Journal of International Economics 96 (2015) 311–322

Contents lists available at ScienceDirect

Journal of International Economics

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

Migration of skilled workers and innovation: A European Perspective

Valentina Bosetti [a,b,c,*], Cristina Cattaneo [a,b], Elena Verdolini [a,b]

[a] Fondazione Eni Enrico Mattei, Italy
[b] Centro Euro-Mediterraneo sui Cambiamenti Climatici, Italy
[c] Department of Economics, Università Bocconi, Italy

ARTICLE INFO

Article history:
Received 18 July 2013
Received in revised form 10 April 2015
Accepted 13 April 2015
Available online 21 April 2015

JEL classification:
F22
J24
O31

Keywords:
Skilled migration
Innovation
Knowledge production function
Europe

ABSTRACT

This paper analyzes the effect of skilled migration on two measures of innovation, patenting and bibliometric data, in a panel of 20 European countries between 1995 and 2008. The empirical findings show that a larger pool of migrants in the skilled professions is associated with higher levels of knowledge creation. Skilled migrants contribute both to the creation of "private" knowledge, measured by the number of patent applications through the Patent Cooperation Treaty, and to more "public" basic research, measured by the number of citations to published articles. This finding is robust, in that it uses both an occupation-based and an education-based index of skilled migration, as well as an instrumental variable estimation accounting for the endogeneity of the skilled migrants indicator and to a number of robustness checks. Our results suggest that policy efforts aiming at attracting skilled migrants to Europe and employing them in skilled professions, such as those put forward in the Europe 2020 Strategy, will indeed foster EU competitiveness in innovation.

1. Introduction

Endogenous growth theory indicates that knowledge formation and the availability of better technologies have important repercussions on productivity and growth (Solow, 1957; Romer, 1990; Aghion and Howitt, 1992; Grossman and Helpman, 1994; Jones, 2009). A core contributor to the knowledge production function is a specialized labor force, namely highly skilled workers engaged in laboratories or in academia (Caballero and Jaffe, 1993; Kerr and Kerr, 2011). In addition to the level of education and the number of workers engaged in research, it is found that diversity in the research team is also a crucial ingredient in the innovation process (Kerr, 2008; Stuen et al., 2012).

This paper combines the literature on innovation and knowledge production with the literature focusing on diversity, migration and productivity. We study how foreign skilled labor contributes to knowledge formation in Europe. This topic has been explored before, but mostly with a focus on the US market. This issue is however crucial in most EU member countries, which were once a "source" of migration, but are now increasingly seen as migration destinations for skilled and unskilled foreign workers (IOM, 2008).[1] The education level of recent flows of migrants has improved considerably over the past decade. Highly-educated foreigners exceed 31 million in the OECD area and account for 45% of the increase in the foreign born population (OECD, 2014). Indeed, attention has been increasingly drawn to the role of highly skilled immigration as a driver of technology development, innovation and economic performance (EMN, 2006; EC, 2007, 2008).

The empirical evidence on whether and how skilled foreigners contribute to European knowledge formation is scarce.[2] Most of the literature on this topic focuses on the USA (Stephan and Levin, 2001; Peri, 2007; Chellaraj et al., 2008; Kerr, 2008; Hunt and Gauthier-Loiselle, 2010; Kerr and Lincoln, 2010; Stuen et al., 2012; Peri, 2012 among others). The few available macro-empirical analyses on European countries have a narrow geographical focus. Niebuhr (2010) shows that ethnic diversity of skilled labor has a positive effect on patent

* Corresponding author at: Corso Magenta 63, 20123 Milan, Italy.
E-mail address: valentina.bosetti@feem.it (V. Bosetti).

[1] In 2008, non-EU migrants to the EU represented around 3.8% of the total population according to the EU Commission. Between 1.5 and 2 million migrants per year have entered the EU since 2002. As of January 2006, 18.5 million non-EU nationals were residents in EU member countries (EC, 2008).

[2] Within the European context, papers have mostly concentrated on the static effect of diversity, namely the effect of migrants on native employment and wages (Dustmann et al, 2008; D'Amuri et al, 2010; Manacorda et al., 2006), the issue of skill-complementarity and task specialization (Cattaneo et al., 2013; D'Amuri and Peri, 2014) and the role of foreigners in fostering trade relations (Iranzo and Peri, 2009).

http://dx.doi.org/10.1016/j.jinteco.2015.04.002
0022-1996/

applications, a common proxy for innovation, in German regions. Ozgen et al. (2011a) generalize these results for 170 NUTS2 regions in 12 Western European countries.[3]

The first aim of this paper is to fill this gap and provide new evidence in support of the positive contribution of skilled foreigners to innovation and knowledge production. We focus on a panel of 20 European countries, which includes historical members of the EU as well as Eastern European economies. The second novelty of our analysis is the use of two different proxies of innovative performance, patent applications and citations, for the purpose of exploring the effect of the foreign skilled labor force on the creation of private and public knowledge, respectively. The final contribution of our analysis is the use of two different dimensions for capturing the skills level of the foreign labor force. In the main part of our analysis we measure the skills level of foreigners by examining their actual occupation. We then check our results by following the general literature, which commonly measures skills by means of the education level of foreigners.

One of the main concerns when estimating the effect of cultural and ethnic diversity on knowledge formation in the current framework is the endogeneity of migration flows. To address this issue appropriately, we employ an "ethnic enclave" instrumental variable. This approach, first suggested by Altonji and Card (1991) and largely used in the subsequent empirical literature (Card and DiNardo, 2000; Card, 2001; Peri and Sparber, 2009; Ottaviano et al., 2013; D'Amuri and Peri, 2014; Ottaviano and Peri, 2012) uses information on the pre-sample distribution of migrants and subsequent flows by area of origin to build imputed shares of migrants for each country.

As in the micro- analyses on this topic, we find a positive synergic interaction of diverse cultures and diverse approaches in problem solving. We show that foreign skilled labor exerts a positive effect on the innovative capacity of the recipient countries both for industrially applicable innovations and for more general abstract knowledge. This positive effect is confirmed independently of whether we measure skill by using a foreigner's education or occupation level. We reinforce the idea that complementarities exist between natives and foreigners. Skilled migrants employed in highly skilled jobs have a positive impact on innovation by increasing researchers' average productivity. The results we present hold true in a series of robustness checks, such as the inclusion of additional control variables, the use of longer lags, the use of different proxies for key explanatory variables, and the exclusion of certain countries from the sample.

The paper is organized as follows: Section 2 reviews the relevant literature; Section 3 presents a model of knowledge production function which highlights the role of diversity and details the methodology and data used in the empirical estimation. Sections 4 and 5 discuss the results and robustness checks, respectively. Section 6 concludes.

2. Literature review

Hicks (1932), Schumpeter (1942) and Schmookler (1966) put forward the crucial hypotheses of induced technical change, creative destruction and the role of supply and demand determinants of innovation. Since then, the literature on knowledge creation and its contribution to growth has been vast. Important shaping forces for innovative capacity are the role of firm size (Cohen and Klepper, 1996), market structure and industry dynamics (Geroski, 1991), market concentration (Arrow, 1962), technological opportunity (Jaffe, 1986) and national innovative capacity (Furman et al., 2002). Great attention has also been given to the market failures which characterize knowledge production, namely the existence of inter-temporal, inter-sectoral and international spillovers (Jaffe, 1986; Coe and Helpman, 1995; Malerba, 1992; Branstetter, 2001; Mancusi, 2008).

Of paramount importance in R&D-based endogenous growth models is the knowledge production function, which is typically a function of the labor force in the research sector and of the available stock of knowledge (Romer, 1990; Aghion and Howitt, 1992; Kortum, 1993; Grossman and Helpman, 1994; Abdih and Joutz, 2006). The larger the pool of researchers, the more innovative the given economy. The larger the knowledge stocks, the bigger the pool of discoveries and ideas that researchers can use to stand "on the shoulder of the giants" (Caballero and Jaffe, 1993), with a positive effect on their productivity.

In addition, many contributions focus on the composition of the research labor force and its impact on knowledge formation. As research problems and technical bottlenecks become increasingly complex, the paradigm of solo geniuses has slowly been replaced by that of large teams and networks, bringing together diverse knowledge and perspectives (Hargadon, 2003; Barabási, 2005; Jones, 2009). Team diversity can take different forms, from the background or ability of workers, to their age or gender, as well as their culture. The positive effect on productivity exerted by differences in ability and knowledge within the team members is rather uncontroversial (Hamilton et al., 2003; Lazear, 1999). The effect of cultural and ethnic diversity is, conversely, more ambiguous. Younglove-Webb et al. (1999) and Katz and Martin (1997) look into academic innovation ability and emphasize the importance of a diverse team and of international collaborations. An investigation into the Rockefeller Institute's scientific successes stresses the positive contribution of permanent foreign staff as well as that of visiting scientists (Hollingsworth and Hollingsworth, 2000). Conversely, Bassett-Jones (2005) and Stahl et al. (2009) argue that team diversity imposes communication costs that might offset the creative benefits induced by complementarities among different team members. Very few studies using micro-data focus on Europe and present conflicting results. In Parrotta et al. (2012), the innovation outcome of a sample of Danish firms increases in the skill and ethnic diversity of the workforce. These findings are, however, in contrast with those of Østergaard et al. (2011), which also focus on a sample of Danish firms, but conclude that ethnic diversity has no impact on innovative activity. Finally, Ozgen et al. (2011b) show that, overall, Danish firms which employ a relatively high share of foreigners are somewhat less innovative, with diversity being associated with higher product innovation only in a subsample of firms.

Building on these micro-founded concepts, the macro literature looks at highly skilled immigration flows and their dynamic implications for the innovative capacity of firms and universities. Results generally show that skilled foreign workers and higher diversity in research personnel are associated with more innovation and patenting activity. Chellaraj et al. (2008), Hunt and Gauthier-Loiselle (2010), Kerr and Lincoln (2010) and Peri (2007), for example, highlight the positive contribution of highly educated foreign-born workers and foreign graduate students to US patenting activities. Many other papers, such as Stuen et al. (2012) and Stephan and Levin (2001) analyze the contribution of foreign-born students and workers using different indicators of research performances, finding a disproportionally positive effect. Kerr (2008) and Kerr (2010) look into some of the mechanisms at the basis of this, such as the bridging role and the greater mobility of skilled workers working abroad.

As with micro-data studies, most of these contributions focus on the USA, where immigrants represent a significant share of highly educated workers.[4] On the contrary, the impact of ethnic diversity on innovation in Europe is under-researched. To our knowledge, Niebuhr (2010) and Ozgen et al. (2011a) constitute the only tests on the effect of ethnic diversity of skilled labor on EU innovation, as measured by patents.

[3] The analysis of Ozgen et al. (2011a) is based on Austria, Belgium, Denmark, France, Western Germany, Ireland, Italy, the Netherlands, Portugal, Spain, Sweden and the UK.

[4] In the USA, 3.2% of the labor force is made up of highly skilled foreign workers (EC, 2007). According to the 2000 Census data, for example, 24% and 47% of the US science and engineering (SE) workforce with bachelors and doctorate degrees are immigrants. The corresponding statistics for the general working population in the USA is 12%.

Both find that ethnic diversity has a positive effect on patenting activities.

3. Methodology

We propose a simple model describing the innovation production function, in line with the R&D-based models presented in Romer (1990) and Grossman and Helpman (1991). In this setup, new ideas, I, are a function of the number of skilled workers employed in the research sector, S and of the average researcher productivity, $\bar{\delta}$.

$$I = \bar{\delta}S. \tag{1}$$

We assume that average productivity per researcher is a function of three key factors. The first is a measure of resources invested in innovative activity A, which is proxied by a cumulative function of past innovation efforts. The higher the A, the higher the historical investment in the production of new knowledge. The stock of knowledge of a given country impacts average productivity through inter-temporal spillovers. Researchers "stand on the shoulders of giants", namely they use previous knowledge as a stepping stone and improve the quality of innovation (Caballero and Jaffe, 1993; Stern et al., 2000).

The second factor affecting average productivity is the number of researchers, S, which captures potential decreasing returns. As the number of researchers in a country increases, negative congestion externalities arise in a given country, the so-called "stepping on toes" effect. Accidental or intentional duplication of efforts thus reduces the average productivity of R&D (Jones and Williams, 2000).

The third factor, which is the core interest of this paper, is an indicator of the share of migrants in the skilled labor forces, D_s and proxies for ethnic diversity. As mentioned above, the role of diversity on innovation is ambiguous. By adding to the pool of skills in destination markets, skilled migrants are likely to positively affect the productivity of natives, as new ideas arise through the interaction of diverse cultures and diverse approaches to problem solving. However, the presence of migrants might also impose higher communication costs. Which effect prevails is a matter of empirical finding.

Hence $\bar{\delta}$ is defined as:

$$\bar{\delta} = (A)^{\alpha}(D_s)^{\beta}(S)^{\vartheta-1} \tag{2}$$

and Eq. (1) becomes:

$$I = (A)^{\alpha}(D_s)^{\beta}(S)^{\vartheta}. \tag{3}$$

Eq. (3) is the basis of the empirical analysis presented in this paper. Our interest lies in the estimation of β, which provides information on the impact of diversity on knowledge production, after checking for other confounding factors.

If we take the natural logarithm of Eq. (3) and explicitly introduce the country and time dimensions, the basic specification for each country i at time t becomes:

$$\ln(I_{i,t}) = \beta_0 + \beta_1 \ln(A_{i,t-1}) + \beta_2 \ln\left(D_{s_{i,t-1}}\right) + \beta_3 \ln(S_{i,\,t-1}) + \mu_t + \mu_i + \varepsilon_{i,t} \tag{4}$$

where μ_t is a set of year dummies; μ_t represents a set of country fixed effects and $\varepsilon_{i,t}$ is an idiosyncratic error term.[5] Eq. (4) is estimated by using an unbalanced panel of 20 European countries from 1995 to 2008.[6] Both the sample of countries and the time period are limited by data availability.

In line with the literature, we use a one-year time lag of the independent variables to account for the time lag between the process of innovation and the codification of tangible or intangible outcomes of this process. In Section 5, we check the results' robustness to the use of higher lags.

Finding a good proxy for innovation, I has been the matter of much debate in the literature. Patent statistics are among the most commonly used. Patents are legal titles protecting a product or a process which are granted by a given patenting authority to the assignee.[7] Notwithstanding some limitations, the use of patent data as a proxy for innovation has been validated by a number of micro- and macro-studies (Griliches, 1990). Patents are linked to the output of the R&D process, and provide information on the number of technological blueprints available in any given market.

The most relevant shortcoming of patents in our context is that not all innovations are patented. Among other reasons, because of the way in which the patent system is constructed, a patent office grants temporary monopoly rights only to inventions which are industrially applicable. However, knowledge and innovation are a much broader concept than the count of patented blueprints. Therefore, a study based only on patents might not successfully capture all outcomes of the innovation process. Hence, to assess the role of diversity on all aspects of innovation production in Europe, we also focus on bibliometric data as an alternative measure of knowledge, one that is more strictly related to basic research.[8] The production of "public" knowledge is a key aspect of a country's innovativeness, one that might have more subtle but long-lasting effects on productivity and competitiveness. Indeed, a number of contributions look into the positive synergies between these public and private forms of knowledge indicators (publications and patents) for specific scientific sectors (Huang and Murray, 2009). The use of two proxies makes it possible not only to check the robustness of our finding, but also to disentangle differences, if any, in the effect of our variable of interest (diversity) on innovations of an inherently different nature.

A matter of concern when using patent and bibliometric data as proxies for innovation is to appropriately account for the quality of the new idea. For this reason, we carefully select two indicators which are closely related to high impact and high quality innovations. With respect to patent data, we exploit the design of the patent system to select our dependent variable. To protect a new idea, innovators can choose between different application "routes", which result in different patent rights. Specifically, an inventor can choose to apply for a patent at a specific national office, effectively gaining patent rights in one single "market", or to apply for patent rights at a "regional" office or through the Patent Cooperation Treaty (PCT), thus eventually obtaining patent rights in more than one country.[9] Among these three options, PCT applications are more costly than applications to regional or national offices,

[5] Focusing on 20 (developed) European countries, we never observed zero patents or citations during our sample period. Taking logs on both sides of Eq. (2) does not result in a loss of observations.

[6] The sample includes: Austria, Belgium, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, the Netherlands, Norway, Poland, Portugal, the Slovak Republic, Spain, Sweden and the United Kingdom.

[7] To be eligible for a patent, an invention (device, process, etc.) needs to be new, susceptible of industrial application and to involve a non-obvious inventive step. To obtain a patent, an inventor has to file an application to a patenting authority. The patent office will check whether the application fulfills the relevant legal criteria and will grant or reject the patent accordingly. The patent ensures the owner the right to assign, or transfer by succession, the patent and to conclude licensing contracts.

[8] See Ducor (2000) for the use of both patents and citations in the definitions of the two faces of a country's knowledge.

[9] These "routes" are generally referred to as the national route, the regional route and the international route. In the national route, the inventor files an application with a national patent office (generally, but not always, the national office of the inventor's country). A second option for inventors is to submit a patent application to a regional office, such as the European Patent Office (EPO), which searches and examines patent applications on behalf of 38 member countries. The EPO grants "European patents", which are valid in all the member states where the holder validates his or her rights. Alternatively, inventors can use the PCT (Patent Cooperation Treaty) procedure, which has been in force since 1978 and is administered by the World Intellectual Property Organization (WIPO). The PCT allows inventors to apply for patent rights in more than one jurisdiction. This is a very popular route among inventors targeting worldwide markets.

and they represent those (higher quality) innovations that the inventor would like to exploit in more than one market (OECD, 2009).[10] We use PCT filings by inventors of each country in year t to measure industrially applicable innovation and thus provide a “quality threshold” which helps to weed out from our sample patents of “lower” quality.[11] Moreover, we count patents by priority date to ensure that each patent application is attributed to the year closest to the actual invention (OECD, 2009). Patent statistics are obtained from the OECD Patent Statistics Database (OECD, 2011). [12]

Regarding bibliometric data, we take into consideration the fact that intangible knowledge has a higher impact the more it is used to build upon in the creation of subsequent knowledge. Based on the assumption that publications which are more frequently cited are those of higher quality and impact, we use the count of aggregate citations. This is a widely used indicator of the impact of a university's (Stuen et al, 2012) or a nation's research output (King, 2004). The number of citations informs us on whether basic knowledge has been useful to other researchers. Focusing on citations is thus tantamount to weighting each country's publications by a measure of quality. Bibliometric data comes from the SCImago Journal & Country Rank (SCImago, 2011). The variable is constructed as the number of citations (excluding self-citations) of all dates received by documents published in a given country in year t.[13] Calculating citations in this way results in a “truncation” of the citation function for the last years in the sample. The statistics are in fact not adjusted for the subsequent pool of possibly citing documents, which is clearly bigger for the older published documents. However, our analysis stops at 2008, while the SCImago aggregate data includes citations to all previous cohorts from articles published as late as 2011. Given that the citation function generally peaks after 4 to 5 years, we believe that using citation counts up to 2008 is reasonable. Moreover, time fixed effects control for different average citation levels received by each cohort of published articles.

The correlation between per capita PCT Patent applications and per capita citations in our sample is reasonably high, namely 0.80. This indicates that countries which are highly productive in patentable knowledge also do well in terms of general and more intangible knowledge (Fig. 1).

Our explanatory variable of interest is the number of foreigners in the highly skilled portion of the labor force over total skilled employment, D_S. To identify top-skilled occupations we use the Standard Classification of Occupations (ISCO-88) of the International Labour Office (ILO, 1990). This classification takes into consideration the kind of work performed as well as the skill embodied in the work (Elias and McKnight, 2001). The occupations are thus grouped according to the similarity of the skills involved in the fulfillment of the tasks and duties of each job. Within ISCO-88, four skill levels are defined. Broadly, the different levels mirror the length of time a person requires to become fully competent in the performance of the tasks associated with his or her job. For a description of the complete classification into the four skill groups, see Table A1 in the online Appendix.

We define “skilled” workers as those workers occupied in the fourth skill group. The fourth skill level requires a college degree or equivalent period of relevant work experience and typically relates to professional occupations and managerial positions in corporate enterprises or the national/local government, such as legislators, senior officials and managers. A breakdown of the foreign population by skills reveals some differences among the various European countries (Table 1). Belgium, Hungary, Poland, Ireland and the UK have the highest share of highly skilled migrants (above 30%); Norway, Portugal, Finland, France, the Slovak Republic and the Netherlands follow (20–30%); finally Austria, Denmark, Spain, Germany, Italy Greece, the Czech Republic and Sweden are in the 10–20% range.[14]

The skills dimension embodied in our measure of diversity is not standard. Conventionally, the literature measures the skills level of foreign workers by using information on their educational attainment, independently of occupational considerations (Borjas, 2003; Card and Shleifer, 2009; Ottaviano and Peri, 2012). On the contrary, we use data on foreigners' occupations to capture more precisely their contribution to the creation of new knowledge. This distinction likely matters more for foreigners than for natives, as the literature shows that skill-mismatch often occurs among migrants (Green et al., 2007). Moreover, the skills classification described above takes into consideration the content of the educational capital embodied in different occupations, since the formal education required to fulfill the tasks and duties associated with a given occupation is one of the dimensions considered for the ISCO-88 aggregation (ILO, 1990). We therefore believe that our approach to measuring skills is more precise. As a robustness check, however, we also show results using educational attainments as the basis for building our proxy for the share of skilled foreigners.

In Fig. 2, in addition to the top skills share of the foreign labor force (the last column of Table 1), we also show the contribution of foreigners to the skilled labor force as well as the contribution of the total number of foreigners to each country's population.[15] In France, Hungary, Ireland, Portugal and the UK, the share of skilled foreigners in skilled labor is more than proportional to the overall foreigner share. In Belgium, the Czech Republic, Finland and Norway the two shares are similar. All other European countries display a share of skilled foreigners lower than the overall foreigners' share, with some cases where the two shares are remarkably different (Austria, Germany and Greece).

The data used to compute this measure of diversity are taken from the EU Labour Force Survey (EU-LFS), which provides information on the nationalities of the respondents, along with their ISCO-88 occupation.[16]

A caveat of the EU-LFS is that it does not cover illegal migration. This limitation should however not be problematic in our context. The component of diversity that affects innovation is provided by highly skilled foreigners, who are most likely employed legally in highly skilled occupations. Highly qualified foreigners illegally entering European countries eventually find low-skill jobs and should not influence the innovation potential of a country.

A second limitation of the dataset is that it does not allow for constructing more sophisticated indexes of diversity, such as the Herfindahl Index. The EU-LFS classifies respondents only in two macro-categories, namely nationals or non-nationals. Details on the main areas of origin for migrants are available only for the last four waves (2004 onward), while detailed country information is never available. We cannot therefore compute a Herfindahl Index without drastically restricting the sample size. We however believe that this is only a minor limitation in our context. Since the share of nationals enters into the traditional computation of the Herfindahl Index, and since in European countries

[10] In 2003 the (estimated) costs of a Euro-PCT (filing through PCT at the WIPO, designating the EPO) averaged around EUR 46,700, while the cost of obtaining a standard Euro-direct patent (direct filing to the EPO or extension of an earlier national patent application) was roughly EUR 30,530 (OECD, 2009).

[11] The use of unweighted “regional” and “international” routes patent data is common in the literature (Crepon and Duguet, 1997; Bottazzi and Peri, 2003; Peri, 2005 among others). An alternative approach would use information on patent family size, number of claims or forward citations to weight patent count data (OECD, 2009). However, the OECD database we have access to does not provide such information (OECD, 2011).

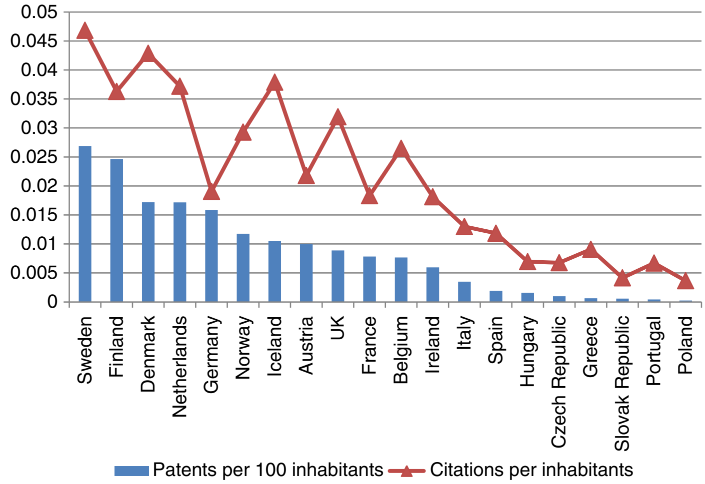
[12] The OECD provides patent statistics which are computed on the basis of the fractional counting method by which for each patent a fraction equal to the share of a country's inventors over total inventors is assigned to each country. The use of count data models, which is common in the literature, is therefore not required.

[13] Both SCImago (2011) and OECD (2011) provide aggregate statistics by country.

[14] The data for Fig. 2 are extracted from the 2000 round of censuses for all countries but Denmark, Finland, Norway and Sweden. For these last four countries, data are taken from population registers. For Iceland neither of the two sources is available.

[15] See note 14.

[16] For most of the countries, the EU-LFS provides information on both the nationality and the country of birth of non-nationals. In this paper we classify foreigners by nationality as this was the most comprehensive information. The EU-LFS has the great advantage of producing highly comparable data for the EU member states, as it is based on a common coding of questions, definitions and classifications of the variables.



Source: OECD Patent Statistics Database (OECD, 2011) and SCImago Journal & Country Rank (SCImago, 2011)

Fig. 1. Patents and citations, average 1995–2008.

foreigners still account for a limited portion of the total population, a diversity measure computed as the ratio of foreigners over the population and the Herfindahl Index is highly correlated.[17] The same consideration holds true if we consider only the skilled portion of the foreign and total population.

Eq. (2) also requires a proxy for the labor force working in the knowledge sector, S, which includes both foreigners and natives. An excellent candidate is the number of employees in technology and knowledge-intensive sectors, which provides information on the size of the research sector. We obtain this variable from the EUROSTAT database (EUROSTAT, 2011). This indicator also serves the purpose of capturing the size of a given economy, thus acting as a scaling factor since our dependent variables are not in per capita terms but in absolute values.

The availability of past knowledge that allows researchers to "stand on the shoulders of the giants" can be measured by cumulative functions of past innovation, A, the patent specification, we therefore use the rich information on past patenting output, which is available since the 1980s (OECD, 2011), as explained more in detail below. In the citation specification this approach is not possible, since data before 1996 are not available. Hence we resort to an input-based measure, using information on yearly total (public and private) intramural R&D expenditure, which we obtain from the EUROSTAT database (see for example Peri, 2005).[18]

The country fixed effects in Eq. (4) capture country-specific financial and macroeconomic shocks and the potential heterogeneity of demand and immigration across country levels. Once these effects are controlled for through the introduction of fixed effects, the remaining variation of immigrants in a cell is assumed to be driven by supply shocks and the OLS estimates should be unbiased.

However, some lingering country-specific demand shocks are potentially in place, calling for an instrumental variable approach. Some unobservables governing the location of foreigners in the different European countries might be correlated with the unobservables governing the evolution of patents or published documents. If migrants, and especially skilled migrants, respond to economic opportunities in destination countries, a non-zero correlation exists between the economic outcomes and the share of (skilled) immigrants, biasing the estimated coefficient associated with such a share. A second source of bias exists due to measurement error in the share of (skilled) foreigners.

We use an instrumental variable approach to address both biases. Altonji and Card (1991) suggest an "ethnic enclave" instrument that has been largely used for migration shares in the subsequent empirical literature (Card and DiNardo, 2000; Card, 2001; Peri and Sparber, 2009; Ottaviano et al., 2013; D'Amuri and Peri, 2014; Ottaviano and Peri, 2012). The instrument is an imputed share of migrants, which nets out the component of migration flows that are attributed to economic opportunities. We use past migration stocks, available with education breakdown in a bilateral form, to compute the instrument (Docquier et al., 2009). To provide further exogeneity to the instrument, the imputed share is computed for the unskilled portion of migration. Unskilled immigrants provide cultural amenities that highly skilled foreigners find attractive, while they do not directly contribute to innovation. Specifically, we select the 1991 stock of unskilled migrants and predict the subsequent stock of low-educated migrants using total yearly immigration flows by area of origin from Ortega and Peri (2009). In agreement with D'Amuri and Peri (2014) we assume a 40% re-emigration rate to net the total gross inflows available. Given areas of origin $n \in N$ and destination countries $i \in N$, we calculate the imputed migrants' shares, $\hat{D}$, for each year t and each destination country i, as the ratio of the imputed stock of unskilled migrants to total imputed unskilled employment as follows:

$$\hat{D}_{i,t} = \frac{F_{i,unskilled,1991} + \sum_{n \in N} \left[\frac{F^n_{i,unskilled,1991}}{F^n_{EU,1991}} \Delta F^n_{EU,t} \right]}{\hat{E}_{i,unskilled,\ t}}$$

[17] The Herfindahl Index, computed after 2004, and the ratio of foreigners to population display a correlation of 0.99.

[18] The stock variable is constructed by applying the perpetual inventory method to each of the two measures of innovative efforts as follows: $A_{i,t} = F_{i,t} + (1-\delta)A_{i,t-1}$. The initial value of the stock is calculated as: $A_{i,t_0} = \frac{F_{i,t_0}}{\bar{g}+\delta}$, where F is the flow of either patents or R&D investment in a given year and country, $\delta = 0.1$ is the depreciation rate set chosen in line with the literature (Keller, 2002) and $\bar{g}$ is the average rate of growth of the flow of innovation efforts for the period between t_0 and $t = 3$, where t_0 is the first year of data availability (Bottazzi and Peri, 2003). This ensures that the choice of the initial value of the knowledge stock has the minimum possible impact on the subsequent levels of the variable. The patent stock is initialized in 1980. Data on GERD R&D expenditures are not available for all countries starting in the same year. We use the first year of data availability to build the initial knowledge stock variable, to limit the measurement error implicit in the rough estimation of the initial A. The base year is as follows: 1981 for Austria, Denmark, Finland, Germany, Greece, Iceland, Ireland, Italy, the Netherlands, Norway, Spain, Sweden and the United Kingdom, 1982 for Portugal, 1983 for Belgium, 1987 for Hungary and Poland, 1991 for France and 1993 for the Czech Republic and the Slovak Republic.

Table 1
Distribution of foreign workers into skill groups (%) — 2001.
Source: data are taken from the 2000 round of censuses. Only for Denmark, Finland, Norway and Sweden are data taken from population registers.

	Skill1	Skill2	Skill3	Skill4
Austria	28.2	46.2	12.4	13.3
Belgium	12.3	47.2	8.9	31.6
Czech Republic	11.8	51.9	17.7	18.6
Denmark	22.6	46.3	14.2	16.9
Finland	14.5	49.6	14.3	21.6
France	11.0	52.9	14.1	22.1
Germany	20.0	56.6	13.3	10.2
Greece	30.3	54.3	4.3	11.2
Hungary	5.8	48.5	13.9	31.8
Ireland	6.8	46.3	8.8	38.1
Italy	22.5	46.2	13.8	17.5
Netherlands	14.2	44.9	15.6	25.3
Norway	11.8	48.4	18.9	20.9
Poland	7.0	48.0	12.3	32.7
Portugal	14.9	50.9	12.8	21.3
Slovak Republic	11.4	45.4	19.3	23.8
Spain	27.4	49.4	7.8	15.5
Sweden	13.3	54.7	13.0	19.0
United Kingdom	10.4	42.3	13.1	34.2

where $F_{i,unskilled,1991}$ is the number of unskilled foreigners in country i in 1991; $F^n_{i,unskilled,1991}$ is the number of unskilled foreigners of area of origin n in country i in 1991. [19] $F^n_{EU,1991}$ is the total number of foreigners from area of origin n in Europe in 1991; $\Delta F^n_{EU,t}$ is the yearly immigration flows to Europe by area of origin n and $\hat{E}_{i,unskilled,t}$ is the unskilled employment, defined as the stock of unskilled natives in each destination country i in 1991, increased by the imputed stock of unskilled migrants in each country.[20]

The advantage of imputed shares is that they are determined only by the initial migration mix by origin and by the variation in flows across origin groups in the different European countries. Given the importance of ethnic networks, migrants tend to settle in established communities of similar origin. Family reunification and ethnic ties are, therefore, the main drivers of country patterns of immigration flows by origin, rather than labor demand conditions. The underlying exclusion restriction for this instrument is that the 1991 settlement of migrants by origin is not correlated with the economic situation after 1996. Moreover, the use of the unskilled share emphasizes the role of geography and taste, and minimizes the role of economic factors that might attract skilled workers.

The primary sources of the 1991 migration stocks are Censuses and Registers. These sources provide highly reliable information on the structure of immigration in all OECD countries. These data should be less affected by sampling errors than survey data and for this reason they adequately address the measurement error bias.

4. Discussion of results

To account for possible serial correlation within countries, the standard errors in all our regressions are clustered at the country level. The small number of clusters in this application implies that an asymptotic refinement through bootstrapping should be implemented. Therefore, we use the wild cluster bootstrap procedure (Cameron et al., 2008; Davidson and MacKinnon, 2010). All tables report both the "plain" clustered p-values (in brackets) and wild cluster bootstrapped p-values for the coefficients of interest (in parenthesis).

4.1. Main specification

The results of the OLS regressions for both the patent and the citation specifications reported in Table 2, columns (1) and (3), indicate that the share of skilled migrants exerts a positive effect on both innovation measures. Specifically, a one percent increase in the share of skilled migrants increases the number of patents by 0.08% and the number of citations by 0.15%, on average and ceteris paribus.[21] All coefficients are significant at least at the 5% level, by using both the clustered and the bootstrapped standard errors. These results show that skilled foreign migrants in highly skilled occupations contribute both to the creation of general "public" knowledge and to the improvements of industrially applicable technologies.

The positive effect of foreign skilled labor on both innovation proxies is confirmed by the 2SLS estimates. The elasticities in the patent and citation equations are 0.89 and 0.63, respectively (columns 2 and 4), with significance level higher than 5%. In both specifications, these 2SLS point estimates are larger (i.e., more positive) that the corresponding OLS estimates. This result is actually in line with that of many other papers that use the ethnic enclave instruments in different settings (among others, Card and DiNardo, 2000; Hunt and Gauthier-Loiselle, 2010; D'Amuri and Peri (2014); Bratti and Conti, 2014). A first explanation for this result is that the 2SLS addresses two sources of bias, moving in opposite directions. On the one hand, a potential measurement error in the statistics of skilled migrations should produce a negative bias in the estimated coefficient. As pointed out by Aydemir and Borjas (2011), the sampling error in the measurements of immigrant supply shift is responsible for a substantial reduction in the estimated impact of migration on wages.[22] On the other hand, the IV strategy controls for the effect of those unobservables governing both the location of foreigners in the different European countries as well as the evolution of patents or published documents. Economic theory traditionally suggests that the latter effect is responsible for an upward bias in the OLS coefficient, as migrants, and especially skilled migrants, respond positively to economic opportunities in destination countries. The increase in the elasticity from the OLS to the 2SLS would indicate that the downward bias due to the measurement error prevails in our case.

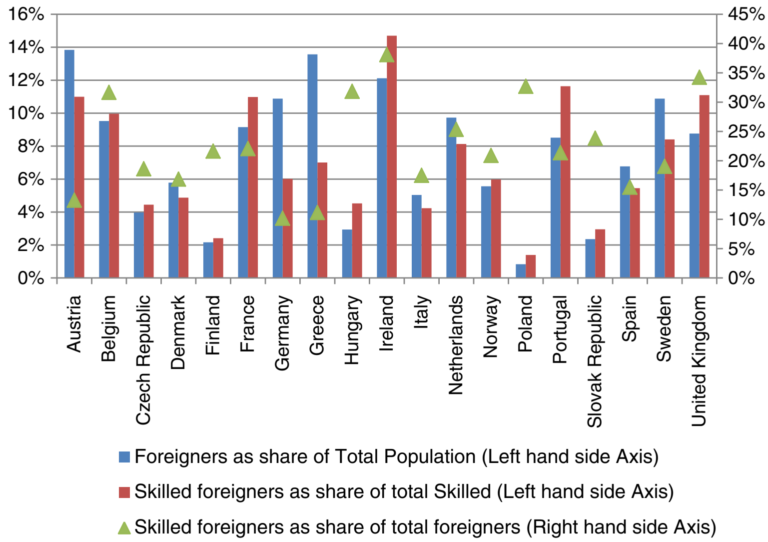
A second explanation for the larger coefficients of the 2SLS with respect to OLS is that, as pointed out in Angrist and Pischke (2009) and summarized in Wooldridge (2010), in the case of heterogeneous treatment effects, the 2SLS estimator consistently estimates a local average treatment effect (LATE). The instrumental variable coefficient hence reflects the effect on innovation of the skilled immigrants whose behavior is more affected by the instrument. On the contrary, the OLS estimates the average treatment effect over the entire population. We recall that our instrument is computed by using information on unskilled enclaves and subsequent unskilled migration inflows, and its main rationale is that unskilled immigrants can provide amenities or cultural ties that channel high-skilled immigration. This instrument emphasizes the role of ethnic ties and taste over that of economic factors that might attract skilled workers specifically. Since the instrument captures the part of migration flows that follows ethnic networks, it may very well be that these immigrants are more productive as the network may have facilitated and optimized their entry into the labor market.

[19] The areas of origin are: Central and South America, Eastern Europe, Middle east central Asia, North Africa, North America, Other Africa, South and Eastern Asia and Western Europe.

[20] Country-of-origin can be tightly linked to country of destination, and this argues against the validity of imputed shares as an instrument. In our case, the use of the unskilled portion of migration and yearly immigration flows to total Europe ($\Delta F^n_{EU,t}$) should provide some confidence in the validity of the instrument. We also tried to exclude own-country flows from the European trend to be even more robust. Unfortunately, this alternative instrument is poorly correlated with the endogenous variable with an F-statistic extremely low.

[21] Diversity may positively contribute to knowledge creation but with diminishing marginal returns. Too much diversity may entail costs from potential conflicts of preferences and hurdles of communication. To test a non-linear impact of diversity on innovation, a squared term was introduced. The resulting coefficient was not statistically significant. This finding may indicate that the level of diversity in Europe is still too limited for detecting an inverted-U shape relationship.

[22] The authors find that allowing for such attenuation bias "can easily double, triple, and sometimes even quadruple the estimated wage impact of immigration".



Note: The data for Figure 2 are extracted from the 2000 round of censuses for all countries but Denmark, Finland, Norway and Sweden. For these last four countries, data are taken from population registers. For Iceland neither of the two sources is available, hence it is missing from the Figure. Skilled foreigners are those occupied as technicians and associate professionals, legislators, senior officials, managers and professionals, according to the Standard Classification of Occupations (ISCO-88).

Fig. 2. Share of foreigners and skilled foreigners (%) — 2001 (national censuses).

Also to be noted is the fact that our analysis results in a large difference in the size of the 2SLS and OLS coefficients. To check the robustness of these findings and explore if they are driven by specific countries, we carry out diagnostic tests on both the first stage and reduced form equations. This detective work has identified Portugal as being highly influential on the 2SLS coefficients. The 2SLS estimates for both patents and citations are halved when Portugal is removed from the sample of countries (Table A2 in the online Appendix).

To explain why Portugal drives the 2SLS estimates upward, we analyzed the scatterplots of the instrument and the dependent variables and noticed that Portugal is among the countries with the steepest fitted lines. The 2SLS coefficient can be expressed as the ratio between the coefficients of the instrument in the reduced form and in the first stage equations. Therefore our results reflect the strong relationship between the innovation proxies and the instrument for Portugal. Portugal experienced a strong growth in both patents and citations, while the share of foreigners (real and imputed) grew less than proportionally. This stylized fact could explain the behavior of Portugal in the empirical estimations.

The exclusion of Portugal does not alter our main findings. The coefficients of skilled foreign labor are lower but still positive and statistically significant, confirming that diversity is beneficial for innovation in Europe (Table A2 in the online Appendix).

If we turn to the other controls, the coefficients of the variable measuring the stock of knowledge in a given country (stock of R&D-expenditures) are in line with our expectations. In the 2SLS specifications, a 1% increase in the knowledge stock is associated with 0.6% and a 0.4% increase in patent applications and citations, respectively, with a level of significance of at least 5% The positive coefficients drive in favor of the "standing on shoulders" assumption: the accumulation of past knowledge benefits the creation of new knowledge. This is consistent with the R&D-based growth models of Romer (1990). The coefficient is however statistically smaller than one, indicating a weaker degree of intertemporal spillovers than that found in Abdih and Joutz (2006).

Conversely, a larger pool of skilled labor is not associated with a significant increase in the productivity of knowledge in our sample, since the estimated coefficients fail to reach the acceptable levels of significance. One possible explanation is that the stock of knowledge already accounts in part for the researchers' population. The input-based measure of R&D includes wages of researchers, while the output-based measure embodies the overall productivity of the innovation sector, which depends on the pool of researchers. We test this hypothesis by running our main specifications without the stock of knowledge variable. The significance of the variable "total number of researchers" improved, but mostly in the OLS specification. One could argue that the productivity of researchers rather than their number is the main driver of innovation. In our case, this is better proxied by the stock of knowledge variable, which is either based on an output measure (patent specifications) or accounts not only for the number of researchers, but also for the quality and money spent in labs, equipment and the like (citation specification).

The first stage estimates for the excluded instrument are reported in Table A3 in the online Appendix, and show that the imputed shares have a positive and significant effect on the actual share of skilled migrants. The size of the F-tests indicates that the instrument is fairly powerful. The statistic is greater than the value suggested by Staiger and Stock (1997) as a rule of thumb to assess the relevance of the instruments.

Our findings confirm the empirical results for the US (Stephan and Levin, 2001; Chellaraj et al., 2008; Kerr, 2008; Hunt and Gauthier-Loiselle, 2010; Kerr and Lincoln, 2010; Stuen et al, 2012; Peri, 2012) as well as the empirical exercises on subsets of European regions (Niebuhr, 2010; Ozgen et al., 2011a). These studies document a positive contribution of the foreign skilled labor force to the production of patented knowledge. We also show that skilled foreigners contribute to the creation of more abstract public knowledge. Our results are consistent with the idea that foreign workers play a positive role by increasing the level of diversity, D_s, and they have a positive effect on the productivity of natives because new ideas are likely to arise through the interaction of diverse cultures and diverse approaches in problem solving. Not only do highly skilled immigrants display high rates of patenting, but they also allow natives to produce greater innovation (Hunt and Gauthier-Loiselle, 2010; Kerr and Lincoln, 2010).

We are aware that the share of skilled migrants, as computed here, does not provide any insight into the "type of diversity". The variety of ethnicities within foreigners, namely the country mix, is likely to be an

Table 2
The effect of skilled foreigners on innovation — OLS and 2SLS.

	Patents		Citations	
	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)
ln(D_s)	0.0766	0.894	0.145	0.631
	[0.0276]	[0.0258]	[0.00279]	[0.00157]
	(0.0378)	(0.0364)	(0.0096)	(0.0388)
ln(A)	0.756	0.557	0.451	0.445
	[0.0259]	[0.0105]	[0.0440]	[0.000282]
	(0.0044)	(0.05099)	(0.0406)	(0.05079)
ln(S)	−0.0923	−0.0256	0.195	0.155
	[0.747]	[0.934]	[0.284]	[0.499]
	(0.84232)	(0.94011)	(0.33017)	(0.57434)
Observations	213	213	213	213
R-squared	0.786	0.481	0.781	0.438
Number of countries	20	20	20	20
F-test 1st stage (standard)		18.64		23.26
F-test 1st stage (bootstrap)		9.56		11.02

Notes: In columns (1) and (2) the dependent variable is the natural logarithm of the patent applications filed under the PCT recorded by priority date; in columns (3) and (4) it is the natural logarithm of the number of citations to publications in year *t*. Country dummies and year dummies are included in all specifications. Clustered p-values are reported in brackets. Wild cluster bootstrapped p-values are reported in parentheses and are performed on 10,000 replications. The excluded instrument in the first stage of the 2SLS is the log of imputed shares.

important driver of economic development. Moreover, some ethnicities, given the language spoken, may better interact with nationals than others, thus providing greater complementarities. The proxy of diversity we constructed on EU-LFS data does not capture such aspects of diversity, since foreigners are only classified as non-nationals in the survey. One source of data that provides detailed bilateral information on origin and destination countries is the OECD International Migration database (OECD, 2013), available since 1990. The limitation of this dataset is that the skill breakdown of movers is not provided. Since our paper focuses on the determinants of knowledge formation, this is a major drawback. The distribution of skilled foreigners by country of origin in the different destinations may not be proportional to the distribution of total foreigners (see Fig. 2). Skilled and unskilled foreigners may also follow different channels of entry, with the latter being more likely to exploit family ties.

Table 3
The effect of diversity on innovation, Herfindahl Index.

	Patents		Citations	
	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)
ln(D_s)	0.053	0.734	0.127	0.670
	[0.220]	[0.0598]	[0.0108]	[0.0171]
	(0.29977)	(0.06859)	(0.0274)	(0.05379)
Ln(Herfindahl)	0.054	−0.840	0.263	−0.617
	[0.877]	[0.410]	[0.457]	[0.439]
	(0.87251)	(0.56854)	(0.50995)	(0.48195)
ln(A)	0.959	0.671	0.452	0.286
	[0.00123]	[0.00981]	[0.0853]	[0.180]
ln(S)	−0.162	−0.022	0.185	0.193
	[0.411]	[0.928]	[0.295]	[0.389]
Observations	183	182	183	182
R-squared	0.825	0.629	0.801	0.442
Number of countries	18	17	18	17
F-test 1st stage		12.8		9.555

Notes: In columns (1) and (2) the dependent variable is the natural logarithm of the patent applications filed under the PCT recorded by priority date; in columns (3) and (4) it is the natural logarithm of the number of citations to publications in year *t*. Country dummies and year dummies are included in all specifications. Clustered p-values are reported in brackets. Wild cluster bootstrapped p-values are reported in parentheses and are performed on 10,000 replications. The excluded instrument in the first stage of the 2SLS is the log of imputed shares.

The only reasonable approach, in our opinion, is to draw on the International Migration database to compute the Herfindahl Index for the total migration population, excluding the share of nationals from the computation. This index is added in the estimation alongside the share of skilled foreigners, computed from the EU-LFS, to capture the variety and the distribution of the nationalities of foreigners. While the share of skilled foreigners captures the density of skilled migrants, the Herfindahl captures the diversity of the total pool of migrants, skilled and unskilled. In this way we offset the limits of a diversity measure which does not account for the variety of ethnicity in the foreign population.

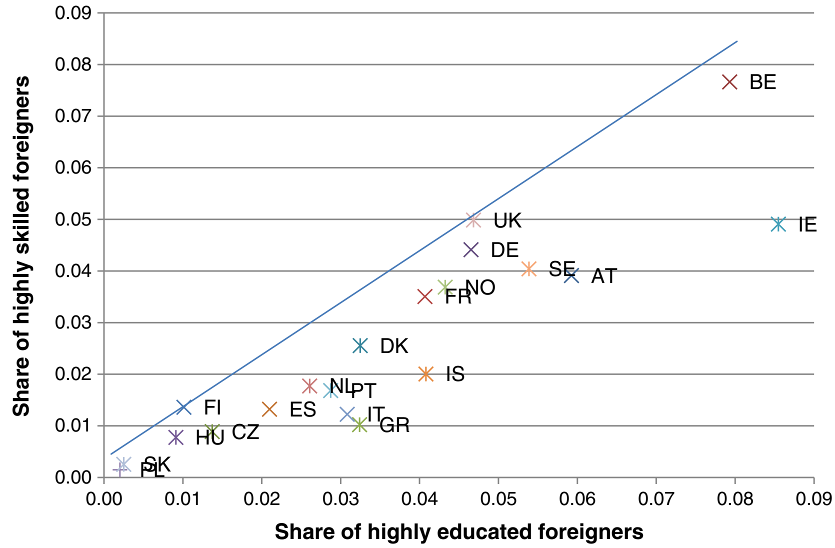
Table 3, columns (1) to (4), reports the empirical findings of the model where the Herfindahl Index is included as a regressor. While the density of skilled foreigners still exerts a positive effect on innovation (which is comparable to the estimates presented in Table 2 and statistically significant at the 10% level for patents and the 5% level for citations), the Herfindahl Index controlling for the diversity of all migrants, is not significant. This is likely due to the limitations linked to the underlying data used in the construction of the index, namely the inclusion of both skilled and unskilled foreigners. For this reason, we only use the share of skilled foreigners as a measure of diversity in the rest of the paper.

4.2. Occupation and education mismatch

As argued above, our measure of skilled foreigners based on occupation rather than education is unusual in the literature but, we believe, more appropriate for capturing the effective contribution of foreigners to the creation of knowledge. The distinction between the two methods of measuring skills with occupation and education is relevant particularly for foreigners, since higher education attainments do not guarantee that migrants are employed in highly skilled occupations. Nonetheless, we test the robustness of our results to the education-based skill metric traditionally used in the literature. We believe that the comparison is beneficial in two ways. First, it helps in understanding if a mismatch in the allocation of skills in the labor market impacts the ability to innovate. Second, it serves to demonstrate whether the use of one measure in place of the other can give rise to conflicting empirical findings.[23]

The correlation between the share of skilled migrants and the share of highly educated migrants is very high (0.90). In Fig. 3, we plot the share of highly educated foreigners against the share of highly skilled foreigners, computed as a mean over the sample period in each country in our sample. Countries below the 45° line are those where there is a mismatch between educational attainments and employment among foreigners. In relative terms, Finland, the United Kingdom, Belgium, the Slovak Republic and Germany are the most virtuous countries, as they show a correspondence between the share of highly educated migrants and that of highly skilled migrants. On the contrary, a gap between the education and the occupation share exists in countries such as Greece, Italy, Iceland, the Czech Republic, Portugal and Ireland. These countries display a disproportionally larger share of highly educated migrants as compared to highly skilled migrants, suggesting a relatively inefficient allocation of qualified migrants in the labor market. Both the high correlation of these two variables and information displayed in Fig. 3 suggest that mismatch is not as large a problem as expected. It should be noted that a large portion of mismatch is not really captured in this dataset on regular immigration, since over-education will disproportionally affect irregular immigrants. Table 4 reports the results of the specification where we use the education-based classification of migrants. The empirical relationship between skilled migrants and innovation is robust to this alternative measure. The estimated coefficients of the 2SLS specifications are positive and statistically

[23] In EU-LFS information on the highest level of education completed is available, codified using the International Standard Classification of Education (ISCED). We define highly educated migrants as those with tertiary education, and compute the share with respect to the highly educated population in a given country.



Source: EU Labour Force Survey (EU-LFS)

Fig. 3. Share of highly educated versus highly skilled foreigners — average 1996–2008.

significant at the 5% level in both the patent and the citation specifications, though the asymptotic refinement makes the coefficient not statistically significant in the citation specification. In the patent specification, the elasticity of diversity computed along the skill dimension is statistically larger than the one computed for the education dimension. This is an interesting result, since it shows that traditional measures of diversity may underestimate the importance of foreigners in the innovation sector of destination countries.

The significant effect of both measures of diversity suggests two main conclusions. First, regardless of where educated migrants are employed, they contribute to the creation of knowledge. The competence acquired through education generates positive externalities that spill over beyond the occupations they are employed in. Second, it indicates that the mismatch in qualification and occupation among migrants is relatively small. This is hardly surprising given the descriptive statistics presented in Fig. 3.

5. Robustness checks

The main specification presented above contains only a limited number of control variables, as derived from the theoretical setting. To avoid misspecification due to omitted variable bias we include some additional controls that may enter a knowledge production function. Throughout this section we refer to results that are reported in the online Appendix.

First, we include a variable proxying for the expected global technology trends. This variable is constructed by calculating the distribution of patents and citations in different technologies/research areas in the base year.[24] The innovation in each technology/research field is then augmented for each period by using the technology/research specific worldwide growth rates. This variable allows us to control for expected innovation due to global trends and initial country allocations. The patent variables can be divided into nine technologies, based on the IPC classification provided in OECD (2011): Human Necessities; Performing Operations/Transport; Chemistry/Metallurgy; Textiles/Paper; Fixed Constructions; Mechanical Engineering/Lighting; Physics and Electricity. Conversely, bibliometric data is divided into four areas, based on the Scopus® Classification: Health, Life Sciences, Physical Sciences and Social Sciences. Results are presented in Table A4 and show that the coefficients on the share of skilled foreigners and on the knowledge stock variable are robust to the inclusion of the expected technology trends in both the patent and the citation equation. Conversely, the coefficients of the trends themselves are insignificant.

Second, we include in our estimation a proxy of international knowledge spillovers. We assume that a preferential channel for knowledge flows across countries is represented by emigrants abroad. They are exposed to the foreign knowledge and at the same time they serve as a channel to their origin country. Kerr (2008), for example, documents that ethnic networks are important for knowledge diffusion from the US. For each country in our sample, we thus calculate the share of nationals in any other country in 1991 based on data from Docquier et al. (2009). We use these shares to weight the knowledge of each destination country. We sum over all destination countries and use this as a measure of country specific foreign knowledge stock. The coefficients associated with this variable are significant in the patent specification, testifying that foreign knowledge positively contributes to the creation of domestic knowledge by allowing researchers to build on previous foreign innovations. Conversely, the variable fails to reach acceptable levels of significance in the citation specification. In both cases, the coefficients on the share of highly skilled migrants and on the own knowledge stock variable are robust.[25]

Third, another source of bias could emerge from geography-related shocks, which can be controlled for by regional-year fixed effects. The limited number of observations in our sample, however, does not allow us to introduce these variables. A potential problem here is that such geographical effects may influence the instrument, as long as the initial distribution of migrants by ethnicity is related to the spatial proximity of source and destination countries. This would adversely affect the validity of the instrumental variable approach. As a robustness check we formed expected migration shares by allocating yearly migration flows from each area of origin to the different destination countries according to a distance metric (rather than an ethnicity metric). In this

[24] The base year is 1996 for the citation variable (first year of data availability) and 1991 for the patent variable (previous years are characterized by a non-ignorable number of zeros in some technological fields).

[25] Two additional controls have been added in the basic specification and are available from the authors upon request. First, in line with Niebuhr (2010), we included a measure of the industrial structure of the countries, computed as the ratio between the manufacturing and the service value added. Second, we also added a control for the size of the country, namely population. Both have a non-significant impact on innovation.

Table 4
The effect of skilled foreigners on innovation: Alternative skill measure — education attainment.

	Patents		Citations	
	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)
$\ln(D_{s\ education})$	0.0967	0.649	0.0522	0.634
	[0.131]	[0.00405]	[0.266]	[0.0187]
	(0.14319)	(0.0334)	(0.26317)	(0.11779)
$\ln(A)$	1.006	0.810	0.469	0.282
	[0.000203]	[9.16e−05]	[0.105]	[0.265]
$\ln(S)$	−0.199	−0.518	0.203	−0.213
	[0.327]	[0.0588]	[0.361]	[0.538]
Observations	207	207	207	207
R-squared	0.805	0.66	0.759	0.286
Number of countries	20	20	20	20
F-test 1st stage		20.31		27.51

Notes: In columns (1) and (2) the dependent variable is the natural logarithm of the patent applications filed under the PCT recorded by priority date; in columns (3) and (4) it is the natural logarithm of the number of citations to publications in year t. Country dummies and year dummies are included in all specifications. Clustered p-values are reported in brackets. Wild cluster bootstrapped p-values are reported in parentheses and are performed on 10,000 replications. The excluded instrument in the first stage of the 2SLS is the log of imputed shares.

distance metric the weights are higher for closer countries and smaller for countries farther away. These counterfactual shares are, by construction, subject to regional issues. A simple correlation test between the ethnic enclave instrument and the counterfactual share indicates that the two variables (ethnicity and distance-based imputed shares) are not correlated (−0.20). Far from being a definitive test, this still suggests that emigration flows that follow an ethnicity metric produce a different distribution of migrants as compared to counterfactual emigration flows that follow a distance metric. Hence, this supports the hypothesis that our original instrument is not subject to shocks that occurred between geographically close areas.

Fourth, we also check the robustness of our findings with respect to different proxies of general "public" knowledge. We use the number of documents published during a specific year and the number of citable documents (articles, reviews and conference papers) as dependent variables in place of the number of citations. The beneficial effect of diversity is robust to these changes, as shown in Table A5. To check if lower income countries drive the result in the unweighted regressions, we use GDP per capita at the beginning of the sample year as weights. The choice of weights is determined by two factors. First, lower income countries may display sudden changes in innovation and in the share of skilled foreigners as they eventually catch up. Second, lower income countries have a smaller pool of migrants and/or might be inherently less efficient in collecting data on migration flows. This would imply that their statistics are more affected by measurement error. The weighted regression on the contrary gives larger weights to richer countries, which have a larger pool of foreigners, more reliable migration statistics and where innovation has traditionally been in place. The estimated coefficients presented in Table A6 are robust to the application of the weights.

Fifth, the theoretical setting suggests that the knowledge production function depends on the labor force in the research sector. Therefore, the diversity variable in this paper is measured by selecting only the skilled portion of the foreign population. However, immigration may boost innovation through indirect channels, such as by lowering the cost of domestic services and therefore by increasing the labor supply of domestic workers, such as native women (Cortes and Tessada, 2011; Barone and Mocetti, 2011; Farre et al., 2011) or by increasing the productivity of native workers, who can select tasks and occupations where they can exploit a comparative advantage (Cattaneo et al., 2013; D'Amuri and Peri, 2014; Peri and Sparber, 2009, 2011). This implied that, not only high-skilled but also foreigners with lower skills can indirectly contribute to innovation. To check whether this is indeed the case, we replace the measure of skilled share by a measure of unskilled share, computed by using information on foreigners employed in the first skill group. Table A7 reports the empirical findings. The coefficient of the share of low-skilled foreigners is statistically significant both in the patent and in the citation specifications. This finding however could be due to an omitted variable bias, as the share of unskilled migrants can proxy for the share of skilled ones. We therefore add the two shares jointly in the specification. While the coefficients of skilled migrants are robust to the inclusion of the unskilled counterpart, the coefficients of the unskilled migrants are now not statistically significant.[26] These findings do not contradict the existing literature, which finds an indirect effect of migration on economic outcomes. While this effect is positive and strong for general economic outcomes, if we limit our attention to innovation outcomes, as in the present case, the direct effect largely prevails over the indirect one.

Sixth, in the specifications presented so far we assumed a limited delay in the response of the dependent variable to changes in the explanatory variables. As an additional check, we have verified the robustness of our findings to the assumption of a slower response of the dependent variable. We rerun the main specification by using a two to four year lag in the controls. The coefficients of the diversity variable and of the stock of knowledge are robust to these different assumptions. In the online Appendix, Tables A8 and A9 report the estimated coefficients.

Finally, in the online Appendix we also present the results of the base regression for different technology subfields, both for patents and citations (Tables A10 and A11). There is some evidence that the effect of migration might be different by sectors, especially in the case of citations. The differential impact of diversity could be attributed to two main reasons. First, the distribution of migrants across sectors in the receiving countries might not be homogenous. As a result, the sectors whose estimated coefficient is positive might be the ones where most of the skilled migrants are employed. Second, knowledge production in a given sector might inherently be associated with higher returns, due for example to higher productivity. Given the aggregate nature of our proxy of diversity, which is not available at the sectoral level but only at the country level, we cannot clearly distinguish between these two effects in the present case.

6. Conclusion

In this paper we propose a simple knowledge production function in which innovation is a function of the stock of knowledge, the number of people employed in the research sector and the share of highly-skilled foreigners. We provide two proxies for the innovative capacity of countries, namely the number of PCT patent applications and the number of citations to published articles. Both are widely adopted measures, the first capturing private, patentable and applied knowledge, the second being a better indicator of general (public) knowledge in a society. In the sample of 20 European countries considered in this study, the two measures are highly correlated.

We show that highly-skilled foreigners have a positive impact on the innovative capacity of the recipient countries both for industrially applicable innovations and for more general abstract knowledge. This evidence extends the existing results to a broader set of countries than previously analyzed and to the use of alternative proxies for innovation. We reinforce the idea that complementarities exist between natives and foreigners. As in the micro-analyses on this topic, we find a positive synergic interaction that might be explained by diversity in cultures and in approaches to problem solving. Skilled migrants employed in highly skilled jobs positively impact on innovation by increasing the

[26] We cannot check the robustness of this result in an IV setting because the presence of two endogenous variables (diversity proxies) and of one instrumental variable (imputed shares) makes the model under-identified.

researchers' average productivity. The results we present hold true in a series of robustness checks, such as the inclusion of additional control variables, the use of longer lags, the use of different proxies for key explanatory variables, and the exclusion of certain countries from the sample.

In this analysis we employ an unconventional skills measure to account for skilled migrants. Rather than measuring education skills, we build our diversity index by using information on the actual occupation of foreign workers. Indeed, one would expect the actual employment of skills to determine an effect on innovation. As a robustness check, we also test whether the effect of highly-skilled migrants is robust to the use of the more traditional proxy based on education data. We find that the elasticities of diversity computed according to the two alternative skill measures are highly comparable in the case of public knowledge (citations). On the contrary, the education-based measure tends to slightly underestimate the contribution of skilled foreigners to the creation of industrially applicable knowledge (patents). Moreover, our estimates are robust to the inclusion of global technology trends, spillovers, geography-related shocks, and to different assumptions regarding the length on the innovation process.

Our results thus shed light on and complement the current debate on the creation of a common EU migration policy framework and the fostering of highly-skilled migration to Europe. We show that indeed the belief that European competitiveness can benefit from attracting highly-skilled migration is founded. An effective allocation of labor resources that reduce the over-qualification of migrants is a pre-condition for reaping higher benefits associated with highly-skilled migration flows. As a result, reforming the system to ease the access and recruitment of highly qualified migration would most likely be associated with significant short-term benefits in the creation of knowledge. In light of these results, schemes such as the EU Blue Card appear as a positive first step in the direction of fostering European innovativeness and competitiveness.

The empirical findings also confirm that the stock of existing knowledge has a positive effect on innovation. This supports the “standing on shoulders” assumption, to the extent that the accumulation of past knowledge increases the creation of new knowledge. Our results suggest that in order to make the EU competitive in the innovation domain, recruitment of highly-skilled migrants is only one of the key ingredients. Investments in R&D are at least as important, and this raises concern over the recent discussion about implementing research budget cuts in both member states and the EU.

Acknowledgments

The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 240895 — project ICARUS “Innovation for Climate Change Mitigation: A Study of Energy R&D, its Uncertain Effectiveness and Spillovers” and under grant agreement no. 308481 (ENTRACTE) “Economic iNsTRuments to Achieve Climate Targets in Europe.” Bosetti is deeply thankful to the Center on Advanced Studies in Behavioral Science at Stanford and IGIER, Bocconi. The authors would like to thank Massimiliano Bratti, Matteo Manera, Giovanni Peri, Silvia Salini and the participants of the XXVII AIEL Conference of Labour Economics, Research Seminar on Global Challenges, CMCC and FEEM Convention, NORFACE Conference, and EEA Conference. The authors are also grateful to the editors and two anonymous referees for their helpful comments and suggestions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jinteco.2015.04.002.

References

Abdih, Y., Joutz, F., 2006. Relating the knowledge production function to total factor productivity. Int. Monet. Fund Staff Pap. 53 (2).
Aghion, P., Howitt, P., 1992. A model of growth through creative destruction. Econometrica Econ. Soc. 60 (2), 323–351.
Altonji, J.G., Card, D., 1991. The effects of immigration on the labor market outcomes of less-skilled natives. In: Abowd, J., Freeman, R.B. (Eds.), Immigration, Trade, and the Labor Market. The University of Chicago Press.
Angrist, J.D., Pischke, J., 2009. Mostly Harmless Econometrics. Princeton University Press.
Arrow, K.J., 1962. The economic implications of learning by doing. Rev. Econ. Stud. 29, 155–173.
Aydemir, A., Borjas, G.J., 2011. Attenuating bias in measuring the wage impact of immigration. J. Labor Econ. 29 (1), 69–112.
Barabási, A.-L., 2005. Network theory—the emergence of the creative enterprise. Science 308 (5722), 639–641.
Barone, G., Mocetti, S., 2011. With a little help from abroad: the effect of low-skilled immigration on the female labour supply. Labour Econ. 18 (5), 664–675.
Bassett-Jones, N., 2005. The paradox of diversity management, creativity and innovation. Creat. Innov. Manag. 14 (2), 169–175.
Borjas, G.J., 2003. The labor demand curve is downward sloping: re-examining the impact of immigration on the labor market. Q. J. Econ. 118, 1335–1374.
Bottazzi, L., Peri, G., 2003. Innovation and spillovers in regions: evidence from European patent data. Eur. Econ. Rev. 47 (4), 687–710.
Branstetter, L., 2001. Are knowledge spillovers international or intranational in scope? Microeconometric evidence from the US and Japan. J. Int. Econ. 53 (1), 53–79.
Bratti, M., Conti, C., 2014. The effect of (mostly unskilled) immigration on the innovation of Italian regions. IZA Discussion Paper n. 7922.
Caballero, R., Jaffe, A., 1993. How high are the giants' shoulders? In: Blanchard, O., Fischer, S. (Eds.), NBER Macroeconomics Annual. 8. MIT Press, Cambridge, Mass, pp. 15–74
Cameron, C., Gelbach, J., Miller, D., 2008. Bootstrap-based improvements for inference with clustered errors. Rev. Econ. Stat. 90 (3), 414–427.
Card, D., 2001. Immigrant inflows, native outflows and the labor market impacts of higher immigration. J. Labor Econ. 19, 22–64.
Card, D., DiNardo, J., 2000. Do immigrant inflows lead to native outflows? Am. Econ. Rev. 90 (2), 360–367.
Card, D., Shleifer, A., 2009. Immigration and inequality. Am. Econ. Rev. Am. Econ. Assoc. 99 (2), 1–21.
Cattaneo, C., Fiorio, C., Peri, G., 2013. What happens to the careers of European workers when immigrants “take their jobs”? IZA Discussion Papers 7282
Chellaraj, G., Maskus, K.E., Mattoo, A., 2008. The contribution of international graduate students to US innovation. Rev. Int. Econ. 16 (3), 444–462.
Coe, D.T., Helpman, E., 1995. International R&D spillovers. Eur. Econ. Rev. 39 (5), 859–887.
Cohen, W., Klepper, S., 1996. Firm size and the nature of innovation. Rev. Econ. Stat. 78 (2), 232–243.
Cortes, P., Tessada, J., 2011. Low-skilled immigration and the labor supply women. Am. Econ. J. Appl. Econ. 3 (3), 88–123.
Crepon, B., Duguet, E., 1997. Estimating the innovation function from patent numbers: GMM on count panel data. J. Appl. Econ. 12 (3), 243–263.
D'Amuri, F., Peri, G., 2014. Immigration, jobs and employment protection: evidence from Europe. J. Eur. Econ. Assoc. 12 (2), 432–464.
D'Amuri, F., Ottaviano, G.I.P., Peri, G., 2010. The labor market impact of immigration in Western Germany in the 1990s. Eur. Econ. Rev. 54 (4), 550–570.
Davidson, R., MacKinnon, J., 2010. Wild bootstrap tests for IV regression. J. Bus. Econ. Stat. 28 (1), 128–144.
Docquier, F., Lowell, B.L., Marfouk, A., 2009. A gendered assessment of highly skilled emigration. Popul. Dev. Rev. 35 (2), 297–321.
Ducor, P., 2000. Intellectual property: coauthorship and coinventorship. Science 289, 873–875.
Dustmann, C., Glitz, A., Frattini, T., 2008. The labour market impact of immigration. Oxf. Rev. Econ. Policy 24 (3), 478–495.
EC European Commission, 2007. Communication From the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. COM (780 final).
EC European Commission, 2008. A Common Immigration Policy for Europe: Principles, Actions and Tools. Communication From the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. COM (359 final).
Elias, P., McKnight, A., 2001. Skill measurement in official statistics: recent developments in the UK and the rest of Europe. Oxf. Econ. Pap. 3, 508–540.
EMN European Migration Network, 2006. Impact of Immigration on Europe's Society. Prepared for the European Commission.
EUROSTAT, 2011. Annual Data on Employment in Technology and Knowledge-intensive Sectors at the National Level (1994–2008, NACE Rev.1.1).
Farre, L., González, L., Ortega, F., 2011. Immigration, family responsibilities and the labor supply of skilled native women. B.E. J. Econ. Anal. Policy 11 (1), 1–46 (Contributions).
Furman, J., Porter, M., Stern, S., 2002. The determinants of national innovative capacity. Res. Policy 21, 899–933.
Geroski, P., 1991. Entry and the rate of innovation. Econ. Innov. New Technol. 1, 203–214.
Green, C., Kler, P., Leeves, G., 2007. Immigrant overeducation: evidence from recent arrivals to Australia. Econ. Educ. Rev. 26, 420–432.
Griliches, Z., 1990. Patent statistics as economic indicator: a survey. J. Econ. Lit. 28 (4), 1661–1707.
Grossman, A., Helpman, E., 1991. Quality ladders in the theory of growth. Rev. Econ. Stud. 58 (1), 43–61.

Grossman, A., Helpman, E., 1994. Endogenous innovation in the theory of growth. J. Econ. Perspect. 8 (1), 23–44.
Hamilton, B.H., Nickerson, J.A., Owan, H., 2003. Team incentives and worker heterogeneity: an empirical analysis of the impact of teams on productivity and participation. J. Polit. Econ. 111 (3), 465–497.
Hargadon, A., 2003. How Breakthroughs Happen: The Surprising Truth About How Companies Innovate. Harvard Business Press.
Hicks, J.R., 1932. The Theory of Wages. Macmillan and Co., London.
Hollingsworth, J.R. and E Hollingsworth (2000) "Major discoveries and biomedical research organizations: perspectives on interdisciplinarity, nurturing leadership, and integrated structure and cultures", mimeo University of Saskatchewan.
Huang, K., Murray, F., 2009. Does patent strategy shape the long-run supply of public knowledge: evidence from human genetics. Acad. Manag. J. 52 (6), 1193–1221.
Hunt, J., Gauthier-Loiselle, M., 2010. How much does immigration boost innovation? Am. Econ. J. Macroecon. 2 (2), 31–56.
ILO, 1990. ISCO-88: International Standard Classification of Occupations. Technical report, ILO Geneve.
IOM, 2008. World Migration 2008, Managing Labour Mobility in the Evolving Global Economy. IOM, Geneva.
Iranzo, S., Peri, G., 2009. Migration and trade: theory with an application to the Eastern–Western European integration. J. Int. Econ. 79 (1), 1–19.
Jaffe, A.B., 1986. Technological opportunity and spillovers of R&D: evidence from firms' patents, profits and market value. Am. Econ. Rev. 76 (5), 984–1001.
Jones, B.F., 2009. The burden of knowledge and the Death of the Renaissance Man: is innovation getting harder? Rev. Econ. Stud. 76 (1), 283–317.
Jones, C.I., Williams, J., 2000. Too much of a good thing? The economics of investment in R&D. J. Econ. Growth 5 (1), 65–85.
Katz, J., Martin, B., 1997. What is research collaboration? Res. Policy 26 (1), 1–18.
Keller, W., 2002. Geographic localization of international technology diffusion. Am. Econ. Rev. 92 (1), 120–142.
Kerr, W., 2008. Ethnic scientific communities and international technology diffusion. Rev. Econ. Stat. 90 (3), 518–537.
Kerr, W., 2010. Breakthrough inventions and migrating clusters of innovation. J. Urban Econ. 67 (1), 46–60.
Kerr, S. And, Kerr, W., 2011. Economic impacts of immigration: a survey. Finn. Econ. Pap. 24 (1), 1–32.
Kerr, W., Lincoln, W., 2010. The supply side of innovation: H-1B Visa reforms and U.S. ethnic invention. J. Labor Econ. 28 (3), 473–508.
King, D.A., 2004. The scientific impact of nations. Nature 430, 311–316.
Kortum, S., 1993. Equilibrium R&D and the patent–R&D ratio: U.S. evidence. Am. Econ. Rev. 83 (2), 450–457.
Lazear, E.P., 1999. Globalisation and the market for team-mates. Econ. J. 109 (454), C15–C40.
Malerba, F., 1992. Learning by firms and incremental technical change. Econ. J. 102 (413), 845–859.
Manacorda, M., Manning, A., Wadsworth, J., 2006. The impact of immigration on the structure of male wages: theory and evidence from Britain. CEP Discussion Paper 754. London School of Economics.
Mancusi, M., 2008. International spillovers and absorptive capacity: a cross-country cross-sector analysis based on patents and citations. J. Int. Econ. 76 (2), 155–165.
Niebuhr, A., 2010. Migration and innovation: does cultural diversity matter for regional R&D activity? Pap. Reg. Sci. 89 (3), 563–585.
OECD, 2009. Patent Statistics Manual. OECD, Paris.
OECD, 2011. Science, Technology and Patents: Patents By Technology. http://stats.oecd.org/ (Accessed December 2011).
OECD, 2013. OECD International Migration Database. OECD, Paris.
OECD, 2014. International Migration Outlook 2014. OECD Publishing.
Ortega, F., Peri, G., 2009. The causes and effects of international migrations: evidence from OECD countries 1980–2005. NBER Working Papers 14833. National Bureau of Economic Research.
Østergaard, C., C., Timmermans, B., Kristinsson, K., 2011. Does a different view create something new? The effect of employee diversity on innovation. Res. Policy 40, 500–509.
Ottaviano, G.I.P., Peri, G., 2012. Rethinking the effect of immigration on wages. J. Eur. Econ. Assoc. 10 (1), 152–197.
Ottaviano, G., Peri, G., Wright, G.C., 2013. Immigration, offshoring and American jobs. Am. Econ. Rev. 103 (5), 1925–1959.
Ozgen, C., Nijkamp, P., Poot, J., 2011a. Immigration and innovation in European regions. IZA Working Paper 5676.
Ozgen, C., Nijkamp, P., Poot, J., 2011b. The impact of cultural diversity on innovation: evidence from Dutch firm-level data. IZA Working Paper 6000.
Parrotta, P., Pozzoli, D., Pytlikova, M., 2012. The nexus between labor diversity and firm's innovation. IZA Discussion paper 6972. IZA Bonn.
Peri, G., 2005. Determinants of knowledge flows and their effects on innovation. Rev. Econ. Stat. 87 (2), 308–322.
Peri, G., 2007. Higher education, innovation and growth. In: Brunello, G., Garibaldi, P., Wasmer, E. (Eds.), Education and Training in Europe. Oxford University Press, Oxford.
Peri, G., 2012. The effect of immigration on productivity: evidence from U.S. states. Rev. Econ. Stat. 94 (1), 348–358.
Peri, G., Sparber, C., 2009. Task specialization, immigration, and wages. Am. Econ. J. Appl. Econ. Am. Econ. Assoc. 1 (3), 135–169.
Peri, G., Sparber, C., 2011. Highly-Educated Immigrants and Native Occupational Choice. Industrial Relations 50 (3), 385–411.
Romer, P., 1990. Endogenous technical change. J. Polit. Econ. 98, 71–102.
Schmookler, J., 1966. Invention and Economic Growth. Harvard University Press, Cambridge, Mass.
Schumpeter, J., 1942. Capitalism, Socialism and Democracy. Harper, New York.
SCImago, 2011. Journal & Country Rank. http://www.scimagojr.com/ (Accessed December 2011).
Solow, R., 1957. Technical change and the aggregate production function. Rev. Econ. Stat. 39 (3), 312–320.
Stahl, G., Maznevski, M., Voigt, A., Jonsen, K., 2009. Unraveling the effects of cultural diversity in teams: a meta-analysis of research on multicultural work groups. J. Int. Bus. Stud. 41 (4), 690–709.
Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. Econometrica 65, 557–586.
Stephan, P.E., Levin, S.G., 2001. Exceptional contributions to US science by the foreign-born and foreign-educated. Popul. Res. Policy Rev. 20 (1), 59–79.
Stern, S., Porter, M., Furman, J., 2000. The determinants of national innovative capacity. National Bureau of Economic Research Working Paper 7876.
Stuen, E., Mobarak, A., Maskus, K., 2012. Skilled immigration and innovation: evidence from enrollment fluctuations in U.S. doctoral programs. Econ. J. 122 (565), 1143–1176.
Wooldridge, J., 2010. Econometric Analysis of Cross Section and Panel Data. 2 edition. MIT Press Books, The MIT Press.
Younglove-Webb, J., Gray, B., Abdalla, C., Thurow, A., 1999. The dynamics of multidisciplinary research teams in academia. Rev. High. Educ. 22 (4), 425–440.