 52(12): 1598–1611.

Martynovich M and Taalbi J (2020) *Related Variety, Recombinant Knowledge and Regional Innovation. Evidence for Sweden, 1991–2010, Papers in Evolutionary Economic Geography, 20.15*. Utrecht: Utrecht University.

Mewes L (2019) Scaling of atypical knowledge combinations in American metropolitan areas from 1836 to 2010. *Economic Geography* 95(4): 341–361.

Montresor F and Quatraro F (2017) Regional branching and Key enabling technologies: evidence from European patent data. *Economic Geography* 93(4): 367–396.

Neffke F, Henning M and Boschma R (2011) How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. *Economic Geography* 87(3): 237–265.

Nelson RR and Phelps ES (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): 1399–1416.

Psacharopoulos G (1994) Returns to investment in education: a global update. *World Development* 22(9): 1325–1343.

Rigby DL, Roesler C, Kogler D, Boschma R and Balland P-A (2022) Do EU regions benefit from smart specialization principles? *Regional Studies* 56(12): 2058–2073.

Romer P (1990) Endogenous technological change. *Journal of Political Economy* 98(5–2): 71–102.

Romer P (1994) The origins of endogenous growth. *Journal of Economic Perspectives* 8(1): 3–22.

Roodman D (2009) How to do Xtabond2: an Introduction to Difference and System GMM in Stata. *The Stata Journal* 9(1): 86–136.

Roper S and Hewitt-Dundas N (2015) Knowledge stocks, knowledge flows and innovation: evidence from matched patents and innovation panel data. *Research Policy* 44(7): 1327–1340.

Rovigatti G and Mollisi V (2018) Theory and practice of total-factor productivity estimation: the control function approach using Stata. *The Stata Journal* 18(3): 618–662.

Saviotti PP and Frenken K (2008) Trade variety and economic development of countries. *Journal of Evolutionary Economics* 18(2): 201–218.

Schatzer T, Siller M, Walde J and Tappeiner G (2019) The impact of model choice on estimates of regional TFP. *International Regional Science Review* 42(1): 98–116.

Siller M, Schatzer T, Walde J and Tappeiner G (2021) What drives total factor productivity growth? An examination of spillover effects. *Regional Studies* 55(6): 1129–1139.

Vogel J (2015) The two faces of R&D and human capital: evidence from Western European regions. *Papers in Regional Science* 94(3): 525–551.

Windmeijer F (2005) A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics* 126(1): 25–51.

Wooldridge JM (2009) On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters* 104(3): 112–114.

Appendix 1

Table 2. Dependent variable: $lnVA_{r,t}$ (2003–2019).

	Rovigatti and Mollisi productivity estimator
$lnL_{r,t}$	0.2162369***
	(0.03462)
$lnK_{r,t}$	0.2702792***
	(0.06254)
Wald test on Constant return to scale	49.61
	[0.000]
Observations	256

Notes: Standard errors (SEs) in parentheses, p-values for tests in square parentheses [p-value]. Estimates are generated in STATA by Rovigatti, Mollisi's (2018) prodest code.
Source: Authors' calculations.
Significance levels: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Table 3. Descriptive statistics.

	Obs	Mean	SD	Max	Min
$grTFP_{r,t}$ (1)	256	0.04	0.03	0.14	–0.02
$lnH_{r,t}$ (2)	256	1.24	0.02	1.31	1.19
$\left(\frac{Patents}{L_{RD}}\right)_{r,t}$ (3)	256	23.49	12.76	72.73	3.10
$\left(\frac{Patents}{L_{RD}} * RelD\right)_{r,t}$ (4)	256	1.78	2.58	15.18	0.0
$\left(\frac{Patents}{L_{RD}} * UnrelD\right)_{r,t}$ (5)	256	1.66	3.55	18.72	0.0
$\left(\frac{Patents}{L_{total}}\right)_{r,t}$ (6)	256	0.11	0.07	0.34	0.01
$\left(\frac{RDexpend}{GDP}\right)_{r,t}$ (7)	256	0.007	0.005	0.024	0.001
$\left(gr\frac{RDexpend}{GDP}\right)_{r,t}$ (8)	256	0.10	0.26	1.32	–0.48
$Variety_{r,t}$ (9)	256	5.24	0.22	5.81	4.92
$Open1_{r,t}$ (10)	256	0.58	0.28	1.58	0.12
$Open2_{r,t}$ (11)	256	0.30	0.13	0.72	0.07
$Open3_{r,t}$ (12)	256	0.29	0.18	0.88	0.05
$share\ empl_{r,t}$	256	0.08	0.06	0.27	0.02
$share\ firms_{r,t}$	256	0.01	0.003	0.01	0.001

Notes:
$\left(\frac{Patents}{L_{total}}\right)_{r,t}$ is the number of patents in relation to the total number of employees in region r in year t;
$\left(\frac{RDexpend}{GDP}\right)_{r,t}$ is total expenditures on the R&D sector to GDP in region r in year t;
$\left(gr\frac{RDexpend}{GDP}\right)_{r,t}$ is the growth rate of R&D expenditures to GDP in region r in year t.
Source: Authors' calculations.

Table 4. Correlation matrix.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
$grTFP_{r,t}$ (1)	1.00													
$lnH_{r,t}$ (2)	−0.08	1.00												
$\left(\frac{Patents}{L_{RD}}\right)_{r,t}$ (3)	−0.19	0.27***	1.00											
$\left(\frac{Patents}{L_{RD}} * RelD\right)_{r,t}$ (4)	−0.08	0.11	0.66***	1.00										
$\left(\frac{Patents}{L_{RD}} * UnrelD\right)_{r,t}$ (5)	−0.05	0.12	0.45***	0.50***	1.00									
$\left(\frac{Patents}{L_{total}}\right)_{r,t}$ (6)	−0.08	0.52	0.54***	0.59***	0.52	1.00								
$\left(\frac{RDexpend}{GDP}\right)_{r,t}$ (7)	0.07	0.72	−0.15	−0.12	−0.08	0.29***	1.00							
$\left(gr\frac{RDexpend}{GDP}\right)_{r,t}$ (8)	−0.12	−0.01	0.03	0.04	0.06	−0.01	0.04	1.00						
$Variety_{r,t}$ (9)	0.24***	−0.07	−0.14	−0.07	−0.09	−0.14	−0.01	0.04	1.00					
$Open1_{r,t}$ (10)	0.09	0.27***	−0.03	0.03	0.06	0.01	0.30***	−0.11	−0.04	1.00				
$Open2_{r,t}$ (11)	0.05	0.23***	0.11	0.09	0.17	0.15	0.12	−0.09	−0.05	0.92***	1.00			
$Open3_{r,t}$ (12)	0.11	0.27***	−0.12	−0.01	−0.03	−0.09	0.40***	−0.10	−0.02	0.96***	0.77***	1.00		
$share\ empl_{r,t}$ (13)	−0.03	0.2***	0.48***	0.78***	0.46***	0.61***	−0.11*	0.03	−0.03	0.01	0.15**	0.16**	1.00	
$share\ firms_{r,t}$ (14)	−0.11*	0.09	0.43***	0.73***	0.34***	0.50***	−0.2***	0.07	−0.02	−0.11*	−0.01	−0.31	0.87***	1.00

Source: Authors' calculations.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 5. Dependent variable: $grTFP_{r,t}$.

Variables	(1)	(2)	(3)	(4)
$grTFP_{r,t-1}$	1.2943***	1.3053***	1.3041***	1.2984***
	(0.0901)	(0.0921)	(0.0920)	(0.1067)
$lnH_{r,t-1}$	0.3447***	0.3486***	0.2808***	0.3237***
	(0.0590)	(0.0613)	(0.0861)	(0.0560)
$lnH_{r,t-1} * \frac{TFP_{r,t-1}}{TFP_{m,t-1}}$	−0.0230**	−0.0247**	−0.0373***	−0.0210***
	(0.0095)	(0.0104)	(0.0087)	(0.0048)
Alternative measures of R&D sector activity				
$\left(\frac{Patents}{L_{RD}}\right)_{r,t-1}$	−0.00013 (0.0002)	–	–	–
$\left(\frac{Patents}{L_{RD}}\right)_{r,t-1}\left(\frac{TFP_{r,t-1}}{TFP_{m,t-1}}\right)$	0.00011 (0.0006)	–	–	–
$\left(\frac{Patents}{L_{total}}\right)_{r,t-1}$	–	−0.0290 (0.0369)	–	–
$\left(\frac{Patents}{L_{Total}}\right)_{r,t-1}\left(\frac{TFP_{r,t-1}}{TFP_{m,t-1}}\right)$	–	0.0507 (0.1193)	–	–
$\left(\frac{RDexpend}{GDP}\right)_{r,t-1}$	–	–	−0.1784 (0.4595)	–
$\left(\frac{RDexpend}{GDP}\right)_{r,t-1}\left(\frac{TFP_{r,t-1}}{TFP_{m,t-1}}\right)$	–	–	1.4610** (0.6272)	–
$\left(gr\frac{RDexpend}{GDP}\right)_{r,t-1}$	–	–	–	−0.0039 (0.0182)
$\left(gr\frac{RDexpend}{GDP}\right)_{r,t-1}\left(\frac{TFP_{r,t-1}}{TFP_{m,t-1}}\right)$	–	–	–	0.0044 (0.0477)
control variable:				
$variety_{it}$	0.0263***	0.0258***	0.0259***	0.0260***
	(0.0034)	(0.0035)	(0.0035)	(0.0038)
$openI_{it}$	0.0088*	0.0085*	0.0121**	0.0088*
	(0.0045)	(0.0045)	(0.0053)	(0.0042)
constant	−0.5716***	−0.5740***	−0.4920***	−0.5467***
	(0.0837)	(0.0878)	(0.1168)	(0.0815)
AR(1)	−2.93	−2.96	−2.96	−2.97
	[0.003]	[0.003]	[0.003]	[0.003]
AR(2)	−0.03	0.01	−0.00	0.04
	[0.974]	[0.989]	[0.999]	[0.970]
Sargan test	0.33	0.32	0.27	0.25
	[0.566]	[0.574]	[0.606]	[0.617]
Hansen test	0.53	0.47	0.42	0.42
	[0.467]	[0.492]	[0.516]	[0.515]
F-test	60.74	50.90	63.62	217.02
	[0.000]	[0.000]	[0.000]	[0.000]
Number of observations	240	240	240	240
Number of instruments	9	9	9	9

Notes: Corrected standard errors (SEs) in parentheses; *p*-values for tests in square parentheses [*p*-value]. We perform the estimations using the xtabond2 command in Stata 17, which implements the system GMM with Windmeijer (2005) correction.
Source: Authors' estimations.
Significance levels: ***$p<0.01$; **$p<0.05$; *$p<0.1$.

Table 6. Moran tests for spatial dependence.

t	*grTFP*	
	Chi^2	*p*-value
2004	1.08	0.5814
2005	1.31	0.5205
2006	1.63	0.4419
2007	2.01	0.3659
2008	1.04	0.5948
2009	2.18	0.3355
2010	0.43	0.8075
2011	3.66	0.1605
2012	0.79	0.6729
2013	3.30	0.1924
2014	1.95	0.3772
2015	1.15	0.5686
2016	3.34	0.1880
2017	0.56	0.7576
2018	2.10	0.3507
2019	1.31	0.5206

Notes: The test was conducted using two spatial weighting matrices: W matrix (neighbourhood), M matrix (distance).
Source: Authors' calculations.

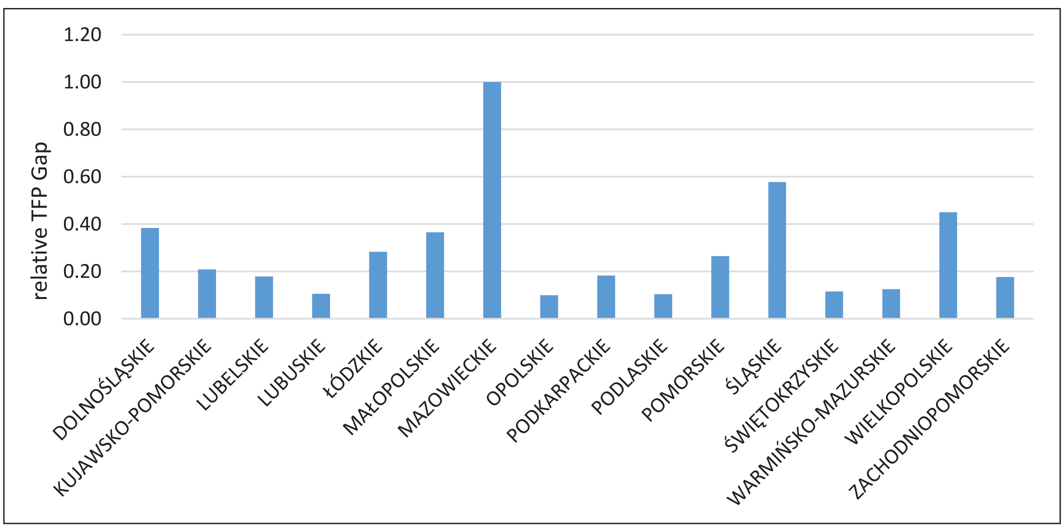

Figure 1. TFP gap to regional technology leader in Poland (average in 2003–2019).