Bayesian modelling of international migration with Labour Force Survey data

Arkadiusz Wiśniowski*

— DRAFT —

Please do not cite without prior permission

March 11, 2013

Abstract

In this paper, a statistical method for estimating international migration flows based on information obtained from Labour Force Surveys (LFS) is developed. The motivation for using LFS data is the general poor quality or absence of data provided by the national statistical institutes. A Bayesian model of flows between two countries is considered, assuming that migration can be measured in the LFS by both sending and receiving countries. The undercount in a particular destination-specific emigration flow is estimated by using the LFS sample of the corresponding receiving country. The model is applied to estimate migration flows from Poland to the United Kingdom in 2002-2008.

1 Introduction

The purpose of this study is to propose a statistical framework for estimating international migration flows based on information obtained from Labour Force Surveys (LFS). The method can be used as a tool for validating and supplementing officially produced statistics.

The motivation for using LFS data is the general poor quality, inconsistency, or absence of data provided by national statistical offices in European countries. Reliable and comprehensive statistics on international migration are required to analyse the reasons and consequences of the movements, as well as to design and implement

*email: A.Wisniowski@soton.ac.uk

fair and effective labour, economic and social policies. As noted by Bilsborrow et al. (1997, p. VI), lack of comprehensive data on migration leads to the analysis '[...] more in terms of impressions than in terms of impact' and to formulation of the policies on the 'shaky basis'.

A simple Bayesian model of flows between two countries is considered, assuming that migration can be measured in the LFS by both the sending and receiving country. Further, the undercount in a particular destination-specific emigration flow is estimated by using the LFS sample of the corresponding receiving country. For illustration, the framework developed in this paper is applied to estimate recent migration flows between Poland and the United Kingdom.

In May 2004, Poland and seven other countries from Eastern and Central Europe joined the EU. That was a symbolic reunion of Central and Eastern Europe with the West. In the aftermath, the divisions have been softened and barriers erected during the Cold War era dismantled. This led to intensification of mobility across Europe and to a substantial increase in migration flows. The total population of the new members was around 73 million people, with Polish population accounting for more than half of this number. The EU enlargement resulted in releasing the labour force from the new member states, posing social and economic challenges in both sending and receiving countries.

2 Migration and migration measurement

Migration can be quantified in various terms and measures which depend on the method and object of data collection. Information collected in the censuses differs from the one recorded in registers or yielded by surveys. As it was mentioned in Section 2, the collected information may concern a migrant or a migration event (Courgeau, 1973). Other differences in the collected data are discussed by, e.g. Rees (1977); Rees and Willekens (1985); Rees (1986); Rees and Woods (1986). Migration can also be quantified in two ways, namely as stocks and as flows, depending on the method applied. Further, two concepts underlie the measurement of flows: (i) events (migrations) or (ii) changes of a status of persons (transitions).

All the data collection methods usually collect information about a person who is migrating or about the event of migration. For example, in censuses there may be a question about place of residence of a respondent five years prior to census date. This information concerns a person. In registers, information about an event of arrival or departure is recorded. This is the main distinction of the data types: events and persons. Events are migrations, that is movements from the origin to the destination country and they involve crossing the administrative border. If the travelling time of a person is neglected, the migration event happens at a specific point in time. The other type is a person, that is a migrant, who is moving from the origin to the destination. A migrant is an actor who performs a migration or many migrations (Rees, 1977).

Flows of migrants is a quantity that is measured over an interval of time, hence it is a measure of flows per unit of time. There are two methods of measuring flows. The first is by counting a number of movements, i.e. all changes of the place of usual residence within a period of time. In that case, if a person relocates a number of times and all of them are migration events. The other way of quantifying flows is to compare usual places of residence at two points in time, one year apart, for instance. The first of the approaches is used in population registers or registers of foreigners. A second one can be used to measure migration flows by comparing the data from, for example, two subsequent censuses. This approach is called a transition approach.

Summarising, stocks are measured in persons at a given time point, say, $t + 1$. Events are migrations counted continuously between two points in time, $(t, t + 1)$, hence, they are measured in migrations per unit of time. Transitions are measured comparing two points in time, $\{t; t + 1\}$ and they concern persons (survivors) per unit of time.

2.1 Quality and availability of the data

There are three distinctive problems concerning the quality of the statistical data on migration, and in particular, migration flows: availability, reliability and comparability (Nowok et al., 2006).

The data are considered available when they are disseminated by the NSI or by some international organisation in a printed or electronic form. Kupiszewska and Nowok (2008) and Nowok (2010) consider the data complete if both pieces of a double-entry matrix on migration are provided. Data availability can be also analysed in the context of the requirements set out in the Regulation (EC) No 862/2007. Availability of the data in the EU countries has been a subject of investigation in projects THESIM – Towards Harmonised European Statistics on International Migration (2004-2005), PROMINSTAT – Promoting Comparative Quantitative Research in the Field of Migration and Integration in Europe (2007-2009) and MIMOSA – Migration

Modelling for Statistical Analyses (2007-2009). THESIM's aim was to assess and document the current state and functioning of migration statistics in 25 EU Member States. The objective of PROMINSTAT was, amongst other things, compilation of the meta-information on the statistical and administrative datasets relevant to study migration in 27 EU Member States plus Norway and Switzerland. MIMOSA's aim was to assess the availability of the data on migration reported by EU countries to Eurostat and develop methods for accommodating for the differences in migration statistics in EU countries.

The data are said to be reliable if they comply with the definition adopted in the statistical institute in a given country. Poor reliability of the statistical data on international migration flows is a well known fact (cf. Bilsborrow et al., 1997; Nowok et al., 2006; Kupiszewska and Nowok, 2008; Fassmann, 2009; de Beer et al., 2010). The main reasons of such a situation are the under-registration of migrants and imperfect data coverage. Reliability of the migration data based on the sample surveys is considered to be low due to the estimation errors resulting from the high volatility of the observed series and insufficiently small sample sizes. It affects particularly the disaggregates of migration flows, such as by age, sex, country of birth.

The data that are unreliable are often incomplete and inconsistent, thus preventing any comparisons between countries or within countries over time. As it is noted by Kupiszewska and Nowok: 'In recent decades there have been many efforts to harmonise migration statistics [...]. However, the results are far from satisfactory' (Kupiszewska and Nowok, 2008). De Beer et al. note even that '[...] the situation today in terms of migration definitions and measurement is not much better than it was [...] 20 years ago' (de Beer et al., 2010). Incomplete and unreliable data and, as a consequence, lack of harmonisation, lie in the data collection mechanisms adopted by particular countries. There are three stages of producing statistics at which the differences between countries can arise (i) collection of raw data in the primary data source, (ii) processing of the collected raw data, (iii) dissemination of the data (Kupiszewska and Nowok, 2006, p. 53).

2.2 Harmonisation attempts

Recently, there have been several attempts on improving the comparability of the data on international migration in the EU. I distinguish two types of harmonisation: (1) legislative, which concerns all changes in the legislation regarding three stages of producing statistics on migration, both on the country and a group of countries level, and (2) data-based, which use the already available data to produce new sets of harmonised statistics. In this section I concentrate on the latter type of harmonisation since the purpose of the model developed in the subsequent sections is a tool for harmonising statistics on flows using not the official data, but the data derived from the LFS.

In the first group, there is the Regulation no. 862/2007 of the European Parliament and of the Council of Europe from July 2007. It provides a set of common definitions and guidance for the collection of the data on migration. This Regulation is rooted in the United Nations *Recommendations on Statistics on International Migration* (1998), which underlies the low quality of migration data and the need for improvement (UN 1998, p. 8). It provides harmonised definition of migration and guidelines for collection and compilation of statistics.

The second group of harmonisation attempts concerns the data that are already collected and disseminated. The first approach, proposed by Lemaître (2005) and Lemaître et al. (2006), is based on selected sources of data on migrants, namely residence and stay permits. Instead of the duration of stay criterion, such as the 12-months one suggested by the UN (1998), they adopt a criterion of a reason for movement to harmonise statistics on migration. These statistics are called 'standardised statistics' and concern permanent immigration of foreigners. The idea of standardised statistics was partially implemented in the International Migration Outlook 2006 (OECD, 2006). This approach has some serious drawbacks. The most important are that it concerns only regulated flows of foreigners and does not capture returning nationals or free flows, such as flows within the EU. It is also impossible to measure emigration (Kupiszewska et al., 2010). Hence, these statistics should be used only as a supplementary source of information on migration.

DeWaard et al. (forthcoming), in their study of migration systems in Europe, use a resampling method and treat the order of countries in the quality ranking as a permutation problem. In this way, they are able to quantify the precision of the estimates for the migration flows aggregated over 2003-2007 period. The methodology proposed by DeWaard et al. (forthcoming) yields similar estimates and patterns of the migration flows in Europe as the ones obtained in the MIMOSA project (de Beer et al., 2010).

Another attempt to estimate a table of comparable flows in Europe, similar to MIMOSA (Poulain, 1993; 1999; and Poulain and Dal, 2008; de Beer et al., 2010), was undertaken by Abel (2009, 2010). Estimation concerns EU-15 countries. Similarly to the estimation procedure in MIMOSA, he uses a constrained optimisation to estimate the correction factors. The methodology applied by Abel allows providing measures of precision for the estimates of the missing flows.

Methodologies applied in the MIMOSA project (Poulain, 1993, 1999; de Beer et al., 2010), DeWaard et al. (forthcoming) and by Abel (2009, 2010) are limited as they do not provide any explanation for the discrepancies between the measures of a flow from the sending and receiving countries. The methodology relies heavily on the benchmark data, which in all cases are figures reported by Sweden. The approach developed by Nowok (2010) provides insights into the differences in migration measures resulting from various duration of stay criteria. The study provides an understanding of the consequences of applying duration criteria of different length in sending and receiving countries, as well as different criteria used for absence and presence, on the measurement of migration and the resulting statistics.

In the project IMEM – Integrated Modelling of European Migration (2009-2012), a Bayesian model for harmonising and correcting the inadequacies in the available data and for estimating the completely missing flows is proposed (Raymer et al., 2010). It provides estimates of the recent international migration flows amongst 31 countries in the EU and EFTA, using the official data published by Eurostat and other national and international institutions. The methodology integrates the data on migration from both sending and receiving countries, covariate information and expert judgement to produce a synthetic data base of migration flows with measures of uncertainty. The IMEM estimates include also interpretable parameters, which handle the data measurement problems: undercount, varying duration of stay criteria, coverage and accuracy of the data collection method.

3 Labour Force Survey (LFS)

LFS is one of the largest and the oldest of surveys carried out by the EU countries. Its primary objective is to obtain information on the labour market across all sectors of economy. In particular, large scale of the survey enables compilation of the detailed statistics on people in employment, unemployment and those economically inactive. The survey targets a sample of individuals or households which is representative for a given country.

The statistical method of measuring migration flows presented in this paper is based on the data obtained from the Labour Force Survey (LFS). Hereinafter in this paper, the acronym LFS will denote the Labour Force Survey in general, without a reference to a particular country, in which it is conducted. Labour Force Survey in Poland will be denoted by its Polish acronym BAEL (*Badanie Aktywności Ekonomicznej Ludności*), whereas Labour Force Survey in United Kingdom will be denoted by BLFS (British LFS).

3.1 Migrants in the LFS

The LFS has never been intended to measure mobility of people. Questionnaires permit statistical inference on both stocks of immigrants and the international flows of migrants. Identification of stocks is possible by using nationality and country or place of birth of an interviewee. The question about the country of residence one year before the survey can be used to identify flows in a transition approach, such as in censuses, by comparing the place of residence at two different points in time. Measurement of flows using transitions means that all movements in between the compared points in time are neglected. The information on changes of all members of the household can be revealed by the surviving immigrant when he or she is present at the time of the interview. Immigrants can be also identified by using questions about the country of origin and year of arrival to the destination country (i.e. the country where LFS is undertaken).

Migration in the LFS is a very rare event. Rareness of migrants in the sample can be treated as a sampling error, that is an error occurring when inference about a population is made using data from a sample. Apart from that, there are at least three other problems which may possibly influence the analysis of the results. These are (1) non-response, (2) using proxies and (3) undercount. Non-response can occur when a respondent cannot be contacted due to various reasons or refuses to answer and it can potentially bias the estimates, especially when a person who does not respond differs from those who do. The second source of potential bias is a proxy. A proxy is an answer to the questionnaire given by another related adult, who is a member of a household, if the selected respondent is not present or cannot answer. Undercount of migrants concerns the respondents who leave a country with the whole household (escape of the sample).

The key idea of this paper is to use complimentary information on migrants that can be obtained from surveys conducted in the sending and receiving country's LFS. In other words, the sampling frame in a receiving country includes (at least in theory) the whole households emigrating from the sending country. This situation is presented in Figure 3.1. The sampling frame of the BAEL captures persons migrating

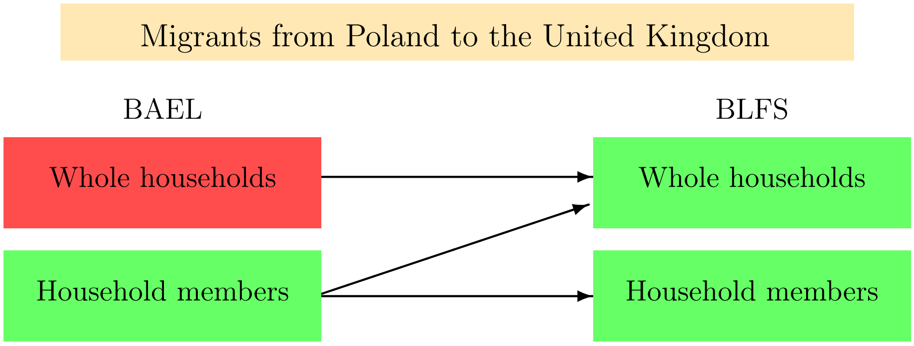


Figure 3.1: Complementarity of the LFS data from Poland and the UK

to the United Kingdom whose household members remain in Poland. The whole households relocating from Poland to the United Kingdom cannot be captured by the BAEL. But these households are included in the sampling frame of the BLFS. In the next two sections the implementation of the BAEL and BLSF and the differences between these two surveys are discussed.

3.2 LFS in Poland (BAEL) and in the United Kingdom (BLFS)

BAEL was launched in Poland for the first time in May 1992. Since the 4th quarter of 1999 it is a continuous and voluntary survey. BAEL covers the population aged 15 and older for those residing in Poland in non-institutional households. All institutional households, such as lodging-houses for employees, student halls of residence, boarding-schools, army barracks, social welfare homes or hospitals are excluded from the sampling frame.

A household is defined as 'a group of relatives or other people living together and maintaining a joint unit [... or a] person not belonging to any household and living and managing the household alone' (Eurostat, 2012, p. 37). If a person is absent for more than three months, he or she is not considered a member of a household. Foreigners who are members of the sampled households and meet the above criteria has been included in the sample since the 1st quarter of 2004 (GUS, 2012). Sampling frame of BAEL is a two-stage stratified probability sampling. All four quarterly samples are divided to 13 weekly elementary samples. In BAEL a rotation scheme is 2-(2)-2. In this scheme a household is interviewed in two consecutive quarters, then it has two quarters break (so called rest period) and then is interviewed again for the next two quarters.

In the United Kingdom, the Labour Force Survey is a part of the Integrated Household Survey (IHS). It was launched in 1973 as a biennial survey. Since 1984 it has been carried out on an annual basis and, from 1992, it has been a continuous, quarterly survey. LFS is voluntary and covers population aged 16 and older. It covers persons living in non-institutional households, and in accommodations of the National Health Service or Hospital Trust. Students living in university accommodation are sample via their parents residing in private households. Communal establishments, such as residential care homes and university halls of residence are excluded from sampling (ONS, 2011, p. 11; Eurostat, 2012).

The target population of the BLFS is based on the resident population, that is persons who consider a sampled address their main address. A private household is defined as one or more persons whose main residence is the same dwelling and who share at least one meal per day or share the living accommodation. There is no temporal criterion that defines place of main residence. It is however assumed, that if a person remains away for more than six months, he or she is no longer a resident (ONS, 2011, p. 9). The rotation scheme in BLFS is 5-, that is a household is interviewed for five quarters without any break and then leaves the sample.

3.3 Data on migrants in BAEL and BLFS

The rotation schemes applied in both BAEL and BLFS permit identification and measurement of both emigrants and immigrants. In practice, the variables that allow such identification are not disseminated, as it is the case in the BLFS. Nevertheless, the statistical method developed in this paper can be applied to estimate one way flows, namely, from Poland to the United Kingdom, in the period 2002-2008. Two definitions of a migrant are considered: (i) long-term and (ii) short-term. The former is based on the criterion that a person stayed in the receiving country for less than 12 months. In the latter, the duration criterion is three months or less. These definitions differ substantially from the definitions recommended by the UN and Eurostat, as well as from the definitions applied in producing official statistics in Poland and the United Kingdom. The BAEL and BLFS definitions imply the maximum duration of stay criteria, not minimum. Hence, it can only be approximately inferred that a person relocated within some period of time, that is, a year or a quarter. It is also unknown, how long a person intends to stay in the United Kingdom.

Table 3.1: Identification of immigrants in BLFS

Sample	2006		2007				2008	
No.	3	4	1	2	3	4	1	2
6		X	X	X	X	X		
7			X	X	X	X	X	
8				**L**	X	X	**S**	**X**
9				**L**	X	X	**S**	**X**
10				**L**		X	**S**	**X**
11				**L**			**S**	**X**
12				**L**			**S**	**X**

3.3.1 Immigrants in the BLFS

First consider measurement of immigrants for 12 months or less from Poland to the United Kingdom in the BLFS sample. For this purpose a question about a country of residence a year prior to the survey can be used. This question is asked only in the 2nd quarter and since 2005. When a person answers that he or she was residing in Poland a year before the survey, then he or she is an immigrant. This method of identification of immigrants is presented in Table 3.1. Bold letters **X** denote the survey of question being asked with the reference point in time denoted by **L**. For years 2002-2004 questions about the country of origin (distinct from the country of birth) and year of arrival to the United Kingdom can be used. Specifically, from the sample in the 1st quarter of a given year, the number of persons who arrived in the previous year is taken as a number of immigrants. These persons have stayed in the United Kingdom for at most one year. The question about the country of origin is not asked in the questionnaire since 2005, hence, there is no consistent method of measuring flows over time.

The short term immigration after 2005 in the BLFS can be identified by using a question of residence three months prior to the survey. In Table 3.1, the observations for 2008 are denoted by **S**. For years 2002-2004, from the sample in the 2nd quarter of a given year a number of persons who arrived in the same year, that is, approximately within three months prior to the survey, is taken.

The definition of an immigrant based on the question about the country of residence a year or three months prior to the survey is based on the transition approach. In this case, it is neglected where a person has been between the time points of the survey and exactly a year before. In the case of deriving counts of migrants by using country of origin and year of arrival, it is deduced that a person has resided in

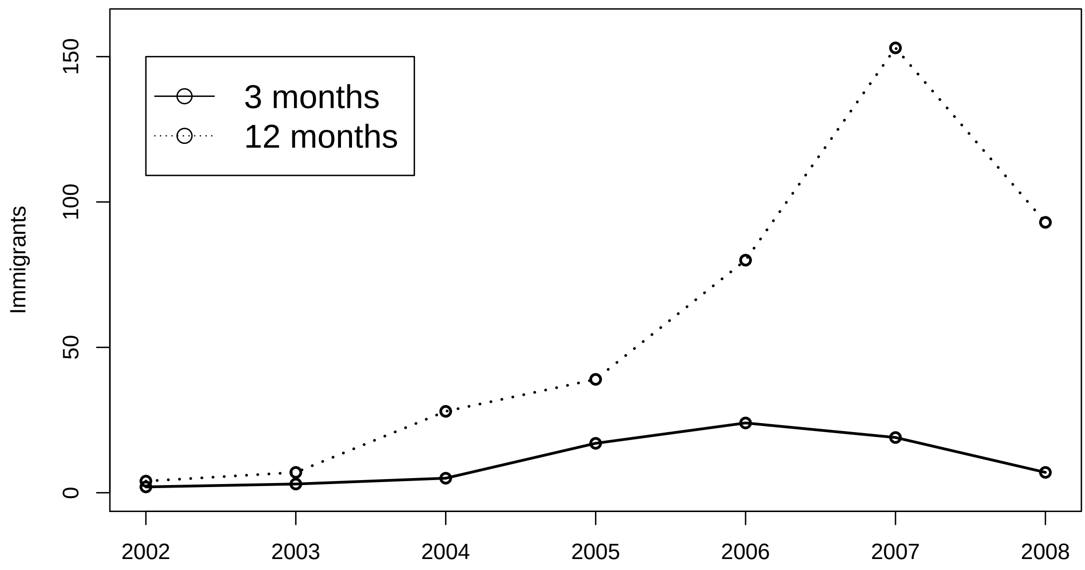


Figure 3.2: Immigrants captured in BLFS

the United Kingdom for one year (or three months in the short-term definition) at most since the arrival. Hence, this definition corresponds with the event approach to measuring flows but, in this case, is identical to the transition approach. For the sake of the availability of the data, this subtle difference between approaches in measuring flows before and after 2005 is neglected and the series obtained from two methods are combined. It seems to be a reasonable assumption that only a negligible number of persons would change a place of usual residence within a period of one year or a quarter.

The number of migrants identified in years 2002-2008 is presented in Figure 3.2. It can be observed that the number of immigrants increases from 4 in 2002 to 153 in 2007 and drops to 93 in 2008 in the 12 months definition, whereas for the three months definition the numbers are 2 in 2002 to 7 in 2008. The sample sizes in this period vary from 140000 in 2002 to around 122000 persons in 2008.

3.3.2 Emigrants in BAEL

Next consider emigration from Poland to the United Kingdom captured in BAEL. Similarly to the immigration in the BLFS, two types of emigrants can be identified: (i) short-term, that is staying for less than three months, and (ii) long-term, that is staying abroad for less than twelve months. In the case of the long-term migration, two methods of identification are proposed. One of these methods is affected by inconsistency with the BLFS counts of immigrants, namely, the definition includes persons that stay abroad for less than twelve months but longer than two or three months. This inconsistency is further explained below.

The first method relies on the question about the country of residence of all household members, both present and absent. Emigrants can be identified by comparing the corresponding quarters of two consecutive years. When a person is present in the household in the first year but absent in the next year, he or she is an emigrant. This situation for identification emigrants in year $t = 2008$ is presented in left hand side of Table 3.2. A letter **O** denotes a sample in which a person is observed, for example in samples No. 33 and 34 of the 4th quarter 2006. Then, if each of these persons is absent in the household in the 4th quarter 2007 (in samples 33 and 34) they are considered emigrants. Analogously, another person is observed in the 1st quarter of 2007 in samples No. 34 and 35 and, in the 1st quarter of 2008, found missing. This method is consistent with the transition approach to measuring migrants. The number of emigrants identified in the sample is presented in Figure 3.3 (dashed line). The respective sample size for a given sample number is around 27000 (i.e. two elementary samples). A decline from 66 emigrants in the 1st quarter 2007 to 40 identified in the 2nd quarter 2008 can be observed.

Table 3.2: Methods for identification of emigrants in BAEL

	Quarters						
Sample	2006		2007				
No.	3	4	1	2	3	4	1
32	X	-	-	X	X		
33	X	**O**	-	-	X	**X**	
34		**O**	**O**	-	-	**X**	**X**
35			**O**	X	-	-	**X**
36				X	X	-	-
37					X	X	-
38						X	X
39							X

	Quarters						
Sample	2006		2007				
No.	3	4	1	2	3	4	1
32	X	-	-	X	X		
33	X	X	-	-	X	X	
34		X	X	-	-	**O**	**X**
35			X	X	-	-	X
36				X	X	-	-
37					X	X	-
38						**O**	**X**
39							X

The second method of identification of emigrants is based on the question of the country of residence of the absent household members and the time of absence, which can be less than a year or more than a year. Should the definition of an emigrant in the BAEL be consistent with the definition of immigrant captured in the BLFS, the former answer is considered. The number of emigrants is obtained

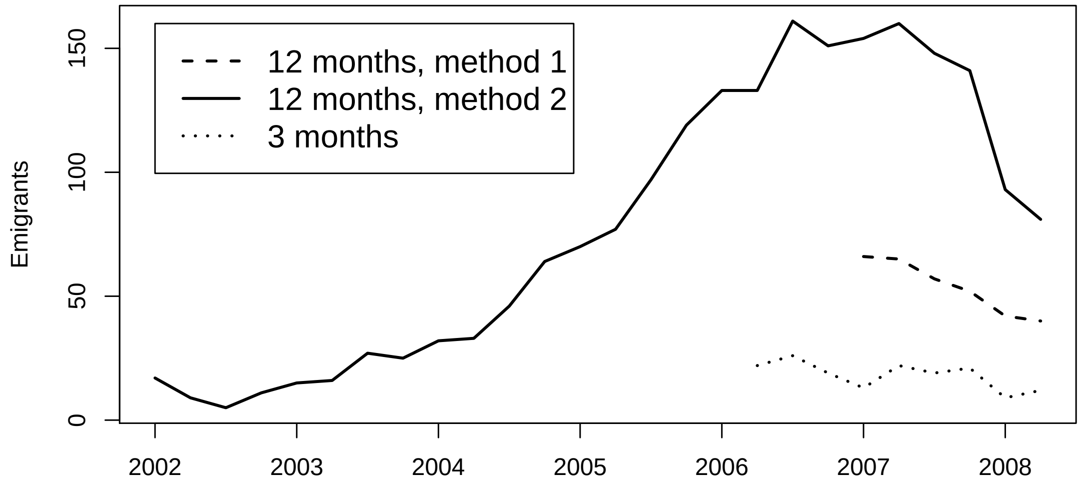


Figure 3.3: Emigrants captured in BAEL, method 2

by counting persons who meet the following two criteria: (i) they are residing in the United Kingdom at the time of a survey and (ii) are absent from the household for less than a year. The underlying 'sample-at-risk', that is the number of persons who could emigrate, can be approximated by taking the sample size of the same quarter in which the absent persons are observed and adding to it the number of all absent persons. An advantage of this method is that it can be used to identify emigrants from Poland to the United Kingdom for 2002-2008. The number of migrants identified in each quarter is presented in Figure 3.3 (solid line). The respective sample sizes range from around 44000 in 2002 to 54000 in 2008 (i.e. four elementary samples). An increasing trend can be observed, especially after Poland joining the EU in 2004. In 2008, the numbers fall sharply, which may be an indication of the crisis in financial markets.

Short-term emigrants in a given year can be identified by comparing country of residence of individuals in two consecutive quarters of the respective samples. For example, for 2008 estimate that refers to the 1 January, samples 34 and 38 can be used. This approach is presented in right hand side of Table 3.2. Due to the short duration criterion, this method of measurement is consistent with both event and transition approach for measuring migration. This method can only be applied since 2007 onwards. An advantage is that it captures the short term migration and can be used as a complimentary observation to the measure of immigrants in the BLFS identified by using three months or less duration criterion.

The dotted line in Figure 3.3 represents the number of short-term migrants counted in a given quarter of 2006 to 2008, i.e., these numbers result from pooling two elementary samples in a given quarter. For example, in the 4th quarter 2007 the number of emigrants is 13 and it has been captured in samples 30 and 34 (see Table 3.2). The overall sample size is around 26000 (two elementary samples). In comparison to the first and second method, the number of migrants identified in this method is relatively small, which is in line with the observations on immigrants yielded by the BLFS.

Identification of emigrants by comparing samples from two subsequent samples (short-term migrants and the first method for the long-term migrants) is the most accurate in reflecting the measurement by using the transition approach to counting flows of migrants. There are two moments in time where presence of persons is compared, as it is the case in censuses. The duration criterion implied by this method is also fully consistent with the one in counting immigrants in the BLFS. An important limitation is the fact, that identification of specific persons in samples before 2006 is impossible due to the coding applied for the persons absent in the household. All above described methods suffer the systematic bias. This bias comprises the households that emigrate as a whole and escape from the sample. Hence, since the interviewers do not 'chase' respondents, no information about country of residence of the persons from these households is provided.

The officially reported population sizes in Poland and United Kingdom, which are used in the subsequent sections of this paper, come from the Eurostat database. For Poland, the reference population size comprises of persons aged 15 years old or more and the sizes range from 31203 thousand people in 2002 to 32.2 million in 2008. For the United Kingdom, the sizes of the reference population (aged 16 years old or more) in the respective years are 47.4 and 49.7 million.

3.4 Using LFS in measuring migration

Since the Labour Force Surveys include questions about the country of birth, nationality, country of origin or year of arrival, they has been utilised to measure migration and analyse labour outcomes of migrants. One of the first attempts to measuring migration using BAEL concerned stocks and was undertaken in the OECD SOPEMI reports, since 1996 disseminated by the Centre for Migration Research (CMR) at the Warsaw University (see, e.g., Okólski, 1996 or Kępińska, 2007).

BAEL results 1999-2008 were used to compile a database of emigrants and returning migrants to Poland (Anacka, 2008). This database was used to assess information about those persons, who stayed abroad for longer than two or three months. The database contains 6338 emigrants and 542 returning migrants for the whole observed period. Anacka (2008) utilises the database based on BAEL to compare characteristics of Polish migrants before and after May 2004, that is the accession of Poland to the EU. To investigate which groups of people are most inclined to emigrate, she uses a migration selectivity index. The results suggest that males, as well as people between 20 and 40 years old, are characterised by the highest propensity to emigrate. It is also found that there was a large increase in migration of persons with higher education after accession to the EU.

In 2008, a special module concerning migration, based on the regulation EC 102/2007 from 2 February 2007, was carried out within the LFS in all EU countries. In Poland, the module was included in BAEL in the 2nd quarter of 2008 and was combined with the survey *Immigrants in Poland*. The questions of the module were incorporated in the BAEL standard survey, hence, all results are subject to limitations as described in previous sections. The sample size comprised 42700 persons. It is found that around 1.1% of respondents considered themselves immigrants, 96% of whom were residing in Poland permanently. This results are in line with the findings obtained in Census 2011. Respondents were also asked if they had ever resided abroad for more than three months. Around 1279 thousand persons answered affirmatively and 453 thousand of them used to be long-term migrants (i.e. they stayed abroad for more than 12 months).

The special module on migration in BAEL was combined with the study *Immigrants in Poland* (GUS, 2008). The main purpose was to survey the institutional households (except for hospitals and other medical institutions, orphanages and other educational institutions, monasteries and asylum seeker centres), as they are not incorporated in the BAEL sampling frame (see section 3.2). In this survey, 10242 immigrants from 131 countries were interviewed. The major limitation of this survey was that the sample was not representative. It has been drawn from the pre-selected households, which were known or assumed to include immigrants. For example, since the majority of the interviews were carried out in the university accommodations, most of the respondents were aged 20-24 years old and were staying in Poland temporarily.

A special module in the BLFS, *The Labour Market Situation of Migrants and their Immediate Descendants*, was carried out by the ONS in 2008 (Khan, 2009). It was aimed at describing the characteristics of immigrants in the United Kingdom, such as reason for coming, social and labour market integration and adaptation or qualifications held. It was found that the main reasons for immigration to the United Kingdom were employment, studying or family reunion or formation.

The results of the module on migration and the usual BLFS sample were used to analyse the employment of foreign workers in the United Kingdom. The analysis concerned period of arrival of immigrants (Khan and Ker, 2009). The definition of a migrant is based on the country of birth or nationality information. The results show that 81% of immigrants born in countries that accessed EU in 2004 arrived between 2004 and 2008 (Khan and Ker, 2009). This is in agreement with the data identified in the samples in the previous section, as well as studies by Anacka (2008) and Grabowska-Lusińska and Okólski (2009). Such a sharp increase in flows from these countries resulted from opening of the labour market in the United Kingdom after accession to EU in May 2004. By using the BLFS question about the month of arrival (introduced in 2008) to the United Kingdom, Khan and Ker (2009) are able to identify flows of migrants by taking 'snapshots' of stocks of foreign born population in each quarter.

The data from the BLFS are used to investigate male and female labour market participation (Khan, 2008a) and earnings of immigrants (Khan, 2008b). Khan (2008a) shows that employment of rate of working age population (that is a population aged 16 to 59 for females and 16 to 64 for males) born in Poland increased from 49% in 1998 to 94% in 2008 for males and from 31% to 79% for females. In 2008, the unemployment and inactivity rates of Polish immigrants were ones of the lowest amongst all foreign born. These results again confirm an increase in the migration driven by employment after EU accession. In the same time, median earnings of immigrants from the member states increased only barely between 1998 and 2008.

The BLFS samples have been extensively used to analyse labour outcomes of immigrants (see, e.g., Blackaby et al., 2005; Dustmann and Fabbri, 2005; Dustmann and Theodoropoulos, 2006; Shields and Wheatley Price, 1998). Drinkwater et al. (2006) use the BLFS to analyse the relative earnings of migrants. They focus on immigrants from Poland and the other 2004 EU accession countries, before and after the EU enlargement in May 2004. In the study, the duration of stay criterion is neglected. It is found that the characteristics of immigrants in the BLFS sample are similar to those reported in the Worker Registration Scheme. The results suggest that the majority of migrants from from Poland are employed in very low paying jobs despite having relatively high levels of education, comparing with the British native employees, as well as the other migrants from new member countries and the rest of the world.

An important study on evaluation of quality and potential comparability of the migration statistics on stocks and flows based on the European LFS surveys was carried out by Martí and Ródenas (2007). They focus on EU-15 countries, that is 15 EU Member States before enlargement in May 2004. The measurement of flows in this study is based on the question about the country of residence a year prior to the survey, whereas stocks are measured by using country of birth of a person. They identify the possible statistical problems that hinder utilising the LFS samples to produce statistics on migration flows. The main obstacles are bias and imprecision.

The main conclusion of the study by Martí and Ródenas (2007) is that the LFS can be used to estimate stocks of migrants in EU-15 countries, but not flows. However, the study seems to neglect the problems with the data collection methods applied across Europe. Another important issue is that it concentrates on immigration and ignores emigration measures. As it was shown in the previous section, emigrants can be identified in the LFS samples. Martí and Ródenas (2007) claim that measures of immigration cannot be perfect as the LFS has not been designed to measure migration. In the next section, a statistical framework for estimating flows from Poland to the United Kingdom by utilising counts of emigrants from the BAEL and counts of immigrants from the BLFS, is provided. The two measures are complementary in a sense that they represent the same migration flow. This complementarity has potential to improve the accuracy and reduce the bias of the estimates.

4 Model of flows

In this section a conceptual framework of a statistical model, which can serve as a tool for measuring migration flows by using data from the LFS, is presented. The adopted perspective is Bayesian. It provides a coherent framework for making statistical inference and permits analysis of the limited number of observations, with the possibility of introducing the expert knowledge by means of subjective a priori distributions. Examples of applications of expert opinion in modelling and forecasting migration see, Bijak, 2010; Bijak and Wiśniowski, 2010; Raymer et al., 2010 or Wiśniowski et al., 2010.

The model is based on the assumption that emigration is measured with bias in the sending countries. As it was described in the previous section, the source of this bias is emigration of the whole households who cannot be identified in the LFS sample in the sending country. By using complimentary observations from the LFS sample of the receiving country, this bias can be corrected and its magnitude estimated.

The model of migration flows is based on the population balance equations (**??**) and (**??**). In Equation (**??**), the population size in year $t+1$ is calculated by using events approach, in (**??**) the transition approach is applied. In this paper an example of two countries is considered. Migration flow from the first one to the second appears in balance equations for each country. Hence, it can be assumed that this flow is measured in both sending and receiving country. This simple example illustrates the idea of using complimentary observations from two countries. Further, an example of flows from Poland to the United Kingdom are estimated by using the model. This particular flow has drawn interest of demographers, migration researchers and economists, especially after the EU enlargement in May 2004.

First, consider $r = 2$ countries and category 0 as the 'rest of the world'. Let $K^{12}_{t,t+1}$ denote the true count of persons who survive in country 2 on 1st January of year t and resided in country 1 on 1st January of year $t-1$. Then, the population size P^1_t in year t in country 1 is calculated as

$$P^1_t = K^{\cdot 1}_{t-1,t} = K^{11}_{t-1,t} + (K^{21}_{t-1,t} + K^{01}_{t-1,t}) + K^{B.1}_{t-1,t} \quad (4.1)$$

$$= K^{1\cdot}_{t,t+1} = K^{11}_{t,t+1} + (K^{12}_{t,t+1} + K^{10}_{t,t+1}) + K^{D1.}_{t,t+1}. \quad (4.2)$$

A dot '$\cdot$' denotes summation over the relevant index, $K^{B.1}_{t-1,t}$ is a count of infants born in all countries since $t-1$ and and surviving in country 1 at t. Analogously, $K^{D1.}_{t,t+1}$ denotes count of deceased between t and $t+1$. Similarly, a population size in country 2 can be calculated as

$$P^2_t = K^{\cdot 2}_{t-1,t} = K^{22}_{t-1,t} + (K^{12}_{t-1,t} + K^{02}_{t-1,t}) + K^{B.2}_{t-1,t} \quad (4.3)$$

$$= K^{2\cdot}_{t,t+1} = K^{22}_{t,t+1} + (K^{21}_{t,t+1} + K^{20}_{t,t+1}) + K^{D2.}_{t,t+1}. \quad (4.4)$$

The counts $K^{12}_{t,t+1}$ in Equation (4.2), which for country 1 represent emigration in year $t+1$, are the same counts as in the balance equation (4.3) but for population P^2_{t+1} at $t+1$, where they represent immigration in year $t+1$.

Next, the following ratios are defined. First, a ratio of counts of migrants from country 1 to country 2, that is $K^{12}_{t-1,t}$, to the population at risk, that is P^1_{t-1},

$$p^{12S}_t = \frac{K^{12}_{t-1,t}}{N^1_{t-1}}. \quad (4.5)$$

Superscript S denotes that the denominator of the ratio is population size of the sending country. This ratio can be interpreted as an emigration rate (see, e.g., Preston et al., 2001). Emigration rate is a popular measure of mobility of people. Second, an immigration ratio, that is a ratio of the count of migrants $K^{12}_{t-1,t}$ to the population size of the receiving country, is defined as

$$p^{12R}_t = \frac{K^{12}_{t-1,t}}{N^2_t}, \tag{4.6}$$

where R in the superscript denotes the receiving country's population size in the denominator. This ratio cannot be interpreted in terms of a rate of immigration. The above immigration ratio can be treated as an empirical measure of probability of immigration to the given country. This ratio can, in theory, take values larger than one. Hence, it has to be assumed that the population size of the receiving country is large enough and always exceeds count of immigrants. This assumption is valid in the cases of Poland and the United Kingdom, which are considered in this paper. It seems to be a highly implausible event that flow of migrants in any direction would exceed the population sizes of these countries.

By using the above emigration rate and immigration ratio, the model of flows can be derived. The first model to consider is a binomial model. It is assumed that the rates defined in Equations (4.5) and (4.6) are the probabilities of drawing a count of migrants from the LFS sample. In the second model, log-normal-Poisson, it is assumed that the counts of migrants in the LFS are realisations of the Poisson distribution. Then, the ratios of these counts to the LFS sample sizes in the sending and receiving country, respectively, are proportional to the true rates defined in Equations (4.5) and (4.6).

4.1 Binomial model

Consider two countries, denoted by country 1 and country 2. Now let k^{12m}_t denote the count of migrants from country 1 to 2 in year t comparing to year $t-1$, observed in the LFS of the sending country, i.e. $m = S$, or in the LFS of the receiving country, i.e. $m = R$. The n^i_t denotes a sample size in the given LFS in country i, $i = 1, 2$. For simplicity of notation it is assumed that $k^{12m}_t \equiv k_t$ unless the full notation is necessary for clarity.

It is assumed that ratios $p^S_t \equiv p^{12S}_t$ as defined in Equation (4.5), and $p^R_t \equiv p^{12R}_t$ as in Equation (4.6), are probabilities of being observed as an emigrant from country 1 to country 2 and as an immigrant to country 2 from country 1, respectively. Then, the likelihood for the flows from country 1 to country 2 can be written as

$$L(k_t^S, k_t^R, n_{t-1}^1, n_t^2; p_t^S, p_t^R, \lambda) \propto \prod_{t=1}^{T} \left(p_t^R\right)^{k_t^R} \left(1 - p_t^R\right)^{n_t^2 - k_t^R} \left(\lambda p_t^S\right)^{k_t^S} \left(1 - \lambda p_t^S\right)^{n_{t-1}^1 - k_t^S}. \tag{4.7}$$

Parameter λ, which can take values in $(0, 1)$, reflects the systematic undercount which results from emigration of the whole households (see Section 3). It is assumed that this parameter reduces the true rate of emigration to match the counts observed in the LFS. In the most simple version of the model, the undercount parameter is assumed to be constant over time. Time dependence in a form of a function of time $f(t)$, as well as additional covariates Y_{tk} that can help explain the undercount, can be introduced as

$$\lambda_{f(t)} = \prod_{k=1}^{K} \phi_k^{Y_{tk}}. \tag{4.8}$$

Since the model considered herein concerns only one direction of migration flow between two countries, the above extension with additional covariates is not further developed in this paper.

From Equations (4.5) and (4.6) it follows directly that the relation between the two probabilities is

$$p_t^{12R} = \frac{N_{t-1}^1}{N_t^2} p_t^{12S}. \tag{4.9}$$

By denoting $N_t = N_{t-1}^1 / N_t^2$, the likelihood in Equation 4.7 can be rewritten as

$$L(k_t^S, k_t^R, n_{t-1}^1, n_t^2, N_t; p_t^S, \lambda) \propto \tag{4.10}$$

$$\prod_{t=1}^{T} \left(N_t p_t^S\right)^{k_t^R} \left(1 - N_t p_t^S\right)^{n_t^2 - k_t^R} \left(\lambda p_t^S\right)^{k_t^S} \left(1 - \lambda p_t^S\right)^{n_{t-1}^1 - k_t^S}. \tag{4.11}$$

By assuming the conjugate beta prior for p_t^S, its posterior characteristics can be easily obtained. The true number of migrants, K_t, can be computed by using Equation (4.5). The full Bayesian model can be written as

$$\begin{aligned} k_t^S &\sim \text{Bin}\left(n_{t-1}^S, \lambda p_t^S\right), & (4.12)\\ k_t^R &\sim \text{Bin}\left(n_t^R, N_t p_t^S\right),\\ p_t^S &\sim \text{Beta}\left(a_t^0, b_t^0\right),\\ \lambda &\sim \mathcal{U}(0, 1), \end{aligned}$$

where a_t^0 and b_t^0 are predefined constants and Bin, Beta and $\mathcal{U}$ denote the binomial, beta and uniform distributions, respectively. The prior density for the undercount λ is uniform over the range $(0, 1)$ reflecting the lack of knowledge about its magnitude.

It can be easily shown that the marginal likelihood for p_t^S in Equation (4.10) has two modes for each t. For example, when the population sizes of the two countries are equal (i.e. $N_t = 1$) and $\lambda = 1$, the likelihood is maximised at 1 and $(k_t^S + k_t^R)/(n_{t-1}^1 + n_t^2)$. Solutions that are larger than, say, a half, are nonsense from the point of view of interpretation. Therefore, it is required that the beta prior density has expected value smaller than a half. Total emigration and immigration rates for countries with large populations rarely exceed 10% of the entire populations (see, e.g. Eurostat migration database or Raymer et al., 2010). Hence, the assumption that the rate of emigration from one country to another is lower than 0.5 is reasonable. Direct derivation of the posterior density for model parameters is not possible, hence, the MCMC methods are required.

An alternative prior density considered is the uniform density for the parameter p_t^S, that is

$$p_t^S \sim \mathcal{U}\left(0, c_t^0\right). \tag{4.13}$$

It seems to be reasonable to assume that some limited number of persons migrate from one country to the other. In the case of Poland to the United Kingdom flow, this limit can be set to, say, a million per year. Then, the hyper-parameter c_t^0 can be computed as $c_t^0 = 10^6/N_{t-1}^1$. This prior reflects the lack of knowledge about the magnitude of the flow up to the predefined limit.

In both of the prior densities considered for the parameter p_t^S, it is possible to use expert opinion about the migration rate. In the case of the beta prior, the quantiles of the distribution can be elicited in a manner similar to Wiśniowski et al. (2012), that is by asking for a range of the number of migrants and certainty attached to this range. Then, the beta density for the emigration rate can easily be derived by dividing the range by the population size in a given year. In the case of the uniform prior, it is sufficient to elicit the upper limit of the migration flow magnitude and transform it to the emigration rate.

4.2 Poisson-log-normal model

In the second model, based on the Poisson and log-normal distributions, it is assumed that the counts of migrants in the LFS samples of sending and receiving countries, that is k_t^S and k_t^R, respectively, are realisations of the Poisson distribution, namely

$$k_t^S \sim \text{Po}\left(\lambda \mu_t^S\right), \tag{4.14}$$

$$k_t^R \sim \text{Po}\left(\mu_t^R\right), \tag{4.15}$$

Parameter λ, analogously as in the binomial model, represents the undercount that results from the emigration of the whole households and their escape from the sample in the LFS. Then, it is assumed that emigration rate defined in Equation (4.5) and immigration ratio in Equation (4.6) are unknown, but their imperfect equivalents are those observed in the LFS samples of the sending and receiving country, respectively, that is

$$\frac{K_t}{N_{t-1}^1} = \frac{\mu_t^S}{n_{t-1}^1} e^{\xi_t^S}, \tag{4.16}$$

$$\frac{K_t}{N_t^2} = \frac{\mu_t^R}{n_t^2} e^{\xi_t^R}, \tag{4.17}$$

where $\xi_t^k \sim N(0, \tau_k)$, $k = S, R$. Note, that here and throughout this paper $N(\mu, \tau)$ denotes the normal density with mean μ and *precision* (inverse variance) τ. The two equations can be written on the logarithmic scale these as

$$\log \mu_t^S = \log K_t + \log M_{t-1}^1 - \xi_t^S, \tag{4.18}$$

$$\log \mu_t^R = \log K_t + \log M_t^2 - \xi_t^R, \tag{4.19}$$

where $M_t^i = n_t^i / N_t^i$ is a known constant.

Prior densities for the model parameters are constructed in a manner that allows comparisons with the binomial model. For the undercount parameter λ a uniform density on the range $(0, 1)$ is assumed as in Equation (4.12). For the true flows K_t two prior densities are assumed: generalised beta and uniform. In the first case the prior is given by

$$K_t \sim \text{gBeta}(a_t^0, b_t^0; 0, N_{t-1}^1), \tag{4.20}$$

where gBeta denotes generalised beta density and the limits for migration flow are set to zero and population size N_{t-1}^1. The uniform prior distribution is given by

$$K_t \sim \mathcal{U}(0, c_t^0 N_{t-1}^1). \tag{4.21}$$

The interpretation of these two priors for K_t and elicitation procedure are similar as in the binomial model, but now it concerns the magnitude of the migration flow rather than the emigration rate p_t^S. For the precision parameters τ_S and τ_R, independent gamma densities $\Gamma(g_1, g_2)$, where g_1 and g_2 are pre-defined constants, are assumed[1]. The same hype-parameters are assumed for two measurements, which can be interpreted as lack of any knowledge about which measurement is more precise, in the sending or receiving country's LFS.

The Poisson-log-normal model seems to reflect the sampling process more realistically than the binomial model, in which the rate of emigration and immigration ratio are underlying the construction of the model. The magnitude of counts of migrants identified in the BAEL and BLFS (see Section 3.3) suggest that the Poisson distribution is adequate for modelling this type of data. The log-normal part permits introduction of additional uncertainty that results from the imperfect measurement, that is the sampling error, as well as the error resulting from the sampling scheme which is inadequate for capturing migrants in the LFS. This error includes differences in timing of measurement, that is the different quarters from which the samples are taken over time (in the BLFS, long-term migrants), as well as between the sending and receiving countries. The timing of measurement concerns also the continuous nature of the survey. As is it not carried out at some given point in time, but continuously over the given quarter, the reference point cannot be exactly determined. Measurement error includes also the effect of The construction of the binomial model does not allow for introducing the additional source of uncertainty. Both models are implemented to estimate the flows of migrants in the subsequent sections. First, the models are applied in the simulation study by using the simulated data on flows. Performance of the models is tested and compared. Then, the models are used to estimated migration flows from Poland to the United Kingdom by using the data obtained from the BAEL and LFS.

4.3 Bayesian inference

The posterior distribution of the parameters of interest is obtained by using the Bayes' theorem (Bayes and Price, 1763; see also Zellner, 1971 or Osiewalski, 2001). According to this theorem the posterior distribution of the model parameters after seeing the data is proportional to a product of the likelihood that describes the model of the phenomenon from which the data are generated, and the prior distributions, which reflect knowledge of a researcher about the model parameters before seeing the data. Posterior characteristics of model parameters are obtained by using Markov

[1]The specification of the gamma density is that the expected value of a random variable $X \sim \Gamma(a, b)$ is $\mathrm{E}(X) = a/b$ and variance is $\mathrm{Var}(X) = a/b^2$, i.e., a is shape and b is a rate parameter.

Chain Monte Carlo (MCMC) simulations. The review of these methods can be found in, e.g., Osiewalski (2001) or Robert and Casella (2004). The simulations have been implemented in R language (R Development Core Team, 2011). Additionally, OpenBUGS (Spiegelhalter et al., 2011) has been used for model development and testing.

The models developed in this section has been tested in the simulation study, where the simulated data are used to obtain the known true values of parameters. Based on this simulation study, the models are applied to estimate the flows from Poland to the United Kingdom in 2002-2008.

5 Estimation of flows from Poland to the United Kingdom

In May 2004, Poland and nine other countries from Eastern and Central Europe joined the EU. The United Kingdom, along with Ireland and Sweden, were the only countries that fully opened their labour markets for the nationals of the new member states. This led to the unprecedented after WWII migration from Poland to the UK (Grabowska-Lusińska and Okólski, 2009). Since the quality of the official statistics on migration flows produced in both countries is doubtful, the models developed in previous section are applied to estimate the number of migrants from Poland to the UK in years 2002-2008.

The weighting procedures are generally applied to ensure that the samples drawn are representative to the population. To produce estimates of migration flows based on the BAEL and BLFS data, weights for observations are not used. This strong assumption results from a number of reasons. First, in producing the data from the BAEL and BLFS the weights in both BAEL and LFS are computed according in different ways, using different stratification variables and techniques. Second, weights available in the BLFS (i.e. counts of immigrants) concern the whole total population, not the population aged 16+ as is constructed the sampling frame. Third, weights for emigrants identified in BAEL are usually unavailable as these persons are missing from the sample and information about them is usually provided by the other members of the household.

5.1 Specification of the prior densities

As in the previous sections, K_t denotes the count of migrants and p_t^S is the emigration rate from Poland to the United Kingdom. Parameter λ, which remains constant over time, measures the undercount of emigrants who left Poland with the whole households. Precisions of the error terms in the Poisson-log-normal model are denoted by τ_S and τ_R for measurement in the sending and receiving countries, respectively.

The prior density for the undercount parameter is a uniform distribution

$$\lambda \sim \mathcal{U}(0, 1),$$

which is non-informative in a sense that it assigns equal probability density to any value from the range $(0, 1)$. For example, high undercount from a range, say, $(0, 0.1)$ is equally probable as low undercount from a range $(0.9, 1)$. Hence, it reflects lack of a priori knowledge about the undercount.

For the sake of brevity, in the case of the Poisson-log-normal model, the generalised beta prior density is reduced to beta, as well as the uniform prior density $\mathcal{U}(0, c_t^0 N_{t-1}^1)$ is reduced to $\mathcal{U}(0, c_t^0)$. Various specifications of the prior densities are considered for the emigration rate p_t^S. They are presented in Table 5.3. The first two specifications

Table 5.3: Specification of the prior densities for the number of migrants.

	Spec.	$E(p_t^S)$	K_{2008}				
			p1%	p25%	median	p75%	p99%
1.	Beta(1, 1999)	0.0005	161	4620	11130	22256	73872
2.	Beta(1, 999)	0.0010	323	9243	22267	44518	147648
3.	Beta(1, 99)	0.0100	3259	93152	223985	446406	1459136
4.	$\mathcal{U}(0, 0.01)$	0.0050	3210	80258	160516	240773	317821
5.	$\mathcal{U}(0, 0.02)$	0.0100	6421	160516	321031	481547	635642
6.	$\mathcal{U}(0, 0.06)$	0.0300	19262	481547	963094	1444640	1906925

are the most conservative in assessing the expected count of migrants. The first specification, Beta(1, 1999), implies a median number of emigrants in 2008 being around 11000 with the 75th percentile being around 22000 persons. The third one, Beta(1, 99), results in the expected rate of emigration being one person out 100 and median count of emigrants being 224000. The specifications of the uniform prior presented in Table 5.3, with c_t^S being 0.01, 0.02 and 0.06, lead to the median counts of migrants of 161000, 321000 and 963000 persons in 2008. All three beta specifications imply that the larger the emigration rate, the lower probability density is. The uniform prior densities assign the same probability density to any value of emigration rate from between a specified range $(0, c_t^0)$. For instance, although the prior $\mathcal{U}(0, 0.02)$ implies the same rate of emigration, as the Beta$(1, 99)$ prior (0.01), the median of the former is around 50% larger than the latter. On the other hand, the 99th percentile of the uniform specification is 636000, but the beta prior is more diffuse with the 99th percentile being 1.46 million.

Following the results of the sensitivity analyses, the assumed prior densities for the precision terms τ_S and τ_R are gamma densities specified as follows

$$\tau_S \sim \Gamma(0.0001, 0.0001), \qquad \tau_R \sim \Gamma(0.0001, 0.0001).$$

These densities are diffuse and reflect the lack of prior information about the precision of the measurement of migration rates in sending and receiving countries.

5.2 Estimation of long-term migration flows

In Table 5.4 the counts of migrants and respective sample sizes from the BAEL, identified by using two methods as described in previous sections, and from the BLFS, together with the population sizes are presented. The first method of identifying emigrants in the BAEL applies only to years 2007 and 2008. Due to scarcity of observations, these data are not used in estimation. In both methods for identification the counts of migrants and sample sizes are the ones from the 1st quarter of a given year. In the case of immigration, for years 2002-2004 the counts come from 1st quarter, for the other years from the 2nd quarter.

Table 5.4: Long-term migrants in BAEL and BLFS.

	BLFS			BAEL (Method 1)		BAEL (Method 2)		
Year	k_t^R	n_t^2	N_t^2	k_t^R	n_t^2	k_t^R	n_t^2	N_t^2
2002	4	138816	47393102			17	60440	30959504
2003	7	133473	47685579			15	60170	31202973
2004	28	128524	48018078			32	59651	31414267
2005	39	126587	48425688			70	58794	31610437
2006	80	124106	48851108			133	57223	31796598
2007	153	123715	49258093	66	27680	154	54893	31967880
2008	93	122049	49681273	42	25955	93	53482	32103119

For both the binomial and Poisson-log-normal model the prior densities described in the preceding section are applied. The number of iterations of the MCMC algorithm is 250 thousand with the burn-in sample of the same size. With thinning of order

$k = 25$ that implies an effective sample size of 10 thousand iterations. Convergence of the algorithm is assessed by a visual inspection of the MCMC samples and statistics based on these samples, such as the running quantiles and auto-correlation functions. In all cases convergence is achieved.

First consider the binomial model with a larger dataset compiled by using the Method 2 (see Table 5.4). The posterior characteristics of the undercount parameter λ, irrespective of the prior density for p_t^S, suggest that there is no undercount resulting from the emigration of the whole households. The posterior median of λ equals in all cases 0.99 and the 1st percentile is 0.97. This results from the fact that the emigration rate in BAEL is larger than the immigration ratio in BLFS. This results suggest that (i) the binomial model or the underlying data are inadequate for estimating migration flows, (ii) the specification of the model is not reflecting the features and differences between the sampling schemes in the United Kingdom and in Poland, or (iii) emigrants from Poland usually leave the other household members at home. The last explanation seems to be rather unrealistic, especially bearing in mind that there are large differences between the sample designs of the LFS in both countries.

Similarly to the undercount parameter, the posterior characteristics of the migration counts K_t are insensitive to the specification of the prior distribution. In the right panel in Table 5.5 the estimates of the counts of migrants for years 2002-2008, based on the prior specification $\mathcal{U}(0, 0.02)$, are presented. It can be observed that the migration flows in 2002 and 2003, that is before accession of Poland to the EU, are relatively much smaller than after accession. They oscillate around 4500 persons. Then the flows increase to reach the maximum in 2007, with median estimate being 73 thousand people. In 2008 flows decrease to 45 thousand in median. The relative uncertainty of the estimates is lower for the larger flows. For example, the ratios of the 75th percentile (third quartile) to median in 2002-2003 are 1.16, while for the largest flows the respective ratios are 1.05 or smaller.

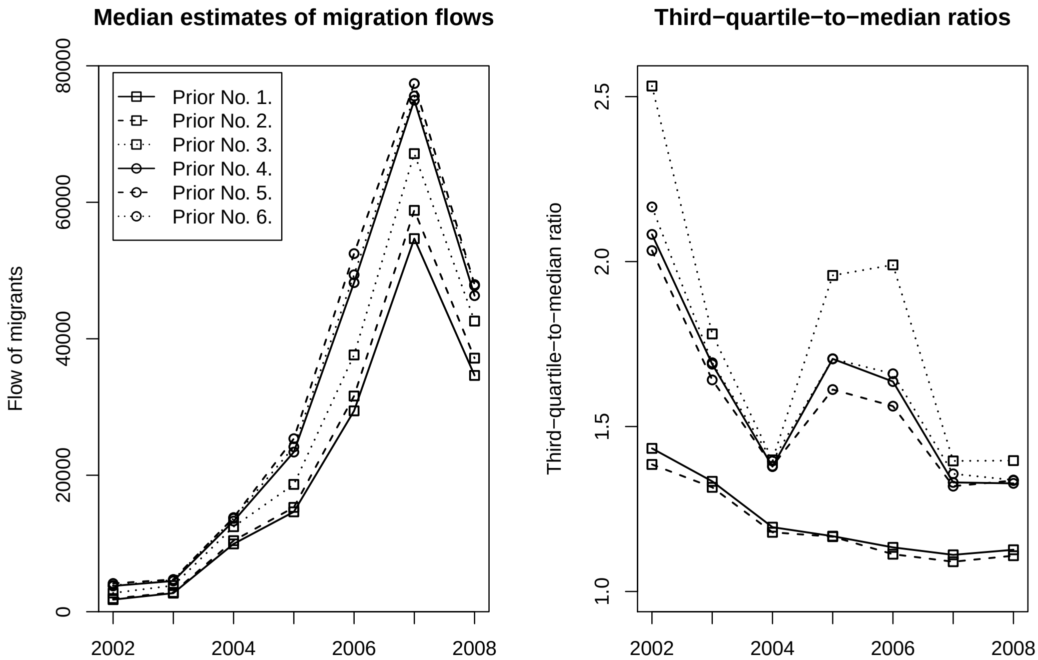
Secondly, the data from the BAEL and BLFS on long-term migrants given in Table 5.4 are analysed by using the Poisson-log-normal model. Selected characteristics of the posterior densities of the migration flow are presented in Figure 5.4. In the left-hand-side plot of the figure medians of the it is observed that the medians of the posterior distributions computed by using prior densities as given in Table 5.3 follow the same time pattern. The results based on the generalised beta prior densities (No. 1-3) are lower than the estimates based on the uniform priors (No. 4-6). In the cases with the generalised beta prior the median estimates of the migration

Table 5.5: Posterior characteristics of the long-term migration flows. Binomial model.

	2002	2003	2004	2005	2006	2007	2008
p1%	2498	2910	9688	19543	42338	62955	37584
p25%	3828	4263	12184	23035	45950	69886	42669
median	4477	4749	13377	24764	48388	73000	44959
p75%	5179	5497	14651	26430	50960	73154	47394
p99%	7139	7577	17096	30757	57169	80302	52845

Note: Method 1 and Method 2 denote the first and second method of identifying emigrants in the BAEL as described in Section 3.3.2.

flows increase with the expected value of the migration rate implied by the prior. When the uniform densities are used, the results seem to remain stable. In year 2007, which was a year of the largest migration flows from Poland to the United Kingdom, the median estimates range from around 54 thousand to 77 thousand persons. In years 2002-2003, that is before the EU enlargement, the median estimates oscillate between 1700 and 4700 persons.



Note: The order of the prior densities is as given in Table 5.3.

Figure 5.4: Posterior characteristics of the migration flows estimated by Poisson-log-normal model with the Method 2 data.

In the right-hand-side plot of Figure 5.4 the ratios of the third quartile to median of the posterior densities for K_t are presented. In general, the relative uncertainty measured by these ratios is decreasing in time. The only exceptions are years 2005 and 2006 in estimates based on the uniform priors and generalised beta No. 3. The uncertainty of estimates based on the beta distributions that imply low emigration rates is considerably lower than in the case of the Beta$(1, 99)$ specification or uniform prior densities.

In the right-hand-side plot of Figure 5.4 it can be also observed, that the uncertainty of the results based on the uniform priors (No. 4-6) is more stable and smaller comparing to the results based on the generalised beta priors (No. 1-3). Therefore, as the final estimates the estimates based on the uniform prior density $\mathcal{U}(0, 0.02)$ are selected. These results are presented in Table 5.6.

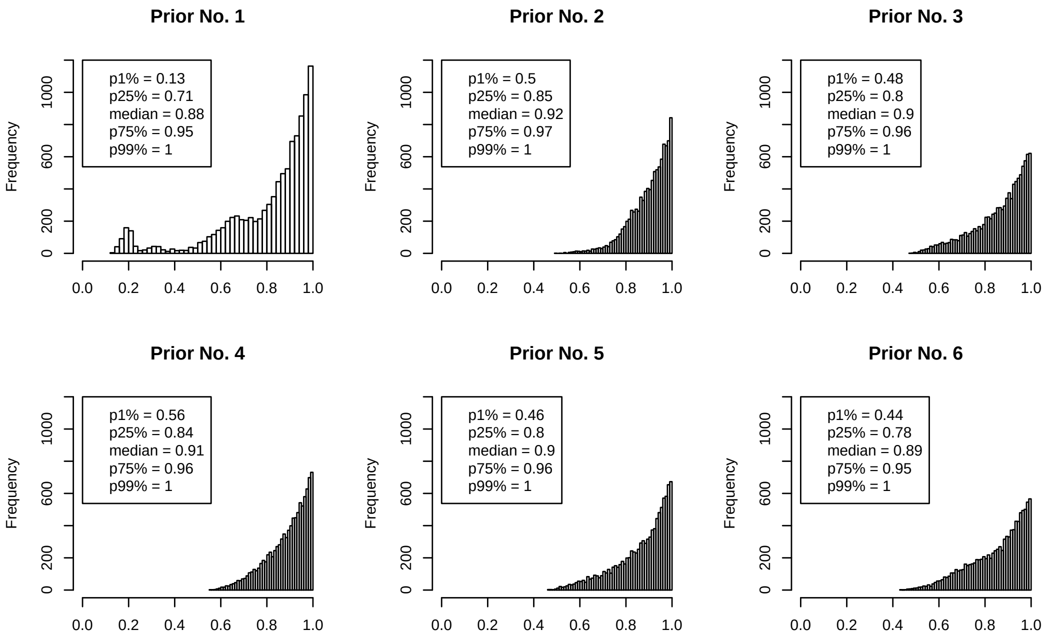
Table 5.6: Posterior characteristics of the long-term migration flows. Poisson-log-normal model.

	2002	2003	2004	2005	2006	2007	2008
p1%	649	1111	5027	8466	17656	30525	18894
p25%	1959	2850	10252	15659	32678	60350	37482
median	4142	4729	13794	25360	52479	77407	47945
p75%	8422	7762	19039	40883	81962	102118	64070
p99%	18208	17884	42019	82600	162792	221805	139330

As far as the undercount is concerned, the results seem to be stable irrespective of the prior density assumed for the emigration rate. The only exception seems to be the posterior density of λ based on the most conservative beta prior specification Beta$(1, 1999)$ (see Figure 5.5). The median estimates of the undercount oscillate around 0.9. This result suggests that the around 10% of emigrants from Poland to the United Kingdom relocated with their whole households. On the other hand, these results are not contradictory to the conclusions drawn from the analysis of the binomial model. The mode of all the posterior distributions in unity suggests that the difference between the counts of migrants in BAEL and BLFS resulting from the undercount is outweighed by the other factors that stem from the differences in the design of both surveys.

5.3 Estimation of short-term migration flows

Short-term migrants identified in the BAEL and BLFS samples are presented in Table 5.7. As it is described in Section 3.3, the duration criterion for these persons requires that they are staying in the receiving country for less than three months. The method of identification of the short-term immigrants in the BLFS is described in Section 3.3.1, whereas the identification of emigrants in the BAEL is described in



Note: The order of the prior densities is as given in Table 5.3.

Figure 5.5: Histograms of the undercount parameter λ. Poisson-log-normal model with the Method 2 data.

Section 3.3.2. Counts of emigrants from the BAEL that are used in the estimation come from the 2nd quarters of 2006-2008. Since the data on emigrants are available only for these three years, the estimation is based on three observations.

Table 5.7: Short-term migrants in BAEL and BLFS.

	BLFS			BAEL		
Year	k_t^R	n_t^2	N_t^2	k_t^R	n_t^2	N_t^2
2002	2	138816	47393102			30959504
2003	3	133473	47685579			31202973
2004	5	128524	48018078			31414267
2005	17	126587	48425688			31610437
2006	24	124106	48851108	22	27770	31796598
2007	19	123715	49258093	22	26808	31967880
2008	7	122049	49681273	12	27022	32103119

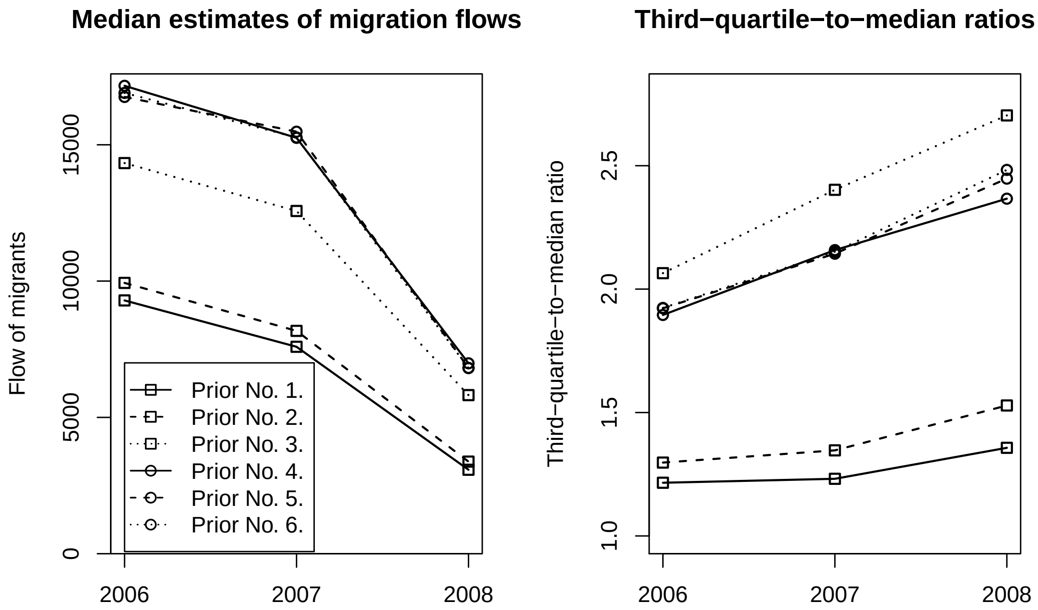
When the binomial model is used to analyse the data, the posterior characteristics of the undercount parameter λ are very much alike the results obtained for the long-term migrants. That is, they suggest that there is no undercount of persons who migrate together with their whole households. The posterior medians of the undercount parameter, in all cases of the prior density for the emigration rate (given in Table 5.3), equal 0.98, with the 1st percentiles being 0.86. The larger uncertainty of λ can be explained by the smaller number of the available observations.

As far as the posterior densities of the migration flows are concerned, they are characterised by a great stability with respect to the specification of the prior distribution for the emigration rate p_t^S. The selected results based on the uniform specification $\mathcal{U}(0, 0.02)$ are presented in Table 5.8. It is observed that the median estimates of migration are decreasing over time from around 14 thousand in 2006 to 6 thousand in 2008. The counts of migrants are considerably lower than the estimates of the long-term migrants for the respective years. The difference is due to the duration criterion. It is expected that the number of persons who reside in the United Kingdom for a at most three months by the time of survey is smaller than the number of persons who reside there for no longer than 12 months. The latter group includes also the former, hence, it should be at least as numerous.

Table 5.8: Posterior characteristics of the short-term migration flows. Binomial model.

	2006	2007	2008
p1%	9618	8610	3434
p25%	12591	11143	5155
median	13913	12438	5954
p75%	15186	13862	6917
p99%	19325	17580	9728

The analysis of data on short-term migration flows with the Poisson-log-normal model yields the results similar in interpretation as for the long-term migration flows. The selected posterior characteristics of the migration flows are presented in Figure 5.6. The median estimates of flows show a downward sloping trend as in the results from the binomial model. Median estimates vary with respect to the specification of the prior distribution. While the only difference between the Beta$(1, 1999)$ and Beta$(1, 999)$ is slightly larger uncertainty of the latter, Beta$(1, 99)$ yields medians by almost 50% larger than the other beta specifications with much larger uncertainty of the estimates – the third quartiles are more than twice larger than medians. On the other hand, the results based on the uniform specifications are stable both in terms of both median estimates and he uncertainty. The uniform-prior-based median migration flows are considerably larger than the ones obtained with the beta priors. The detailed posterior characteristics of migration flows for the specification $\mathcal{U}(0, 0.02)$ are presented in Table 5.9. The median estimates of flows are larger than in the binomial model by around 3000 persons for 2006 and 2007, and by around 900 persons for 2008. The posterior distributions in the Poisson-log-normal have very heavy right tails. The 99th percentiles exceed 200 thousand persons for 2006 and 2007. In the case of the binomial model that percentiles are around 17 to 19 thousand, respectively.



Note: The order of the prior densities is as given in Table 5.3.

Figure 5.6: Posterior characteristics of the short-term migration flows estimated by Poisson-log-normal model.

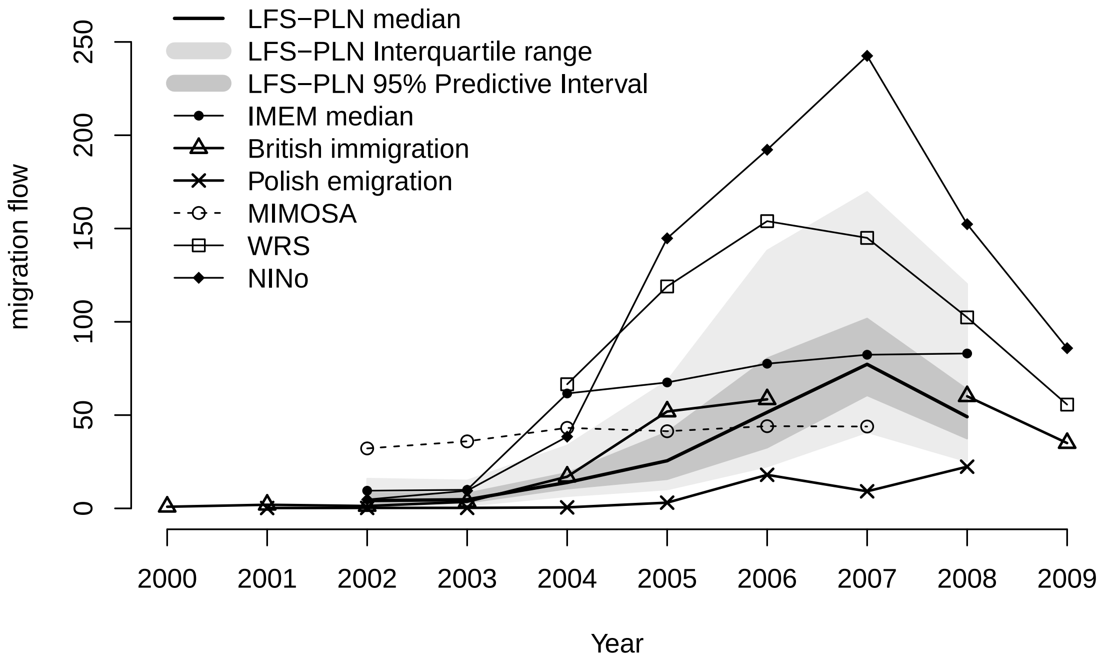
Table 5.9: Posterior characteristics of the short-term migration flows. Poisson-log-normal model.

	2006	2007	2008
p1%	2755	2270	822
p25%	9648	7861	3125
median	16901	15269	6809
p75%	32515	32853	16907
p99%	224418	236351	114237

The undercount estimates are insensitive to the specification of the prior for the emigration rate. The median estimates of the undercount are around 0.78. This result suggests that around 22% of migrants arrived to the United Kingdom with their whole households and have not been captured in the BAEL in Poland as emigrants. A rather surprising characteristic is that the histograms of the undercount parameter for short-term migration flows are more diffuse than for the long-term migration flows estimated by using only two observations from 2007 and 2008 in the Method 2 data. This result is likely caused by the larger variability of the data on short-term migrants.

5.4 Comparison of the results with the reported statistics

The officially reported statistics on flows from Poland to the United Kingdom relate to the intended permanent migrations in the Polish register, and to the actual migrations for more than 12 months in the British statistics. This results in the large discrepancies in the observed numbers. In Figure 5.7, emigration reported by Polish CSO is denoted by a solid line with crosses, while the LTIM figures from the British ONS – by a solid line with triangles. For the latter, data for 2007 are missing.



Source: LFS-PLN and IMEM data are based on own elaboration; British and Polish migration data come from Eurostat statistics database, WRS data come from the Home Office website (`http://www.homeoffice.gov.uk/science-research/`) and concern only the registrations for the first job of each applicant; NINo data come from the Department for Work and Pensions website (`http://www.dwp.gov.uk/research-and-statistics/`); MIMOSA estimates received after personal communication with their author James Raymer.

Figure 5.7: Comparison of the results from the Poisson-log-normal model (LFS-PLN), IMEM, MIMOSA and the reported flows from Poland to the United Kingdom (in thousands).

We observe that the Polish emigration is significantly smaller than the British immigration. There were 22 thousand emigrations officially recorded by Polish CSO in 2004-2006, comparing to 127 thousand immigrations measured by the British IPS. In 2006, the number of emigrants reported in the Polish statistics is considerably larger comparing to 2007. This figure could be presumably attributed to the requirement of deregistration from the Polish register in order to avoid double taxation in both countries (Journal of Laws [Dz. U.] 2006 No. 226 item 1646). In 2007 an agreement about avoidance of double taxation came into effect (Journal of Laws [Dz. U.] 2006 No. 250 item 1840), thus, the number of deregistrations fell considerably.

MIMOSA estimates fail to capture the effect of the EU accession. Estimated total flows between 2004 and 2006 is around 129 thousand, which is just as the reported figures by the British IPS. However, similar annual flows are estimated for years prior to the EU accession. Moreover, the level of annual migration flows is lower than median respective total flows of 210 thousand from the IMEM model. Both MIMOSA and IMEM estimates of flows from Poland to the United Kingdom are not comparable with the LFS-based estimates, as they concern flows for 12 months or more.

Another reported statistics to compare with come from the WRS (see joint report of Home Office Border and Immigration Agency, Department for Work and Pensions, HM Revenue and Customs and Communities and Local Government, 2009). We observe that the WRS numbers are significantly larger than the estimates based on the LFS. However, the WRS is a biased source as it does not include spouses of the workers, the self-employed or those have been working legally in the United Kingdom for 12 months. For instance, the estimated proportion of the self-employed immigrants from the A8 countries is 14% (Pollard et al., 2009). The registration related to employment and not to the immigration or emigration event. Moreover, the information collected upon registration concerned the nationality, not country of previous residence.

The results of the survey carried out by Pollard et al. (2009, p. 18) suggest that around 42% of Polish migrants who have worked in the United Kingdom after 2004 and returned to Poland never registered at the WRS. The results of other surveys mentioned by Pollard et al. (2009) suggest that the A8 immigrants who did not register at the WRS constituted around 25-33% of the WRS figures. Pollard et al. (2009) also estimate that the total number of migrant workers from A8 countries who arrived between May 2004 and December 2007 to around 1 million.

The similar conclusions as for the WRS data can be drawn about the data on the National Insurance Numbers (NINo). The NINo data are even larger than the WRS (except for 2004), reaching their peak of 250 thousand people in 2007. The time pattern of these data follows the one observed in the LFS data-based estimates. The NINo data can, though, be imprecise, mainly due to the postponement of registration and only single recording of a given person. The data are disaggregated by nationality, hence, they can also include the persons who resided in the United Kingdom without permission prior to the EU enlargement. On the other hand, it is estimated that 13% of A8 nationals have not obtained their NINo (Pollard et al., 2009, p. 19).

According to the estimates quoted in Grabowska-Lusińska and Okólski (2009), 1.1 million persons left Poland to the other EU countries within 24 months after the EU accession. They migrated predominantly to the United Kingdom, Ireland and Germany. These results are rather weakly supported by the results obtained from the IMEM and LFS-based models, as well as the data from the WRS and NINo sources, where the number of Polish applicants were 340 thousand and 366 thousand. Even the estimated 1 million of migrants from A8 countries by Pollard et al. (2009) does not seem to support these calculations, as this number concerns all new member countries and relates to migrations between 2004 and December 2007.

6 Conclusions

The model developed in this paper provides a framework for estimating migration flows between two countries with the Labour Force Survey data. Various methods of identification of migrants in the samples are presented and applied to the Polish and British LFS. In the context of missing or unreliable data available from the national statistical offices, as well as the requirement of the Eurostat to provide harmonised statistics on migration, this tool can be applied to produce, harmonise or enhance the statistics on migration flows. An advantage of the models is a great simplicity and the fact that the LFS is carried out in all member states of the EU, EFTA and candidate countries.

The Bayesian approach used for analysing the models offers a flexibility of handling limited number of observations available from the LFS samples. It also provides a coherent framework for integrating data and the expert opinions in form of the prior distributions.

The estimated total flows from Poland to the United Kingdom after the EU enlargement is around 230 thousand people. In 2002 and 2003 there were only about 9 thousand migrants in total. This number concerns migrants aged 16+, who at the time of survey were residing in Britain for less than 12 months. In the Polish official statistics, the number of persons who permanently emigrated from Poland was around 50 thousand (500 migrations in 2002-2003). The officially estimated immigration of persons for 12 months or longer in the United Kingdom was around 250 thousand, comparing to 5000 in 2002-2003. However, the data reported by the other British sources, such as Workers Registration Scheme and National Insurance Numbers, are larger. Since the reported statistics and the estimates from the model concern persons for whom duration criteria are different, the numbers are not comparable.

The model has some shortcomings. First, the produced flows may be underestimated which results from the fact that the model does not reflect the differences between the design of the LFS in countries, various non-response rates or proxies in the responses. Second, the estimates rely on the population at risk, which, in this case, is the population of Poland as reported by the registers. But if we believe that the number of people going from Poland to the UK is different from the reported one, then it obviously affects the underlying population size. Third, the population at risk used for the estimates differs from the one in the LFS sampling scheme as, e.g., Polish LFS is not carried out in the institutional households. Hence, the reference population should be lower than the officially reported but it is difficult to estimate the total number of people residing in these households.

The method developed in this paper can be extended in various ways. The most obvious is adding more countries to the matrix of flows. Then, it would be possible to include additional explanatory variables for the counts of migrants and the undercount parameters. The undercount can also depend on time. Another option is to treat the undercount as a random effect, the precision of which depends on the exogenous covariates. These covariates can include the non-reponse, refusal or proxies rates, the aggregate information about the migrants or the LFS design in a given country. Explanatory variables for the migration model can draw from experiences in projects such as MIMOSA or IMEM and works by Jenissen (2006), Abel (2010) or Raymer et al. (2011).

Summarizing, the application of the binomial and Poisson-log-normal model to the data obtained from the BLFS and BAEL leads to two conclusions: (i) estimated migration flows seem to be realistic and reflect the general patterns observed in the alternative sources of data on migration, and (ii) the assumption that immigration are captured without bias in the receiving country's LFS may not be valid.

References

Abel, G. J. (2009), International Migration Flow Table Estimation, PhD thesis, University of Southampton.

Abel, G. J. (2010), Estimation of International Migration Flow Tables in Europe, *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 173(4), 797-825.

Bell M, Blake M, Boyle P, Duke-Williams O, Rees P, Stillwell J, Hugo G (2002) Cross-national comparison of internal migration:issues and measures. Journal of the Royal Statistical Society: Series A 165:435-464.

Bilsborrow R.E., Hugo G., Oberai A.S. and H. Zlotnik (1997) International Migration Statistics. Guidelines for improving data collection systems. International Labour Organization. United Nations, Geneva

Bilsborrow R (2007) Surveys of International Migration: Issues and Tips. Sixth Coordination Meeting On International Migration, New York, 26-27 November 2007

Brown S. K. and F.D. Bean (2005) International Migration, in: Poston D. L. and Micklin M. (eds.) Handbook of Population. XIII ed., Springer Handbooks of Sociology and Social Research.

Carletto C. and A. de Brauw (2007) Measuring migration using household surveys, Operational Note No. 2, Migration Operational Vehicle, The World Bank

Castles S and M.J. Miller (1998) The Age of Migration. International Population Movements in the Modern World, 2nd Edition, Basingstoke, UK

Courgeau D (1973) Migrants et migrations. Population 28:95-129. Also in English: Migrants and migrations. Population, Selected Papers (1979), 3:1-35

De Beer J. (2007) Forecasting international migration: time series projections vs argument based forecasts. In: Raymer J. and F. Willekens (eds.) International Migration in Europe: Data, Models and Estimates. London, Wiley

DeWaard, J., Kim, K., and Raymer, J. (forthcoming), Migration Systems in Europe: Evidence from Harmonized Flow Data, *Demography.*

Dumont J.-C. and G. Lemaître (2008) Counting foreign-born and expatriates in OECD countries: a new perspective, in: Raymer J. and F. Willekens (eds.) International Migration in Europe: Data, Models and Estimates. London, Wiley

Eurostat, (2005) Consumers in Europe, Facts and Figures. Data 1999-2004, Office for Official Publications of the European Communities, Luxembourg.

EU (n.d.) Statistics Explained: Glossary. European Commission, Eurostat. http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Category:Glossary, accessed in April 2012

Gelman A. (2006) Prior distributions for variance parameters in hierarchical models. Bayesian Analysis, 1, Number 3, pp. 515-533.

Grabowska-Lusińska and Okólski M (2009) Emigracja ostatnia? [Last emigration?] Scholar, Warsaw.

Herm, A. (2006). Recommendations on international migration statistics and development of data collection at an individual level. In M. Poulain, N. Perrin, and A.

Singleton (eds.), Towards the Harmonisation of European Statistics on International Migration (THESIM), Chapter 2, pp. 77 - 107. Louvain-La-Neuve, Belguim: UCL-Presses Universitaires de Louvain.

Hoinville G. and Jowell R. (1978) Survey research practice, Heinemann Educational Books, London

IOM (2011) International Migration Law No.25 - Glossary on Migration, International Organization for Migration, 2nd Edition, Geneva

Jennissen, R. (2004), *Macro-Economic Determinants of International Migration in Europe*, PhD Thesis, Rijksuniversiteit Groningen.

Ker D., Zumpe J., Blake A. (2009) Estimating International Migration: An exploration of the definitional differences between the Labour Force Survey, Annual Population Survey, International Passenger Survey and Long-Term International Migration. Office for National Statistics. United Kingdom.

Knight, P. T. (2002) Small-Scale Research. Pragmatic Inquiry in Social Science and the Caring Professions. SAGE Publications Ltd, London

Kraler , A. (2006). Recommendations on international migration statistics and development of data collection at an individual level. In M. Poulain, N. Perrin, and A. Singleton (Eds.), Towards the Harmonisation of European Statistics on International Migration (THESIM), Chapter 2, pp. 77 - 107. Louvain-La-Neuve, Belguim: UCL-Presses Universitaires de Louvain.

Kupiszewska D (2009) Country Report Poland, Promoting Comparative Quantitative Research in the Field of Migration and Integration in Europe (PROMINSTAT), May 2009.

Kupiszewska, D. and B. Nowok (2008). Comparability of statistics on international migration flows in the European Union. In: Raymer J. and F. Willekens (eds.) International Migration in Europe: Data, Models and Estimates. London, Wiley

Kupiszewska D., Kupiszewski M., Martí M., Ródenas C. (2010) Possibilities and limitations of comparative quantitative research on international migration flows, Promoting Comparative Quantitative Research in the Field of Migration and Integration in Europe (PROMINSTAT), Working Paper No. 04.

Lemaître G., Liebig T. and C. Thoreau (2006) Harmonised statistics on immigrant inflows – preliminary results, sources and methods, Organisation for Economic Co-operation and Development, Paris

Martí M. and C. Ródenas, (2009) Estimating False Migrations in Spain, Population-E, 64(2), pp.361-376.

Massey, D.S., J. Arango, G. Hugo, A. Kouaouci, A. Pellegrino and J.E. Taylor (1998) Worlds in Motion: Understanding International Migration at the End of the Millennium. Oxford: Clarendon Press.

Missiakoulis S. (2010). Cecrops, King of Athens: the First (?) Recorded Population Census in History. International Statistical Review 78 (3): 413-418.

Nowok, B., D. Kupiszewska, and M. Poulain (2006). Statistics on international migration flows. In M. Poulain, N. Perrin, and A. Singleton (Eds.), Towards the Harmonisation of European Statistics on International Migration (THESIM), Chapter 8, pp. 203-233. Louvain-La-Neuve, Belguim: UCL-Presses Universitaires de Louvain.

Nowok B. (2010) Harmonization by simulation: a contribution to comparable international migration statistics in Europe. PhD thesis, University of Groningen

OECD (2006) International Migration Outlook 2006, Organisation for Economic Cooperation and Development Publishing, Paris

ONS (2011) Labour Force Survey User Guide – Volume 1: Background and Methodology, Office for National Statistics, Version 1.0 August 2011

ONS (2012) Migration Statistics Improvement Programme Final Report. Office for National Statistics. UK

ONS (n.d.) Reconciliation of ONS estimates: Comparisons of combined IPS (long and short term migration) estimates with administrative data sources. Office for National Statistics. United Kingdom.

Osiewalski J. (2001) Ekonometria bayesowska w zastosowaniach [Bayesian Econometrics in Applications], Economic Univeristy in Cracow, Poland.

Poulain, M. (2006). Registration of the Resident Population. In M. Poulain, N. Perrin, and A. Singleton (Eds.), Towards the Harmonisation of European Statistics on International Migration (THESIM), Chapter 2, pp. 109 - 131. Louvain-La-Neuve, Belguim: UCL-Presses Universitaires de Louvain.

Raymer J, de Beer J, van der Erf R (2011b) Putting the pieces of the puzzle together: Age and sex-specific estimates of migration amongst countries in the EU/EFTA, 2002-2007. European Journal of Population, 27, (2), 185-215.

Raymer J. and P.W.F. Smith (2010) Modelling migration flows, Journal of the Royal Statistical Society: Series A (Statistics in Society), Vol. 173, No. 4., pp. 703-705

Raymer J, Forster JJ, Smith PWF, Bijak J, and A Wiśniowski (2012) Integrated Modelling of European Migration: Background, specification and results. NORFACE MIGRATION Discussion Paper No. 2012-04

Rees, P. H. (1977) The measurement of migration from census data and other sources. Environment and Planning A 9, no. 3: 247-272.

Rogers, A. 2008. "Demographic Modeling of the Geography of Migration and Population: A Multiregional Perspective." Geographical Analysis 40:276-296.

Singleton Ann, Audrey Lenoël, Olga Gora and Lynnmarie Sardinha (2010a) Country Report Ireland, Promoting Comparative Quantitative Research in the Field of Migration and Integration in Europe (PROMINSTAT), April 2010.

Swanson, D.A. and Stephan, G.E. (2004). A Demography Time Line. In The Methods and Materials of Demography, 2nd ed. Eds. D. Swanson and J.S. Siegel. Amsterdam: Elsevier Academic Press, 9-42.

United Nations (1998) Recommendations on statistics of international migration. Statistical Papers Series M, No. 58, Rev.1, Department of Economic and Social Affairs, Statistics Division, United Nations, New York.

United Nations (2008) Principles and Recommendations for Population and Housing Censuses, Revision 2, Department of Economic and Social Affairs, Statistics Division, United Nations, New York.White M.J. and D. P. Lindstrom (2005) Internal Migration, in: Poston D. L. and Micklin M. (eds.) Handbook of Population. XIII ed., Springer Handbooks of Sociology and Social Research.

Wright J.D. and Marsden P.V. (2010) Survey Research and Social Science: History, Current Practice and Future Prospects. In: Marsden P.V. and Wright J.D., Handbook of survey research, 2nd Edition, Emerald Group Publishing, Bingley.

Willekens F (1982) Identification and measurement of spatial population movements. In: A National Migration Survey, Manual X: Guidelines for Analysis, United Nations ESCAP: New York, 74-97

Willekens F. (2008) Models of migration: observations and judgements, in: Raymer J. and F. Willekens (eds.) International Migration in Europe: Data, Models and Estimates. London, Wiley

Wiśniowski A, Keilman N, Bijak J, Forster JJ, Smith PWF, Christiansen S and Raymer J (2012) Augmenting migration statistics with expert knowledge. NORFACE MIGRATION Discussion Paper No. 2012-05

Zlotnik H. (1987) The concept of international migration as reflected in data collection systems. International Migration Review 21: 925-946.

Zlotnik, H. (1992) Empirical Identification of International Migration Systems. Pp. 19-40 in International Migration Systems: A Global Approach. Edited by M.M. Kritz, L.L. Lim and H. Zlotnik. Oxford: Clarendon Press.