Measuring Researcher Mobility

A Comparison of Different Datasets and Methods with an Empirical Application of Micro-Data for the United States and Germany

Daniel Schiller und Alexander Cordes

Paper to be presented in the OECD Blue Sky Forum 2016

August 2016

Corresponding author:

Prof. Dr. Daniel Schiller

Ernst Moritz Arndt University Greifswald

Institute of Geography and Geology

Makarenkostraße 22

17487 Greifswald, Germany

Tel. +49 3834 86-4524

Fax. +49 3834 86-4481

E-Mail: daniel.schiller@uni-greifswald.de

Abstract

Researcher mobility has received increasing attention with regard to theoretical and methodological issues. While significant progress has been made in both spheres, a unified methodological approach is still to be developed. However, such a widely accepted and applicable methodology would be the prerequisite for empirical results on researcher mobility that can be used to inform policy makers. The aim of this contribution is twofold. First, an overview of the pros and cons of different methods to measuring research mobility is provided. Second, microdata are used as an empirical showcase for yielding evidence for the largest economy of the world and the main hub for research mobility (United States) and the largest economy in the European Union (Germany) from existing datasets.

0 Introduction

New knowledge and technologies are not the result of an anonymous production process. They are rather the outcome of individual scientific and inventive personalities. These persons possess extraordinary knowledge within their field of research, which they have gained over a long time span and which can only be transferred to others at high costs. These key figures act as knowledge-spillover-agents, because they contribute to diffuse implied knowledge and transfer scientific knowledge into new technologies due to their spatial mobility and presence at specific locations (Schiller and Revilla Diez 2010).

Highly qualified people are the most mobile population group worldwide. In Western Europe seven percent of this population group have migrated to another OECD-country, whereas other qualification groups only make up two percent. Simultaneously a third of immigrants in OECD-countries are highly qualified, whereas their stake in the working population accounts to only ten percent (Docquier and Marfouk 2004). Scientist within this group are very mobile. Ioannidis (2004) has shown that a third of top-level researchers in the USA were born abroad and in other European countries similar values are achieved (Germany: 19%, UK: 24%). Yet, only the USA have obtained a net surplus of most-cited researches, whilst Germany on basis of comparison of birth- and working place, has registered a net loss of 28% (Schiller and Revilla Diez 2008). Regarding highly qualified (with third-level education) on a whole, Germany is a net importer (OECD 2008).

The first part of the paper summarizes results from a feasibility study on different data sources and methods for measuring researcher mobility. Existing datasets and methods to generate new datasets are compared based on a set of criteria. The datasets are analysed with regard to the possibility to gather the following kinds of information: identification of researchers, coverage of immigration, emigration and remigration of researchers (number and lengths of stays), reproduction of individual careers, identification of workplaces, research performance, individual research networks, motives and barriers for researcher mobility, individual characteristics (age, qualification, subject area, social and family background).

The feasibility study comprises the following datasets and methods: microdata based on labour force surveys and other labour market related datasets, publication and patent data, national researcher registers and CV databases, graduate students and PhD holder panels, immigration data, professional social networks, and use of national language processing (NLP) for the generation of datasets on researcher mobility.

When taking all criteria into account, the use of microdata from labour force surveys and big data applications which gather publicly available information in an automated way via NLP seems to be the most promising methods for future studies on researcher mobility. It is shown that assessments based on the isolated use of publication or patent data for measuring researcher mobility are to a certain degree inadequate because of a lack of information on the nationality of researchers and their career path prior to the first publication or patent. However, publication and patent records can be combined with other CV information generated via NLP applications mentioned above.

Based on the feasibility study, the second part of the paper puts one of the promising methods into practice. Microdata from labour force surveys in the United States and Germany for different years are pooled. The dataset is prepared for a comparative and longitudinal description of researcher mobility patterns. In particular, problems related to different classifications of occupation are solved and researchers are identified in a similar way in both countries.

The results include a measurement of the share of (mobile) researchers among the total labour force and the share of researchers among the mobile population in both countries. For both indicators, the share of mobile researchers is much higher in the United States than in Germany. While mobility results on an increasing share of researchers in the United States, the qualification structure of German immigrants is lower than in the resident population.

However, Germany has caught up during the last years if data for different time periods are compared. Reasons for that might include changes in immigration laws (more restrictive in the United States, less restrictive in Germany) and the affectedness of the two countries by the global economic and financial crisis. Further differentiations within the data set are provided for regional distribution within the countries, occupations, years of immigration, and estimated retention rates. Finally, the potential for applying this method to other countries is discussed.

The findings of the paper are relevant for policy makers in OECD countries in at least three ways. First, the general overview of different methods for measuring researcher mobility can be used as a starting point for further development of the most promising datasets and methods. It has the potential to assist science and innovation policy in the preparation of pilot applications. Second, experiences from the pilot application to microdata from the United States and Germany can be used for implementing similar studies in other OECD countries. Third, comparative empirical evidence on the main hub for researcher mobility in the world (United States) and the largest economy in Europe (Germany) is relevant for designing policies to attract mobile researchers.

1 Comparison of datasets and methods

2.1 Properties of a reporting system

2.1.1 Definition of researchers and mobility

At the beginning of the research of researcher mobility, the terms researcher and mobility have to be defined. The Frascati-Manual offers an internationally recognised definition for researchers, which can be operationalised with the International Standard Classification of Occupations (ISCO). Within the feasibility study only international mobility was investigated. Furthermore it is recommended not to take mobility of under six months into account and to focus the research on short-term (six months to one year), mid-term (one to three years) and long-term mobility (more than three years).

2.1.2 Relevant contents

To give a complete answer to the content issues mentioned in the introduction, it would be necessary to collect a variety of personalised data. However, it is not to be expected to find a data source or dataset which covers all the needed information to the same extent. The following overview should be seen as a wish list, which will be fulfilled by the considered methods more or less to the same extent:

- Identifying researchers and mobility by the means of the above mentioned definitions
- Outline of individual careers
- Type, amount and time frames of mobility
- Work spaces (spatial and institutional)
- Individual research performance and networking
- Research performance: publications, patents, acquiring of third party funds, etc.
- Networking: co-publications, co-patents, membership in research associations, etc.
- Research performance and networking of the home and hosting institution of the mobile researcher
- Motives and barriers of mobility
- Further structural, personal characteristics
- Age, qualification within field of research, sex, social and family background, migrant background, etc.

To show the migration balance (how many researchers immigrate, emigrate, return?), the first part would be sufficient. For the analysis of migration patterns, further information to career paths is necessary. If interdependences of migration and research performances (individual and/or institutional) should be investigated, more information to these aspects are necessary.

2.1.3 Methodological requirements

Studies in the past have often used surveys to gain answers to specific questions about mobility behaviour. For this purpose researchers were identified through databases (e.g. University lecturer lists or researchers funded within mobility programmes) and ask to participate in the survey. The great advantage of this method is, that the information needed for the above mentioned questions can be gained easily. Especially qualitative assessments to motives and barriers can be gathered through this method. Along with the advantages, this method also faces disadvantages referring to methodological aspects and the required resource input. Worth mentioning here are arguments which counter a follow up of surveys within the feasibility study:

- Relatively high resource input for implementing the survey or very low response rare. In 2006 the RESCAR-Project of the EU reached a response rate of only 13%. The MORE-Project registered an even lower response rate of merely 11%.
- Low response rate query the representativeness of the method. It can be assumed, that surveys about mobility are more frequently answered by mobility affine researchers, which leads to overestimation of mobility.
- Due to the necessary limitation of the population to a specific researcher population, a simultaneous registration of emigrated, immigrated and remigrated researchers is not possible

In the past, researchers which have participated in funding programmes for international mobility, such as DAAD (DAAD and HIS 2012), Alexander von Humboldt-Foundation, or members of networks supporting returns, such as GAIN (Engin and Reifenberg 2012), have been frequently questioned. An advantage to this approach, that German researches abroad can be reached. To gain statements on mobile researchers through funded mobile researches lead to two problematic reasons. On the one hand, it is unclear how high the share of funded mobility is within the researcher mobility as a whole. On the other hand, funded mobility does not differ from non-funded mobility. The former is mostly fixed within the time frame of the funding. Furthermore the aim to return is a condition fixed by the funding programme. The GAIN-network aims at researchers willing to return to Germany, making this an unsuitable database for statements of the return rates of all German researchers

Because of content-related distortions, methodological problems and great resource input, the mentioned questioning method for a reporting system of researchers can be regarded as a wrong approach. The methodological demand of this feasibility study is to consider methods which provide an automatic or automated data supply or use already existing personalised micro data of mobility as well as research and innovation performance. The requirements for the suitability of such data sources or datasets must provide a precise definition of the research object "researcher", information of the above mentioned topics, a great amount of case number for analysing different features and must be intertemporally and interregionally comparable.

2.2 Assessment of German data sources

Detailed information on the assessment of German data sources can be found here (in German):

- Feasibility study on measuring researcher mobility:

  http://www.e-fi.de/fileadmin/Innovationsstudien_2014/StuDIS_9_2014.pdf

- Analysis of microdata for Germany and the United States:

  http://www.e-fi.de/fileadmin/Innovationsstudien_2014/StuDIS_10_2014.pdf

The applied data sources, which potentially contain information about researcher mobility, include:

- Regular collected statistics containing individual information (micro census, socio-economic panel (SOEP))
- Process data from employment administration (*integrated employment biographies* (IEB) of the federal labour office)
- Login data of the immigration offices *(central register of foreigners)*
- Panel studies of career development of university alumni and doctoral candidates (e.g alumni register of HIS, ProFile-doctoral candidate panel of iFQ) and
- The recent efforts of the science council implementing a data set "research", which should lead to exchange of available data in a standardized format

The discussion of German datasets and sources regarding researcher mobility has shown that there is information regarding partial aspects of the phenomenon, but that there is no dataset which collects all information needed for the questions mentioned in the introduction. One major problem lays within the restriction of investigating only German datasets, which only partly illustrates an international phenomenon. Emigrating Germans are not represented to satisfaction in any of the datasets. Of the German datasets, the dataset of the *Mikrozensus* is the best for the researcher mobility. As a representative survey of the German population, it fulfills most likely the criteria of a non-distorted data acquisition of the phenomenon for all employment fields and groups. Furthermore, by using the *Mikrozensus* an international comparison with methodological similar surveys, e.g. employees data acquisition of Germany researcher mobility in the main target country, is possible.

2.3 Assessment of international data sources

International data sets regarding mobility of German researchers are very suitable to investigate outflow in Germany and to set up international comparisons. Most promising data sets are (i) the use of publication databases, (ii) international comparisons of national census data and labour data, as well as (iii) automated processes generating career paths of individual researchers based on information freely available on the internet and techniques of National Language Processing. The other examined international datasets neither offer sufficient information on the topic, nor a large enough number of cases (international researcher data bases), nor do they correspond to the exact definition of researchers (international alumni and doctoral candidate survey).

2.3.1 Publication data

The affiliation of authors provided by publication data can be used for analyzing researcher mobility. If an author has published two papers under different affiliation, it can be assumed that he has moved from one institution to another. Starting point of such an analysis is a base population of researchers, i.e. all researchers that were mentioned in a given time period in a certain publication database. Publications are merged for each author and the mobility path is reconstructed based on the affiliations provided. Based on this information, the following questions can be answered:

- Outmigration: How many researchers, who have published a paper in country x in the base year/period, have left this country later on? How big is there share among all researchers, who have published papers in country x in the base year/period?
- Inmigration: In which countries have researchers, who have published a paper in country x in the base year/period, been active previously? How big is there share among all researchers, who have published papers in country x in the base year/period?
- Remigration: How many researchers have published papers in country x in the base year/period and in the most recent year/period, but have publications in at least one other country in the meantime?

An analysis of this kind has been published by Moed et al. (2013). In Germany, the Fraunhofer Institute for Systems and Innovation Research (ISI) is working on similar analysis within the framework of the Competence Centre for Bibliometrics[1]. Both, groups of researchers are using the SciVerse Scopus database. It offers a better aggregation of publications for individual authors and includes more publications than the Web of Science database. Thus, more information on affiliations can be retrieved.

Using publication data faces some important limitations. A main problem is related to the fact that the first location of the researchers is often mistakenly interpreted as his home country. It is, however, only the country in which the respective researcher has published his first paper. The resulting distortion of migration flows can be illustrated by a publication analysis carried out by OECD (2013). In this report, the migration balance of the United States with China is negative. This result can be explained by the high number of Chinese PhD students in the United States who publish their first paper in the US and move to China later on. By the publication method, these researchers are counted as US researchers moving to China. The potential of the publication method could be improved, if it was combined with the extraction of information from CVs, e.g. about the birthplace or institutions at which education was received.

Additional limitations arise from different publication cultures among subjects. In particular, the information density is highly reduced for researcher in humanities. The same holds true for

[1] http://www.forschungsinfo.de/Bibliometrie/en/index.php?id=home

researchers in the private sector who do not publish results of their work in scientific journals on a regular basis. The use of affiliation also does not provide information about the length of stay. Visiting fellowships cannot be distinguished from more permanent forms of mobility. In addition, due to delays in the process of publication (e.g. due to lengthy review processes or a backlog of accepted papers) the order of affiliations that becomes visible in publication data does not necessarily correspond with the real migration path. Thus, the number of remigrants could be overestimated.

Despite these shortcomings, the publication method is in principle a very promising way to extract information about the location of scientists over a long period over time. There is a lot of potential to combine the method with the automated reconstruction of careers paths from other sources, e.g. CVs, see chapter 2.4.

2.3.2 Census data

Recourse on census data and employment surveys make it possible to internationally compare and to implement an in-depth analysis of the extent and structure of researcher mobility. Hereby German micro census data can be combined with the relevant surveys of other countries. The main focus of the survey should be the main target countries of German researcher mobility, which samples are large enough for analysis of sub-populations. The data of the American Community Survey (ACS) and in parts the employment surveys of the UK and France are suitable. The required analyses can be implemented on short-term notice and, due to the annual performance of the surveys, offer the best potential to be the basis of a report system.

An application of census data to measuring researcher mobility in Germany and the United States will be discussed in further detail in chapter 3 and 4.

2.4 Case Study: Natural Language Processing (NLP)

Natural Language Processing (NLP) is the attempt to gain, save and process relevant information of (un-) structured texts. Websites, online data bases and PDF-Documents represent the basis of this procedure. This procedure is applicable on analyses of researcher mobility, because a vast amount of information of the person concerned is freely accessible online. The individual research performances within publically funded research are made public rather often. Personal information about researchers and their abode can be found on institute websites, personal homepages, open source CVs, public profiles on social networks, biographies in publications (especially open source papers) and research projects, media reports as well as several electronic documents online. Information about researchers in private industry is more restricted, due to the frequent secrecy of their work. However, it is assumed, that information can be made available.

To access the problems of researcher mobility NLP-Methods can be used to gain personal information on work locations in the course of a career. The acquired data can be applied for the analysis of mobility behaviour. Hence this process can be used to apply the CV-analysis, which

has gained on importance within the last years (Cañibano und Bozeman 2009), on a more comprehensive data base.
The challenge of such research is the implementation of a largely automated generation of such data sets. Following steps are advisable:

Step 1: Establishment of a reference data base

At first it has to be determined of what kind of people mobility tracks should be collected. This first step is of great importance, because it is here where the basic population for the research is fixed. German sources can be used to investigate the mobility behaviour of researches living in Germany. To gain an insight into the international status quo, it is advisable to call on international data sources, too (e.g. researcher register of other countries or subject-specific international associations).

To limit the research on to certain science field is sensible, due to the use of publication data bases, such as Web of Science or Scopus, to identify researches of a specific field. Further advantages of limiting the research to subjects, is that one gains a complete overview of subpopulations of researchers at home and abroad and minimizes the problem of identifying researchers with the same name (homonyms). For narrowing the research down to research fields, subjects with publications listed on Web of Science and Scopus represent a good match, which excludes humanities based on specific technologies non-interdisciplinary humanities. At last, the decision which reference data base to use, is made through the availability of data bases with names of researches referring to the research issue.

The consideration of names for the choice of the reference data base is not sufficient, due to the lack of differentiation of homonyms. Ideally, additional information such as current work place, selected prior work placements / employers or publications should be known.

Step 2: collection of electronic documents of persons in the reference data base

In this step, freely available online documents and further information from specific data bases of the people listed in the reference data base is collected. It is not necessary to limit the search to specific documents. The search of documents is conducted with a purpose-built Webcrawler, which is attached to the Google search and performs a quest with the Brute-Force-Method, meaning different search combination are used and each time a specific percentage of the most relevant data of the Google search of the relevant person are set aside. Due to the intensive mathematics the search takes longer, but offers a more extensive and complete overview. This method guarantees recording all of the information available to the public, including public profiles on social networks (e.g. those mentioned in chapter 4.1.4, like LinkedIn and ResearchGate).

Additionally, a search in secured data bases (e.g Scopus) can be conducted. GoogleScholar is not suitable, because it prevents a search function with Webcrawlers. One is conscious of the value of the collected data in GoogleScholar and not inclined of cooperating and interested in gaining data of other researchers. For the research which is in need of the researcher names, a cooperation with social networks is not realistic due to privacy settings. A limitation to public accessible information is necessary. Recently Facebook has interpreted its data using methods of network analysis, in order to evaluate how many contacts are necessary to get in touch with every Facebook-user (Ugander et al. 2011, Backstrom et al. 2012). With a result of 4,74 degrees of separation, the information was only personal interest of the company.

Step 3: Identifying mobility traces in the documents

This step, representing the core of the NLP-Method, focuses on identifying mobility traces, meaning passages relating to information regarding mobility, and to separate this information from other. There are several methods to identify the required information (towns, countries, organisations, time). Already during step 2 it is possible to reduce the text volume by making out specific expressions and syntax and saving that information. Thus, a prior reduction of the text volume accelerates the semantic classification.

The saved text volume is analysed with suitable programmes (e.g Python Packet Natural Language Toolkit NLTK), classifying sentences grammatically and syntactically. First a simplification and normalisation is undertaken, e.g. replacing plural with singular (so called Stemming) and deleting articles and expletives (applying so called stopwords). The simplified data set is classified by grammar rules. Existing dictionaries to filter time and location information are used to extract mobility related information, saving it under the researcher's name, so to process it further during the following step.

Step 4: Evaluation of mobility traces

This step is the technically most demanding one: the semantic analysis based on classified mobility related information, which grades and associates the information automatically. A possibility to analyse CVs, other relatively systematic information and to link publication dates with personal location information is to conduct a search for time information laying close to location information. The programme provides a scoring probability which can be used on automated decision rules.

How to adjust the semantic analysis, will reveal itself during the conduct of the research. It is necessary to train the programme with trial data sets. First the analysis is applied on categories of persons with known information, e.g. by being familiar with the résumé. The algorithms will be improved until all necessary information can be gathered automatically with a high reliability.

The feasibility of this approach was recently conducted on test data sets by Uruguayan researchers (García Flores et al. 2012). A reliability of 80% was achieved for mobility relevant information. The reliability is depends heavily on the nature of the data source. Résumés or public LinkedIn profiles can be analysed with a scoring probability of 100%; other documents such as continuous texts on institute website or media reports on individual researcher can decrease the quota slightly. At this point rules can be applied, which statements of location at a certain time are more important than others.

Step 5: Creation of career paths of persons in the reference data base

In this last step, the career paths of persons in the reference data base is created using the mobility related information gained. The aim is to receive a full résumé including location and time slots from studies to recent activity. This database can be analysed by statistical methods.

Final assessment

On condition, that the portrayed method produces reliable results, it offers the great chance to implement a new innovative open access database. García Flores et al. (2012) have proven the feasibility of the procedure. An application on larger researcher population would a first. An automated implementation of a database on careers of German researchers is desirable since the accessibility of such information is hardly given. A database based on the NLP-Method can not only be used for the research on researcher mobility, but also on different research concerning scientific careers. Because of identifying persons and categorising personal career paths, a link to information from other sources is possible.

Due to the innovative character of the method, there are risks concerning the project. It is important, that the crawling process, meaning automatic screening of information online, is not disturbed during data collection. GoogleScholar prohibits such screening in general. Other databases require either access authorization (e.g. Web of Science) or an Offline screening (e.g. Scopus). Coding of the algorithms for the analysis present another challenge. Despite heterogenic sources and partially unstructured syntax, it is necessary to filter the necessary information accurately. Coding problems and insufficient computing capacity can occur, because the methods can lead to long computing time, when large text volume is involved. These problems cannot be anticipated, they occur during the research process, e.g. during trial of the programmes with trial datasets. In the field of informatics, similar applications have been implemented successfully in the last years. Hence case-related and creative solutions for this research method should be manageable, if needed.

2 Micro data analysis for Germany and the United States

3.1 Introduction

Following the preceding findings micro data analysis of census data is supposed to be one adequate method to depict researcher mobility. Moreover, we employ micro data of two countries, the US and Germany for different reasons. First of all, we are challenged to deal with different classifications of economic activity and occupations and to define a common base. Secondly, both countries differ in the institutional settings of their research sectors as well as their general immigration conditions. These basic differences increase the interest in their relative attraction potential. Finally, the geographical location of the US on the one hand and Germany on the other hand each is predetermined to attract people from different parts of the world and we are able to trace favorite destinations of mobile researchers.

This chapter first describes the data used, the German Microcensus (*Mikrozensus, MZ*) and the American Community Survey (*ACS*). What follows is our suggestion of a unifying operationalisation of sectors and occupations. Then we conduct empirical analyses for each of both countries separately, leading us to some concluding remarks about the empirical results as well as a summary of our attempts to provide a methodogical example for micro data analysis of researcher mobility by using census information.

3.2 Data

The data used in this study is publicy available. Usage of the German Microcensus (*Mikrozensus*) requires a scientific purpose, a formal obligation based on the Federal Statistics Law, and a small fee for every survey year employed. Access to the American Community Survey is provided via the IPUMS-USA database[2] (Ruggles et al. 2010) which allows an individual composition of the data needed and its instant download.

3.2.1 German Microcensus

The *Mikrozensus* is a survey of approximately one percent of the total German population. Participation is obliged by law. Several individual and household related characteristics are asked, i. e. the family context, income, education, training and labor market participation. Data access is generally provided via remote data access at the research data centre (*Forschungsdatenzentrum*) of the Federal Statistical Office (*Statistisches Bundesamt*) as well as via the usage of scientific use files – the latter being also used in this study. Our analysis covers the years 2006 to 2010 but descriptive results are mostly reported for 2010.

We reduced the sample to concentrate our analysis on employees following the Eurostat definition of persons not in education. From the immigrant population we excluded persons who achieved their highest educational level after moving to Germany. Thus, only immigrants are considered aiming at utilizing their already existing skills. This is especially important when regarding researcher mobility in order to avoid mixing it up with students from abroad.

3.2.2 American Community Survey (ACS)

The ACS supplements the Current Population Survey (CPS) which only serves for providing general statistical population numbers. It also covers a larger sample size of about 3,000,000 in contrast to the CPS (60,000). This also allows for more detailed analysis of smaller subsamples, e. g. for regional perspectives, which was the original intention, but also regarding skill, occupation or migration background.

In contrast to the *Mikrozensus*, the ACS does not contain information on the point of time when the highest educational level has been achieved. For reasons of comparability only persons at the age of 28 and older are regarded assuming that they already hold their highest degree when immigrating to the US. The higher sample size, however, allows us to report countries of origin.

2 https://usa.ipums.org/usa/index.shtml, accessed on August 28, 2013.

3.3 Operationalisation

3.3.1 Scientific Occupations

The selection of the occupations of interest is confronted with two different classifications applied which are the ISCO-88 in Germany on the one hand, and the 2010 Census Code in the US on the other hand. Since the ACS provides the most detailed – 4-digit – occupational level, we took this as a starting point for a subsequent delineation of the German occupational classification. The advantage was to identify scientific occupations by their name. It is often too difficult to define a certain group of occupations using less detailed levels since one only hardly knows the weight of the single potentially interesting occupations within the larger level aggregate. This problem is especially relevant for scientific occupations. [3]

In general, we select occupations which are expected not only to require an academic education but also to comprise tasks related to research and development (R&D). We neglect medical occupations, first, due to some restrictions for persons with foreign degrees, and second, the large number of persons involved in practical medical jobs, e. g. in hospitals, rather than in R&D related tasks.

In a second step, after identifying potential scientific occupations in the US classification, the corresponding occupations in the German classification are then found via a common table of occupations translating one classification to another. Finally, we were able to exclude occupations which usually encompass educational levels lower than an academic degree: This concerns occupations belonging to the ISCO-88 group number 3 whose titles also suggest scientific tasks but only refer to technicians and similar educations.

Table 1 shows the final selection of occupations for both classifications.

Table 1: Selected scientific occupations

ISCO-88
• 21: Physical, mathematical and engineering science professionals
• 221: Life science professionals
• 231: College, university and higher education teaching professionals
• 244: Social science and related professionals
2010 Census Code
• 1000-1240: computer and mathematical occupations ohne: computer support specialists (1050)
• 1300-1560: architecture and engineering occupations ohne: marine engineers and naval architects (1440), drafters (1540), engineering technicians, except drafters (1550), surveying and mapping technicians (1560)
• 1600-1965: life, physical, and social science occupations ohne: technicians (1900-1965)
• 2200: postsecondary teachers

Quelle: International Standard Classification of Occupations, 1988 version. American Community Survey and Puerto Rico Community Survey, 2011 Code List. – Own compilation.

[3] See chapter 2.1.1 and for a more detailed discussoin Schiller (2014, p. 5).

3.3.2 Research Sector

Besides the occupation-related definition of scientific employment, we also provide an industry-related perspective. By focusing on universities and R&D services it will be assured that – similar to the occupations approach – the scientific content is of substantial importance within the statistical aggregates chosen. Although employment in both industries also comprises other tasks, e. g. administration, it is assumed that it still markedly higher than in other industries which are defined by other core activities, such as production. Practically, the two classifications in Germany (*Wirtschaftszweigsystematik (WZ) 2008*) and in the US (2007 Census Code and NAICS Code) are highly comparable. However, some national particularities are not possible to disentangle within the R&D services industry. In Germany extra-university research institutes such as the *Max Planck* or the *Fraunhofer* institutes are wide-spread while in the US private, for-profit R&D services are more present.

Table 2: Selected scientific industries

	USA			Deutschland	
Forschungssektor	Bezeichnung	2007 Census Code	2007 NAICS Code	Bezeichnung	WZ 2008
Hochschulen	Colleges and universities, including junior colleges	7870	6112, 6113	Tertiärer und post-sekundärer, nicht tertiärer Unterricht	854
FuE-Dienstleister	Scientific research and development services	7460	5417	Forschung und Entwicklung	72

Source: Klassifikation der Wirtschaftszweige (WZ), Ausgabe 2008. American Community Survey and Puerto Rico Community Survey, 2011 Code List. – Own compilation.

4 Empirical analysis of researcher mobility

4.1 Germany

Foreign employment in scientific occupations in Germany amounts to 135,000 immigrants (**Table 3**). Regarding single occupations, more than the half of that number works as architects, engineers or alike (67,000). Computer scientists on the one hand and social scientists and alike (excluding social workers) on the other hand share a similar number of 24,000 and 23,000 immigrant workers, respectively. Related to total immigrant employment, the share of immigrant scientists amounts to 3.6 percent which is half the share compared to non-immigrant German workers of which 7.2 percent are employed in scientific occupations. From the perspective of employment in scientific occupations only, immigrants make out 5.5 percent. This share is much lower than in non-scientific occupations where one of ten workers (10.8 percent) migrated to Germany after completing their education.

In the scientific industries total employment amounts to 780,000 persons which is thus a more narrow definition than by means of the occupations approach comprising 2.5 m persons. Within the scientific industries, R&D services (600,000) are more than three times the size of the university sector (180,000). Within the scientific industries immigrant workers have a higher importance than regarding scientific occupations: 9.1 percent compared to 5.5 percent reported above. There is also a slight difference between both industries. The immigrant share is significantly higher in the R&D services (9.9 percent) than in the university sector (6.2 percent).

Table 3: Scientific employment by occupation and industry in Germany

occupation / industry	German		Immigrants			total
	absolute (1,000s)	structure (%)	absolute (1,000s)	structure (%)	share (%)	absolute (1,000s)
occupations						
non-scientific occupations	**30,235**	**92.8**	**3,662**	**96.4**	***10.8***	**33,990**
Physicists, chemists and related professionals	65	0.2	3	0.1	*4.8*	68
Mathematicians, statisticians and related professionals	13	0.0	1	0.0	*6.6*	14
Computing professionals	456	1.4	24	0.6	*5.0*	481
Architects, engineers and related professionals	1,081	3.3	67	1.8	*5.8*	1,151
Life science professionals	59	0.2	3	0.1	*5.2*	63
College, university and higher education teaching professionals	113	0.3	13	0.3	*10.2*	126
Social science and related professionals	549	1.7	23	0.6	*4.0*	573
scientific occupations total	**2,336**	**7.2**	**135**	**3.6**	***5.4***	**2,478**
total	**32,571**	**100.0**	**3,797**	**100.0**	***10.4***	**36,468**
industries						
non-scientific industries	**32,274**	**97.8**	**3,807**	**98.2**	***10.5***	**36,179**
research services	543	1.6	60	1.5	*9.9*	604
universities	169	0.5	11	0.3	*6.2*	181
scientific industries	**712**	**2.2**	**71**	**1.8**	***9.0***	**785**
Zusammen	**32,986**	**100.0**	**3,878**	**100.0**	***10.5***	**36,964**

Source: Mikrocensus 2010 (Scientific Use File). Own calculations.

Besides the occupational assignment, the *Mikrozensus* 2007 also contains questions about the predominant tasks. One of the 20 items is related to research and design. The share of workers in scientific occupations occupied with research and design amounts to 30.2 percent (**Table 4**). In other occupations the share is 2.0 percent. More interesting, immigrant workers in scientific occupations are more frequently concerned with research tasks than native workers by 10 percentage points (39.7 percent compared to 29.8 percent). This pattern is found for any of the scientific occupations. The difference is mostly pronounced for university teachers who obviously migrate especially for research purposes (38.5 percent) compared to their non-immigrant colleagues being involved into teaching and administration to far higher extent (research tasks 14.6 percent).

Table 4: Share of research and design tasks by occupation and industry in Germany (2007)

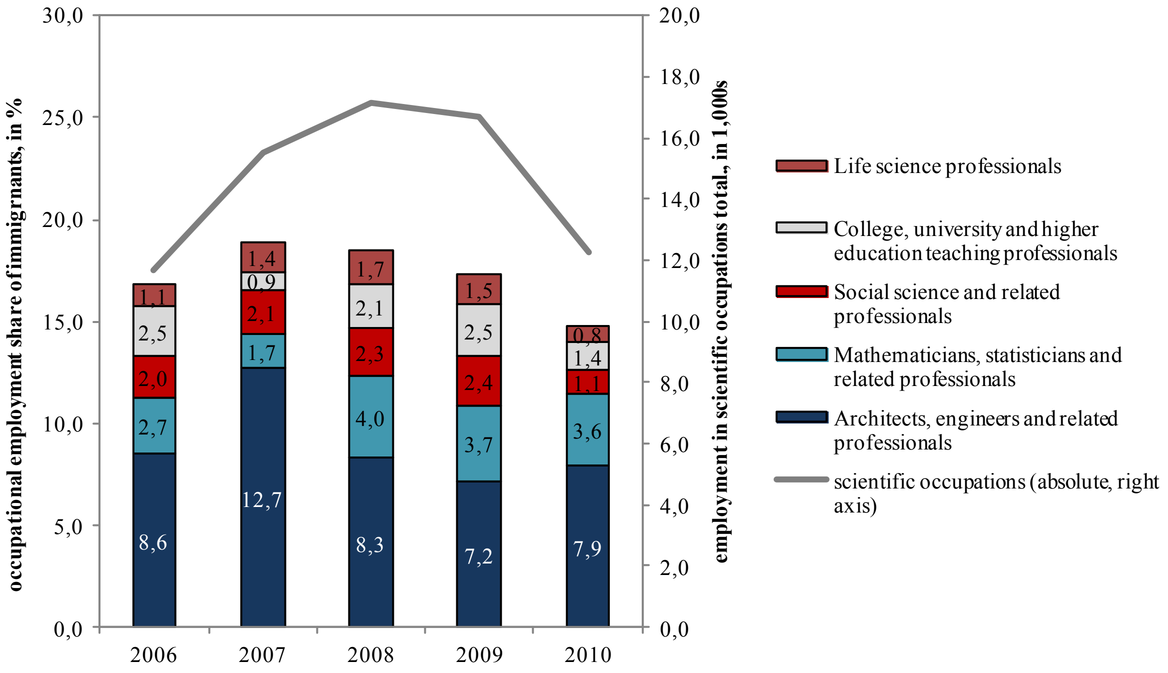
occupation	non-immigrants	immigrants	total
Physicists, chemists and related professionals	47.9	69.2	49.5
Mathematicians, statisticians and related professionals	12.4	-	12.4
Computing professionals	26.4	37.3	26.9
Architects, engineers and related professionals	44.2	46.8	44.3
Life science professionals	30.1	-	32.4
College, university and higher education teaching professionals	14.6	38.5	16.9
Social science and related professionals	4.0	9.1	4.2
scientific occupations total	29.8	39.7	30.2
non-scientific occupations total	2.1	0.9	2.0
total	4.0	2.0	3.8

Comment: „-“ no report due to small number of observations.
Source: Mikrocensus 2007 (Scientific Use File). Own calculations.

The recent development of migration in scientific occupations and industries is analysed based on the survey question of moving to Germany within the last 12 months in each of the waves. Between the period 2007 to 2010 the share of scientific occupations within the freshly arrived working migrants continuously fell from 18.8 percent in 2007 to 14.7 percent in 2010 (**Figure 1**). The development of the absolute number of migrants in scientific occupations basically describes an inverted u-shaped curve with a peak in 2008 (17,000) and markedly lower numbers at the margins (about 12,000 in 2006 and 2010, respectively). The decline especially hit the architecture and engineering occupations which dropped by 4.4 percentage points to 8.3 percent in 2010. In contrast, migration in occupations related to natural sciences, mathematics and computer science slightly increased and then stabilized at this higher level of 3.6 percent within total in-migrating employment of the year 2010. The remaining scientific occupations were of lower importance for immigrant workers in these years. In 2010, social science occupations had a share of 1.1 percent, university teachers 1.4 percent, and 0.8 percent concerning biological science occupations, respectively.

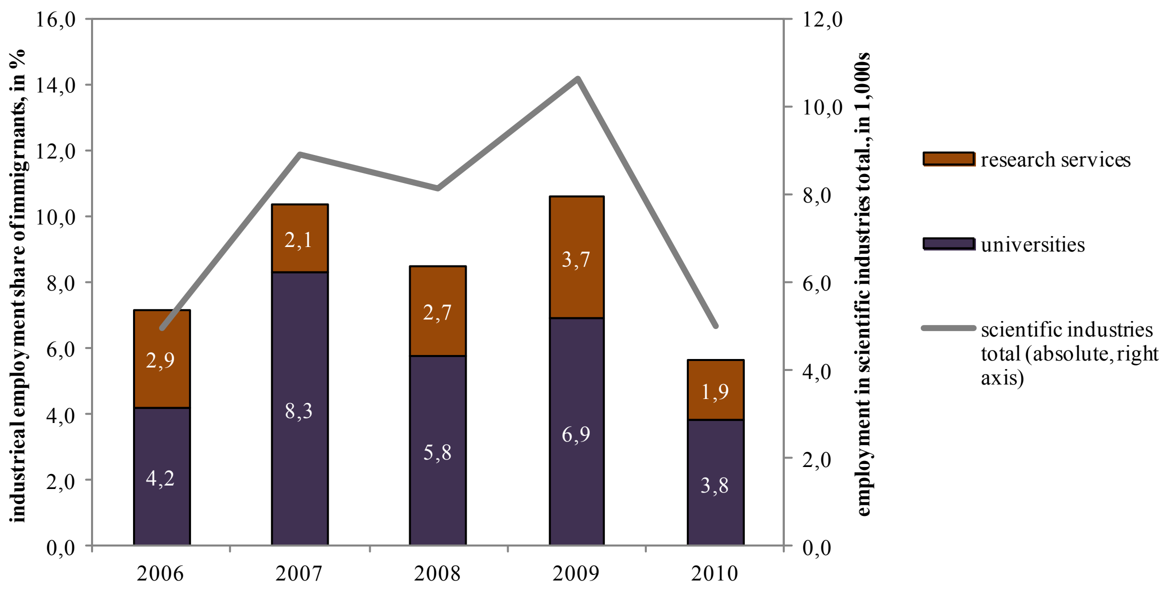
Migration dynamics in the scientific industries is volatile as well. The highest share is reported for the years 2007 and 2009 with slightly more than 10 percent (**Figure 2**). More recently (2010), the employment numbers make out about 6 percent or less. In absolute terms, the maximum was reached in 2009 with a total of 11,000 new migrant workers in scientific industries and again the lowest numbers at the margins with less than half that size (5,000). Although the preceding analysis showed that R&D services are larger than the university sector, the latter is twice important for recent in-migration. With shares varying from 4 to 8 percent universities are more frequent the first employer in Germany than R&D services are (2 to 4 percent).

Figure 1: In-migration in scientific occupations 2006 to 2010



Source: Microcensus 2006 to 2010 (Scientific Use Files). – Own calculations.

Figure 2: In-migration in scientific industries 2006 to 2010



Source: Microcensus 2006 to 2010 (Scientific Use Files). – Own calculations.

To sum up, there is a tendency of migration into scientific occupations to decline. Moreover, circular migration is generally supposed to play an important role when comparing the share of in-migration in single years with up to 20 percent to the migrant's share of total employment in scientific occupations which amounts to 3.6 percent only. This might also concern the scientific industries which exhibit a total migrant share of 1.8 percent but in more recent single years of 6 to 10 percent.

4.2 United States

Total employment in scientific occupations in the US amounts to 5.6m persons (**Table 5**). Nine of ten are natives or born abroad of American parents. Immigrants in scientific occupations have an employment share of 541,800 or 9.6 percent. The most important countries of origin are India (2.3 percent) and China (1.7 percent). Each of these two exhibits a higher number of former inhabitants now working in the US than Canada, Middle and South America together (1.3 percent). From Europe, 22,000 UK citizens represent 0.4 percent and the number from Germany amounts to 13,000 or 0.2 percent.

Table 5: Share of immigrants in scientific occupations

origin	total	non-scientific occupations total	scientific occupations total	computer and mathematical occupations	architecture and engineering occupations	life, physical, and social science occupations	postsecondary teachers	Insgesamt	non-scientific occupations total	scientific occupations total	computer and mathematical occupations	architecture and engineering occupations	life, physical, and social science occupations	postsecondary teachers
total	101,306	95,658	5,649	2,379	1,502	774	994	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Born abroad of American parents	1,035	955	80	38	19	10	13	1.0	1.0	1.4	1.6	1.3	1.3	1.3
US	93,182	88,155	5,027	2,122	1,359	668	877	92.0	92.2	89.0	89.2	90.4	86.3	88.3
immigrants total	7,090	6,548	542	219	124	95	104	7.0	6.8	9.6	9.2	8.3	12.3	10.4
Middle America + US Insular Areas	2,705	2,676	29	10	10	4	6	2.7	2.8	0.5	0.4	0.6	0.5	0.6
Canada	135	115	19	6	5	4	5	0.1	0.1	0.3	0.3	0.3	0.5	0.5
South America	643	616	27	7	7	6	8	0.6	0.6	0.5	0.3	0.5	0.7	0.8
Germany	69	56	13	3	3	3	4	0.1	0.1	0.2	0.1	0.2	0.4	0.4
France	29	23	6	1	1	1	2	0.0	0.0	0.1	0.1	0.1	0.2	0.2
UK	143	121	22	7	6	4	4	0.1	0.1	0.4	0.3	0.4	0.5	0.4
other Northern Europe	43	37	7	2	2	2	1	0.0	0.0	0.1	0.1	0.1	0.2	0.1
other Central Europe	31	25	6	1	1	1	2	0.0	0.0	0.1	0.1	0.1	0.2	0.2
South Europe	79	69	10	2	2	3	4	0.1	0.1	0.2	0.1	0.1	0.3	0.4
other Europe	288	265	24	6	8	4	5	0.3	0.3	0.4	0.3	0.5	0.5	0.5
other USSR/Russia	241	205	36	17	8	6	5	0.2	0.2	0.6	0.7	0.5	0.7	0.5
Near East	195	175	20	6	5	3	6	0.2	0.2	0.4	0.3	0.4	0.4	0.6
Africa	393	365	28	9	5	5	9	0.4	0.4	0.5	0.4	0.4	0.6	0.9
China	461	366	95	31	20	26	18	0.5	0.4	1.7	1.3	1.3	3.4	1.8
India	520	392	128	86	19	12	11	0.5	0.4	2.3	3.6	1.3	1.6	1.1
other Asia, Oceania, other World	1,114	1,042	71	22	21	13	15	1.1	1.1	1.3	0.9	1.4	1.7	1.5

Source: ACS. Own calculations.

In non-scientific occupations, the share of foreign born workers is lower – by about 3 percent. Two thirds of this difference are caused by the much higher engagement of Middle Americans in non-scientific rather than scientific occupations. On a lower level the reverse is true for Canadians who are three times as likely to work in scientific occupations (employment share of 0.3 percent compared to non-scientific occupations with 0.1 percent). Similar patterns are visible for most other countries except Africa or other European and Asian regions than they are distinguished in this table.

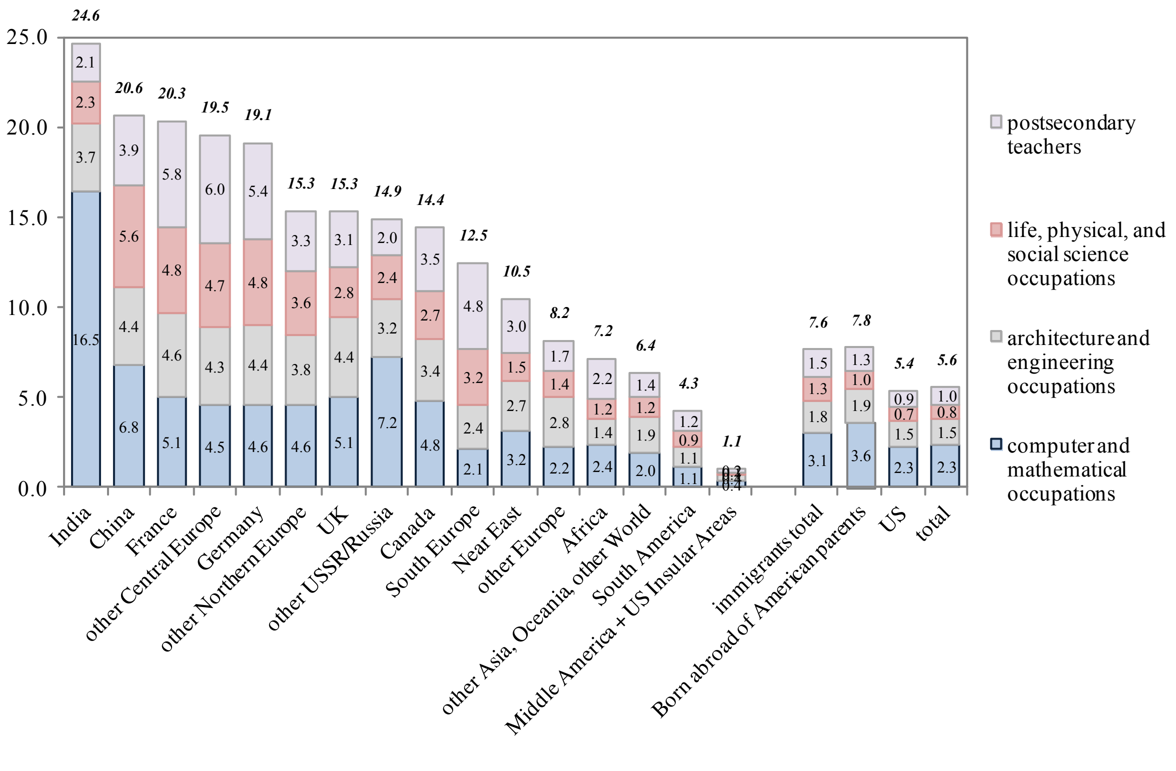
Significant contributions from immigrants to single occupations are found in the life, physical, and social science occupations totaling 12,3 percent, followed by the university teachers with 10.4 percent. Mathematical and computer related occupations as well as architecture and engineering occupations exhibit comparably lower shares which are, however, still higher than on the average over all immigrant employment (7.0 percent)

Considering the structure of immigrant's occupations the large inflow from Indian scientists especially contributes to mathematical and computer related occupations which make out 16.5 percent of all Indian workers. All other scientific occupations are practiced like in other immigrant groups. Also Russians are relatively specialized in mathematical and computer related occupations although these' occupations share is less than half of that compared to Indians (7.5 percent). For the whole of the US workers, however, these occupations are of minor importance (2.3 percent).

This pattern also holds when looking at scientific occupations together. On average 7.6 percent of working immigrants are employed in this segment, while for the native working population the corresponding value is only 5.4 percent. Immigrant scientists are thus more involved in scientific tasks yet not contributing markedly since they only shift the total share of scientific occupations to merely 5.6 percent.

Remarkably, though, are the large shares of immigrants in scientific occupations from China and the important European countries of origin. These people are concerned with scientific tasks with shares between 15 and 20 percent. Regarding the structure of occupations more in detail, most of them are more equally distributed over the four larger aggregates distinguished (postsecondary (mostly university) teachers; life, physical and social sciences; architecture and engineering occupations; mathematical and computer sciences).

Figure 3: Occupational structure by country of birth



Source: ACS. Own calculations.

Defining scientific employment by the industrial assignment, the science sector comprises 2.3m workers or 2.1 percent of total employment. The majority stems from universities (1.9m, 1.8 percent) whereas research services represent only 0.3 percent (340,000 persons). Compared to Germany, the related service industry is thus of minor importance.

The differences between US-Americans' and immigrants' concentration on each of the two industries are almost negligible. The former are employed by 1.8 percent in universities and 0.3 percent in the research service industry. Workers born outside the US are marginally more frequently employed with 0.2 percentage points in both industries. Taking the other perspective, where the workers are from, 180,000 immigrants in the research sector represent a moderately higher share of 8.0 percent than in the economy as a whole (7.0 percent).

Table 6: Origin of workers in scientific industries in the US

country / region of birth	universities	research services	research industries total	other	total	universities	research services	research industries total
	absolute (in 1,000s)					share on total (in %)		
total	1,932	342	2,274	101,306	105,854	1.8	0.3	2.1
Born abroad of American Parents	23	5	28	1,035	1,090	2.1	0.4	2.6
US	1,762	301	2,064	93,182	97,309	1.8	0.3	2.1
Zuwanderer insg.	147	36	182	7,090	7,454	2.0	0.5	2.4
Middle America + US Insular Areas	17	2	19	2,705	2,743	0.6	0.1	0.7
Canada	6	1	8	135	150	4.1	1.0	5.0
South America	11	2	13	643	669	1.6	0.3	1.9
Germany	4	1	5	69	80	5.1	1.7	6.8
UK	5	2	7	143	157	3.3	1.3	4.6
France	2	0	2	29	33	5.3	0.6	5.9
other Northern Europe	2	1	2	43	48	3.9	1.2	5.1
other Central Europe	2	1	3	31	36	5.4	1.8	7.2
South Europe	4	1	5	79	89	4.7	0.9	5.6
other Europe	7	2	9	288	306	2.3	0.6	2.9
other USSR/Russia	8	3	10	241	262	2.9	1.0	3.9
Near East	6	1	7	195	210	3.0	0.5	3.5
Africa	12	2	14	393	421	2.8	0.5	3.3
China	25	8	33	461	528	4.8	1.6	6.3
India	14	4	18	520	555	2.5	0.7	3.2
other Asia, Oceania, other World	22	5	27	1,114	1,168	1.9	0.4	2.3

Quelle: ACS. – Berechnungen des NIW.

Again, Asian workers from China (33,000), India (18,000) and other countries are the largest group among immigrants in the science segment. German workers amount only to a small number (5,000) although they are strongly concentrated on the research sector representing 6.8 percent. A similar pattern is found for other Central European countries (7.2 percent). The university sector is an employer for about 5 percent of immigrants from these countries. Regarding the research services, the share varies over all countries of origin between 0.5 and 1.5 percent.

4.3 Summary of the empirical analysis

Germany and the US, both heavily relying on the economic revenues of research and innovation, lay different emphasis on the contribution of immigrants in the corresponding key segments of the labor market. The most striking difference between the two countries becomes visible when comparing scientific with non-scientific occupations. While in Germany the share foreign-born workers is twice the size in non-scientific occupations than in scientific ones, it is found vice versa in the US, where the immigrants share is higher in scientific occupations. In the same vein, the share of immigrant workers in scientific occupations in Germany (3.6 percent) is much lower than in the US (7.6 percent). Moreover, in recent years (until 2010) immigration in scientific occupations even declined.

Table 7: Synopsis of selected results for Germany and the US

indicator	Germany	US
Share of foreign-born workers on total employment by occupational group / industry	non-scientific occupations: 10.8 percent scientific occupations: 5.5 percent research services: 9.1 percent	non-scientific occupations: 6.8 percent scientific occupations: 9.6 percent research services: 8.0 percent
Share of workers in scientific occupations, by origin	natives: 7.2 percent immigrants: 3.6 percent	natives: 5.4 percent immigrants: 7.6 percent
Scientific occupations with highest shares of foreign-born workers	University teachers, mathematicians and related	life, physical, and social science occupations, postsecondary teachers

Source: Own compilation.

In Germany, the share of foreign-born researchers is the highest among university teachers (10.2 percent), which, however, is still less than in non-scientific occupations (10.8 percent). In absolute terms, architects and engineers are the predominant group representing almost half of all foreign-born scientists in Germany. For immigrant workers in the US, in contrast, life, physical and social science occupations exhibit slightly higher shares (12.3 percent) than compared to the postsecondary teachers whose share is of almost equal size as in Germany (10.4 percent). This last results also indicates that the gains from immigration in the US especially benefit scientific employment in the private sector.

Literature

Aghion, P.; Bloom, N.; Blundell, R.; Griffith, R.; Howitt, P. (2005): Competition and Innovation: An Inverted-U Relationship, in: The Quarterly Journal of Economics, 120(2): 701-728.

Aghion, P.; Dewatripont, M.; Du, L; Harrison, A.; Legros, P. (2012): Industrial Policy and Competition. NBER Working Paper No. 18048.

Anger, C.; Plünnecke, A. (200): Signalisiert die Akademikerlücke eine Lücke bei den hochqualifizierten? - Deutschland und die USA im Vergleich, in: IW-Trends 3/2009: 19-31.

Baycan-Levent, T.; Nijkamp, P. (2009): Characteristics of migrant entrepreneurship in Europe, in: Entrepreneurship & Regional Development, 21(4): 375-397.

Borjas, G. (1987): Self-Selection and the Earnings of Immigrants, in: American Economic Review, 77: 531-553.

Cordes, A. (2012): Projektionen von Arbeitsangebot und -nachfrage nach Qualifikation und Beruf im Vergleich. Studien zum deutschen Innovationssystem Nr. 3-2012.

Foley, C. F.; Kerr, W. R. (2013): Ethnic Innovation and U.S. Multinational Firm Activity, in: Management Science, 59(7): 1529-1544.

Honig, B.; Drori, I.; Carmichael, B. (2010): Transnational and Immigrant Entrepreneurship in a Globalized World. University of Toronto Press: Toronto.

Hunt, J. (2013): Are Immigrants the Best and Brightest U.S. Engineers? NBER Working Paper No. 18696.

Kerr, W. (2008): Ethnic Scientific Communities and international Technology Diffusion, in: Review of Economics and Statistics, 90(03): 519-537.

Malchow-Møller, N.; Munch, J.R.; Skaksen J.R. (2011): Do foreign experts increase the productivity of domestic firms? IZA Discussion Paper No. 6001.

Moed, H.F.; Aisati, M.; Plume, A. (2013): Studying scientific migration in Scopus. Scientometrics 94(3): 929-942.

Mundra, K. (2012): Immigration and trade creation for the US: The role of immigrant occupation. IZA Discussion Paper No. 7073.

Nathan, M. (2013): The Wider Economic Impacts of High-Skilled Migrants: A Survey of the Literature. IZA Discussion Paper No. 7653.

OECD (2002): Frascati Manual. Proposed Standard Practice for Surveys on Research and Experimental Development. 6th edition. Paris: OECD Publishing.

OECD (2008): The Global Competition for Talent: Mobility of the Highly Skilled. Paris: OECD.

OECD (2013): Researchers on the Move. in: OECD Science, Technology and Industry Scoreboard 2013, p. 132-133.

Page, S. (2007): The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools and Societies. Princeton University Press: Princeton.

Papademetriou, D. G.; Terrazas, A. (2009): Immigrants in the United States and the Current Economic Crisis. http://www.migrationinformation.org/usfocus/display.cfm?ID=723, zuletzt besucht am 23.08.2013.

Parey, M.; Waldinger, F. (2010): Studying Abroad and the Effect on International Labour Market Mobility: Evidence from the Introduction of Erasmus, in: The Economic Journal, 121(551): 194-222.

Ruggles, S.; Alexander, J. T.; Genadek, K.; Goeken, R.; Schroeder, M. B.; Sobek, M. (2010): Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]. Minneapolis: University of Minnesota.

Sachverständigenrat deutscher Stiftungen für Integration und Migration (2013): Erfolgsfall Europa? Folgen und Herausforderungen der EU-Freizügigkeit für Deutschland. Jahresgutachten 2013 mit Migrationsbarometer.

Saxenian, A.-L. (2006): The New Argonauts: Regional Advantage in a Global Economy. Harvard University Press: Cambridge MA.

Saxenian, A.-L.; Sabel, C. (2008): Venture Capital in the 'Periphery': The New Argonauts, Global Search and Local Institution-Building, in: Economic Geography, 84(4): 379-394.

Scellato, G.; Franzoni, C.; Stephan, P. (2012): Mobile Scientists and International Networks. NBER Working Paper No. 18613.

Schiller, D. (2014): Machbarkeitsstudie „Forschermobilität und Innovation“. Studien zum deutschen Innovationssystem (*im Erscheinen*).

Stephan, P.; Levin, S. (2001): Exceptional Contributions to US Science by the foreign-born and foreign-educated, in: Population Resarch and Policy Review, 20: 59-79.

Zucker, L. G.; Darby M. R. (2007): Star Scientists, Innovation and Regional and National Immigration. NBER Working Paper No. 13547.