DISCUSSION PAPER SERIES

IZA

IZA DP No. 9062

How Do Native and Migrant Workers Contribute to Innovation? A Study on France, Germany and the UK

Claudio Fassio
Fabio Montobbio
Alessandra Venturini

May 2015

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

How Do Native and Migrant Workers Contribute to Innovation? A Study on France, Germany and the UK

Claudio Fassio
CIRCLE, Lund University and BRICK, Collegio Carlo Alberto

Fabio Montobbio
University of Turin, CRIOS, Bocconi University and BRICK, Collegio Carlo Alberto

Alessandra Venturini
University of Turin, MPC, European University Institute and IZA

Discussion Paper No. 9062
May 2015

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 9062
May 2015

ABSTRACT

How Do Native and Migrant Workers Contribute to Innovation? A Study on France, Germany and the UK *

This paper uses the French and the UK Labour Force Surveys and the German Microcensus to estimate the effects of different components of the labour force on innovation at the sectoral level between 1994 and 2005. The authors focus, in particular, on the contribution of migrant workers. We adopt a production function approach in which we control for the usual determinants of innovation, such as R&D investments, stock of patents and openness to trade. To address possible endogeneity of migrants we implement instrumental variable strategies using both two-stage least squares with external instruments and GMM-SYS with internal ones. In addition we also account for the possible endogeneity of native workers and instrument them accordingly. Our results show that highly-educated migrants have a positive effect on innovation even if the effect is smaller relative to the positive effect of educated natives. Moreover, this positive effect seems to be confined to the high-tech sectors and among highly-educated migrants from other European countries.

JEL Classification: O31, O33, F22, J61

Keywords: innovation, migration, skills, human capital

Corresponding author:

Alessandra Venturini
University of Turin
Department of Economics
Via Po, 53
10124 Torino
Italy
E-mail: alessandra.venturini@unito.it

* We would like to thank, for their extraordinary research assistance, Metin Nebiler, Stella Capuano, Laura Bartolini, Davide Cannito, Marco Le Moglie, Sona Kantalaryan; for discussion and suggestions Herbert Brucker, Christian Dustmann, Alessandro Sembenelli, Paula Stephan, Cristian Bartolucci, Agnese Romiti, and Giovanni Facchini. We thank the financial support of the Migration Policy Center, EUI. This paper is a joint research between the Migration Policy Center, EUI and the University of Turin. The paper has benefited from comments and suggestions received during the following seminar and conference presentations: XXIX AIEL Conference, Pisa, September 2014; 5th ZEW/MaCCI Conference on the Economics of Innovation and Patenting, Mannheim, June 2014; Seminar at the Department of Spatial Economics, VU University, Amsterdam, January 2014; NORFACE-TEMPO Conference, Nottingham, September 2013; 3rd SEEK Conference, Mannheim, April 2013; and Seminar at the Department of Economics, University of Trieste, March 2013. The usual disclaimers apply.

1. Introduction

Over the years many researchers have claimed that skilled migration could increase innovation and productivity. Skilled immigrant workers contribute directly to research activities and may: provide complementary skills to natives; generate lower costs; and enhance critical mass and specialization of tasks within the firm. Most of the evidence using individual data focuses on skilled workers in Science & Engineering (S&E) in the United States where, according to the 2000 census, immigrants were 24% of the U.S. S&E workforce with a bachelor's degree and 47% of workforce with doctorates (Kerr and Lincoln, 2010; Hunt, Gauthier-Loiselle, 2010; Chellaraj *et al*., 2008; No and Walsh, 2010; Stephan and Levin, 2001). In parallel, macro evidence at country level tends to confirm the view that the share of immigrants in the total population has a positive effect on the level of Total Factor Productivity (Ortega and Peri, 2014; Alesina et al.2013).

In Europe evidence is more nuanced. Micro evidence on individual inventors shows that immigrant inventors outperform natives in terms of number of patent applications only in some countries (Breschi *et al*. 2014; Zheng and Ejermo, 2015). Macro studies, using regions as a unit of analysis, show controversial evidence: some studies show a positive effect for the share of skilled migration on innovation (Bosetti *et al*. 2012 for EU countries; Gagliardi, 2011 for UK), while other studies do not find this positive effect (Ozgen *et al*. for EU regions , 2012; Bratti and Conti, for Italy 2014).

The impact of migration on innovation and productivity is a key policy question in Europe where economic growth is slow and concerns have been expressed about sluggish improvements in tertiary education and innovation activities (European Commission, 2012). The future innovation capacity of Europe is also affected by the ageing population (Prskawetz and Lindh, 2006), long-term below replacement fertility and, finally, a continuous rise in life expectancy[1]. Lack of skills and an ageing labour force could hamper competitiveness and slow down the process of economic recovery. If the overall characteristics of European labour force seem problematic, it is important to understand whether migration can stimulate innovation and growth. Given the proliferation in recent years of economic studies on knowledge creation and innovation, it is surprising that there is still scant evidence, in particular for Europe, on the relationship between the different characteristics of the labour force and the rate and trajectories of technological innovation.

In order to address this issue, this paper estimates an innovation production function in 16 manufacturing industries (two digit level of the NACE classification) in France, Germany and UK, 1994-2005. In order to extend the analysis beyond the UK and the US this paper analyses the three largest European countries in terms of population and GDP. In addition the UK, France and Germany are the three European countries with the longest tradition of migrant employment in their labour markets. Our paper measures innovation using patents (weighted with forward citations) applied at the European Patent Office. The characteristics of the labour force are based on the Labour Force Surveys in France and the UK and the Microcensus in Germany.

The paper estimates an innovation production function similar to Furman *et al*. (2002) which contains all the different components of the labour force (age, level of education, ethnicity). We control for the existing stock of knowledge, R&D expenditures, and openness to trade. The paper marks an advances on previous works of research in at least three respects.

[1] In France the young (below 15 years old) are 1.35 of the retired (more than 65 years old), while in Germany and the UK the size of the young is smaller than the size of the older (respectively 0.85 and 0.89) (Eurostat, 2012).

First, it fully controls for the different characteristics of the labour force, in particular level of education and age and not only for the country of origin of migrants. The level of education measures the human capital of the worker and his or her ability to learn and its propensity to innovate. So we can compare, for the UK, Germany and France, the effects of skilled and unskilled migrants. Low-skilled migration could affect technological adoption decisions and investments in physical capital (Lewis 2011; Bratti and Conti, 2014). It is, therefore, relevant to consider the impact of both skilled and unskilled migration. In addition recent literature shows that the risk propensity (which is strongly correlated to the propensity to innovate) and the depreciation of human capital vary strongly with age (Prskawetz and Lindh, 2006). Moreover, our empirical specification allows us to directly compare the contribution of migrant and native workers to innovation along these different lines.

Second, we identify the effect of migration on innovation at the level of industry. This provides an improvement and a complementary view relative to the existing literature[2]. There is, in fact, a vast literature showing that sectors differ substantially in terms of innovation and R&D intensity. Recent papers studying the impact of migration on patents take a regional or provincial perspective (Ozgen *et al*. 2012; Alesina *et al*. 2013; Bosetti *et al*. 2012; Bratti and Conti 2014). However, it is difficult to get away from the problem that industries vary dramatically in the production of patents and that an empirical strategy based on regions and provinces as a unit of analysis is not able to provide information on whether immigrants (in particular skilled immigrants) are really employed in the patenting sectors[3]. In addition, if migration tends to concentrate in specific fields of activities[4], aggregate effects on innovation and productivity might be related not only to migration flows but also to the sectoral composition of the economy.

Finally, this paper addresses a number of econometric issues. Demand pull effects on migration at industry level require appropriate instruments. Moreover, there might be a set of additional unobserved factors that affect both patent production and migration at the industry level. Also, the use of Labour Force Surveys can generate measurement errors. Our identification strategy (employing longitudinal data at industry-country level) is based on two different instrumental variable strategies: the first relies on the adaptation of the common procedure used in the literature, devised by Card (2001); the second exploits the availability of internal instruments, that is lags in the endogenous variables (system-GMM: Blundell and Bond, 1998). Both bring with them advantages and drawbacks: the use of external instruments need specific behavioural assumptions, which might or might not apply. On the contrary the use of internal instruments is better suited for large samples with a high number of observations.

Our paper shows that highly-skilled migration has a positive effect on innovation. At the same time it finds that the effect is smaller relative to that of skilled natives (about one third). This is a warning flag because if skilled immigrants displace skilled natives the aggregate effect on innovation

---

[2] An attempt in this direction can be found in European Commission (2009) with an analysis restricted to migrants' share.

[3] A different and complementary approach has explored how ethnic diversity affects innovation and productivity growth finding in most cases a positive effect (Niebuhr, 2010; Bosetti *et al.* 2012; Bratti and Conti 2014; Ozgen *et al.* 2011; Alesina *et al.* 2013).

[4] For the UK see for instance Dustmann *et al.* (2003), where the concentration of different ethnic migrants in different sectors is displayed in Table 3.3 pag.31. For Germany see Fertig and Schmidt (2001), while for France see Constant (2005).

could be negligible. In addition, the positive effect for migrants seems to be confined to the high tech sectors and to skilled migrants from other European countries[5].

The paper proceeds in Section 2 by positioning our paper in the context of the available empirical literature. Section 3 explains data and methodology and defines the knowledge production function that is used to model the innovative output. This depends upon the different characteristics of the labour force, controlling for the usual determinants of innovation activities. In Section 3 we, also, explain our identification strategy. Section 4 describes the data and discusses our empirical results, while Section 5 offers a conclusion.

2. The related literature on migration and innovation

The recent literature has paid a great deal of attention to migration as a potential determinant of innovation and productivity. Most studies have focused on the role of skilled migrants, since these are more likely to have an effect on innovation. A number of studies have focused especially on the role of graduates, inventors and scientists in Science and Engineering (S&E) disciplines, often taking advantage of micro data on individuals: the results point to a general positive effect of highly-skilled immigrants on a number of innovation measures such as patents, citations or scientific publications. Kerr and Lincoln (2010) study the link between patents and a special US visa policy (H-1B), which favours the entrance of foreign workers in S&E, noting a positive effect of migration on the overall production of patents in US cities. Hunt and Gauthier-Loiselle (2010) find as well that an increase in the share of tertiary educated migrants in the US increases the number of patent applications. In both cases the positive effects are strongly driven by the high share of highly-skilled immigrants in S&E disciplines. Chellaraj *et al.* (2008) show that the presence of foreign graduate students in US universities has a significant and positive impact on both future patent applications and future patents awarded to university and non-university institutions. Similar results are provided by No and Walsh (2010), who find that among the respondent of a survey of US-based inventors the share of non-US-born among the leading inventors is disproportionally high. Stephan and Levin (2001) focus on scientists in the US and find an over-representation of foreign-born scientists among the individuals the make exceptional contributions to Science and Engineering.

While these studies provide very accurate evidence on the positive effect of skilled migration on innovation outcomes, their results are often limited to the subset of skilled immigrants in S&E disciplines. Therefore, the external validity of the results is quite low and might not be sufficient for the implementation of migration policies that, instead, necessarily concern a wider range of migrants. Moreover, the results are limited to the US. Only recently some studies have provided initial evidence on European countries using individual data on inventors. Breschi *et al.* (2014) employing data on patent application at the EPO found that for some European countries immigrant inventors show a very high patent productivity. However, their results do not hold for all European countries. Zheng and Ejermo (2014) using individual data on Swedish foreign-born inventors confirm that the positive effect of skilled migrants in Europe is less clear-cut, since they find that in Sweden immigrant inventors do not outperform natives in terms of the number of patent applications submitted.

Another perspective on the link between migration and innovation is provided by the literature, which uses aggregate data at the regional or country level: these studies adopt a more comprehensive approach, not only S&E and inventor migrants are considered, but also other types of

[5] Unfortunately we are unable to distinguish between former member of the EU (15) and the new accession countries (EU12). Note that the Former Yugoslav countries are included.

skilled and sometimes unskilled migrants. [6] In this case the effect of migration is not only tested on the number of patents and citations, but also on other proxies of innovativeness such as Total Factor Productivity (TFP) or the introduction of innovations by firms. Ortega and Peri (2014) implement a cross-country analysis using a sample of 188 countries and find a positive elasticity in the share of immigrants (regardless of their skill level) over the total population on the level of TFP. In this stream of literature there is also much more evidence for European countries. Bosetti *et al.* (2012), using a panel of twenty European countries, find that skilled migrants contribute positively to the number of patents and citations of scientific publications. Gagliardi (2011) finds that the share of skilled migrants within a UK province has a positive impact on the innovative performances of firms in that specific province. In most of these studies not only the share of (skilled) migrants is considered, but also their degree of diversity in terms of countries of origin. Alesina, Harnoss and Rapoport (2014) using cross-country data find a positive effect for diversity, especially among highly-skilled migrants. Ozgen *et al.* (2012) find that in a sample of 170 European regions the share of immigrants does not lead to a higher number of patent applications, while the diversity of the countries of origin leads, indeed, to more patents. Niebhur (2010) finds, meanwhile, that the diversity of the migrant population (especially of highly-skilled immigrants) has a positive effect on the level of patent applications among German regions. However, not all studies find a positive effect of immigrants on the innovativeness of regions, especially in countries in which skilled migration is not a common phenomenon. Using data on Italian provinces Bratti and Conti (2014) do not find that skilled migration has any effect on patent production. They find, instead, a negative and significant effect of un-skilled migration on innovation.

These studies allow us to broaden the focus of analysis from S&E immigrants to a wider set of skilled (and unskilled) migrants. However they all adopt a geographical approach, according to which the effect of migrants is measured on the innovative performance of the country/region/province in which they are resident. This methodology carries the risk of overlooking an important confounding factor represented by the sectoral specialization of each geographical unit. Indeed, it is well known that the pace of innovation is strongly technology-specific and varies dramatically across sectors (Breschi *et al.* 2000). Immigrants might be attracted to regions in which the growth of high-technology sectors is very strong and, therefore, also the number of patents is growing steadily. However, a geographical approach cannot distinguish between the effect of immigrants that directly contribute to innovation because they work in innovative sectors, and the, let's say, 'spurious' effect of immigrants that merely work in complementary sectors in regions with a high growth in innovative sectors. In this respect a sectoral approach like the one we are implementing in this study seems better suited to measuring the effect of immigrants on innovation.

Another confounding factor that has not been considered in the current literature, but that is likely to affect the overall contribution of immigrant workers to innovate is their age profile. The human capital theory (Becker, 1975) shows that, at the end of the education period, workers reach their maximum productivity, which depreciates as their career goes on. This result can be imputed to the decline in cognitive abilities for older individuals, as stated by Oberg (1969), Jones (2010) and Fargues and McCornick (2013). Schubert and Andersson (2013) using matched employer-employee data for Sweden find that the overall employees' age has a negative impact on innovation outcomes. Considering that immigrants are on average younger than natives, not controlling for the age effect can induce an overestimate in immigrants' role in innovation. Finally, the role of low or medium educated (low-skilled) migrants in the innovation process has not been explored in depth. Only Bratti and Conti (2014) find that, in Italy, low-skilled immigrants contribute negatively to the number of

[6] In this case skilled migrants can be either migrants with tertiary education level (most frequent case) or migrants employed in highly-skilled occupations.

patent applications, since in some middle or low technology intensity sectors also non tertiary educated immigrants might contribute to the innovation performances of firms.

3. Model, Methodology and Data

3.1 Model

Unlike the previous literature that uses mostly country, regions or provinces, our unit of analysis is the manufacturing sector. Our empirical model adapts Furman *et al.* (2002) that studies the innovative capacity of countries. According to standard endogenous growth models (Romer, 1990) the rate of technological progress is given by:

$$\dot{A}_t = \delta(A_{t-1}^{\beta}\ \mathrm{H}_{t-1}^{\gamma}) \quad (1)$$

The sustainable rate of technological progress at time *t* (Å$_t$) depends upon the stock of accumulated knowledge A_{t-1} and by an ideas generation input (H_{t-1}), which operates according to a standard Cobb-Douglas production function. This particular specification assumes some complementarity between inputs, so that the marginal impact on innovation of the inputs increases in the level of all of the other factors. Our analysis is performed at industry level and therefore expanding Eq. (1) we obtain:

$$\dot{A}_{it} = \delta(A_{it-1}^{\beta}\ R\&D_{it-1}^{\gamma}\ L_{it-1}^{\varphi}\ \mathrm{X}_{it-1}^{\theta}) \quad (2)$$

We test whether the annual flow of patents (Å$_{i,t}$) (weighted by citations) in year *t* and sector *i* is explained by lagged yearly expenditures in Research and Development ($R\&D_{i,t-1}$) and a lagged measure of the openness to trade of a specific sector ($X_{i,t-1}$) that is the volume of exports plus imports per unit of production in sector *i* at time *t-1*. The annual number of patents being an annual flow, following equation (2), we also control for the stock of patents in the previous year ($A_{I,t-1}$). $A_{I,t-1}$ measures the stock of prior ideas and prior research. Note that if the coefficient of A is positive this means that the stock of prior ideas increases patent productivity (this is also called the “standing on the shoulders of giants” effect), but if the coefficient is negative it would indicate that new inventions are becoming increasingly difficult. The main focus of the paper is on the role of human resources in innovation. We use the lagged human capital characteristics ($L_{i,t-1}$) in that specific sector *i*. It is important to underline that we decompose the human capital variable by age, education and ethnicity. In doing so we assume imperfect substitutability of different labour factors as in Ottaviano and Peri (2012). It is important to remark that unlike Furman *et al.* (2002) we are also interested in the role of workers without tertiary education.

The dependent variable is the number of forward citations received by the patents in the four years after the application date.[7] We model our production function as a Cobb Douglas and we take logs to estimate the elasticity of each of the different inputs. We lag each independent variable by one year[8] as follows:

[7] We use the number of forward citations received by each patent, instead of the simple number of patents, in order to select only patents with economic value (for a thorough explanation see Section 3.3)

[8] We acknowledge that the lag could be longer, but considering that we are using the priority date of patents, and that the R&D and labour force time series are quite persistent, we believe that one lag is a correct compromise in order to maintain a sufficient number of observations.

$$\ln \dot{A}_{it} = d + \beta \ln A_{it-1} + \gamma \ln R\&D_{it-1} + \sum_k \phi_k \ln L^k_{it-1} + \theta \ln X_{it-1} + \alpha_i + \lambda_t + \varepsilon_{it} \quad (3)$$

The employment L is divided into k different components, according to ethnicity, education and age; a_i is the time-invariant fixed effect of each sector, λ_t denotes a common time trend (that we proxy with time dummies) and ε_{it} is the idiosyncratic shock occurring at time t in sector i. The analysis covers seventeen industries (two digits NACE) in the manufacturing sector, from 1994 to 2005 and three countries: France, Germany and the UK. As a consequence subscript i refers to the country-sector pair, which is our observational unit in the panel. Table (1) provides a precise list with the definition of variables.

3.2 Identification strategy

In order to estimate equation (3) we need to address a number of econometric issues that might affect our coefficients of interest. Our main concerns are directed towards the correct identification of the effect of labour variables, and in particular of migrant workers, on patent production. A first problem is related to the fact that the decision to move to a specific country is, in most of the cases, a strategic decision that depends on the specific dynamics of the sectors in which migrants will work. In other words a sector that is expanding and that needs additional manpower will attract workers both from inside and from outside the country. This demand-pull effect, if not accounted for, is likely to affect our estimates, because current and past shocks of the dependent variable might be correlated with our variables of interest. Moreover, it is likely that patent productivity shocks in a given sector have differentiated effects according to workers' skills and education. Indeed, an increase in the overall number of patents in a sector indicates a gradual shift of firms towards higher levels of technological sophistication. According to the vast literature on biased technological change (Acemoglu, 2002) technical change is more likely to exert a positive effect on the demand for educated workers, while it might have a negative effect on the demand for unskilled ones. In this respect the choice to lag by one year all the independent variables in equation (3) represents a first step in addressing this problem. But it is not likely to solve it completely.

A second problem is generated by other unobserved factors, which might affect both patent-productivity at the sectoral level and the decision of migrants to move to a specific national sector. For example a high-tech multinational that starts a green field investment in a given country is likely to affect both the production of new patents in a given sector and the flow of skilled migrants that come to work in that same sector. Again these factors would lead to problems of omitted variables bias due to both time-invariant and time-varying unobserved heterogeneity. Finally the last problem is related to the existence of possible measurement errors in the number of migrant workers. The use of Labor Force Surveys data should allow us to take into account sampling errors, through the use of population weights. However, the probability of incurring random measurement errors in national statistics on the labor force is significant, especially for data on migrant workers,. This might lead to attenuation bias problems in the estimation of the coefficients of interest.

We address these issues in the following way. Our starting point is a fixed-effects Ordinary Least Squares estimation that accounts for time-invariant unobserved heterogeneity denoted by a_i in equation (3). However the fixed effects estimator is consistent under the unrealistic assumption of strict exogeneity between the covariates and the sector-specific idiosyncratic productivity shock ε_{it}. This means that the independent variables must be uncorrelated with past, present and future shocks of the dependent variable (Chamberlain, 1982; Griliches and Mairesse, 1998). While we can easily assume that the labour variables are uncorrelated with future shocks, a past shock in patent productivity will typically affect the levels of employment in the following periods and possibly also in that same time-period. This is particularly important for migrant workers, but it might also affect the behaviour of native workers. As shown by Wooldridge (2002, p.301) the bias of the fixed effects estimator when strict exogeneity does not hold might be quite large, especially when time series are persistent, as is often the case for aggregated labor variables time series[9]. Wintoki *et al.* (2012) focus specifically on the direction of the bias of the fixed effects estimator when strict exogeneity is violated and find that when the explanatory variable is negatively correlated with past values of the dependent variable the fixed effects estimator will have an upward bias. A positive correlation of the explanatory variable with past shocks of the dependent variable will, meanwhile, lead to a downward bias in the fixed effects estimator. In the case of patents the demand for educated workers is positively correlated with past shocks of patent productivity, while the opposite might occur for unskilled workers. Therefore, we expect a downward bias of the fixed effects estimator for educated workers and, possibly, an upward bias for unskilled workers.

In addition fixed effects estimators fail to account for the unobserved factors that might occur during the period of observation (as in the example of multinationals' brand new investments) and which might also induce a bias in the coefficients of interest.

In order to address these problematic issues related to the use of fixed effects estimators we implement two different instrumental variable strategies: the first relies on the use of external instruments, according to a common procedure used in the literature and first devised by Card (2001), while the second exploits the availability of internal instruments, that is lags in endogenous variables. We implement both strategies since they have advantages and drawbacks: the use of external instruments is well suited to our empirical setting, but it relies on specific behavioural assumptions by individuals which may or may not apply. On the contrary the use of internal instruments does not require specific assumptions. Rather, it is better suited to large samples with a high number of observations.

<u>External instruments</u>

Our first instrumental variable strategy relies on the well-known identification strategy first implemented by Card (2001). He addresses the potential endogeneity of the flows of migrants with respect to the economic conditions of the geographical areas in which they would migrate. This methodology takes advantage of the fact that migrants of a certain nationality tend to move to locations where other people of the same nationality had already settled. Therefore, by using the original distribution of nationalities at the beginning of the period of observation and the exogenous migration flows, it is possible to create fictional flows of migrants to be used as external instruments. This is possible because these flows are strongly correlated with the endogenous stocks of migrants, but at the same time because they are also uncorrelated with the shocks of the dependent variable. For our empirical design we adapt this instrumental variables (IV) methodology substituting sectors for geographical areas. In other words, we do not exploit the fact that migrants tend to move to areas where people of their same nationality are already settled. Rather, we take advantage of the fact that migrants often work in the same economic activities in which their compatriots are already active. The validity of this identification strategy rests on the hypothesis that the network effect, or better the effect of the "migratory chain" on the new inflows of migrants is not only limited to location effects: these produce a concentration of migrants in the same area (as in the original Card model). Rather, the "migratory chain" extends also to the sector of employment. Indeed the community of origin acts as a placing agency, reducing the cost of finding a job in the sectors in which the migrants from a specific country of origin are already concentrated (Ellis and Wright, 1999; Strom *et al.*, 2013). Frequently job engagement is already found before the arrival of the co-nationals.

---

[9] While for weakly dependent time series the bias of fixed effects is of order T^{-1} and hence it can be minimal for sufficiently long time series, if processes are very persistent (close to unit root AR(1) processes) the bias instead is independent of T and therefore can be relevant (Wooldridge, 2002).

For each of our migration-related variables we implement the following strategy in order to create the fictional levels of migrants workers in each sector. Sticking to the original notation of Card (2001), for each destination country (France, Germany and the UK), we compute the flow M_{ot} of new migrants from a specific area of origin (we use eight large geographic groups[10]) *o* in year *t*. Then for each destination country we computed the distribution of migrant workers from a specific area of origin in the different sectors of the economy at the beginning of our period of observation.[11] For each sector and for each area of origin we calculated the share λ_{oj}, where *j* indicates the sector in which they are active:

$$\lambda_{oj} = \frac{Mig_{oj94}}{Mig_{o94}}$$

In order to distinguish between skilled and unskilled migrants we calculated for each year *t* the fraction τ_{ogt} of all new immigrants from a specific country of origin *o* that have a specific type *g* of education (either high or middle-low education) as follows:

$$\tau_{ogt} = \frac{\Delta Mig_{ogt}}{\Delta Mig_{ot}}$$

For each sector *j* in each destination country, the fictional flow of new migrants from a specific country of origin *o* with education *g* is equal to:

$$\Delta Mig\_instr_{ojgt} = M_{ot} * \lambda_{oj} * \tau_{ogt}$$

These fictional flows of new migrants are aggregated over countries of destinations (differentiated by the two types of education) to obtain the fictional stocks of total migrants of a specific type of education in sector *j* at time *t*. These new stocks are used as external instruments for the real stocks of high and middle-low educated migrants in equation (3) in an IV setting with a two-stage least squares estimator. If our hypotheses hold these fictional stocks should be correlated with the actual stocks of migrants in each sector; but at the same time they should not be correlated with the patent shocks.

Internal instruments

Our second instrumental variable strategy relies instead on the use of internal lags in the endogenous variables as suitable instruments: we use the Blundell and Bond GMM-SYS estimator (1998). The GMM-SYS estimator accounts for the violation of the strict-exogeneity condition, which can greatly affect the reliability of fixed effects estimates. Indeed the GMM-SYS allows for sequential exogeneity, i.e. the explanatory variables need to be uncorrelated only with future shocks in the dependent variable, that is a much more plausible assumption. Moreover, differently from the exactly-identified Card-like IV strategy based on external instruments, the GMM-SYS estimator allows us to test for the exogeneity of the instruments, since the use of several lags in the endogenous variables allows for an over-identified specification. Finally the GMM-SYS allows us to

[10] Following D'Amuri and Peri (2014) we use the following eight zones of origin: Africa, North America, Central and South America, Middle East and Central Asia, South and Eastern Asia, Eastern Europe, Western Europe, and Oceania.

[11] This corresponds to 1994 for France and the UK and 1996 for Germany.

instrument, as well, the labor variables that measure native workers, since these variables are, also, likely to be endogenous.[12]

In equation (3) we consider the labour variables (both migrants and natives) as *endogenous*, that is, correlated with past and present values of the error term, while we consider all the other control variables as strictly exogenous. We will then estimate equation (3) instrumenting the endogenous variables L^k with their own lags (in differences and in levels). A possible shortcoming of the GMM-SYS estimator is that it is better suited for large samples of individuals, while in our sample the number of sectors in the three countries is limited. This fact may lead to the problem of instruments over-fitting (Roodman 2009), due to the high number of instruments with respect to the number of observations, which decreases the reliability of the Hansen test on the exogeneity of the internal instruments. For this reason, in our estimates, we limit, as much as possible, the number of lags used as instruments, employing only those that are most informative. Furthermore, we implement the procedure suggested by Roodman (2009) in order to reduce the overall number of instruments, by collapsing, into one single instrument, all the lags used as instruments, in order to reduce the proliferation of instruments.

Finally the adoption of internal instruments is also able to address the problems related to measurement errors. Indeed, if measurement error is free of serial correlation (and we believe this would be the case in our context), the panel dimension of the data deals with attenuation bias, precisely because it provides internal instruments. Griliches and Hausman (1986) show that the use of fixed effects (within estimators) can amplify the problems due to measurement error in panel studies. They also show that the best strategy to overcome this problem, more than IV strategies based on external instruments, is the use of internal instruments.

3.3 Data

We take advantage of an original dataset which combines data on innovation, as proxied by patents, and information on the characteristics of the labour force (migration, age and education) at the sectoral level. Measuring innovation and technical change is a daunting challenge since innovation is a multi-faceted phenomenon and knowledge creation does not always leave a paper trail. One of the most popular indicators of innovation is the number of patents applications at industry or country level (e.g. Furman *et al.* 2002)[13]. We use patent applications at the European Patent Office (EPO) because we analyse three European countries. In addition international patent applications at the EPO are costly and, therefore, we select inventions with relevant market potential (Deng, 2007)[14].

---

[12] The choice to use the GMM-SYS estimator instead of the Arellano and Bond GMM-DIF (1991), which also uses lags of the endogenous variables as suitable instruments, is motivated by the fact that labour variables are usually quite persistent. When time series are persistent the GMM-SYS specification is to be preferred because it not only uses lagged levels of variables as instruments for the equation in differences - which have very a low predictive power in the case of persistency - but also lagged differences for the equation in levels, which, instead, have a good explicative power in cases of persistency (Blundell, Bond, 1998).

[13] Patent data are typically considered an important indicator of innovation activity and they are extensively used in the economic literature. They provide valuable information on the technological activities of inventors and companies across countries in specific technological fields for long time series (Pavitt, 1985; Grupp, 1990 and Griliches, 1990). The economic literature has validated the use of patents showing that there is a high level of correlation with R&D activities at the firm level (Griliches, 1990) and that patents are a good proxy for the technological effort of companies and non-firm organizations aiming to create new products and processes.

[14] Patent indicators have many limitations that have to be taken into account. Many inventions are not patented. Even if patents are increasingly used by companies, the evidence provided by many surveys of R&D managers indicates that, in many sectors, patents are not considered the major source of profit from new

Finally we use an international patent office to offer a homogeneous database which allows cross-country comparisons and that is less distorted by country-specific institutional or policy changes.

The technological and economic value of patents varies considerably and many patents have low economic and technological value, while a few of them are extremely valuable. Patent citations are then used to correct this problem and to measure the economic and technological value of a patent[15]. For all three countries patents and patent citations are derived from PATSTAT (see Appendix B), which provides full data about patents registered at the EPO and the citations received. The conversion of the International Patent Classification to NACE sectors is provided by Schmoch *et al.* (2003). Patents are assigned to countries using the address of the inventors and fractional counting.

The information concerning human capital (level of education, country of origin, age) was retrieved through the aggregation at the sectoral level of micro data from the national Labour Force Surveys for the UK, France and from the Microcensus in Germany. Appendix B provides an extensive description of the data. R&D expenditure and trade data by sectors are provided by the STAN database (OECD). The list of the countries of origin used in the paper is in Appendix C.

4. The empirical analysis

4.1 Descriptive Evidence

Table (2) and (3) display the main characteristics of the database in the three countries in two sub-periods at the beginning and at the end of the period considered (1994-2005): the number of patents and citations *per* worker, the share of immigrants and the share of worker aged 35 years or younger (40 for Germany). Table (2) refers to all sectors, Table (3) shows the data just for high-tech sectors (See the Table A1 in the Appendix for the classification of sectors). Patents and patent citations *per* employee are higher in the high tech sector. The number of citations decreases substantially in the second period due to the obvious right-end truncation bias[16]. In the manufacturing sectors considered the share of young workers remarkably decreases over time, mainly because of the decreasing share of young workers among the non-tertiary educated. The share of tertiary educated, instead, is on the rise particular in the UK and France.

The overall share of immigrant workers in manufacturing sectors is falling slightly in Germany, where it is about 12% of the overall employment; on the contrary, in the UK and France the share of immigrants increases, respectively, from 6% to almost 8% and from 2% to 4%. Note that the share of tertiary-educated immigrants is growing in all countries: in the UK, where it already accounted for 1.2% of the labour force in 1994-1996, it doubles during the period of observation and reaches 2.4%

---

products and processes (e.g. Cohen *et al*., 2000). This depends upon the nature of the technologies. As a consequence, companies have a significantly different propensity to patent across different sectors of economic activity. Finally, like R&D measures, patents tend to be a better proxy for the technological activities of large firms. Small firms tend to have a lower propensity to patent because – all other things being equal – the use of intellectual property requires high fixed costs of implementation and scale (Bound *et al.* 1984, Patel and Pavitt, 1994). It follows that the size distribution of firms may have an important effect on the aggregate count of patents at the national level.

[15] In the literature expenditure on research and development (R&D) is often used as a proxy of innovation potential. However, in our analyses R&D is used as an input in the production of innovation; in addition we use R&D to identify the high-tech and low-tech sectors, following the standard OECD classification (Hatzichronoglou, 1997).

[16] See Bacchiocchi and Montobbio (2012) for the analysis of the truncation bias in patent citations in different patent offices.

in 2003-2005. Also in France, where the shares of highly-educated migrants are substantially lower (0.3% in 1994-1996), the percentage doubles reaching 0.7% in 2003. In Germany the growth is slightly less high (from 0.7% to 1.1%).

Table (2) shows an increase in the number of EU27-nationals immigrants in France and the UK. In the UK this is primarily due to the growth of tertiary educated EU27-nationals (mainly young highly-educated immigrants from Eastern Europe[17]). In Germany instead the share in EU2-nationals is quite stable over time, but there is an increase in the share of tertiary educated EU-nationals.

In Table (4) we show the number of patents and citations *per* employee, as well as the share of immigrants, broken down by sectors. The Table highlights once more the great heterogeneity in the production of patents at the sectoral level: high tech sectors like Office, Accounting and Computing Machinery display more than 10 patents for 1000 workers, compared to 0.2 in the Textile sector. The share of immigrant workers is high in the Textile and Automotive sector, but it mainly consists of low and middle educated workers. On the contrary tertiary-educated immigrants are more numerous in Office, Accounting and Computing Machinery, as well as in the Chemicals and Pharmaceutical sectors and Radio, Television and Communication Equipment. The share of European Union workers is quite constant across all sectors (around 3-4% of the labor force); on the contrary, the share of tertiary-educated EU nationals is substantially higher in all high tech sectors.

4.2 Econometric Results

Table (5) reports the descriptive statistics for each of the variable used in the estimations. We have sixteen two-digit sectors for twelve years in France (1994-2005) fourteen two-digit sectors for twelve years in UK (1994-2005) and sixteen two-digit sectors for ten years in Germany (1996-2005)[18].

As a preliminary result in Table (6) we display a set of simple pairwise relationships between our innovation measures and several measures of the size of our units of observation *i* (country-sector pairs). In our econometric analysis of equation (3) individual fixed effects will capture the average size of the dependent variable (number of citations): however it is also important to clarify the unconditional correlation between these different measures of size and the level of innovation activities. As measures of size we consider: value added, total employment, tertiary educated employment, total level of R&D and the accumulated stock of knowledge. The R^2 indexes in Table (6) show that individually, each measure explains between 7% and 88% of the overall variation in the

---

[17] It must be stressed that the Czech Republic, Estonia, Cyprus, Latvia, Lithuania, Hungary, Malta, Poland, Slovakia and Slovenia entered the European Union on the 1 May 2004.

[18] For the UK we lack data on R&D expenditures in two sectors (Manufacture of wood products and cork; Manufacture of paper and paper products) therefore we can only apply our model to fourteen sectors. Our original sample consists hence of 520 observations: 192 observations in France, 168 observations in UK and 160 observations for Germany. In the estimation we use one year lag and therefore we lose sixteen observations in France and Germany and fourteen in the UK (46 overall), which correspond to the first year of each time-series. Furthermore, in France, in the first years of observation in some sectors with a low number of employees (Wood and products of wood and cork, Paper and paper products, Office Accounting and Computing Machinery) there are no foreign workers at all, so we cannot retrieve information on the average age of foreign workers: therefore, we lose fifteen observations in France. This also happens (for only one observation) both in UK and Germany. Overall, and obviously discounting the 'lost' observations, we have 161 observations for France, 143 observations for Germany and 153 observations for the UK, which sums up to 457 observations that will be used in our estimates.

number of citations. Value added and total employment account for a very low fraction of the cross-sectoral variance of innovative activities. The scale of tertiary educated employment, R&D and knowledge stock have, as expected, a much higher explicative power. Scale dependent variables related to R&D efforts can explain a substantial portion of the innovative output. In parallel the estimated coefficients have values that range from 0.64 to 1.46, suggesting the existence of decreasing or close to constant returns-to-scale. The only exception is the size of the tertiary educated employment, (higher than 1). However, these coefficients (and in particular the one on employment) provide little intuition on how they drive innovation activities. On top of these scale-dependent effects, the question remains whether it is possible to disentangle a separate and quantitatively significant impact on innovation of the different components of the labour force.

In Table (7) we turn to the estimation of equation (3) using data from all countries, including time dummies to account for the common time trend. The dependent variable is the number of citations received by the patents applied at the EPO in the four years after the application. All variables are in logs and each covariate is included with a lag of one year in order to partially reduce the problems linked to reverse causality. Our specifications include controls for openness to trade, expenditures in R&D, and the cumulated stock of patents.

In Table (7) we measure the effect of all the labour force together and then we distinguish between tertiary educated and low-middle educated workers. The GMM-SYS estimators in columns (2) and (4), which properly account for the possible endogeneity of the labour force, show that the coefficient of all those employed is negative and significantly different from zero. In column (4), when we distinguish between high and low educated workers however, we find that, as expected, the two have differentiated effects on citations: tertiary educated workers display positive and significant effects, while middle-low educated workers have a negative and significant effect. The negative sign of the total labour force is, therefore, due to a composition effect, since middle and low educated workers represent the majority of total employment. With respect to the other control variables in the model, that we treat as strictly exogenous, the results show a negative effect on the average age of workers, especially for non-educated workers, and positive and significant coefficients for R&D expenditures and the stock of knowledge: the openness to trade is, meanwhile, negative and significant. The AR(1) and AR(2) tests confirm the goodness of our model specification, since they show that there is no residual serial correlation in the error term of the model. Moreover, the heteroskedasticity-robust Hansen test accepts the null-hypothesis of strength and exogeneity of the lagged instruments in use.

As a benchmark in Table (7) we also report the coefficients obtained with fixed effects estimators in columns (1) and (3) to check whether the direction of the bias of these estimators is consistent with our expectations. Woolridge (2002) and Wintoki *et al*. (2012) analyse the direction of the bias in fixed effects estimates in which the explanatory variable is correlated with past shocks of the dependent variable. According to Woolridge (2002) and Wintoki *et al*. (2012) we should expect a downward bias for educated workers (positive correlation with past shocks) and an upward bias for unskilled ones (negative correlation with past shocks). Indeed, looking at the results of column (5) we find that the fixed effects estimator displays a downward bias in the coefficient of educated workers, with respect to the GMM-SYS estimates, and an upward bias in the coefficient of non-educated workers.

So far our results confirm that human capital quality is a key variable influencing innovation performances. However, our aim is to, also, check the differentiated contribution of native and immigrant employees, controlling for their education.

In Table (8) we distinguish between the native and immigrant workforce and within each of these subsets we discriminate between tertiary educated and non-tertiary educated employees. Our specifications include time dummies and all the additional controls (R&D expenditures, stock of citations, openness to trade): furthermore we also check here for the effect of age distinguishing between the average age of each of the four identified groups of workers (highly-educated natives, highly-educated immigrants, low educated natives and low educated immigrants): all the coefficients of the control variables are reported in the Appendix in Table (A2). In Table (8), instead, our focus is on the estimated coefficients of the labour variables. In column (1) we report the coefficients obtained with a fixed effects estimator: the estimated coefficients of the labour variables are likely to be affected by endogeneity, therefore we report them only as a benchmark with respect to the results obtained through the use of external and internal instruments. In columns (2) and (3) we show the results of a Two Stage Least squares (2SLS) instrumental variable estimation in which we use, as external instruments, the fictional stocks of high and low educated migrants, following our modified version of Card (2001). In order to understand properly how the instrumental strategy works we first instrument only non-educated migrants with the fictional stocks of non-educated migrants and then we instrument only highly-educated migrants with the fictional stocks of highly-educated migrants. As we will show this we use this strategy because the behavioural assumption at the basis of the instrumental variable strategy does not work in the same way for both types of migrant workers: more specifically we find that the Card-like instruments works only with middle and low educated workers.

The results in column (2), in which we instrument only middle-low educated migrants, show that this category of migrants has a negative and significant effect. The non-instrumented highly-educated migrants have, on the other hand, a positive coefficient, of the same magnitude as the one obtained in the fixed effects specification. As in Table (7) when we instrument non educated workers we find the coefficient becomes even more negative, in line with the hypothesis of an upward bias in fixed effects estimates (Wintoki *et al.*, 2012). Natives are, instead, never significant, whether they are educated or not. When we look at the first stage statistics in the lower part of Table (8) we see that the external instrument is a good predictor of the levels of non-educated migrants: though, since we are in an exactly-identified specification, we cannot test for the exogeneity of the instrument. The Angrist and Pischke F-statistics show that the instrument is not weak.[19] In column (3), instead, we adopt the same specification, but this time we instrument the highly-educated migrants with their fictional stocks: in this case the predictive power of the instrument is extremely low, contrary to the case of low educated migrants we cannot rely on this identification strategy for this category of migrant workers. The Angrist and Pischke F-statistics is equal to 1.25, showing that the instrument is extremely weak, while the Hausman test rejects the hypothesis of exogeneity of the instrumented variable: though note that the Hausman test is not reliable with very weak instruments. We interpret these results as an empirical test of the behavioural assumptions behind our estimation strategy: while for low-educated workers it seems that the presence of immigrants from a certain country in a given sector attracts new migrants from abroad to the same sector, in the much more recent and lower scale case of highly-educated workers this is not the case. For highly-skilled migrants the market signals are more efficient that ethnic networks in helping co-ethnics find jobs.

Conversely the sectoral choice of highly-educated workers, with specialized skills, is not affected by the sectoral specialization of the workers from the same country of origin. Indeed, Card's strategy is originally devised to account for the behaviour of mainly low-skilled migrants entering the United States (in Card's study immigrants were mostly Hispanics from Mexico and South America and had, on average, two or three of education less than natives).

To address endogeneity we also chose to implement a GMM-SYS estimator. First we address a number of issues related to the correct choice of the lag specification of the variable. Indeed since

[19] The Hausman test on the endogeneity of the instrumented variables cannot reject the null-hypothesis of exogeneity, though the p-value of the test is relatively low, which casts some doubts on the real exogeneity of the variable.

our data has a limited number of observational units (country-sector pairs) and a quite large number of years, we are very parsimonious with respect to the number of lags used as instruments, to avoid the problem of over-fitting instruments (Roodman, 2009). Moreover we test whether the Blundell-Bond (1998) GMM-SYS estimator is more appropriate than the Arellano-Bond (1991) estimator. This is the case if the specification of equation (3) in levels with the lagged instruments in differences works better than the specification in differences with the lagged levels of the instruments. Therefore, in Table (9) we test the predictive power of lagged levels and lagged differences of each of the labour variables. In the upper panel of Table (9) we present the first stage results of a fixed effects 2SLS estimation of equation (3) in levels, in which we instrument separately each of the four labour variables with their lagged differences. On average the results show that lagged differences have a good predictive power, as shown by the significance of the coefficients. However, we find that for educated migrants the first, second and third lagged difference can be used as suitable instruments, while for non-educated migrants only the second and third lagged differences are relevant. When we check for educated natives we find that only the first lagged difference is significant, while for non-educated natives the first and second lagged differences are important. In the lower panel of Table (9) we check whether when we transform equation (3) in first-differences, lagged levels of the endogenous labour variables are good instruments. In line with our expectations we find that lagged levels are not sufficiently powerful instruments for the variables in differences, due to the persistency of the labour variables (Blundell and Bond, 1998): almost all the lagged levels are insignificant in the first stage, with the exception of educated natives, in which, instead, the one and two-years lagged levels are significant. These results confirm that the GMM-SYS specification is legitimised by the relevance of lagged differences as instruments for the equation in levels.

On the basis of the findings of Table (9) we estimate equation (3) with a GMM-SYS in which we use, as instruments, only the lags that are found to be useful for each labour variable. We avoid using more than two lags for each of the variables in order not to use too many lags. Moreover, we apply the procedure suggested by Roodman (2009) which collapses instruments in order to further decrease the number of instruments.[20] In Column (4) of Table (8) we present the results of a GMM-SYS in which only migrants are instrumented with their own suitable lags. The results are quite in line with those obtained in column (3), when we instrumented only non-educated migrants. Highly-educated migrants are always significant and their coefficient increases by 80%. Low educated migrants have, instead, a negative coefficient, whose size is comparable with the one found in the external-instruments 2SLS estimation. However, in this case it is not significant. The AR(1) and AR(2) tests on the presence of serial correlation in the error terms show that the specification chosen is correct, while the Hansen test accepts the null hypothesis that instruments are strong and orthogonal to the error term. Since we have good reasons to believe that natives may not be strictly exogenous, in column (5) we also instrument highly-educated natives and low-educated natives with their own lags. In this specification the coefficients of the migrants (high and low educated) are consistent with the previous specification in column (5): highly-educated migrants have a positive and significant coefficient, while non educated migrants have a negative and significant coefficient. The results change, instead, for what concerns the native labour force, consistent with our expectations. Now highly-educated natives are positive and significant, while non educated native display a negative and significant effect: again when we endogenize the labour variables we find that, with respect to the fixed effects results, the coefficient for educated workers increases in size, while it decreases for unskilled ones. In the Appendix in Table (A2) we display the coefficients of the variables controlling for the effect of age, distinguishing by education level and ethnicity: the results

[20] In Table (7) we show one-step standard errors, since in small samples with a large number of instruments (due to a large T) standard errors in two-step GMM tend to be severely downward biased (Windmeijer, 2005). However, when we implemented the two-step procedure the results of the GMM-SYS estimation were largely unaffected.

show significant coefficients only for the middle-low educated natives, which display a negative effect on innovation.

Our results show that highly-educated migrants have a positive effect on innovation in the three European countries analysed, but their effect is smaller than educated natives: it stands, in fact, at only one third that of natives. A 1% increase in the number of highly-educated natives leads to a 0.3% increase in the number of citations, whereas a 1% increase in the number of highly-educated migrants leads to slightly less than 0.1% increase in citations.

Heterogeneous effects

The effects that we have identified for the foreign labor force might differ according to specific characteristics of migrants (e.g. country of origin) or to the specific sectors in which they are employed. A typical drawback of our use of international educational standards to classify foreign workers is the great heterogeneity in terms of the quality of higher education degrees in different countries of the world. Indeed, it is likely that graduates from different countries will also display different levels of skills according to the quality of their national higher education system (of course this problem is much less relevant for the native educated labour force). If this is the case it is not correct to pool graduates from very different countries together[21]. We address this issue distinguishing between migrants coming from European and extra-European countries. We believe this distinction should allow for a lower degree of heterogeneity, at least as far as concerns European foreign workers, since most countries in Europe have gone through an important process of convergence in the organization of their educational system (see in particular the Bologna process). Recent works (Moguérou, Di Pietrogiacomo, 2008; Breschi *et al.*, 2014) also show that skilled migration in Europe consists of European citizens moving from one country to another, exploiting their right to move freely across European borders[22].

In Table (10) we distinguish between workers whose country of origin is a European country and workers who come from outside the Europe[23]: we do this both for tertiary-educated workers and for workers without a tertiary degree. We estimate the model both through OLS estimators with fixed effects and with the GMM-SYS estimator used in the previous section, in order to account for the possible endogeneity of each component of the labor force. In the Appendix in Table (A3) we report all the coefficients of the variables in our model, while in Table (10) we only report the coefficients of interest. The OLS and GMM-SYS results in columns (1) and (2) of Table (10) are qualitatively similar. However, the GMM coefficients are slightly larger and more significant than the OLS ones. Both results show that educated foreign workers have a positive effect on innovation, but that in the case of European educated workers the effect is larger than for non-European ones: indeed the positive coefficient of the former is more than twice as large as the coefficient of the latter. When, instead, we focus on the effect of non-tertiary educated foreign workers distinguishing between European

[21] In principle this problem could also affect non-tertiary educated foreign workers, since in this case, too, the skill level of workers might depend on the level of development of their own national below-tertiary education system, (i.e. technical and professional schools).

[22] This is a key difference with respect to other countries such as the United States, where migrants come from very world regions (Latin and Central America, as well as Asia and Europe).

[23] In our analysis the set of European countries includes also some countries which are not inside the European Union, such as Norway, Switzerland, Bosnia-Herzegovina, Serbia, Montenegro and Albania. We made this decision as some national statistical offices aggregated workers coming from a contiguous set of countries, so in some cases we could not distinguish between, say, a Slovenian (inside the European Union) and a Bosnian (outside the European Union).

and non-European ones, we find that only non-European workers display a negative effect on patent citations, while the effect of European foreign workers is not significantly different from zero.

Another potential limitation of our baseline specification has to do with the specific sector of activity in which foreign workers are employed. Most of the literature that analysed the effect of migration on innovation has found a positive effect for tertiary educated foreign workers in high-tech occupations, especially in the US (Kerr and Lincoln, 2010; Hunt and Gauthier-Loiselle, 2010). Therefore, it is important to check whether in the three European countries that we analyse the positive effect of educated migrants on innovation is mainly related to high-tech sectors, as in the US, or whether it is homogenous across all sectors. Also, the contribution of middle-low educated foreign workers might differ according to the type of sector in which they are employed, whether high or low-tech. In order to take these heterogeneous effects into account we classify sectors following the usual OECD classification (Hatzichronoglou, 1997)[24] in high-tech and low-tech sectors. In order to obtain comparable results with the previous specifications in Table (8) we keep the same number of observations and interact educated migrants with two dummies. A high-tech dummy which is equal to 1 for the observations belonging to high-tech sectors or that is otherwise zero. A low-tech dummy which is equal to 1 for the observations belonging to low-tech sectors or that is otherwise zero. In column (3) and (4) of Table (10) we present a new specification, which is equal to the one presented in Table (8). Now, though, educated migrants are interacted with the two dummies, to check for differentiated effects. In column (4) the GMM-SYS estimates show that the positive effect of educated migrants is positive and significant only in high-tech sectors: moreover, its coefficient is 40% higher than the coefficient found for the total economy in Table (8). In column (6) we, instead, interact non-educated migrants with the technological dummies. We find that when we adopt the GMM-SYS estimator the negative effect of non-educated migrants is stronger in low tech sectors.

5. Concluding comments

In this article we estimate the effect of the employment of native and migrant workers on innovation using the French and UK Labour Force Surveys and the German Microcensus; these were, then, merged at the sectoral level, with the European patents and citations database (PATSTAT). This paper adds to the existing literature in a number of respects.

It complements previous results on migration and innovation that are based on regional and national approaches whose results are often driven by the specific sectoral distribution of the migrants by country of origin. It studies the impact on innovation of the labour force in three large European countries, expanding the evidence available for the US and, finally, compares the impact of migrants and natives workers.

Using an innovation production function we control for age, level of education, countries of origin, R&D, knowledge stock and openness to trade. Identification is based on two different instrumental variable strategies: the first extends the common procedure of Card at sectoral level (2001); the second exploits the availability of internal instruments (GMM-SYS). The second seems more appropriate and tackles the issue of the endogeneity of both migrant and native workers.

[24] More specifically, due to the lack of three-digit sectoral disaggregation in our database, our high-tech sectors correspond to the set of OECD medium-high tech and high tech, while low-tech sectors correspond to the set of low-tech and mid-low tech sectors.

We show that a highly-educated labour force has a positive impact on innovation. This holds not only for the aggregate number of highly-educated workers but also, with a smaller coefficient, for migrants. In particular a 1% increase in the number of educated natives leads to a 0.3% increase in the citation-weighted number of patents, a 1% increase in the number of highly-educated migrants leads to a slightly less than 0.1% increase in the citation-weighted number of patents. This paper also shows that the effect of migrants varies according to sector and that highly-educated migrants have a positive effect on innovation in particular for high-tech sectors. The effect of low-skilled migrants is, meanwhile, negative in both sectors but more negative in low tech sectors.

The positive effect is stronger for European migrants than for third-countries nationals but in general it is rather small. This raises the question of the appropriate policies to adopt to favour European innovation. Migration policies could favor the entrance of potential workers with education in science, technology, engineering, or mathematics (STEM), or with advanced degrees, such as masters' degrees and doctorates in these areas. National policies – at least, in the Netherlands, Sweden and the UK - and the European Blue Card (Blue Card Directive, 2009) –used extensively in Germany- try to facilitate the recruitment of highly-skilled workers. They should probably focus more, though, on highly educated workers or on foreign students already in destination countries. Our results also suggests that investment in the education of native students could have a positive effect on innovation activities, particularly in high tech sectors.

7. References

Acemoglu, D., (2002), Directed Technical Change, *Review of Economic Studies*, 69(4), pp. 781-809.

Alesina A and Harnoss J, Rapoport H., (2013), Birthplace Diversity and Economic Prosperity, *NBER Working Papers 18699*, National Bureau of Economic Research, Inc.

Arellano, M. and S. Bond. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *Review of Economic Studies*, 58, 277-297.

Bacchiocchi E., Montobbio F. (2010), International knowledge diffusion and home-bias effect. Do USPTO & EPO patent citations tell the same story?, *Scandinavian Journal of Economics*, 112 (3), pp. 441-470

Becker G., (1975), *Human Capital. A Theoretical and Empirical Analysis with Special Reference to Education*, University of Chicago Press, Chicago.

Blundell, R. and S. Bond (1998), Initial Conditions and Moment Restrictions in Dynamic Panel Data Models, *Journal of Econometrics*, 87, 115-143.

Bosetti V., Cattaneo C., Verdolini E., (2012), Migration, Cultural Diversity and Innovation: A European perspective, *FEEM Working Paper*, 2012.069

Bound, J., Cummings, C., Griliches, Z., Hall, B., Jaffe, A., (1984), Who does R&D and who patents? in Griliches, Z. (ed.) "R&D, Patent and Productivity", *National Bureau of Economic Research,* University of Chicago Press, pp. 21-54.

Bratti M. and Conti C., (2014), The Effect of (Mostly Unskilled) Immigration on the Innovation of Italian Regions, *Institute for the Study of Labor (IZA) DP No. 7922*.

Breschi, S., Malerba F., and L. Orsenigo, (2000), Technological Regimes and Schumpeterian Patterns of Innovation, *The Economic Journal*, 110 (463), 388–410.

Breschi, S., Lissoni, F. Tarasconi, G., (2014), Inventor Data for Research on Migration and Innovation: A Survey and a Pilot, *WIPO Economic Research Working Papers n.17*, World Intellectual Property Organization, Geneva.

Card, D. (2001). Immigrant Inflows, Native Outflows, and the Local Market Impacts of Higher Immigration, *Journal of Labor Economics*, 19(1), 22-64.

Chamberlain, G. (1982) Multivariate regression models for panel data, *Journal of Econometrics*, 18, 5-46.

Chellaraj, G., Maskus, K. and A. Mattoo, (2008), The contribution of skilled immigrations and international graduate students to U.S. innovation. *Review of International Economics* 16, 3, 444–62.

Cohen, W. M., Nelson, R. R. and J. P. Walsh, (2000), Protecting their intellectual assets: appropriability conditions and why U.S. manufacturing firms patent (or not), *NBER Working Paper 7552*.

Constant A., (2005), Immigrant Adjustment in France and Impacts on the Natives in K.F. Zimmermann (ed.), "European Migration: What Do We Know", *Oxford, OUP*, and *IZA DP. 2004/8063.*

D'Amuri, F. and Peri, G., (2014), Immigration, Jobs and Employment Protection: Evidence from Europe before and during the Great Recession, *Journal of the European Economic Association*, 12(2), 432–464.

Deng, Yi, (2007), Private value of European patents, *European Economic Review*, 51, 1785–1812.

Dustmann C., Fabbri F., Preston I., Wadsworth J., (2003), Labour market performance of immigrants in UK labour market, *Home office on line report 05/03*.

Ellis, M., and Wright, R., (1999), The Industrial Division of Labor among Immigrants and Internal Migrants to the Los Angeles Economy, *International Migration Review*, 33(1), pp. 26-54

European Commission (2012); Progress towards the common European objectives in education and training (2010/2011) - Indicators and benchmarks. http://ec.europa.eu/education/lifelong-learning-policy/indicators10_en.htm

Fargues Ph., McCormick A., (2013), Ageing of skills and complementary immigration in the EU, 2010 - 2025 *EUI/RSCAS Working Papers 2013/81*.

Fertig, M., Schmidt C., (2001), First- and Second-Generation Migrants in Germany - What Do We Know and What Do People Think, in R. Rotte, P. Stein (eds.), “Migration Policy and the Economy: International Perspectives”, *IZA DP2001/286*.

Furman J., Porter M.E., Stern S., (2002), The determinants of national innovative capacity, *Research Policy* 31, 899–933.

Gagliardi L., (2011), Does Skilled Migration Foster Innovative Performance? Evidence from British Local Areas, *SERC Discussion papers 97*.

Griliches, Zvi, (1990), Patent statistics as economic indicators: A survey, *Journal of Economic Literature* 28, 4, 1661–1707.

Griliches, Z. and Hausman, J. A., (1986), Errors in variables in panel data*, Journal of Econometrics*, 32, 93-118.

Griliches, Z. and J. Mairesse, (1998), Production functions: The search for identification, Chapter 6, in S. Strom (ed.), Econometrics and Economic Theory in the 20th Century, Cambridge: Cambridge University.

Grupp, H., (1990), Technometrics as a Missing Link in Science and Technology Indicators, in J. Sigurdson (Ed.), “Measuring the Dynamics of Technological Change”, Pinter, London, pp. 57–76.

Hatzichronoglou, T. (1997), Revision of the high-technology sector and product classification, *STI Working Papers, OECD/GD*, 97, 216.

Hunt, J., and Gauthier-Loiselle M., (2010), How much does immigration boost innovation?, *American Economic Journal: Macroeconomics*, 2(2), 31-56.

Jones B., (2010), Age and Great Invention, *The Review of Economics and Statistics*, 92, (1), 1-14.

Kerr W., Lincon W.F., (2010), The Supply Side of Innovation: H-1B Visa Reforms and U.S. Ethnic Invention, 28(3), 473-508, *Journal of Labour Economics*.

Lewis, E., (2011), Immigration, Skill Mix, and Capital Skill Complementarity, *The Quarterly Journal of Economics*, 126(2), 1029-1069.

Moguérou P., Di Pietrogiacomo M.P., (2008), Stock, Career and Mobility of Researchers in the EU, *JRC Scientific and Technical Reports*, European Commission.

Niebuhr A., (2010), Migration and innovation. Does Cultural Diversity matter for regional R&D Activity?, *Papers in Regional Sciences*, 89(3), 563-585.

No, Y. and Walsh, J. P., (2010), The importance of foreign-born talent for US innovation, *Nature biotechnology*, 28(3), 289-291.

Oberg W., (1969), Age and achievement- And the technical man, *Personnel Psychology*, 13 (3), 245-259.

Ortega, F. and G. Peri, (2014), Openness and income: The roles of trade and migration, *Journal of International Economics*, 92(2), 231-251.

Ottaviano GM., Peri G., (2012), Rethinking the Effects of Immigration on Wages, *Journal of the European economic association*, vol.10(1), pp.152-197.

Ozgen, C., Nijkamp, P., Poot, J., (2012), Immigration and Innovation in European Regions, in P. Nijkamp, J. Poot J and M. Sahin (eds.) “Migration Impact Assessment: New Horizons”, Edward Elgar.

Patel, P. and K. Pavitt, (1994), The Nature and Economic Importance of National Innovation Systems, *STI Review*, No. 14, OECD, Paris.

Pavitt, K., (1985), Patent Statistics as Indicators of Innovative Activities: Possibilities and Problems. *Scientometrics*, 7, 77–99.

Prskawetz A., and Lindh T., (2006), The Impact of Population Ageing on Innovation and Productivity Growth in Europe, in *Research report 28* by Prskawetz A., Mahlberg B., Skirbekk V., Freund I., Winkler-Dworak M., Vienna Institute of Demography, Austrian Academy of Sciences, and Lindh T., Malmberg B., Jans A-C., Nerdstrom O., Andersson F., Institute for Future Studies, Stockholm.

Roodman, D. (2009), A Note on the Theme of Too Many Instruments, *Oxford Bulletin of Economics and Statistics*, 71, 135–158.

Romer, P. M., (1990). Endogenous Technological Change, *Journal of Political Economy*, 98(5), 71-102.

Schmoch U., Laville F., (2003), Linking Technologies to industrial Sectors. *Final report to the European Commission*, DG Reaserch.

Stephan P., Levin S., (2001), Exceptional contributions to US science by the foreign-born and foreign-educated, *Population Research and Policy Review*, 20(1), 59-79.

Strom S., Venturini A., Villosio C., (2013), Wage assimilation: migrants versus natives, internal migrants versus foreign migrants, *MPC-RSCAS 2013-30*.

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

Wintoki, M.B., Linck, J.S., Netter, J.M., (2012), Endogeneity and the dynamics of internal corporate governance, *Journal of Financial Economics*, 105, 581–606.

Wooldridge, J.M.,(2002), *Econometric Analysis of Cross Section and Panel Data*, The MIT Press, Cambridge.

Zheng, Y. and Ejermo, O. (2015), How Do the Foreign-born Perform in Inventive Activity? Evidence from Sweden, *Circle Papers in Innovation Studies*, 2015/9.

TABLES

Table 1. Description of variables

VARIABLE	DESCRIPTION	SOURCE
Logcit	Log of the 4-years citation-weighted patents.	PATSTAT – EPO Database
R&D	log of R&D expenditures (in PPP 2005 dollars)	OECD, BERD-SAN Database
Stock_cit	log of the stock of citations-weighted patents, created using a perpetual inventory method (depreciation rate set at 15%)	PATSTAT – EPO Database
Open	log of the openness to trade. (Import + Export)/Value added	OECD-STAN database
E	log of total employment	Labour force Surveys for France and UK. Microcensus for Germany
E_Tedu	log of employees with tertiary education. In the UK we consider as tertiary-educated those workers that left school when they were older than 21 years old. In France we consider tertiary educated those workers who obtained a degree that is beyond that of the "baccalaureat general". In Germany is tertiary education.	Labour force Surveys for France and UK. Microcensus for Germany
E_noTedu	log of employees without tertiary education	Labour force Surveys for France and UK. Microcensus for Germany
E_Tedu_nat	log of native employees with tertiary education	Labour force Surveys for France and UK. Microcensus for Germany
E_Tedu_imm	log of immigrant employees with tertiary education. In each of the Labour Force Surveys (and the Microcensus for Germany) we considered as immigrant/foreigner any worker whose nationality is different from that of the country in which he or she is working.	Labour force Surveys for France and UK. Microcensus for Germany
E_noTedu_nat	log of native employees without tertiary education	Labour force Surveys for France and UK. Microcensus for Germany
E_noTedu_imm	log of immigrant employees without tertiary education	Labour force Surveys for France and UK. Microcensus for Germany
E_Tedu_mig EU	log of immigrant employees with tertiary education holding the nationality of a European country	Labour force Surveys for France and UK. Microcensus for Germany
E_Tedu_mig NOEU	log of immigrant employees with tertiary education holding the nationality of a country outside	Labour force Surveys for France and UK. Microcensus for Germany

	Europe	
E_noTedu_mig EU	log of immigrant employees without tertiary education holding the nationality of a European country	Labour force Surveys for France and UK. Microcensus for Germany
E_noTedu_mig NOEU	log of immigrant employees without tertiary education holding the nationality of a country outside Europe	Labour force Surveys for France and UK. Microcensus for Germany
Avg_age	log of the average age of the total employment	Labour force Surveys for France and UK. Microcensus for Germany
Avg_age_Tedu	log of the average age of employees with tertiary education	Labour force Surveys for France and UK. Microcensus for Germany
Avg_age_nat	log of the average age of the native employees	Labour force Surveys for France and UK. Microcensus for Germany
Avg_age_Tedu_nat	log of the average age of native employees with tertiary education	Labour force Surveys for France and UK. Microcensus for Germany
Avg_age_imm	log of the average age of the immigrant employees	Labour force Surveys for France and UK. Microcensus for Germany
Avg_age_Tedu_imm	log of the average age of immigrant employees with tertiary education	Labour force Surveys for France and UK. Microcensus for Germany

Table 2. Descriptive statistics.

	UK		FRANCE		GERMANY	
	1994-1996	2003-2005	1994-1996	2003-2005	1996-1998	2003-2005
Patents/Citations (per 1000 employee)						
Patents	0.91	1.64	1.42	2.09	2.28	3.08
Citations	1.54	0.73	2.04	0.78	3.21	1.54
Share of young workers	0.44	0.35	0.40	0.37	0.41	0.34
Tertiary-educated	0.05	0.07	0.08	0.11	0.04	0.03
Non-tertiary-educated	0.39	0.28	0.33	0.27	0.37	0.31
Share of tertiary educated	0.08	0.14	0.14	0.20	0.10	0.11
Share of immigrants	0.064	0.079	0.027	0.042	0.127	0.121
Tertiary-educated	0.012	0.024	0.003	0.007	0.008	0.011
Non-tertiary-educated	0.052	0.055	0.024	0.035	0.110	0.103
EU nationals	0.021	0.024	0.012	0.020	0.062	0.063
EU nationals tertiary-educated	0.004	0.007	0.002	0.004	0.004	0.006

We classify as "young" workers that are younger than 35. See Table (1) for a precise definition of "tertiary-educated workers" and "immigrant workers".

Table 3. High Tech sectors

	UK		FRANCE		GERMANY	
	1994-1996	2003-2005	1994-1996	2003-2005	1996-1998	2003-2005
Patents/Citations (per 1000 employee)						
Patents	1.67	2.86	2.79	4.23	3.74	4.88
Citations	2.98	1.31	4.20	1.62	5.53	2.47
Share of young	0.45	0.34	0.38	0.37	0.41	0.34
Tertiary-educated	0.06	0.09	0.11	0.14	0.05	0.05
Non-tertiary-educated	0.38	0.26	0.27	0.23	0.35	0.30
Share of educated	0.12	0.18	0.21	0.28	0.15	0.16
Share of immigrants	0.061	0.078	0.021	0.035	0.118	0.113
Tertiary-educated	0.016	0.030	0.004	0.011	0.012	0.016
Non-tertiary-educated	0.045	0.048	0.017	0.024	0.098	0.090
EU nationals	0.020	0.024	0.011	0.018	0.060	0.060
EU nationals tertiary-educated	0.005	0.009	0.003	0.008	0.006	0.008

We classify as "young" workers that are younger than 35. See Table (1) for a precise definition of "tertiary-educated workers" and "immigrant workers" and Table A1 in the Appendix for the definition of high-tech sectors.

Table 4. Patents and migrant shares by sector

		Patents/Citations (per 1000 employee)		Immigrants				
Industry	ISIC REV. 3.1	*Patents*	*Citations*	*Share of immigrants*	*Tertiary educated*	*Non-tertiary educated*	*EU nationals*	*EU nationals tertiary educated*
Food Products, Beverages And Tobacco	15-16	0.12	0.09	0.07	0.01	0.06	0.03	0.003
Textiles And Textile Products, Leather And Footwear	17-19	0.21	0.16	0.12	0.01	0.11	0.04	0.003
Wood And Products Of Wood And Cork	20	0.14	0.06	0.05	0.00	0.05	0.03	0.002
Pulp, Paper, Paper Products, Printing And Publishing	21-22	0.38	0.38	0.08	0.01	0.07	0.04	0.004
Chemicals And Pharmaceuticals	24	4.63	5.86	0.06	0.02	0.04	0.03	0.008
Rubber And Plastics Products	25	1.54	1.12	0.08	0.01	0.08	0.03	0.002
Other Non-Metallic Mineral Products	26	1.11	0.90	0.06	0.01	0.05	0.03	0.003
Basic Metals	27	0.68	0.42	0.09	0.01	0.08	0.03	0.002
Fabricated Metal Products, exc. Machinery. and Equip.	28	0.56	0.38	0.07	0.01	0.07	0.03	0.002
Machinery And Equipment, Nec	29	2.90	2.20	0.06	0.01	0.05	0.03	0.004
Office, Accounting And Computing Machinery	30	10.57	7.12	0.08	0.04	0.04	0.03	0.017
Electrical Machinery And Apparatus, Nec	31	1.73	1.33	0.07	0.01	0.05	0.03	0.005
Radio, Television And Communication Equipment	32	6.80	6.52	0.07	0.02	0.04	0.03	0.010
Medical, Precision And Optical Instruments	33	6.10	5.58	0.06	0.01	0.04	0.03	0.006
Motor Vehicles, Trailers And Semi-Trailers	34	1.63	1.90	0.10	0.01	0.08	0.04	0.004
Other Transport Equipment	35	0.79	0.48	0.05	0.01	0.04	0.02	0.006

Table 5. Descriptive statistics

Variable	Mean	Std. Dev.	Min	Max	Observations
logcit	5.355	1.607	0.405	8.677	457
R&D	20.219	1.525	16.132	23.374	457
open	-0.154	0.606	-1.412	1.631	457
stock_cit	7.821	1.527	2.943	10.739	457
E	12.461	0.657	9.834	14.052	457
E_Tedu	10.379	0.703	8.503	12.030	457
E_noTedu	12.268	0.725	8.849	13.826	457
E_Tedu_nat	10.280	0.723	8.439	11.957	457
E_noTedu_nat	12.192	0.707	8.849	13.717	457
E_Tedu_mig	7.697	0.995	4.691	9.826	457
E_noTedu_mig	9.445	1.161	4.940	11.889	457
E_Tedu_mig EU	6.360	2.141	0	9.210	457
E_Tedu_mig NOEU	6.453	2.471	0	9.302	457
E_noTedu_mig EU	8.537	1.435	0	11.289	457
E_noTedu_mig NOEU	8.720	1.722	0	11.238	457
avg_age	3.681	0.033	3.546	3.764	457
avg_age_Tedu	3.644	0.073	3.483	3.859	457
avg_age_Tedu_nat	3.643	0.076	3.476	3.867	457
avg_age_Tedu_mig	3.649	0.125	2.996	4.135	457
avg_age_noTedu	3.686	0.037	3.544	3.787	457
avg_age_noTedu_nat	3.687	0.037	3.545	3.783	457
avg_age_noTedu_mig	3.706	0.093	3.332	4.060	457

We have 16 two-digit sectors for 12 years for France (1994-2005), 14 two-digit sectors for 12 years for the UK (1994-2005) and 14 two-digit sectors for 10 years for Germany (1996-2005). Our original sample, thus, consists of 520 observations: 192 observations in France, 168 observations in the UK and 160 observations for Germany. Because of the one year lag chosen for our estimation, we lose 16 observations in France and Germany and 14 in the UK (46 overall), which correspond to the first year of each time-series. Furthermore, especially in France, in the first years of observation for some small and high tech sectors there were no foreign workers at all, so we can’t retrieve information on the average age of foreign workers: therefore we lose those observations (15 observations in France). This also happens, for only one observation, both in the UK and Germany. Net of these missing observations, overall we have 161 observations for France, 143 observations for Germany and 153 observations for the UK, which sums up to 457 observations that are used in our estimates.

Table 6. Scale effects. Bivariate OLS regressions of innovation on size variables.

Dep. Variable: Logcit $_{i,j,t,}$	*Value added only*	*Total Employment only*	*Tertiary educated employment only*	*R&D only*	*Patent Citations Stock & Trend*
ln(Value Added)$_{i,j,t}$	0.64***				
	(0.30)				
E $_{i,j,t}$		0.66**			
		(0.35)			
E_Tedu $_{i,j,t}$			1.46***		
			(0.20)		
R&D $_{i,j,t}$				0.80***	
				(0.08)	
Stock_citations $_{i,j,t}$					0.95***
					(0.03)
YEAR					- 0.12
					(0.01)
R-squared	0.1	0.07	0.36	0.56	0.88

The Table displays the coefficients of bivariate OLS regressions. The dependent variable is Logcit $_{i,j,t}$, where *i* indicates country, *j* indicates sector and *t* denote the year. All independent variables are on logs. YEAR indicates a time trend. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table 7. Skills and Age.

Variables	(1) OLS	(2) GMM-SYS	(3) OLS	(4) GMM-SYS
E_{t-1}	-0.180	**-0.441****		
	(0.167)	**(0.222)**		
E_Tedu_{t-1}			-0.015	**0.468****
			(0.091)	**(0.220)**
E_noTedu_{t-1}			-0.123	**-0.886*****
			(0.185)	**(0.323)**
avg_age_{t-1}	-2.429**	-3.526***		
	(1.088)	(1.236)		
$avg\_age\_Tedu_{t-1}$			-0.400	-0.187
			(0.664)	(0.471)
$avg\_age\_nToedu_{t-1}$			-2.133**	-2.880**
			(0.980)	(1.272)
$R\&D_{t-1}$	0.301***	0.305***	0.289***	0.261***
	(0.084)	(0.076)	(0.083)	(0.092)
$stock\_citations_{t-1}$	0.134	0.183**	0.131	0.133
	(0.160)	(0.072)	(0.163)	(0.085)
$open_{t-1}$	-0.587***	-0.920***	-0.552***	-0.878***
	(0.195)	(0.258)	(0.185)	(0.303)
Constant	9.458*	17.213***	9.526*	16.314***
	(4.940)	(5.561)	(5.440)	(5.924)
time effects	YES	YES	YES	YES
fixed effects	YES	YES	YES	YES
Observations	457	457	457	457
number of id	46	46	46	46
R-squared	0.790	-	0.791	-
AR(1) p-value		0.002		0.003
AR(2) p-value		0.546		0.508
Hansen test		0.649		1.828
Hansen test p-value		0.723		0.767

The dependent variable is the log of citations received in the last 4 years. All variables are in logs and are 1-year lagged. In columns (1) and (3) OLS estimators are implemented. In columns (2) and (4) (one-step) robust GMM-SYS estimators are used. All models include time, country and industry dummies. In the GMM models the endogenous variables are *E, E_edu, E_noedu*. All the other variables are considered as exogenous. The collapse option (Roodman, 2009) is implemented for the GMM-style instruments in levels and in differences, only the first two lags of the endogenous variables are used. Standard errors are clustered at the country-industry level, *** p<0.01, ** p<0.05, * p<0.1.

Table 8. Skills, Age and Ethnicity

	(1) OLS	(2) IV -CARD	(3) IV -CARD	(4) GMM-SYS	(5) GMM-SYS
Variables	*all exog*	*E_noedu_im endog*	*E_edu_im endog*	*E_noedu_im & E_edu_im endog*	*all labor endog*
E_Tedu_mig$_{t-1}$	**0.036***	**0.037***	-1.551	**0.067***	**0.089****
	(0.021)	**(0.022)**	(1.362)	**(0.038)**	**(0.045)**
E_noTedu_mig$_{t-1}$	-0.048	**-0.212****	0.081	-0.170	**-0.338***
	(0.052)	**(0.103)**	(0.162)	(0.283)	**(0.201)**
E_Tedu_nat$_{t-1}$	-0.013	0.045	0.173	-0.082	**0.292****
	(0.083)	(0.099)	(0.329)	(0.107)	**(0.141)**
E_noTedu_nat$_{t-1}$	-0.129	-0.100	0.981	0.199	**-0.752****
	(0.195)	(0.186)	(1.044)	(0.270)	**(0.354)**
Other controls	YES	YES	YES	YES	YES
time effects	YES	YES	YES	YES	YES
fixed effects	YES	YES	YES	YES	YES
First stage		*lognosk_im*	*logsk_im*		
E_Tedu_mig_card$_{t-1}$		-	0.085		
		-	(0.076)		
E_noTedu_mig_card$_{t-1}$		**0.857*****	-		
		(0.147)	-		
Angrist-Pischke F test of excl. instr:		33.97	1.25		
p-value		0.000	0.269		
Hausman endog. test p-value		0.142	0.060		
Observations	457	451	448	457	457
Number of id	46	45	45	46	46
R-squared	0.795	-	-	-	-
num. instruments				44	47
AR(1) p-value				0.002	0.001
AR(2) p-value				0.758	0.578
Hansen test p-value				0.170	0.426

The dependent variable is the log of citations received in the last 4 years. All variables are in logs and are 1-year lagged. All models include additional control variables, as displayed in Table (A2). All models include time, country and industry dummies. In column (1) the OLS estimator is implemented. In columns (2) and (3) two-stage least squares estimators are implemented. In the panel below first-stage coefficients are reported. The Angrist-Pischke test of excluded instruments reports the probability that excluded instruments in columns (2) and (3) are weak, the Hausman test reports the probability that the instrumented variables are endogenous. In columns (4) and (5) (one-step) robust GMM-SYS estimators are used. The collapse option (Roodman, 2009) is implemented for the GMM-style instruments in levels and in differences. In column (4) the endogenous variables are *E_Tedu_mig*, *E_noTedu_mig*. On the basis of the results in Table (9) *E_Tedu_mig* is instrumented with one and two year lags, while *E_noTedu_mig* is instrumented with two and three years lags. In column (5) the endogenous variables are *E_Tedu_mig*, *E_noTedu_mig*, *E_Tedu_nat*, *E_noTedu_nat,*. On the basis of the results in Table (9) both *E_Tedu_nat* and *E_noTedu_nat* are instrumented with one year lags, while *E_Tedu_mig*, *E_noTedu_mig* are instrumented with the same lags as in column (4). In columns (4) and (5) all the additional control variables are considered as exogenous. Standard errors are clustered at the country-industry level, *** p<0.01, ** p<0.05, * p<0.1.

Table 9. First-stage on the lag specification

	(1)	(2)	(3)	(4)
		specification in levels		
Regressors	*Tedu_natives*	*Tedu_migrants*	*noTedu_natives*	*noTedu_mig*
Δx_{t-1}	**0.275*** **	**0.170** **	**0.410*** **	0.101
	(0.078)	**(0.082)**	**(0.153)**	(0.069)
Δx_{t-2}	0.033	**0.211*** **	**0.368** **	**0.218*** **
	(0.087)	**(0.072)**	**(0.155)**	**(0.078)**
Δx_{t-3}	0.051	**0.181*** **	0.145	**0.183*** **
	(0.090)	**(0.054)**	(0.147)	**(0.056)**
F-statistics	3.939	5.847	4.171	4.802
Hausman test p-value	0.400	0.846	0.317	0.005
Hansen test p-value	0.466	0.985	0.850	0.100
obs	304	296	304	300

	(1)	(2)	(3)	(4)
		specification in first differences		
Regressors	*Tedu_natives*	*Tedu_migrants*	*noTedu_natives*	*noTedu_mig*
x_{t-2}	**-0.184*** **	0.009	-0.052	0.124
	(0.062)	(0.078)	(0.076)	(0.087)
x_{t-3}	**0.196** **	-0.010	0.087	-0.082
	(0.077)	(0.085)	(0.117)	(0.075)
x_{t-4}	-0.006	-0.048	-0.013	0.047
	(0.061)	(0.0666)	(0.077)	(0.064)
F-statistics	3.54	0.411	2.974	1.628
Hausman test p-value	0.558	0.860	0.088	0.933
Hansen test p-value	0.657	0.422	0.149	0.014
obs	302	295	302	299

The estimates in the upper panel report the results from a First-stage instrumental variable estimation of equation (3) in levels. Each of the columns reports the results of first stage estimates in which only one of the endogenous variables is instrumented with its own lags in differences. The estimates in the lower panel report the results from a First-staqe instrumental variable estimation of equation (3) in differences. Each of the columns reports the results of first stage estimates in which only one of the endogenous variables is instrumented with its own lags in levels. The F-statistics refer to the first-stage estimation. The Hansen test reports a test of over-identifying restrictions on the goodness of the instruments (the null-hypothesis is that instruments are valid). The Hausman test checks for the exogeneity of the instrumented variable in equation (3), the null hypothesis is that the instrumented regressor is exogenous. Standard errors are clustered at the country-industry level, *** p<0.01, ** p<0.05, * p<0.1.

Table 10. Heterogeneous effects by country of origin and type of sector

VARIABLES	(1) OLS	(2) GMM-SYS	(3) OLS	(4) GMM-SYS	(5) OLS	(6) GMM-SYS
	by country of origin		*by sector of activity*			
E_Tedu_im EU$_{t-1}$	0.018	**0.036*** **				
	(0.011)	**(0.010)**				
E_Tedu_im NOEU$_{t-1}$	0.004	**0.014***				
	(0.007)	**(0.008)**				
E_noTedu_im EU$_{t-1}$	-0.013	0.004				
	(0.018)	(0.028)				
E_noTedu_im NOEU$_{t-1}$	**-0.039****	**-0.067****				
	(0.017)	**(0.028)**				
*E_Tedu_im$_{t-1}$*hitech sectors*			0.031	**0.128****		
			(0.034)	**(0.059)**		
*E_Tedu_im$_{t-1}$*lowtech sectors*			0.041	0.039		
			(0.027)	(0.067)		
*E_noTedu_im$_{t-1}$*hitech sectors*					-0.037	-0.099
					(0.055)	(0.113)
*E_noTedu_im$_{t-1}$*lowtech sectors*					-0.064	**-0.478****
					(0.082)	**(0.235)**
Other labor variables	YES	YES	YES	YES	YES	YES
Other controls	YES	YES	YES	YES	YES	YES
time effects	YES	YES	YES	YES	YES	YES
fixed effects	YES	YES	YES	YES	YES	YES
Observations	457	457	457	457	457	457
Number of id	46	46	46	46	46	46
R-squared	0.798	-	0.795	-	0.795	-
Num. instruments		54		51		51
AR(1) p-value		0.003		0.002		0.002
AR(2) p-value		0.897		0.828		0.625
Hansen test		4.771		7.020		7.164
Hansen test p-value		0.965		0.724		0.710

The dependent variable is the log of citations received in the last 4 years. All variables are in logs and are 1-year lagged. All models include additional control variables, as displayed in Table (A3). All models include time, country and industry dummies. In columns (1), (3) and (5) OLS estimators are implemented, while in columns (2), (4) and (6) (one-step) robust GMM-SYS estimators are used. In the GMM-SYS estimates all labor variables are considered as endogenous, while the control variables (R&D, openness to trade and the stock of citations) are considered as exogenous. The collapse option (Roodman, 2009) is implemented for the GMM-style instruments in levels and in differences, only the first two lags of the endogenous variables are used. Standard errors are clustered at the country-industry level, *** p<0.01, ** p<0.05, * p<0.1.

Appendix A

Table A1. Definition of high tech and low tech sectors

Low tech	
15-16	Food products, beverages and tobacco
17-19	Textiles, textile products, leather and footwear
20	Wood and products of wood and cork
21	Paper and paper products
25	Rubber and plastics products
26	Other non-metallic mineral products
27	Basic metals
28	Fabricated metal products, except machinery and equipment
High tech	
24	Chemicals and chemical products
29	Machinery and equipment, nec
30	Office, accounting and computing machinery
31	Electrical machinery and apparatus
32	Radio, television and communication
33	Medical, precision and optical instruments
34	Motor vehicles, trailers and semi-trailers
35	Other transport equipment

Table A2. Full coefficients of the regressions in Table (7)

Variables	(1) OLS *all exog*	(2) IV -CARD *E_noedu_im endog*	(3) IV -CARD *E_edu_im endog*	(4) GMM-SYS *E_noedu_im & E_edu_im endog*	(5) GMM-SYS *all labor endog*
E_Tedu_mig$_{t-1}$	**0.036***	**0.037***	-1.551	**0.067***	**0.089****
	(0.021)	**(0.022)**	(1.362)	**(0.038)**	**(0.045)**
E_noTedu_mig$_{t-1}$	-0.048	**-0.212****	0.081	-0.170	**-0.338***
	(0.052)	**(0.103)**	(0.162)	(0.283)	**(0.201)**
E_Tedu_nat$_{t-1}$	-0.013	0.045	0.173	-0.082	**0.292****
	(0.083)	(0.099)	(0.329)	(0.107)	**(0.141)**
E_noTedu_nat$_{t-1}$	-0.129	-0.100	0.981	0.199	**-0.752****
	(0.195)	(0.186)	(1.044)	(0.270)	**(0.354)**
avg_age_Tedu_mig$_{t-1}$	0.021	-0.010	-0.381	-0.044	-0.101
	(0.113)	(0.106)	(0.513)	(0.187)	(0.187)
avg_age_noTedu_mig$_{t-1}$	0.283	0.402	-0.359	0.235	0.315
	(0.198)	(0.286)	(0.878)	(0.162)	(0.214)
avg_age_Tedu_nat$_{t-1}$	-0.341	-0.421	-2.235	0.126	-0.292
	(0.639)	(0.647)	(2.196)	(0.424)	(0.444)
avg_age_noTedu_nat$_{t-1}$	-2.502**	-2.348**	-1.732	-2.011**	-4.170**
	(1.016)	(1.059)	(3.453)	(0.851)	(1.652)
R&D$_{t-1}$	0.279***	0.291***	0.907	0.234***	0.312***
	(0.081)	(0.082)	(0.644)	(0.064)	(0.105)
stock_citations$_{t-1}$	0.118	0.089	-0.287	0.275***	0.160
	(0.170)	(0.175)	(0.433)	(0.076)	(0.098)
open$_{t-1}$	-0.526**	-0.576***	-0.247	-0.607***	-1.017***
	(0.197)	(0.199)	(0.425)	(0.217)	(0.304)
time effects	YES	YES	YES	YES	YES
fixed effects	YES	YES	YES	YES	YES
First stage		*lognosk_im*	*logsk_im*		
E_Tedu_mig_card$_{t-1}$		-	0.085		
		-	(0.076)		
E_noTedu_mig_card$_{t-1}$		0.857***	-		
		(0.147)	-		
Angrist-Pischke F test of excl. instr:		33.97	1.25		
p-value		0.000	0.269		
Hausman endog. test p-value		0.142	0.060		
Observations	457	451	448	457	457
Number of id	46	45	45	46	46
R-squared	0.795	-	-		

num. instruments	44	47
AR(1) p-value	0.002	0.001
AR(2) p-value	0.758	0.578
Hansen test p-value	0.170	0.426
E_Tedu_mig $_{t-1}$, *lags used (1, 2)*		
Hansen test excluding group: chi2	0.793	0.306
Difference (null H = exogenous): chi2	0.096	0.530
E_noTedu_mig $_{t-1}$, *lags used (2, 3)*		
Hansen test excluding group: chi2	0.166	0.534
Difference (null H = exogenous): chi2	0.213	0.274
E_Tedu_nat $_{t-1}$, *lags used (1, 1)*		
Hansen test excluding group: chi2		0.296
Difference (null H = exogenous): chi2		0.629
E_noTedu_nat $_{t-1}$, *lags used (1, 1)*		
Hansen test excluding group: chi2		0.516
Difference (null H = exogenous): chi2		0.287

The dependent variable is the log of citations received in the last 4 years. All variables are in logs and are 1-year lagged. All models include additional control variables, as displayed in Table (A4). All models include time, country and industry dummies. In column (1) the OLS estimator is implemented. In columns (2) and (3) two-stage least squares estimators are implemented. In the panel below first-stage coefficients are reported. The Angrist-Pischke test of excluded instruments reports the probability that excluded instruments in columns (2) and (3) are weak, the Hausman test reports the probability that the instrumented variables are endogenous. In columns (4) and (5) (one-step) robust GMM-SYS estimators are used. The collapse option (Roodman, 2009) is implemented for the GMM-style instruments in levels and in differences. In column (4) the endogenous variables are *E_Tedu_mig*, *E_noTedu_mig*. On the basis of the results in Table (10) *E_Tedu_mig* is instrumented with one and two year lags, while *E_noTedu_mig* is instrumented with two and three years lags. In column (5) the endogenous variables are *E_Tedu_mig*, *E_noTedu_mig*, *E_Tedu_nat*, *E_noTedu_nat,*. On the basis of the results in Table (10) both *E_Tedu_nat* and *E_noTedu_nat* are instrumented with one year lags, while *E_Tedu_mig*, *E_noTedu_mig* are instrumented with the same lags as in column (4). In columns (4) and (5) all the additional control variables are considered as exogenous. Standard errors are clustered at the country-industry level, *** p<0.01, ** p<0.05, * p<0.1.

Table A3. Full coefficients of the regressions in Table (9)

VARIABLES	(1) OLS	(2) GMM-SYS	(3) OLS	(4) GMM-SYS	(5) OLS	(6) GMM-SYS
E_Tedu_im EUt-1	0.018	**0.036***				
	(0.011)	**(0.010)**				
E_Tedu_im NOEUt-1	0.004	**0.014***				
	(0.007)	**(0.008)**				
E_noTedu_im EUt-1	-0.013	0.004				
	(0.018)	(0.028)				
E_noTedu_im NOEUt-1	**-0.039***	**-0.067***				
	(0.017)	**(0.028)**				
*E_Tedu_imt-1*hitech sectors*			0.031	**0.128***		
			(0.034)	**(0.059)**		
*E_Tedu_imt-1*lowtech sectors*			0.041	0.039		
			(0.027)	(0.067)		
*E_noTedu_imt-1*hitech sectors*					-0.037	-0.099
					(0.055)	(0.113)
*E_noTedu_imt-1*lowtech sectors*					-0.064	**-0.478***
					(0.082)	**(0.235)**
$E\_Tedu\_nat_{t-1}$	-0.009	0.169	-0.012	0.248	-0.008	0.202
	(0.091)	(0.140)	(0.084)	(0.152)	(0.080)	(0.197)
$E\_noTedu\_nat_{t-1}$	-0.108	-0.439*	-0.131	-0.727**	-0.121	-0.863***
	(0.203)	(0.231)	(0.198)	(0.320)	(0.206)	(0.294)
$E\_Tedu\_mig_{t-1}$					0.034*	0.054
					(0.020)	(0.048)
$E\_noTedu\_mig_{t-1}$			-0.048	-0.222**		
			(0.052)	(0.099)		
$avg\_age\_Tedu\_nat_{t-1}$	-0.324	-0.051	-0.341	-0.176	-0.344	0.084
	(0.631)	(0.482)	(0.639)	(0.448)	(0.639)	(0.573)
$avg\_age\_noTedu\_nat_{t-1}$	-2.585***	-3.354***	-2.506**	-3.701***	-2.499**	-3.143*
	(0.950)	(1.116)	(1.022)	(1.409)	(1.020)	(1.733)
$avg\_age\_Tedu\_mig_{t-1}$	0.270	0.145	0.285	0.253	0.282	0.257
	(0.195)	(0.168)	(0.199)	(0.194)	(0.201)	(0.229)
$avg\_age\_noTedu\_mig_{t-1}$	0.013	-0.039	0.022	-0.071	0.016	-0.110
	(0.110)	(0.146)	(0.117)	(0.162)	(0.117)	(0.193)
$R\&D_{t-1}$	0.261***	0.234***	0.280***	0.278***	0.278***	0.303**
	(0.080)	(0.072)	(0.082)	(0.097)	(0.082)	(0.118)
$stock\_citations_{t-1}$	0.123	0.199**	0.120	0.166*	0.115	0.165
	(0.165)	(0.078)	(0.169)	(0.088)	(0.171)	(0.108)
$open_{t-1}$	-0.576***	-0.768***	-0.524**	-0.984***	-0.526**	-1.132***
	(0.203)	(0.241)	(0.196)	(0.299)	(0.197)	(0.321)
Constant	10.555**	15.454***	10.082*	16.660**	10.047*	14.792*
	(5.153)	(5.119)	(5.162)	(6.752)	(5.229)	(8.486)
time effects	YES	YES	YES	YES	YES	YES
fixed effects	YES	YES	YES	YES	YES	YES
Observations	457	457	457	457	457	457

	(1)	(2)	(3)	(4)	(5)	(6)
Number of id	46	46	46	46	46	46
R-squared	0.798		0.795		0.795	
num instruments		54		51		51
AR(1) p-value		0.003		0.002		0.002
AR(2) p-value		0.897		0.828		0.625
Hansen test		4.771		7.020		7.164
Hansen test p-value		0.965		0.724		0.710

The dependent variable is the log of citations received in the last 4 years. All variables are in logs and are 1-year lagged. All models include additional control variables, as displayed in Table (9). All models include time, country and industry dummies. In columns (1), (3) and (5) OLS estimators are implemented, while in columns (2), (4) and (6) (one-step) robust GMM-SYS estimators are used. In the GMM-SYS estimates all labor variables are considered as endogenous, while the control variables (R&D, openness to trade and the stock of citations) are considered as exogenous. The collapse option (Roodman, 2009) is implemented for the GMM-style instruments in levels and in differences, only the first two lags of the endogenous variables are used. Standard errors are clustered at the country-industry level, *** p<0.01, ** p<0.05, * p<0.1.

Appendix B - Data description

Patents data come from the PATSTAT-KITES database

PATSTAT (EPO Worldwide PATent STATistical Database) is a patent database, run by the European Patent Office (EPO) developed in cooperation with the World Intellectual Property Organisation (WIPO), the OECD and Eurostat. PATSTAT provides raw patent data coming from around 90 patent offices worldwide, including, of course, the most important and largest ones such as the European Patent Office (EPO) and the United States Patent and Trademark Office (USPTO). The data set includes the full set of bibliographic variables concerning each patent application. PATSTAT is provided in a raw format. Data coming from PATSTAT has, therefore, been thoroughly elaborated by KITES (Bocconi University: http://db.kites.unibocconi.it/) to produce a clean and harmonized database. Data processing consisted mainly in a thorough work of cleaning and standardization rough information provided by the EPO. The aggregation of patent technological classifications (so called IPC classes) into NACE Rev. 1 fields follows Schmoch *et al.* (2003)[25]

UK Labour Force Survey

The British Quarterly Labour Force Survey (QLFS) is a quarterly sample survey of households living at private addresses in Great Britain. The QLFS is conducted on a quarterly basis and aims to obtain a sample of around 60,000 households every quarter. Since 1992 respondents are interviewed in five successive waves, thus approximately a fifth of the sample in each quarter will contain individuals from each of the five waves. Every quarter one wave of approximately 12,000 leaves the survey and a new wave enters. The rotational element to the QLFS creates an 80 percent overlap between quarters and thus 20 percent of the sample enter and exit the survey each quarter.

The survey contains data on among other variables: employment and self-employment; full-time and part-time employment; second jobs; average age; economic activity; occupations and industry sectors and education.

French Labour Force Survey

The French Labour Force Survey was launched in 1950 and applied in 1982 as an annual survey. Redesigned in 2003, the survey is a continuous survey providing quarterly results. The survey covers private households in metropolitan France. It includes a part of the population living in collective households, persons who have family ties with private households. Participation in the survey is compulsory. The resident population comprises persons living in the French metropolitan territory.

The household concept used is that of the ‘dwelling household’: a household means all persons living in the same dwelling. It may consist of a single person or of two families living under the same roof.

The survey provides longitudinal data on households and individuals. Persons average aged fifteen years or over are interviewed. Data refer to the number of persons who were working during the survey week including employees, self-employed as well as family workers. Data include persons who have a job but are not at work due to illness (less than one year), vacation, labour dispute, educational leave, etc.

[25] ftp://ftp.cordis.europa.eu/pub/indicators/docs/ind_report_isi_ost_spru.pdf

German Microcensus

The Microcensus provides official statistics for the population and the labour market in Germany. The Labour Force Survey of the European Union (EU Labour Force Survey) forms an integral part of the Microcensus. The Microcensus supplies statistical information in a detailed subject-related and regional breakdown on the population structure, the economic and social situation of the population, families, consensual unions and households, on employment, job search, education/training and continuing education/training, the housing situation and health. The German Microcensus includes 1% of the resident population in the former West Germany, and is a large, representative, random sample containing comprehensive information on individual and household characteristics.