

Martin Kahanec and Klaus F. Zimmermann. Migration and Globalization: Challenges and Perspectives for the Research Infrastructure, in: German Data Forum (Ed.), Building on Progress: Expanding the Research Infrastructure for the Social, Economic, and Behavioral Sciences, Vol. 2, Budrich UniPress Ltd., Opladen and Farmington Hills (MI), 2011, 689-701. IZA Discussion Papers, No. 3890, 2008. RatSWD Working Paper No. 51, 2008.

## 1.1 Migration and Globalization

*Martin Kahanec and Klaus F. Zimmermann\**

---

\* The views expressed are the authors' alone and do not necessarily correspond to those of the German Data Forum (RatSWD). We thank Uwe G. Rehfeld and the participants to the Second Workshop of the German Data Forum (RatSWD) for helpful comments.

Contact:

Klaus F. Zimmermann  
IZA, P.O. Box 7240  
53072 Bonn  
Germany  
e-mail: Zimmermann[at]iza.org

## Abstract

The international migration of people is a momentous and complex phenomenon. Research on its causes and consequences requires sufficient data. While some datasets are available, the nature of migration makes it complicated to use them for scientific research. There is currently no existing dataset that effectively captures international migration trajectories. To alleviate these difficulties, we recommend: (1) the international coordination of data collection methodologies and standardization of immigrant identifiers; (2) a longitudinal approach to data collection; (3) the inclusion of adequate information about relevant characteristics of migrants in surveys, including retrospective information; (4) minimal anonymization; (5) immigrant boosters in existing surveys; (6) the use of modern technologies and facilitation of Data Service Centers; and (7) making data access a priority of data collection.

Keywords: migration, immigrants, data collection, data access, data infrastructure

## 1. Introduction

The international migration of people lies at the core of the ongoing process of globalization. People migrate to improve their economic prospects, ensure a more secure living environment, reunite with their family members, or avoid persecution in their country of origin. For these and other reasons 3 percent of the world's population found themselves on an international migration trajectory in 2005. Since a large proportion of these migrants head towards developed countries, the share of international migrants in these countries reached as much as 9.5 percent in 2005.<sup>1</sup> These individuals experience not only important economic and social consequences of their move, but also psychological ones. Migration may involve a new job with higher pay, the loss of old social ties and the establishment of new ones, as well as the psychological costs of missing the homeland.

Migration, however, does not only affect the fates of those who are directly involved. Various effects emerge at the interface of migrant and native populations. Immigrants may bring with them new cultures or preferences, they may compete for certain jobs and create others, or claim publicly financed social security benefits. More broadly, migrants contribute to a more efficient allocation of resources and often become a driving force behind knowledge transfer and technological advancement. All these effects have repercussions for the native population, who may react to migrant inflows not only in determining their current actions, but also in making

---

<sup>1</sup> See United Nations, Department of Economic and Social Affairs, World Migrant Stock: The 2005 Revision Population Database.

long-term investment plans, such as those concerning education. Finally, natives may view immigrants positively or negatively and shape their attitudes accordingly.

Migration is a dynamic phenomenon involving many twists and turns. Driven by a multitude of possible reasons, migrants may move temporarily or permanently, transnationally and nationally, individually or in groups, return to their countries of origin or migrate to another country, or move between two or more countries in a circular way. The complex underlying processes driving migration and its effects have attracted significant and growing attention from scientists. Chiswick (1978) and Borjas (1985) have pioneered the scientific work on immigrant adjustment in host societies, highlighting the significance of the host country experience and stressing the importance of cohort effects, country of origin, religion, education, as well as a number of demographic characteristics such as age and gender. From a different perspective, research on the migration decision has been inspired and advanced by Harris and Todaro (1970), Becker (1964), Mincer (1978), and Borjas (1985). Immigrant self-selection discussed by Borjas (1987) and Chiswick (1999) implies the need for specific techniques (Heckman 1979) to consistently evaluate causal mechanisms behind immigrant adjustment.

The impact of immigration on the host labor market has been modeled by Chiswick, Chiswick, and Karras (1992) and Chiswick (1998). A large body of empirical literature, summarized by Kahanec and Zimmermann (2009), provides mixed evidence on the sign and determinants of these effects on wages and employment.<sup>2</sup> More recently, the roles of intermarriage (Meng and Gregory 2005), citizenship (Bratsberg et al. 2002), social networks (Munshi 2003), and attitudes (Bauer et al. 2000; Kahanec and Tosun 2009) pertaining to immigrant adjustment have received significant attention. The concept of ethnic identity has been extended by Constant and Zimmermann (2008), who elaborate on how attachment to the country of origin and the host country affect immigrant adjustment.

Although measuring the effects of migration is by no means an easy or straight-forward job, migration is a phenomenon that undoubtedly affects the well-being of the whole society and as such, has become an important and sensitive policy issue. Questions about the labor market consequences of migration, immigrant adjustment in host societies, and welfare competition have received particularly significant policy attention.

Understanding the causes and effects of international migration flows requires a sound and in-depth analysis. The need for such analysis is most conspicuous in the study of causal relationships, as these are difficult to establish empirically and their misrepresentation compromises both scientific and policy analyses. In fact, it may lead to incorrect policy recommendations,

---

2 The evidence on migration effects in source countries is mainly related to remittances (e.g., Barham and Boucher 1998), and wage and employment effects (Brückner 2007).

which may in turn lead to unpredictable consequences or even effects contrary to those intended. Since such analysis is impossible without high-quality data, such data are indispensable for policy analysts as well as scientists.

## 2. Relevant and available data

Despite the general scarcity of migration data, scientists and analysts have been able to use some existing survey or administrative datasets as well as small-scale dedicated survey data to study migration issues. While these datasets have facilitated valuable research, missing variables, excessive anonymization, and flaws in data collection design often compromise scientists' efforts to broaden and deepen our knowledge of migration causes and effects. In this section we focus on some large-scale datasets collected at the European level, since they, in contrast to small-scale surveys, have an intrinsic potential to provide the necessary transnational, longitudinal, and systematic data collection framework.

There are four extensive datasets that in some dimension provide coverage of European migration trajectories: the European Community Household Panel (ECHP), the EU Statistics on Income and Living Conditions (EU-SILC), the EU Labour Force Survey (EU-LFS), and the OECD/SOPEMI (*Système d'Observation Permanente des Migrations*) dataset. Each of these datasets contains information about demography, labor force participation, employment, unemployment, self-employment, and educational attainment of immigrants. In addition, the European Social Survey (ESS) covers people's attitudes toward immigrants as well as their voting preferences, thus addressing migration indirectly.

Table 1 depicts the character of these datasets, highlighting some of their strengths and weaknesses. We can identify at least three major gaps in the available data.<sup>3</sup> First, these datasets provide either none or only a very limited account of migration trajectories. Transnational migration trajectories may involve simple or repetitive moves between two or more countries with temporary periods of residence of varying length as well as permanent moves. It is almost impossible to track such trajectories – with all their spells, stops, and circularities – within Europe or between Europe and third countries. In particular, little or no information is available on migrant experience prior to their arrival in the country of current residence or their intentions for further moves. Secondly, the data typically permit determining immigrant status based on an individual's citizenship and country of origin,

<sup>3</sup> See also the discussion in Bauer and Zimmermann (1998).

neglecting the large groups of people with an immigrant background who are native citizens or those with dual citizenships. Finally, anonymization often renders any valuable analysis impossible, for example when immigrants from very different origins (e.g., Zimbabwe and Japan) are grouped into one category (e.g., non-EU).

### 3. Data access issues and needs

Inadequate access to existing datasets is one of the most limiting factors for scientific and policy analysis. Due to restrictive data access policies, a lack of interest on the side of the officials responsible, misinterpreted data protection rules, or simply the lack of adequate data access infrastructure, the use of datasets for scientific and policy purposes is, in general, severely limited. Since migration is, by definition, a transnational and dynamic phenomenon (i.e., involving one-way as well as repeat, sequential, and circular movement between more countries), its proper analysis requires a combination of information from multiple countries and across multiple periods.<sup>4</sup> Therefore, restrictions on data access and a lack of coordination in establishing access rules are particularly detrimental to the analysis of migration issues. Below we list some of the most pressing issues that obstruct availability of data for migration analysis and determine the needs concerning collection of adequate data on migration.

One of the main problems is that identifying and defining migrants in the existing datasets is not a trivial matter. The migration background, a foreign origin (foreign birthplace), citizenship, or ethnicity can be used to determine whether or not someone is an immigrant. Unfortunately, only a subset of this information, if any, is available in existing datasets. Only rarely can one identify first, second, and further generations of immigrants, citizens and non-citizens, and distinguish immigrants of different origin and ethnicity.

It is even more seldom that it is possible to obtain information that can be used to characterize migration trajectories. Perhaps with the exception of length of stay in the host country, pre-migration experience, tracking of all migration moves, or migration trajectories of family members (spouses) are rarely available. While the lack of data describing migration trajectories of those who make more frequent, possibly circular, moves is a general problem, it is particularly problematic in the case of highly skilled migrants, since these are the most fluid and mobile segment of the migrant population.

---

4 Bauer, Pereira, Vogler and Zimmermann (2002) have merged Portuguese data and German data on Portuguese migrants to be able to compare migrants in the sending and a receiving country. See Crul and Vermeulen (2006) for another project in this spirit.

Other relevant and often missing information include language, religion, and attachment to the host society and the country of origin.

A further and related problem is that the effects of out-migration are hard to capture, since we typically do not observe people who leave or record their characteristics (they do not de-register and are in a different country when data are collected). In fact, this deficiency creates problems for the analysis of the entire population as well, since it compromises the representativeness of datasets. For example, according to the Weekly Report of the German Institute for Economic Research ("Wochenbericht des DIW") (2008, 382), doubts have arisen in Germany as to whether the official census statistics still represented the actual reality of the German population. Since the German national census data has only been based on registers since 1987 – which depend on the proper *registration* and *deregistration* of people – those who leave the country and do not deregister are erroneously counted. An example of the magnitude of the measurement error which can result from failing to track out-migration of those who have not deregistered was revealed in a clean-up of the data from the German Central Register of Foreigners (*Ausländerzentralregister*) in 2004, which showed that the official census statistics had overstated the number of foreigners in Germany by about 600,000.

Another problem with the current situation is that most datasets are representative of the total population and contain a limited number of observations. While this is not necessarily a problem in other contexts, in the context of migration it often implies insufficient samples of the immigrant population. In addition, many datasets are cross-sectional and thus do not capture the dynamic nature of migration. In particular, the snapshot picture that such datasets provide can only capture the most recent move and cannot distinguish some important effects, such as those of host country experience and immigrant cohort on immigrant adjustment.

Finally, knowledge of migration intentions and reasons, and their relationship to actual migration decisions is indispensable for predicting future migration flows as well as for understanding the social and economic outcomes of migrants in the host societies. Precise estimates of the directions and characteristics of such flows are crucial for designing effective and efficient immigration policies, for instance in the context of EU enlargement. The intention to stay, namely, whether migrants perceive their situation as temporary or come to settle in the host country permanently, carries important consequences for their labor market behavior and thus the effects they have on the host economy. Similarly, migrants who come for economic reasons and those who come as refugees or asylum seekers have very different labor market opportunities as well as intentions in the host country.

These issues concerning the availability of and access to adequate migration data define the primary needs concerning the collection of relevant

data. In particular, collected data should properly identify migrants and people with immigrant backgrounds, and contain sufficient samples of migrants. They should cover (transnational) migration trajectories and, in particular, capture pre-migration experience and out-migration, as well as measure intentions and reasons for migration.

#### 4. Future developments and challenges

The enlargement of the European Union and the concurrent expansion of the European Economic Area as well as the persistent economic and social hardship and insecurity in large parts of the world will continue to fuel substantial international flows of people. High-quality data are and will remain a key ingredient to understanding the causes and effects of these migration flows. Given the traditional prominence of quantitative techniques in economics and the growing emphasis on such techniques in other social sciences, especially sociology, we can project increasing demand for such data among scientists in the future. This demand will be further intensified by the increasing need for well-founded policy analysis at the European and national levels. Another contributing factor may be the business sector, which may seek to exploit the potential benefits from precise information about their current and potential customers.

The provision of high-quality migration data is in general insufficient, although it has somewhat improved over the last decade or two. This improvement has been enabled by the emergence of advanced information and data management technologies that can facilitate a wide access to existing datasets. This development particularly relates to a group of international institutions that have started to provide online access to some of their datasets (European Union, World Bank, ILO,<sup>5</sup> UN) as well as private and non-governmental organizations (IZA)<sup>6</sup> that use innovative technologies to promote access to own and third-party datasets. While some improvements have been made at the national level, government institutions still lag behind in data access provision. More recently, some remarkable developments have taken place, however, involving a partnership between public and non-governmental or private institutions aiming at a wider dissemination of valuable data collected by public institutions. For instance, the International Data Service Center of the IZA offers onsite computing via ultra thin access, and remote computing by means of a remote computing solution (JoSuA)

---

<sup>5</sup> International Labour Organization.

<sup>6</sup> Institute for the Study of Labor (IZA, *Forschungsinstitut zur Zukunft der Arbeit*).

facilitating the use of scientific use data of the German Federal Statistical Office.<sup>7</sup>

These positive developments should not hide the difficult reality of migration research and analysis as concerns data availability. Besides the various difficulties that migration researchers face regarding identification of migrants in existing datasets as well as a lack of relevant information about them, virtually no existing dataset has the necessary transnational and longitudinal perspective to capture complete migration trajectories. Thus, the key challenge in this respect is to track migrants and their migration experience as they move internationally. The associated practical challenge is to coordinate data collection methodologies across Europe and, even more complicated, between Europe and third countries.

## 5. Conclusions and recommendations

This essay summarizes some of the key problems and challenges related to the availability of data for the study of migration issues. Having considered the long-standing as well as more recent developments in migration research, it is clear that access to data of good quality, harmonized across time and countries, is one of the main bottlenecks hindering advances in our understanding of the causes and effects of migration. To alleviate this problem, there are a number of policy tools that may help.

First, coordination of data collection methodologies and standardization of immigrant identifiers across the EU would facilitate international comparability. It is necessary to harmonize data collection methods so that migration trajectories in Europe-wide datasets can be observed. In particular, unique individual identifiers need to be traceable across European countries. An open method of coordination, transparent indicators, benchmarking, and an efficient exchange of best practices seem to be the way to go in this regard. This also involves merging datasets transnationally and across time, including proper harmonization and linking of data, records, and topics. In particular, given the advancement in data management and storage techno-

7 The International Data Service Center of the IZA, one of the Data Service Centers facilitated by the "KVI Commission," offers an integrated service which consists of a metadata portal and a remote computing solution. The International Data Service Center of the IZA's metadata service comprises a detailed, in depth, searchable and standardized information and documentation service on a growing number of datasets currently in the areas of employment and wages, education and training, and demographics and migration. The International Data Service Center of the IZA remote computing solution, known as JoSuA, facilitates usage of restricted datasets bridging the otherwise wide gap between legal constraints and scientific freedom without violating the former or constraining the latter. For further details see Schneider and Wolf (2008).

logies, this objective entails not only prospective but also retrospective harmonization and merging of datasets as well, involving digitalization of old datasets whenever necessary.

Second, whenever possible, a longitudinal approach should be adopted, both to facilitate the separation of spurious effects driven by unobserved cross-sectional variation from true causal relationships as well as to capture the dynamic nature of migration. In this regard, one should also consider extending selected existing cross-sectional datasets by surveying the covered individuals in one or more additional waves.

Third, adequate information about the relevant characteristics of migrants – experience in the host society (years since migration), country of origin, citizenship, ethnicity, language, religion, attachment to the host society and the country of origin, and migration intentions and reasons – is requisite. For example, it is of key importance to distinguish temporary and permanent migrants as well as economic migrants from those that come as refugees or asylum seekers, or as tied movers. Retrospective questions in survey questionnaires are necessary to track migrants' pre-migration experience (i.e., experience prior to the last observed move).

Fourth, anonymization should be limited to the smallest possible degree. As an option, alternative anonymization procedures could be applied to the same dataset, allowing two or more versions to be accessible to the researcher, each facilitating research on different research questions.

Fifth, immigrant boosters in existing surveys with a well-defined control group would facilitate sufficient immigrant sample sizes.

Sixth, online Data Service Centers, data registers and metadatabases can provide an invaluable service to the research community. In fact, the Internet is itself becoming a rich source of data and a tool to collect new data that still needs to be properly exploited.

Seventh, the use of modern data information technologies should be promoted to facilitate the collection, management, and storage of good quality data, as well as, importantly, enabling access to it. As part of this objective, the creation of Data Service Centers facilitating prudent access to such data is desirable.

Finally, the facilitation of data access for researchers should be embraced as one of the main objectives of data collection. Adequate effort by all the actors involved is necessary not only to facilitate knowledge about migration as such, but also, to the extent that suitable policies are adopted, to improve the welfare of substantial numbers of people who are directly or indirectly affected by migration.

Table 1: Datasets

Dataset	Type	Years	Countries	Measures of immigrant status	Weakness (selected)
ECHP	Longitudinal	1994 - 2001	EU 15	Year of arrival (region/country), country of birth, first and second citizenship, mother tongue	No information on immigrant experience prior to his/her arrival to the country of present residence.
EU-SILC	Longitudinal	2004 - 2006	Until 2004 EU15 Since 2006 EU25	Country of birth, citizenship (first)	Anonymization leading to a mixing of immigrants from very different origins.
EU-LFS	Survey	1983 - 2006	BE, CZ, DK, DE, EE, GR, ES, FR, IE, IT, CY, LV, LT, LU, HU, MT, NL, AT, PL, PT, SI, SK, FI, SE, UK, BG, RO (+HR, TR, IS, NO, CH)	Nationality (citizenship), years of residence, country of birth (anonymized), country of residence one year before the survey	Anonymization leading to a mixing of immigrants from very different origins.
OECD/SOPEMI	Macro-data	1983-2008	OECD	Stocks of foreign nationality and foreign born populations, country of birth, flows of foreign-born workers	While the dataset provides aggregate data, no information about the individual characteristics of migrants is available.
ESS	Cross-sectional	2002, 2004, 2006	EU25	Voting preferences, attitudes toward immigrants and ethnic minorities	The cross-sectional nature of the dataset does not capture the dynamic nature of migration.

## References:

- Barham, B. and Boucher, S. (1998): Migration, Remittances and Inequality: Estimating the Net Effects of Migration on Income Distribution. In: *Journal of Development Economics* 55 (2), 307-331.
- Bauer, T./Lofstrom, M. and Zimmermann, K.F. (2000): Immigration Policy, Assimilation of Immigrants, and Natives' Sentiments towards Immigrants: Evidence from 12 OECD Countries. *Swedish Economic Policy Review* 7 (2), 11-53.
- Bauer, T. and Zimmermann, K.F. (1998): Causes of International Migration: A Survey. In: Gorter, C./Nijkamp, P. and Poot, J. (Eds.): *Crossing Borders: Regional and Urban Perspectives on International Migration*. Aldershot.
- Bauer, T./Pereira, P.T./Vogler, M. and Zimmermann, K.F. (2002): Portuguese Migrants in the German Labor Market: Performance and Self-Selection. *International Migration Review* 36 (2), 467-491.
- Becker, G.S. (1964): *Human Capital. A theoretical and empirical Analysis*. New York.
- Borjas, G.J. (1985): Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants. *Journal of Labor Economics* 3 (4), 463-489.
- Borjas, G.J. (1987): Self-Selection and the Earnings of Immigrants. *American Economic Review* 77 (4), 531-553.
- Bratsberg, B./Ragan, J.F. and Nasir, Z.M. (2002): The Effect of Naturalization on Wage Growth: A Panel Study of Young Male Immigrants. *Journal of Labor Economics* 20 (3), 568-597.
- Brücker, H. (2007): Migration after the EU's Eastern Enlargement: Who Wins, Who Loses? [Paper presented at Second IZA Migration Workshop: EU Enlargement and the Labor Markets, Bonn 7.-8. September 2007].
- Chiswick, B.R. (1978): The Effect of Americanization on the Earnings of Foreign-born Men. *Journal of Political Economy* 86 (5), 897-921.
- Chiswick, B.R. (1998): The Economic Consequences of Immigration: Application to the United States and Japan. In: Weiner, M. and Hanami, T. (Eds.): *Temporary Workers or Future Citizens? Japanese and U.S. Migration Policies*. New York.
- Chiswick, B.R. (1999): Are Immigrants Favorably Self-Selected? *American Economic Review* 89 (2), 181-185.
- Chiswick, C.U./Chiswick, B.R. and Karras, G. (1992): The Impact of Immigrants on the Macroeconomy. *Carnegie-Rochester Conference Series on Public Policy* 37 (1), 279-316.
- Constant, A. and Zimmermann, K.F. (2008): Measuring Ethnic Identity and Its Impact on Economic Behavior. *Journal of the European Economic Association* 6 (2-3), 424-433.
- Crul, M. and Vermeulen, H. (2006): Immigration, education and the Turkish second generation in five European nations: a comparative study. In: Parsons, C.A. and Smeeding, T.M. (Eds.): *Immigration and the Transformation of Europe*. Cambridge.
- Harris, J.R. and Todaro, M.P. (1970): Migration, Unemployment and Development: A Two-Sector Analysis. *The American Economic Review* 60 (1), 126-142.

- Heckman, J.J. (1979): Sample Selection Bias as a Specification Error. *Econometrica* 47 (1), 153-161.
- Kahanec, M. and Tosun, M. (2009): Political Economy of Immigration in Germany: Attitudes and Citizenship Aspirations. *International Migration Review* 43 (2), 263-291.
- Kahanec, M. and Zimmermann, K.F. (2009): International Migration, Ethnicity and Economic Inequality. In: Salverda, W./Nolan, B. and Smeeding, T.M. (Eds.): *Oxford Handbook on Economic Inequality*. Oxford.
- Meng, X. and Gregory, R.G. (2005): Inter-marriage and economic assimilation of immigrants. *Journal of Labor Economics* 23 (1), 135-175.
- Mincer, J. (1978): Family Migration Decisions. *Journal of Political Economy* 86 (5), 749-773.
- Munshi, K. (2003): Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market. *The Quarterly Journal of Economics* 118 (2), 549-599.
- Schneider, H. and Wolf, C. (2008): Die Datenservicezentren als Teil der informationellen Infrastruktur. In: Rolf, G./Zwick, M. and Wagner, G.G. (Eds.): *Fortschritte der informationellen Infrastruktur in Deutschland. Festschrift für Johann Hählen zum 65. Geburtstag und Hans Jürgen Krupp zum 75. Geburtstag*. Baden-Baden.
- Weekly Report of the German Institute for Economic Research (Wochenbericht des DIW) Nr. 27-28/2008, DIW Berlin.