NORFACE MIGRATION

**NORFACE MIGRATION** Discussion Paper No. 2012-05

# Augmenting migration statistics with expert knowledge

Arkadiusz Wisniowski, Nico Keilman, Jakub Bijak, Solveig Christiansen, Jonathan J. Forster, Peter W.F. Smith and James Raymer

www.norface-migration.org

# Augmenting migration statistics with expert knowledge

Arkadiusz Wiśniowski*, Nico Keilman†, Jakub Bijak*,
Solveig Christiansen†, Jonathan J. Forster*
Peter W.F. Smith* and James Raymer*

May 22, 2011

Paper prepared for IMEM Workshop, Chilworth, 25 - 27 May 2011

**Abstract**

International migration statistics vary considerably from one country to another in terms of measurement, quality and coverage. Furthermore, immigration tend to be captured more accurately than emigration. In this paper, we first describe the need to augment reported flows of international migration with knowledge gained from experts on the measurement of migration statistics, obtained from a multi-stage Delphi survey. Second, we present our methodology for translating this information into prior distributions for input into the Integrated Modelling of European Migration (IMEM) model, which is designed to estimate migration flows amongst countries in the European Union (EU) and European Free Trade Association (EFTA), by using recent data collected by Eurostat and other national and international institutions. The IMEM model is capable of providing a synthetic data base with measures of uncertainty for international migration flows and other model parameters.

## 1 Introduction

In order to fully understand the causes and consequences of international movements in Europe, researchers and policy makers need to overcome the limitations of the various data sources, including inconsistencies in availability, definitions and quality. In this paper, we describe the obtainment and development of expert-based prior distributions for use in the Integrated Modelling of European Migration (IMEM) model, which has been developed to estimate international migration flows between the 31 countries in the European Union (EU) and the European Free Trade Association (EFTA). The model both harmonises and corrects for inadequacies in the available data and estimates the completely missing flows with the aim of providing the best statistics possible with measures of uncertainty. The IMEM model is framed within a Bayesian statistical setting.

Bayesian statistical methods are particularly adept at handling data from different sources and are ideal for situations in which the data are inadequate or missing

*Southampton Statistical Sciences Research Institute, University of Southampton
†Department of Economics, University of Oslo

because additional expert information can be included in the form of distributions reflecting the expert beliefs and judgements. The resulting estimates are based on distributions from the combination of expert beliefs and other available information, including all relevant data sources and covariate information. These distributions can also be used to quantify the uncertainty in the estimates, providing governments and planning agencies valuable information to design their policies directed at supplying particular social services or at influencing levels of migration.

The structure of this paper is as follows. First, we describe our attempts to augment reported flows of international migration with knowledge gained from experts on the measurement of migration statistics, obtained from a multi-stage Delphi survey. Second, we present our methodology for translating this information into prior distributions for input into the IMEM model. The paper ends with a summary.

# 2 Obtaining Expert Information

In the IMEM project, we have designed a migration model where expert opinion can be conveniently incorporated and estimates and measures of precision efficiently computed. The Bayesian approach permits expert opinion to be combined with the data to strengthen the inference. The Bayesian approach also facilitates the combination of multiple data sources, with their differing levels of error, as well as prior information about the structures of the migration processes, into a single prediction with associated measures of uncertainty. Given the substantial inconsistencies in reported migration data (see, e.g., Poulain et al. 2006), the elicitation of expert opinion concerning various aspects of model specification and data are critical for the success of the project.

The elicitation of prior information involves specifying the quality of data sources, the differences in definitions and the role of the explanatory variables. Some information is elicited from external experts, while other information is provided by team members. The experts are asked to rate the credibility they give to different types of data from different types of data collection mechanisms (e.g., survey versus register) and emigration versus immigration. Further, the experts are asked about rates of bias (e.g., systematic undercount) in the migration flows. Each expert is asked to supply a set of distributions representing their beliefs about certain model parameters. The totality of expert opinions can then be combined into a single set of distributions, allowing for the introduction of yet another source of uncertainty, related to the heterogeneity of experts.

To keep it under control, a multi-stage process of extraction (elicitation) of expert judgement is used, within a Delphi survey framework, whereby the expert opinions are allowed to converge towards a common consensus. However, expert knowledge about data collection systems is heterogenous. Hence, in order to obtain an overall assessment of the systems across Europe and not to reduce the uncertainty regarding their characteristics, we are not aiming at convergence of the expert opinions. The purpose of the adaptation of the Delphi survey in this study is to provide a convenient framework for exchange of opinions and views.

## 2.1 Delphi technique

The Delphi technique is a method used to obtain information from a group of experts in order to make judgments and forecasts, when extensive or reliable data in the field of enquiry is not available (Rowe and Wright, 1999). It was first developed by the RAND Corporation for US military use in the 1950s.

The elicitation of expert opinions takes the form of an anonymous questionnaire with multiple rounds where the experts report their subjective beliefs. Between rounds, experts are given feedback informing them of the answers in the preceding round and arguments given in support of these answers. The experts then complete the next round of the survey where they are free to alter their previous answers in the face of the new information provided by the feedback.

According to Rowe and Wright (2001) the Delphi technique is most reliable where there are between 5 and 20 respondents who are experts in the field of enquiry and there is heterogeneity among the experts. The questions should be long enough to contain the relevant information but not cause information overload. The answers given in the final round of the survey are then used as input into the model. These answers are usually weighted equally but they can also be weighted differently if knowledge about the respondents' expertise gives good reason to do so.

Evaluations have shown that the answers from the final round Delphi surveys are more accurate than set-ups using only one expert, traditional groups or single-round questionnaires (Rowe and Wright, 2001). By using an anonymous questionnaire instead of a group meeting, one avoids group pressure and some individuals who dominate the group. The Delphi method may also lead to better results because the experts think more carefully when responding when they know that their answers will be given as feedback to other experts.

The Delphi technique was used in the IDEA (Mediterranean and Eastern European Countries as new immigration destinations in the European Union) project where the goal of the a Delphi survey was to provide qualitative input to a forecasting model (Wiśniowski and Bijak, 2009, Bijak and Wiśniowski, 2010) and the MIGIWE (Migration and Irregular Work in Europe) project where a Delphi survey was employed to elicit expert knowledge on irregular foreign employment in Austria following the 5th Enlargement of the EU (Jandl et.al, 2007).

## 2.2 Constructing a questionnaire

The elicitation process involved 11 external experts. The online questionnaire was pre-tested by additional two external experts and two IMEM team-members. The survey was preceded by an invitation letter, in which its aim and the purpose of the project were explained.

We asked the experts to give their opinion about how specific measurements of international migration deviates from a benchmark. As the benchmark, we adopted the United Nations definition of a long term migrant:

> *A person who moves to a country other than that of his or her usual residence for a period of at least a year (12 months), so that the country of residence effectively becomes his or her new country of usual residence. From the perspective of the country of departure, the person will be a long-term emigrant and from that of the country of arrival, the person will be a long-term immigrant.* (UN,1998).

The place of usual residence is in the same UN publication defined as *The country in which a person lives, that is to say, the country in which he or she has a place to live where he or she normally spends the daily period of rest. Temporary travel abroad for purposes of recreation, holiday, visits to friends and relatives, business,* *medical treatment or religious pilgrimage does not change a person's country of usual residence.* This definition of place of usual residence does not have an explicit time dimension. In the UN recommendation for population and housing census (2008) place of usual residence is however defined as:

> *The place at which the person has lived continuously*:
>
> a) *for most of the last 12 months (that is, for at least six months and a day) not including temporary absences for holidays or work assignments or intends to live for at least 6 months, or*
>
> b) *for at least the last 12 months not including temporary absences for holidays or work assignments or intends to live for at least 12 months.*

We haven chosen not to define explicitly place of usual residence, as the country of residence will be the same in almost all cases, whether one uses one or the other option from the UN census definition.

In theory, the UN definition we have adopted includes undocumented ('illegal') migrants. In practice, the migration statistics in most countries do not cover undocumented migrants. When we refer to the UN definition as the benchmark, we do however include undocumented migrants.

### 2.2.1 Round 1 questionnaire

The final questionnaire in Round 1 consisted of a definition of a long-term migrant according to the UN definition discussed above (UN, 1998) and 14 questions grouped into four sections. Each section contained a specific type of questions and an open question, in which experts were allowed to express their comments or arguments related to their answers. In all questions, experts were asked to provide their answers in terms of a range of percentages, which concerned various phenomena depending on a question (explained in detail below), and to state how certain they were about a given range. The levels of certainty that could be chosen were 50%, 75%, 90% and 95%. The experts were also given a choice to state a different percentage. Sections A, B and C were restricted to intra-EU/EFTA migrants, Section D concerned rest of world migration to and from the EU/EFTA countries. At the end, experts were allowed to provide general comments or suggestions, as well as to ask questions of their own. The Round 1 questionnaire is attached in the Appendix.

The undercount of migration between EU / EFTA countries and from / to the rest of the world were the focus of Sections A (Questions 1-3) and D (Questions 12-14), respectively. Here, experts were asked to provide their judgements and uncertainty regarding the lowest and highest percentages of the undercount of emigration and immigration in the published statistics. The difficult part was to get the experts to consider a non-specific European country with a good population register and migration definitions corresponding exactly with the United Nations (1998) recommendation. In other words, we wanted the experts to think of migration collection systems rather than specific country experiences.

The focus of Section B (questions 4-6) concerned the duration of stay in the definition of migration. Different timing criteria are used by different countries and we wanted to get a sense for how this might affect the relative levels of reported migration. In Question 4, experts were asked how much, in percentage terms, the level of migration would be for a duration of stay criterion of six months instead of 12 months. Question 5 asked for the difference between three and six months criteria. Answers were provided as ranges of percentage with assigned levels of certainty.

Finally, the questions in Section C were aimed at obtaining opinions about the accuracy of population registers in measuring migration. Experts were asked to consider registers in which there was no systematic bias and with random factors being the main source of error. In Questions 7 and 8, experts were asked to provide their beliefs and certainty regarding published statistics being within an interval from minus 5% to plus 5% compared to the true total level of emigration and immigration, respectively.

### 2.2.2 Feedback to experts and Round 2 questionnaire

The questionnaire in Round 2 consisted of the same set of questions as Round 1. As feedback from the first round, the second round questionnaire included tables with the (anonymous) answers given by the participating experts to each question in the first round and some arguments supporting their answers. The experts also had the possibility to look at graphical representation of their individual answers, such as those shown in Figure 1. In this figure a beta density with proportional quantiles was applied to represent expert answers about undercount of emigration. Details on how these have been computed are given in Section 3.

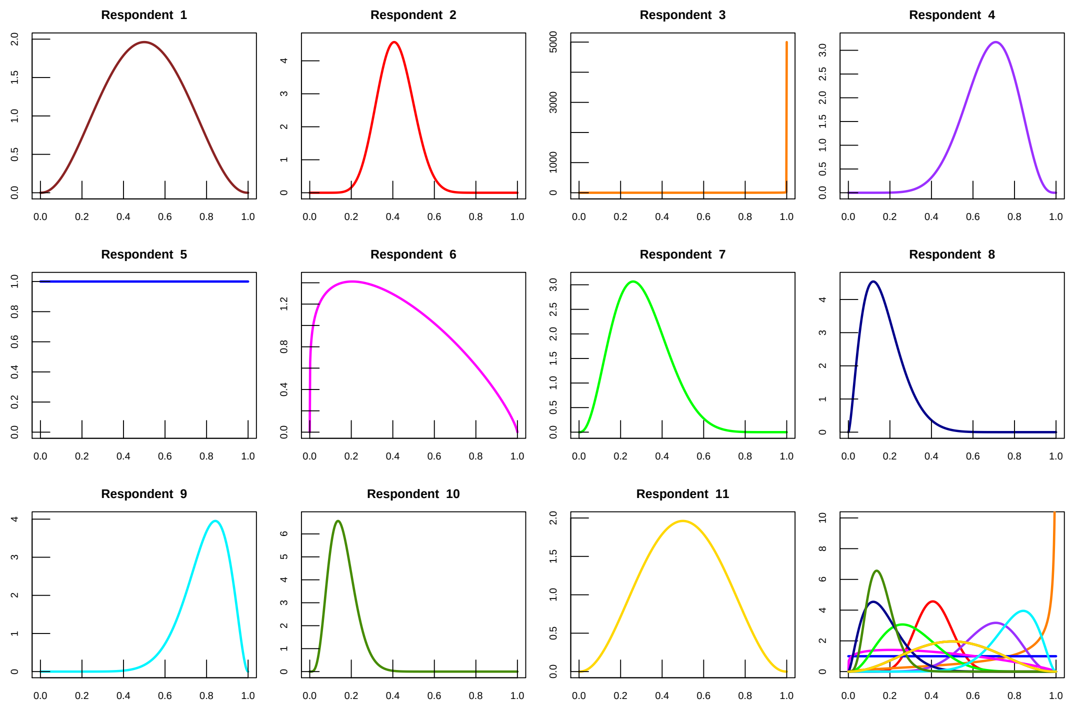


Figure 1: Graphical representation of expert answers from Round 1, undercount of emigration

In Round 1, a few of the experts gave answers to some of the questions on undercount which lay outside the 0-100% range, making interpretation difficult. For information on how we have treated these answers please see Section 3.2.1. Those experts who in the first round had provided answers outside the 0-100% range were also contacted in order to confirm that this interpretation of their answers was feasible. In the second round, we specifically stressed that the answer to some of the questions must lie in the interval 0-100%.

All the respondents from Round 1, except one, also took part in Round 2 of the questionnaire, which means that 10 experts took part in the whole two round Delphi process. Of these, eight chose to change their answers to one or more of questions in Round 2. Further information about the changes in the experts' opinions between the two rounds can be found in the subsections discussing undercount of emigration and immigration (Section 3.2.2), overcount due to duration of stay (Section 3.3.2) and accuracy (Section 3.4.2), respectively.

# 3 Translating the Expert Information into Prior Distributions

In this section, we explain how the opinions and judgements obtained in the first and second round of the Delphi survey, described in the previous section, were translated into prior distributions for the IMEM model parameters. First, we describe the IMEM model in general terms. The model is presented in detail in Raymer et al. (2011). Second, we present our methodology for converting the expert judgements into prior distributions for the model parameters addressing undercount, duration of stay and accuracy.

In general, constructing of of the prior density based on expert answers was a three-step process. First, having obtained the raw answers to a given question about some parameter, denote it as $\theta$, we identified the distribution $f$, which in our opinion reflected expert judgements about the $\theta$ most appropriately. Second, we constructed such a distribution $f_i(\theta)$ for each of our experts, $i = 1, \ldots, n$. The last, third step, was to combine together all individual representations into a single prior density, which ultimately was incorporated in the model as a prior. In order to achieve that, an unweighed mixture density, denoted by $p$, was applied.

$$p(\theta) \sim \frac{1}{n} \sum_{i=1}^{n} f_i(\theta). \tag{1}$$

The mixture prior was used for model testing with both Round 1 and Round 2 results of the Delphi questionnaire. We think that the mixture of individual densities was a proper choice for a prior as the expert opinions were heterogenous. Thus all different and sometimes opposing assessments could be fed into the model. Using smoothing techniques or fitting a parametric distribution to the expert answers could be an alternative option for priors elicitation, yet we believe these methods reduce the amount of information carried by an individual expert. Another option would be to perform Bayesian model averaging over models with each single expert prior as a separate input.

## 3.1 The IMEM model for observations and measurement

The data of interest can be conveniently expressed in a two-way contingency table or matrix showing the origin-to-destination flows with the cell counts corresponding to the number of migrants in a specified period. We observe counts (flows) $z_{ijt}^{k}$ from country $i$ to country $j$ during year $t$ reported by either the sending $S$ or receiving $R$

country, where $k \in \{S, R\}$. These flows can be represented by matrices $Z_t^S$ and $Z_t^R$:

$$Z_t^S = \begin{pmatrix} 0 & z_{12t}^S & z_{13t}^S & \cdots & z_{1nt}^S \\ z_{21t}^S & 0 & z_{23t}^S & \cdots & z_{2nt}^S \\ z_{31t}^S & z_{32t}^S & 0 & \cdots & z_{3nt}^S \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z_{n1t}^S & z_{n2t}^S & z_{n3t}^S & \cdots & 0 \end{pmatrix} \quad Z_t^R = \begin{pmatrix} 0 & z_{12t}^R & z_{13t}^R & \cdots & z_{1nt}^R \\ z_{21t}^R & 0 & z_{23t}^R & \cdots & z_{2nt}^R \\ z_{31t}^R & z_{32t}^R & 0 & \cdots & z_{3nt}^R \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z_{n1t}^R & z_{n2t}^R & z_{n3t}^R & \cdots & 0 \end{pmatrix}.$$

The interest of this research is to estimate a matrix $Y_t$ of true migration flows with unknown entries:

$$Y_t = \begin{pmatrix} 0 & y_{12t} & y_{13t} & \cdots & y_{1nt} \\ y_{21t} & 0 & y_{23t} & \cdots & y_{2nt} \\ y_{31t} & y_{32t} & 0 & \cdots & y_{3nt} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{n1t} & y_{n2t} & y_{n3t} & \cdots & 0 \end{pmatrix}.$$

For all $i$, $j$ and $t$, we assume that $z_{ijt}^k$ follows a Poisson distribution

$$z_{ijt}^S \sim \text{Po}(\mu_{ijt}^S), \tag{2}$$
$$z_{ijt}^R \sim \text{Po}(\mu_{ijt}^R). \tag{3}$$

In our model, $y_{ijt}$ is a true flow of migration from country $i$ to country $j$ in year $t$. It includes migration flows to and from rest of world (category $i = 0$). In terms of measurement, true flows are consistent with the United Nations (UN, 1998) recommendation for long-term international migration.

The two measurement error equations are

$$\log \mu_{ijt}^S = \log y_{ijt} + \psi_i - \log\left(1 + e^{-\kappa_i}\right) + \varepsilon_{ijt}^S, \tag{4}$$
$$\log \mu_{ijt}^R = \log y_{ijt} + \gamma_j - \log\left(1 + e^{-\kappa_j}\right) + \varepsilon_{ijt}^R, \tag{5}$$

where we assume $\varepsilon_{ijt}^S \sim \mathcal{N}(0, \tau_i^S)$ and $\varepsilon_{ijt}^R \sim \mathcal{N}(0, \tau_j^R)$. The precisions (reciprocal variances) of the error terms depend on whether the data are captured by sending or receiving countries. Thus we take

$$\tau_i^S = t_{c(i)}^S, \tag{6}$$
$$\tau_j^R = t_{c(j)}^R, \tag{7}$$

where $c(i)$ denotes the type of collection system (e.g., population register or survey). For the moment, $c(i)$ is the same for all countries. The accuracy is only distinct for emigration and immigration.

The differences in duration of stay criterion, which depend on the reporting country, and the effect of undercount are captured by the parameters $\psi_i$ and $\gamma_j$,

$$\psi_i = \begin{cases} \delta_1 + \log \lambda_1 & \text{if duration is 0 months} \\ \delta_2 + \log \lambda_1 & \text{if duration is 3 months} \\ \delta_3 + \log \lambda_1 & \text{if duration is 6 months} \\ \log \lambda_1 & \text{if duration is 12 months} \\ \delta_4 + \log \lambda_1 & \text{if duration is permanent} \end{cases}, \tag{8}$$

$$\gamma_j = \begin{cases} \delta_1 + \log \lambda_2 & \text{if duration is 0 months} \\ \delta_2 + \log \lambda_2 & \text{if duration is 3 months} \\ \delta_3 + \log \lambda_2 & \text{if duration is 6 months} \\ \log \lambda_2 & \text{if duration is 12 months} \\ \delta_4 + \log \lambda_2 & \text{if duration is permanent} \end{cases}. \tag{9}$$

Finally, the $\kappa_i$ parameter is a normally distributed country-specific random effect

$$\kappa_i \sim \mathcal{N}(\nu_i, \zeta_i),$$

where $\nu_i = \nu_{m(i)}$ is a group-specific mean and $\zeta_i = \zeta_{m(i)}$ is a group-specific precision and $m(i)$ denotes a type of coverage assumed for country $i$. For the time being, there are two coverage types, that is, $m(i) \in \{\text{standard}, \text{excellent}\}$. The logistic transformation of $\kappa$ in Equations 4 and 5 ensures that the function is bounded within a range $(0, 1)$ on the linear scale. It can be interpreted in terms of the differences in coverage with respect to the UN definition of migration.

For the migration to and from the rest of world there is only one equation per outflow and inflow, respectively, i.e.,

$$\begin{aligned} \log \mu^S_{i0t} &= \log y_{i0t} + \psi_i + \varepsilon^S_{i0t}, \quad \text{for all } i \text{ and } t & (10) \\ \log \mu^R_{0jt} &= \log y_{0jt} + \gamma_j + \varepsilon^R_{0jt}, \quad \text{for all } j \text{ and } t, & (11) \end{aligned}$$

All other parameters remain same as described above, except for $\psi_i$ and $\gamma_j$, which are defined as in Equations 8 and 9 with $\lambda_1$ and $\lambda_2$ replaced with $\lambda_3$ and $\lambda_4$, respectively. Note, that in the measurement of the flows to and from the rest of world we assume a perfect coverage for all countries, i.e., there are no country-specific random effects.

## 3.2 Undercount of Emigration and Immigration

### 3.2.1 Prior construction method

In the first and fourth section of the Delphi questionnaire, experts were asked to provide answers to the following question about undercount of migration within Europe and to and from the rest of world (Round 1 questionnaire is provided in Appendix A):

> a) By how many per cent do you expect that emigration (or immigration) flows are undercounted in the published statistics, as compared to the true total level of emigration (immigration)? Please provide a range in percentages.
>
> b) Approximately, how certain are you that the true undercount will lie within the range that you provided above?

Let $P_1$ and $P_2$ denote the lower and upper percentages stated by an expert about undercount and $c$ denote the certainty about the range $(P_1,\ P_2)$. The underlying assumption regarding undercount is that a number $P \in [0, 1] \times 100\%$, which is

$$(1 - P)y = z, \tag{12}$$

where $y$ are true flows and $z$ are reported flows. Then $(1 - P)$ can be interpreted as a fraction of the true flow which is captured in the reported data. Note, some of the answers provided by experts, especially for the first round questionnaire, were not clear in terms of their interpretations of the question. We believe that some experts had difficulty understanding our questions or expressing their beliefs in statistical terms. For example, one respondent in both rounds gave answers for Question 12 of 300% and 350% for lowest and highest percentages, respectively, despite that there in Round 2 was a line stating that the percentage should lie within 0-100% range. The same respondent's answers to Question 1, which contained the same notion of undercount, were 4% and 8%. Another expert provided all values larger than 100%. This suggests that the undercount was understood as how many times larger are the true flows in comparison to the reported data, that is,

$$y = (1 + a)z. \tag{13}$$

Hence, if an expert provided at least one number $a$ not falling into a range $[0, 1]$, both answers were treated according to the latter interpretation and recomputed to be $P = 1 - 1/(1 + a)$.

To convert the experts' answers into priors for the IMEM model parameters, we needed to first identify probability distributions which would both accurately reflect their beliefs and work well with the model framework. We considered three densities: piece-wise uniform, logit-normal and beta. These densities were chosen because they could be constrained to values between zero and one, they were flexible in terms of shapes, their parameters were easily to calculate, and they were easy to implement in the overall model. Truncated distributions, such as normal or log-normal, were considered but rejected as they were difficult to handle in the computations.

Note, that identifying a probability distribution to reflect someone's opinion is an extremely difficult task. The best option would be to ask an expert to draw a distribution. However, this would require such an expert to be trained in statistics, which was not the case for our study. Furthermore, the drawn density would have to be usable in computations. Since experts may not agree with our interpretation of their judgements, we utilise a multistage Delphi approach. After the first round of questions, experts were provided with the densities resulting from our interpretation and parameterisation of their answers, as well as the anonymous results from other experts in the study. This allowed them to reconsider and revise their opinions.

To illustrate the differences between these different densities, consider the four expert answers to Question 1 set out in Table 1. The answers can be interpreted as follows. A given expert, say, Respondent 2, believes that the emigration flows in the published statistics are undercounted by $P_1 = 30\%$ to $P_2 = 50\%$, compared to the true level of emigration. Respondent 2 also believes that this range is true with a probability $c = 75\%$. If $c = 100\%$, then the expert would have to be perfectly sure about the range he or she provided[1]. It means that what we observe in the data constitutes only 50% to 70% of the true flows (from $(1 - P)$). According to this interpretation, Respondent 4 believes that the reported flows of emigrants are only 4% to 8% smaller than the true level of emigration, which is a precise range, but his or her certainty is only 5%. It should be intuitive that the wider the range of undercount, the larger the certainty should be. Note that in Round 1 of the Delphi survey, almost all answers were consistent with this rule. For the questions concerning undercount, only one expert indicated relatively large range with a small level of certainty. This led to some computational and interpretation problems.

For the case of the piecewise uniform densities, the computation was straightforward. We assumed that the certainty level $c$ provided by a given respondent corresponded with the probability mass between $P_1$ and $P_2$. The remainder, $(1 - c)$, was proportionally distributed between $[0, P_1]$ and $[P_2, 1]$. Thus, the quantiles of the

[1] We believe it is obvious that a statement about undercount being between 0% and 100% should be provided with 100% certainty.

Table 1: Experts answers to question 1 - undercount of emigration

| Respondent | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Lowest percentage, $P_1$ | 20 | 30 | 50 | 4 |
| Highest percentage, $P_2$ | 80 | 50 | 90 | 8 |
| Certainty, $c$ | 90 | 75 | 90 | 5 |

Source: Delphi survey

resulting piecewise uniform density were

$$q_1 = \frac{(1-c)P_1}{1+P_1-P_2},$$
$$q_2 = \frac{(1-c)(1-P_2)}{1+P_1-P_2}. \tag{14}$$

The resulting piecewise uniform densities, after transformation into undercount using Equation 12, are presented in the first row of Figure 2.

In the case of the logit-normal density, it was assumed that

$$\begin{cases} \mu + \sigma\Phi^{-1}(q_1) = \frac{\log(P_1)}{1-\log(P_1)} \\ \mu + \sigma\Phi^{-1}(q_2) = \frac{\log(P_2)}{1-\log(P_2)}. \end{cases} \tag{15}$$

Two specifications of $q_1$ were considered. In the first one, the probability mass $c$ lies between $P_1$ and $P_2$ and the remainder, $(1-c)$, symmetrically distributed between $[0, P_1]$ and $[P_2, 1]$:

$$q_1 = \frac{1-c}{2},$$
$$q_2 = \frac{1+c}{2}. \tag{16}$$

The second specification is based on quantiles as in the piecewise uniform approach (Equation 14). The resulting densities (after transformation using Equation 12) for these two approaches are shown in second and third row, respectively, in Figure 2.

Finally, two sets of quantiles were also considered for the beta distribution. The hyperparameters $\alpha$ and $\beta$ of the beta density were computed by solving a set of two equations

$$\begin{cases} F_b^{-1}(P_1, \alpha, \beta) = q_1 \\ F_b^{-1}(P_2, \alpha, \beta) = q_2. \end{cases} \tag{17}$$

This was achieved by minimising to zero the following expression

$$\min_{\alpha,\beta} \left\{ \sum_{i=1}^{2} \left( F_b^{-1}(P_i, \alpha, \beta) - q_i \right)^2 \right\}, \tag{18}$$

where $q_1$ and $q_2$ were either symmetrically (Equation 16) or proportionally (Equation 14) distributed. Vector $(1, 1)$ was used as a starting point. The densities obtained for the four example experts are presented in Figure 2 in the fourth and fifth rows for symmetric and proportional quantiles, respectively.

From all of the approaches considered to translate and represent the subjective expert opinions, the beta density with proportional quantiles was ultimately chosen. Piecewise uniform was rejected because it produced crude results (see, e.g., row

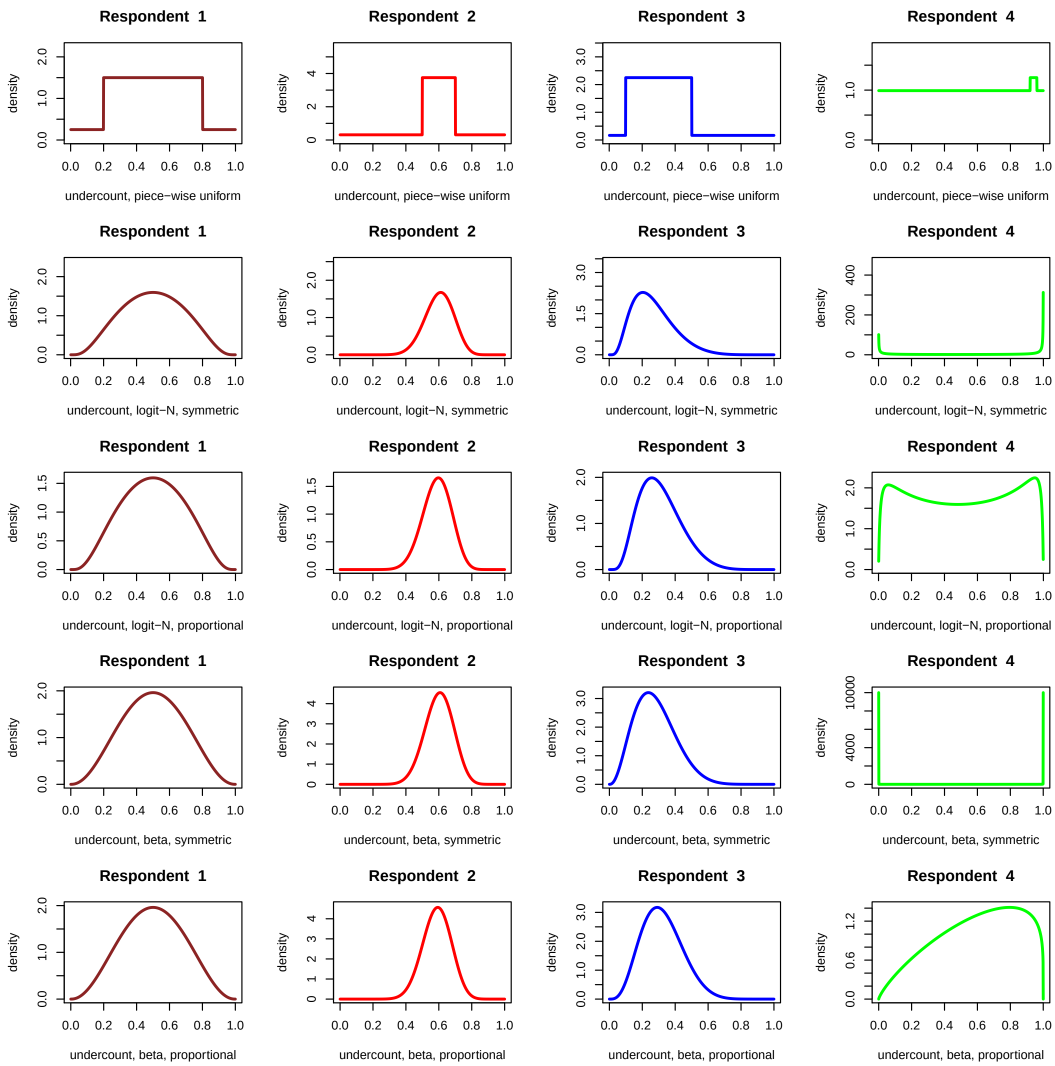


Figure 2: Densities for four experts with various specifications

1, column 2 in Figure 2). The logit-normal and beta distributions with symmetric quantiles also tended to yield unintuitive shapes, especially in cases where experts assigned more certainty to regions close to 0% or 100% undercount. Such a case is represented by Respondent 4 in Figure 2. Both symmetric approaches (logit-normal and beta in rows 2 and 4, respectively) are bimodal with most of the probability mass assigned close to 0 and 1, which was considered to be rather implausible representation of expert's opinion. The proportional logit-normal approach also resulted in a bimodal density and was rejected[2].

In Figure 2, the results for Respondents 1, 2 and 3 are presented. They show that the shapes are similar in all five densities. For Respondent 4, the different shapes occur because all numbers given were very close to 0%. This situation also occurred for other respondents, albeit with some providing answers very close to 100%.

### 3.2.2 Expert answers and resulting prior densities

The raw answers (in terms of proportions), provided by the experts to the question about the migration undercount within EU and EFTA countries, are presented in

[2]Depending on relative sizes of $\mu$ and $\sigma$ logit-normal distribution has one or two modes, see Johnson (1949, pp. 158-159).

Table 2 for emigration and Table 3 for immigration. For the emigration undercount we observe that two respondents did not change their opinions between Round 1 and Round 2, while three increased their confidence. Some of the experts provided wide percentage spans with large confidence (e.g. respondents 1, 4, 10, 11), while some gave a comparatively narrow range with lower certainty (respondents 2, 6 or 9). Respondent 3 provided a percentage range exceeding the envisaged 0-100% range, with a relatively small confidence in it. Hence, we interpreted it as the undercount given in Equation 13 and transformed it accordingly.

In the Round 2 answers, we observe that only one expert lowered certainty about the given percentage. In the case of immigration undercount six respondents left their answers unchanged, one of them increased the confidence. One respondent decreased certainty providing wider range of the undercount.

Figures 3 and 5 present the Round 1 and 2 expert answers transformed into beta densities with proportional quantiles (described in previous section), for emigration and immigration undercount, respectively. These individual curves were then used to construct a mixed prior densities in Figure 4 and in Figure 6. Note that these mixture priors reflect the undercount as it was included in the model, that is they represent the value $(1 - P)$ from Equation 12.

The prior for emigration undercount, based on answers from Round 1 (Figure 4), is weakly informative in a sense that there is no clear region of undercount that would be indicated by the majority of experts to be most plausible. The density has four modes. Mean undercount is 52% with standard deviation 27%. It means that the observed flows, in the eyes of the experts, on average constitute 52% of the true unobserved flows. Round 2 prior is unimodal, with mean 56% and standard deviation 22%. Unimodality and lower spread in the second round suggests there has been some convergence of the answers.

Comparing priors of the immigration undercount we observe a shift of the probability mass from the region of a very high undercount (near 0) to the values suggested by the majority of experts, that is around 60-80%. The Round 1 prior mean is 68% with standard deviation 25%, in the second round these changed to 72% and 18%. Again, the three modes of the Round 1 prior were exchanged by a unimodal density in Round 2, which is a sign of convergence in judgements.

The overall large spread (large standard deviation and a relatively 'flat' shape of the distribution) of the mixture densities reflects the heterogeneity of expert judgements about the undercount. It may also stem from different experiences of the experts with migration statistics. Some of them, in the open ended questions, indicated that they did not have enough expertise in the data collection systems across whole domain of countries considered in the model. Thus, they based their opinions on the systems known best to them. Moreover, they were pointing out to the differences across countries in Europe which may have contributed to the flatness of the mixture density.

The expert assessment of the undercount of migration from and to the rest of world is more ambiguous than in the case of the intra European migration. Tables 4 and 5 present Rounds 1 and 2 answers. For both emigration and immigration, four experts stood by their first round answers, two reduced their confidence and changed the undercount range. Note, that for computations answers of respondents 3 and 6 were transformed to represent undercount given in Equation 12.

The transformation of subjective opinions into individual densities is presented in Figures 7 and 9 for emigration and immigration, respectively. The resulting mixture priors for rest of world undercount are presented in Figure 12 and Figure 14. For

Table 2: Experts answers concerning undercount of emigrants

| Resp | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Round 1 | | | | | | | | | | | |
| LP | 0.2 | 0.3 | 0 | 0.5 | 0.1 | 0.04 | 0.1 | 0.01 | 0.8 | 0.05 | 0.2 |
| HP | 0.8 | 0.5 | 10 | 0.9 | 0.3 | 0.08 | 0.4 | 0.3 | 0.95 | 0.2 | 0.8 |
| Cert | 0.9 | 0.75 | 0.5 | 0.9 | 0.2 | 0.05 | 0.75 | 0.9 | 0.5 | 0.75 | 0.9 |
| Round 2 | | | | | | | | | | | |
| LP | 0.25 | 0.3 | 0.1 | NA | 0.1 | 0.04 | 0.2 | 0.01 | 0.5 | 0.5 | 0.3 |
| HP | 0.75 | 0.5 | 1 | NA | 0.3 | 0.08 | 0.5 | 0.5 | 0.75 | 0.9 | 0.9 |
| Cert | 0.9 | 0.75 | 0.5 | NA | 0.5 | 0.05 | 0.5 | 0.9 | 0.75 | 0.9 | 0.9 |

LP - Lowest proportion, HP - Highest proportion, Cert - Certainty

Source: Delphi survey

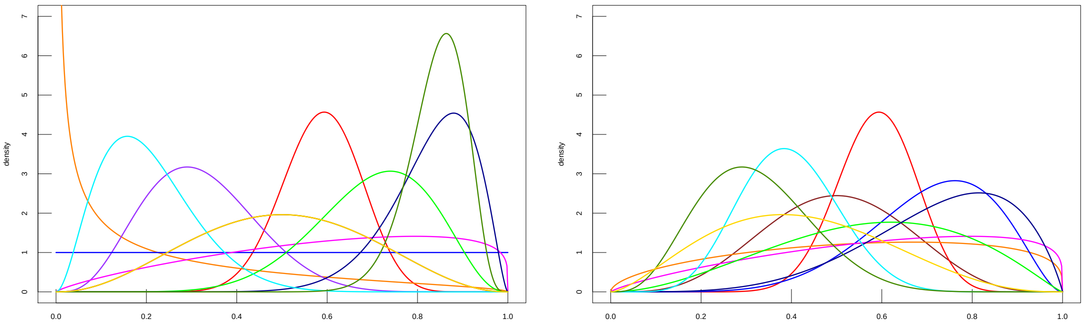

Figure 3: Expert answers transfomed to densities for undercount of emigrants, Rounds 1 (left) & 2 (right)

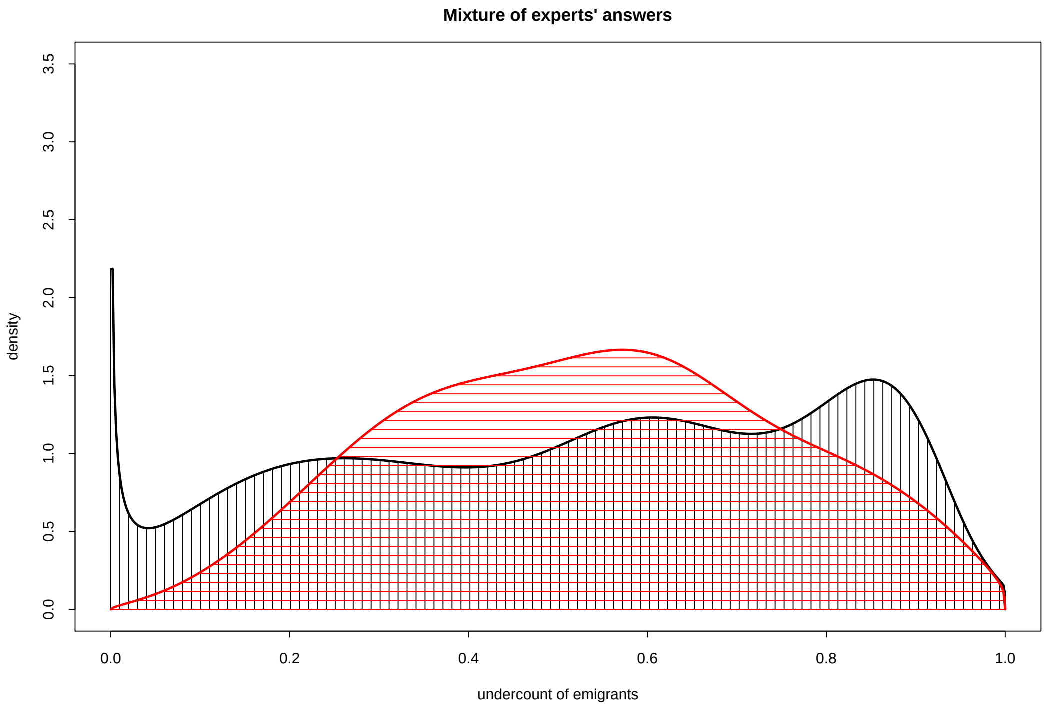


Figure 4: Mixture prior densities for undercount of emigrants, Rounds 1 (vertical) & 2 (horizontal)

Table 3: Experts answers concerning undercount of immigrants

| Resp | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Round 1 | | | | | | | | | | | |
| LP | 0.1 | 0.1 | 0 | 0.2 | 0.1 | 0.04 | 0.1 | 0.01 | 0.1 | 0.02 | 0.1 |
| HP | 0.5 | 0.3 | 10 | 0.6 | 0.3 | 0.08 | 0.2 | 0.15 | 0.2 | 0.1 | 0.5 |
| Cert | 0.9 | 0.9 | 0.5 | 0.9 | 0.2 | 0.05 | 0.75 | 0.9 | 0.9 | 0.75 | 0.9 |
| Round 2 | | | | | | | | | | | |
| LP | 0.1 | 0.1 | 0.1 | NA | 0.1 | 0.04 | 0.1 | 0.01 | 0.1 | 0.2 | 0.2 |
| HP | 0.5 | 0.3 | 1 | NA | 0.3 | 0.08 | 0.3 | 0.15 | 0.2 | 0.6 | 0.6 |
| Cert | 0.9 | 0.9 | 0.5 | NA | 0.5 | 0.05 | 0.5 | 0.9 | 0.9 | 0.9 | 0.9 |

LP - Lowest proportion, HP - Highest proportion, Cert - Certainty

Source: Delphi survey

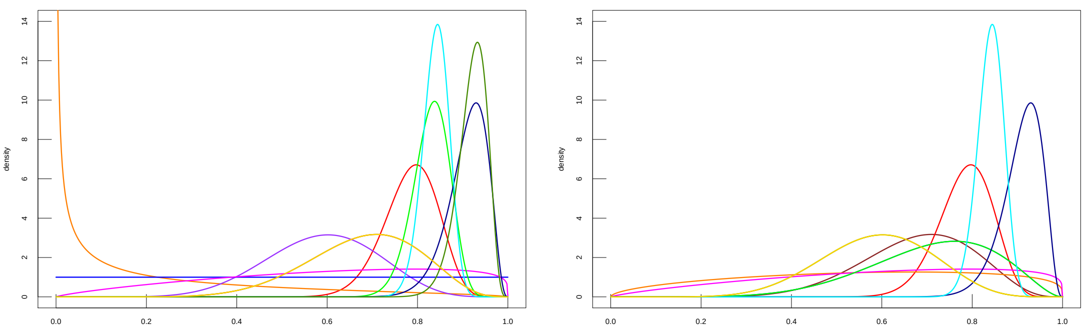

Figure 5: Expert answers transfomed to densities for undercount of immigrants, Rounds 1 (left) & 2 (right)

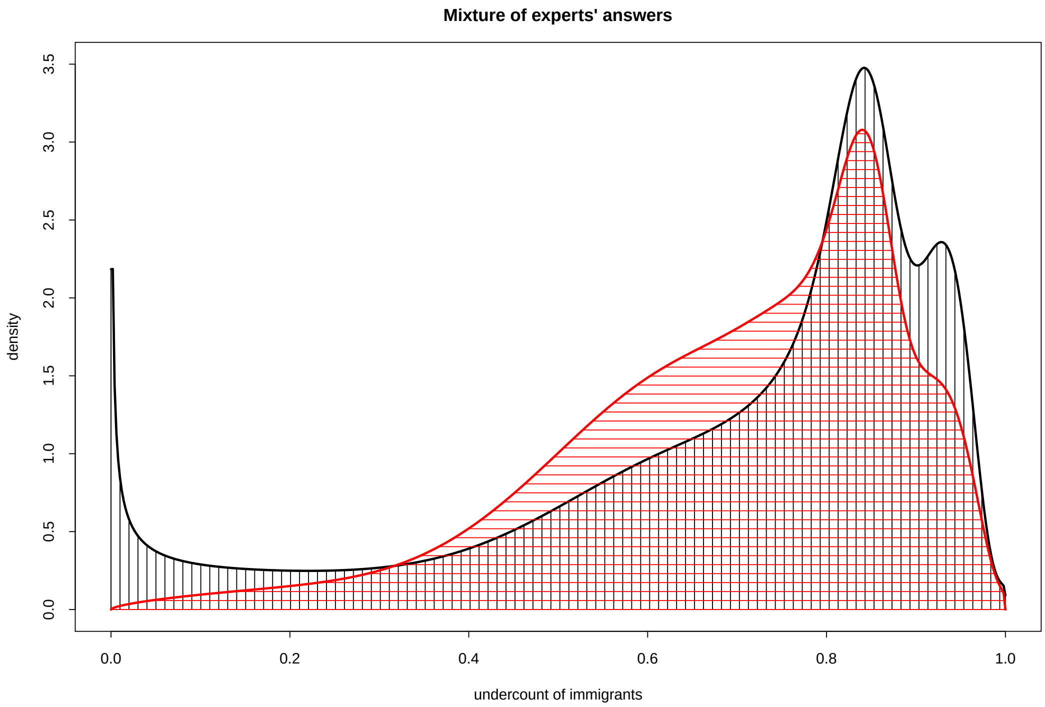


Figure 6: Mixture prior densities for undercount of immigrants, Rounds 1 (vertical) & 2 (horizontal)

Table 4: Experts answers concerning undercount of emigrants to rest of world

| Resp | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Round 1 | | | | | | | | | | | |
| LP | 0.2 | 0.3 | 0.1 | 0.3 | 0.4 | 3 | 0.1 | 0.01 | 0.8 | 0 | 0.3 |
| HP | 0.8 | 0.5 | 1 | 0.7 | 0.7 | 3.5 | 0.4 | 0.1 | 0.95 | 0.1 | 0.9 |
| Cert | 0.9 | 0.75 | 0.25 | 0.75 | 0.5 | 0.5 | 0.75 | 0.9 | 0.5 | 0.95 | 0.75 |
| Round 2 | | | | | | | | | | | |
| LP | 0.25 | 0.3 | 0.1 | NA | 0.4 | 3 | 0.2 | 0.01 | 0.3 | 0.5 | 0.4 |
| HP | 0.75 | 0.5 | 1 | NA | 0.7 | 3.5 | 0.6 | 0.3 | 0.5 | 0.8 | 1 |
| Cert | 0.9 | 0.75 | 0.25 | NA | 0.5 | 0.5 | 0.5 | 0.9 | 0.75 | 0.75 | 0.75 |

LP - Lowest proportion, HP - Highest proportion, Cert - Certainty

Source: Delphi survey

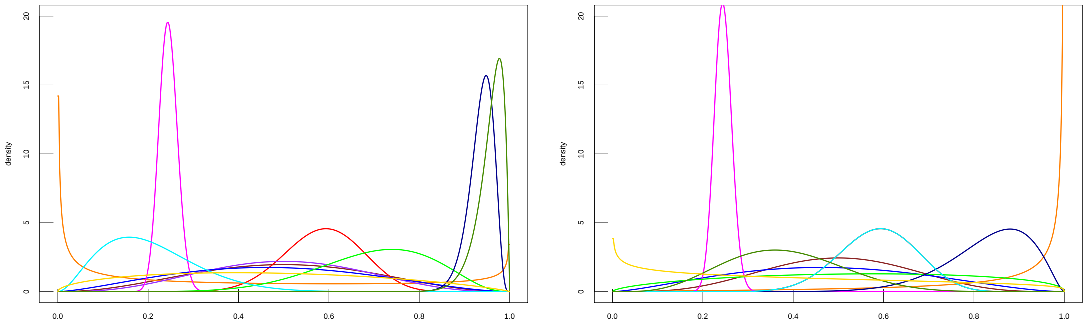

Figure 7: Expert answers transformed to densities for undercount of emigrants to rest of world, Rounds 1 (left) & 2 (right)

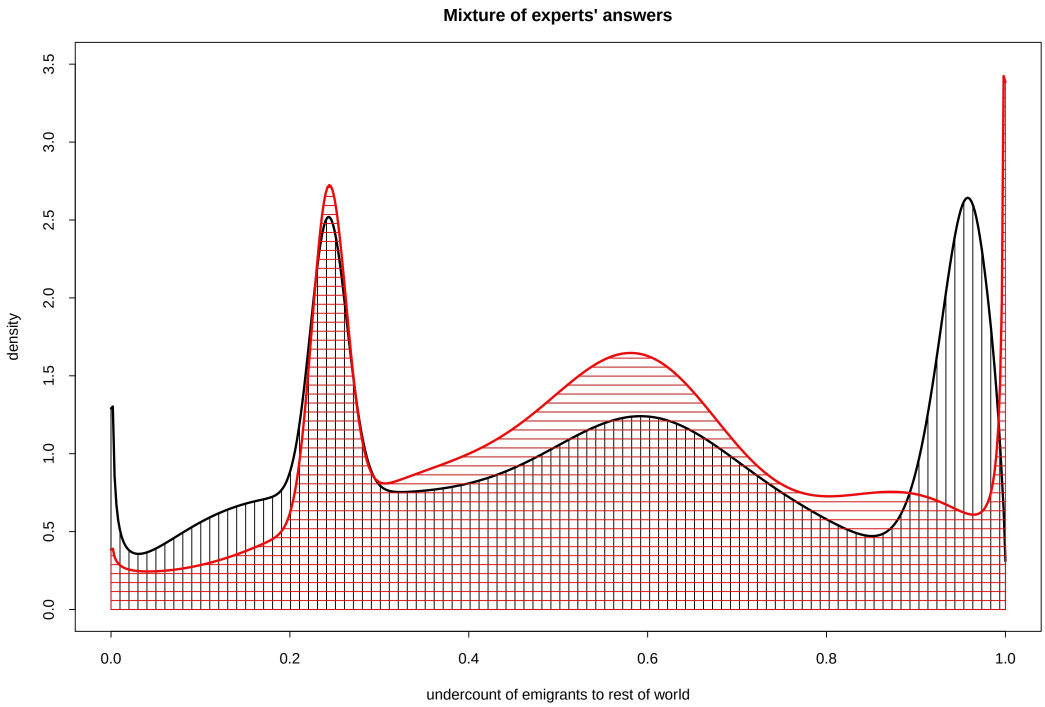


Figure 8: Mixture prior densities for undercount of emigrants to rest of world, Rounds 1 (vertical) & 2 (horizontal)

Table 5: Experts answers concerning undercount of immigrants from rest of world

| Resp | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Round 1 | | | | | | | | | | | |
| LP | 0.1 | 0.1 | 1 | 0.2 | 0.4 | 2 | 0.1 | 0.01 | 0.3 | 0 | 0.1 |
| HP | 0.5 | 0.3 | 10 | 0.6 | 0.7 | 2.5 | 0.3 | 0.2 | 0.6 | 0.25 | 0.5 |
| Cert | 0.9 | 0.9 | 0.25 | 0.75 | 0.5 | 0.5 | 0.75 | 0.9 | 0.75 | 0.95 | 0.75 |
| Round 2 | | | | | | | | | | | |
| LP | 0.1 | 0.1 | 0.5 | NA | 0.4 | 2 | 0.1 | 0.01 | 0.2 | 0.3 | 0.2 |
| HP | 0.5 | 0.3 | 1 | NA | 0.7 | 2.5 | 0.4 | 0.4 | 0.5 | 0.6 | 0.6 |
| Cert | 0.9 | 0.9 | 0.25 | NA | 0.5 | 0.5 | 0.5 | 0.9 | 0.75 | 0.75 | 0.75 |

LP - Lowest proportion, HP - Highest proportion, Cert - Certainty

Source: Delphi survey

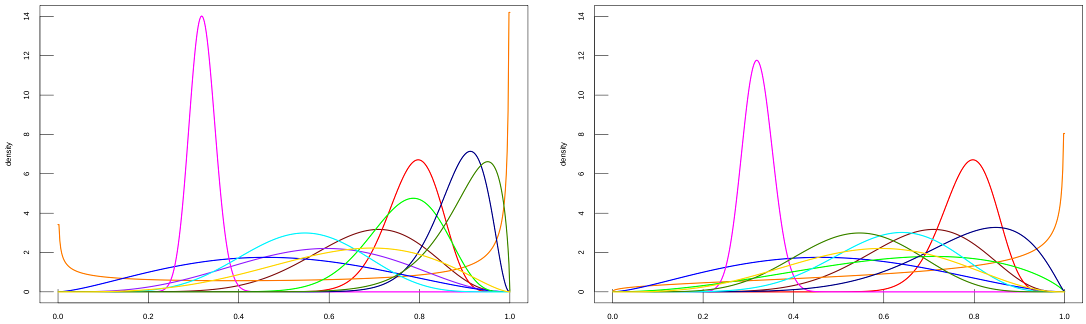

Figure 9: Expert answers transformed to densities for undercount of immigrants to rest of world, Rounds 1 (left) & 2 (right)

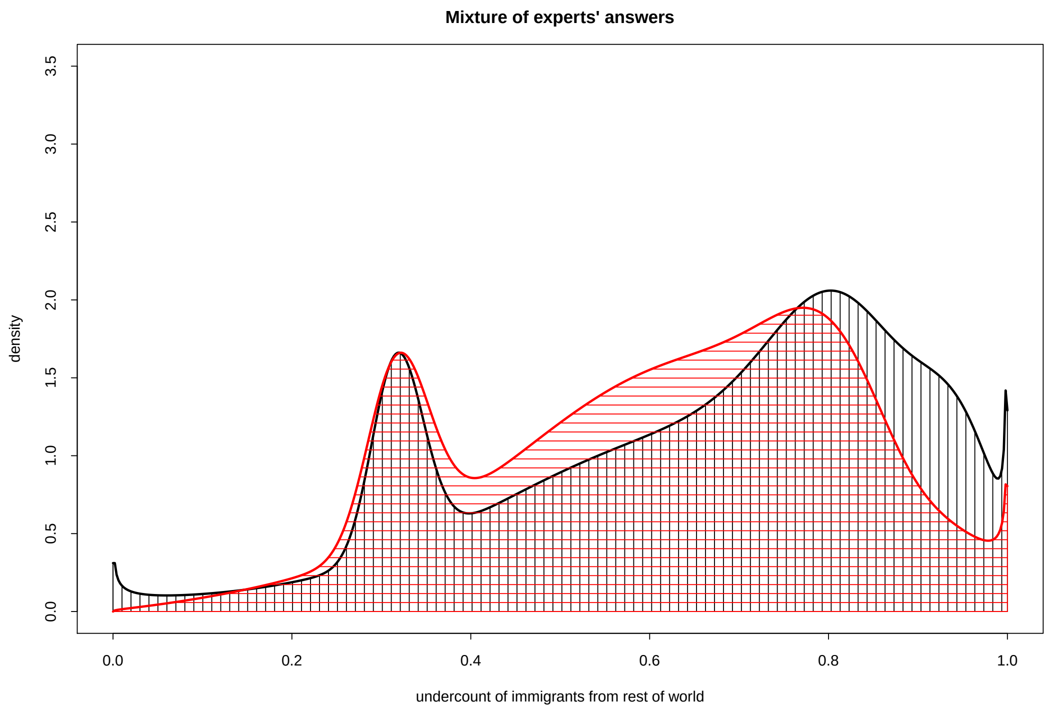


Figure 10: Mixture prior densities for undercount of immigrants from rest of world, Rounds 1 (vertical) & 2 (horizontal)

emigration both first and second round mixtures have four modes, in the Round 2 density, two of them are on the boundaries 0% and 100%. Two middle modes are around 25% and 60%. The overall mean changes from 56% in Round 1 to 54% in Round 2, with standard deviations 28% and 25%, respectively. Assessment of the immigration undercount is similar. The mode in 0% disappears after the second round and the probability mass concentrates more in the middle (40-80%) but the mixture density is still trimodal. The means of Rounds 1 and 2 are 63% and 61% with standard deviations 24% and 21%, respectively.

The consensus among experts concerning the undercount of rest of world flows has not been reached. Respondents, in comments and rationale for their answers, pointed out that the data on non EU citizens are in general better captured, due to more requirements for them, than the data on nationals or other EU citizens. This would reduce the undercount. On the other hand, including the undocumented migrants has a reverse effect and blurs its evaluation. Some experts commented that the difference between the measurement of intra and extra European migrants should not be significant.

## 3.3 Overcount due to duration of stay

### 3.3.1 Prior construction method

Duration of stay parameters capture the effect of the particular duration criterion applied in a given country. We assumed that the shorter duration of stay was, the more migrants were recorded, that is

$$y_p > y_{12} > y_6 > y_3 > y_0,$$

where subscript of the true flow $y$ denotes the duration criterion applied (permanent, 12 months, six months, three months and no time limit, respectively).Our benchmark criterion was 12 months, following the UN definition described in Section 2.2. The overcount of migrants, due to the different duration criterion in the reported data $z$, could be expressed by a factor $e^{\delta_k}$ in equation

$$y_{12} = e^{\delta_k} z.$$

The question in the Delphi study about the overcount was formulated as follows:

a) By how many per cent do you expect that the level of migration with the SIX (THREE) MONTH criterion is higher than with the 12 (SIX) MONTH criterion? Please provide a range in percentages.

b) Approximately, how certain are you that the true value will lie within the range that you provided above?

The experts provided lower and upper percentages of the overcount, denoted as $P_1$ and $P_2$, and $c$, that is the certainty about the range $(P_1,\ P_2)$. Percentage $P > 0$ provided by experts represented the duration overcount in following way:

$$y_b = (1 + P)y_a, \tag{19}$$

where $a$ was a shorter duration criterion than $b$. Then we assumed that the overcount due to using six months criterion instead of 12 months was captured by parameter $1 + P = \exp(d_3)$, $d_3 > 0$, so that

$$y_{12} = e^{d_3} y_6.$$

Similarly, we defined the overcount of migrants measured using 3 months criterion compared to 6 months to be reflected in parameter $\exp(d_2)$, $d_2 > 0$, which could be written as

$$y_6 = e^{d_2} y_3.$$

Then, the effect of using 3 months criterion compared to 12 months was

$$y_{12} = e^{d_2+d_3} y_3.$$

For permanent duration, which was captured by parameter $\delta_4$, the scaling factor was

$$y_{12} = e^{-d_4} y_p,$$

where $d_4 > 0$. That formulation led to the following constraints imposed on duration parameters $\delta_k$:

$$\begin{aligned}
\delta_1 &= d_1 + d_2 + d_3, \\
\delta_2 &= d_2 + d_3, \\
\delta_3 &= d_3, \\
\delta_4 &= -d_4.
\end{aligned}$$

We further assumed that each $d_k$ followed a log-normal distribution. Then the parameters of each expert-specific density for $\delta_k$, $k = 1, 2, 3$ could be calculated by solving a set of equations

$$\begin{cases} \mu + \sigma\Phi^{-1}(1/2 + c/2) = \log\log(1 + P_1) \\ \mu - \sigma\Phi^{-1}(1/2 + c/2) = \log\log(1 + P_2). \end{cases} \tag{20}$$

For $\delta_4$ the resulting set of equations was

$$\begin{cases} \mu + \sigma\Phi^{-1}(1/2 + c/2) = \log\log(1 + P_1)^{-1} \\ \mu - \sigma\Phi^{-1}(1/2 + c/2) = \log\log(1 + P_2)^{-1}. \end{cases} \tag{21}$$

We also considered an alternative construction of the prior. Let us define the duration overcount similarly as in Equation 19, that is

$$y_b = (1 + d_k) y_a, \tag{22}$$

where $d_k > 0$, $k = 1, \ldots, 4$ were overcount factors. Then the parameters $\delta_k$ could be expressed as

$$\begin{aligned}
\delta_1 &= \log(1 + d_1) + \log(1 + d_2) + \log(1 + d_3), \\
\delta_2 &= \log(1 + d_2) + \log(1 + d_3), \\
\delta_3 &= \log(1 + d_3), \\
\delta_4 &= -\log(1 + d_4).
\end{aligned}$$

Then, we assumed that $d_k$ were log-normally distributed with parameters derived from a set of equations:

$$\begin{cases} \mu + \sigma\Phi^{-1}(1/2 + c/2) = \log(P_1) \\ \mu - \sigma\Phi^{-1}(1/2 + c/2) = \log(P_2), \end{cases} \tag{23}$$

where values $P_1$ and $P_2$ with certainty $c$ were elicited from the experts for each of $d_k$, $k = 1, \ldots, 4$.

The resulting mixture densities for $\delta_k$ were very similar in both approaches and they lead to very similar posteriors. In the end we decided to use the first approach for our computations.

Table 6: Experts answers concerning duration overcount, 12m vs. 6m criterion

| Resp | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Round 1 | | | | | | | | | | | |
| LP | 0.1 | 0.1 | 0.3 | 1 | 0.2 | 0.35 | 0.2 | 0.05 | 0.2 | 0.05 | 0.1 |
| HP | 0.4 | 0.25 | 1 | 3 | 0.4 | 0.65 | 0.4 | 0.15 | 0.4 | 0.15 | 0.3 |
| Cert | 0.9 | 0.5 | 0.05 | 0.5 | 0.3 | 0.4 | 0.5 | 0.75 | 0.5 | 0.75 | 0.75 |
| Round 2 | | | | | | | | | | | |
| LP | 0.1 | 0.1 | 0.3 | NA | 0.2 | 0.35 | 0.15 | 0.05 | 0.2 | 0.1 | 0.2 |
| HP | 0.4 | 0.25 | 1 | NA | 0.4 | 0.65 | 0.4 | 0.15 | 0.4 | 0.2 | 0.4 |
| Cert | 0.75 | 0.5 | 0.25 | NA | 0.3 | 0.4 | 0.5 | 0.75 | 0.5 | 0.75 | 0.75 |

LP - Lowest proportion, HP - Highest proportion, Cert - Certainty

Source: Delphi survey

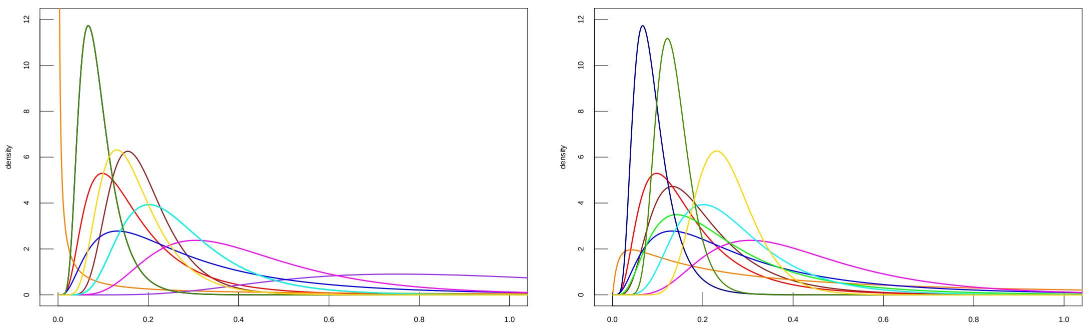

Figure 11: Expert answers transformed to densities for duration overcount, 6m vs. 12m, Rounds 1 (left) & 2 (right)

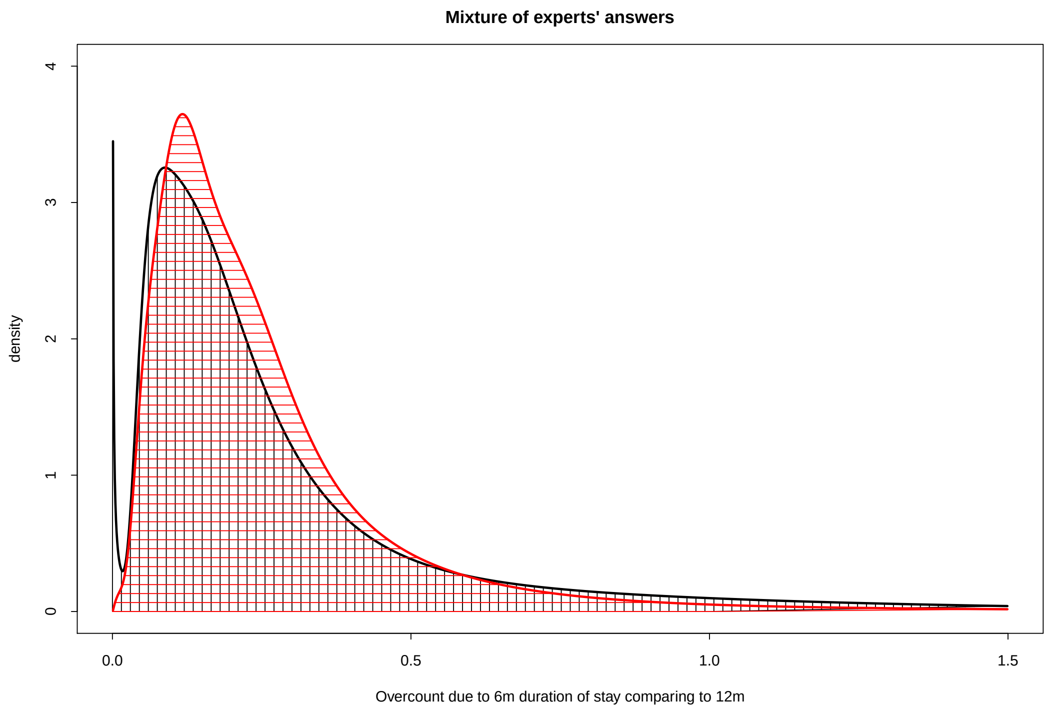


Figure 12: Mixture prior densities for duration overcount, 6m vs. 12m, Rounds 1 (vertical) & 2 (horizontal)

### 3.3.2 Expert answers and resulting prior densities

Tables 6 and 7 present the expert opinions concerning the overcount of migration due to different duration of stay criteria. In the comparison of the 12 months and 6 months criteria, seven respondents remained with their first round answers, one of them reducing certainty. In the answers concerning 6 months and 3 months criteria, four experts left their answers unchanged. Only one of respondents increased his or her confidence.

The representations of individual expert answers are shown on a log scale in Figures 11 and 13. Logarithmic scale was used due to the computational problems. This means that the curves represent expert answers translated into densities for parameters $\delta_k$, not overcount factors $e^{\delta_k}$.

When we compare the mixture prior densities (Figures 12 and 14) resulting from two rounds of questions about the overcount due to different duration criteria, we observe two important changes between Round 1 and Round 2. In both 12-6 and 6-3 months comparisons, the expert whose answer was contributing to the mode 0% changed his or her judgement. Due to a comparatively small confidence given by Respondent 3 in Round 1, the mixture is a fat-tailed distribution. Hence, computing means is problematic because of the numerical problems. The medians of the distribution on the log scale were 0.53 and 0.20 for 12-6 and 6-3 months overcount, respectively. The Round 2 results are no longer fat-tailed; on the log scale the means are 0.68 and 0.35, while the medians are 0.50 and 0.20, respectively for experts' assessment of the 12-6 months and 6-3 months duration differences.

One of the experts stated that these percentages of overcount may vary a lot across countries, mainly due to the under registration of short-term movements. Another expert pointed out that some registers are able to provide statistics on migration flows with different duration criteria (e.g. Austria and the Netherlands).

## 3.4 Accuracy

### 3.4.1 Prior construction method

The question regarding accuracy of data collection appeared to be the most challenging for the experts to answer. It was asked in the third section of the Delphi questionnaire.

> a) For EMIGRATION (IMMIGRATION), how probable do you think it is that the published statistics are within an interval from minus 5% to plus 5% compared to the true total level of emigration? (If it helps think of how often the annual published statistics are within this interval during a period of 100 years). Please provide a range in percentages.
>
> b) Approximately, how certain are you that the true value will lie within the range that you provided above?

The interpretation of the question in brackets was provided to help respondents understand the notion of the accuracy. In the preamble to the question (see Appendix A) it was also explained that accuracy should be assessed assuming there were no biases in the measurement.

To transform experts' answers into priors for the precision of the random terms in the measurement equations, we assumed that the error $\xi$ in

$$z = y \times \xi, \tag{24}$$

Table 7: Experts answers concerning duration overcount, 6m vs. 3m criterion

| Resp | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Round 1 | | | | | | | | | | | |
| LP | 0.2 | 0.1 | 0.5 | 1 | 0.1 | 0.4 | 0.2 | 0.1 | 0.4 | 0.05 | 0.2 |
| HP | 0.6 | 0.25 | 1.5 | 3 | 0.2 | 0.7 | 0.4 | 0.3 | 0.65 | 0.15 | 0.5 |
| Cert | NA | 0.5 | 0.05 | 0.5 | 0.3 | 0.4 | 0.5 | 0.75 | 0.5 | 0.75 | 0.75 |
| Round 2 | | | | | | | | | | | |
| LP | 0.2 | 0.1 | 0.5 | NA | 0.1 | 0.4 | 0.2 | 0.05 | 0.3 | 0.1 | 0.3 |
| HP | 0.6 | 0.25 | 1 | NA | 0.2 | 0.7 | 0.5 | 0.15 | 0.5 | 0.3 | 0.6 |
| Cert | 0.75 | 0.5 | 0.25 | NA | 0.3 | 0.4 | 0.5 | 0.75 | 0.5 | 0.75 | 0.75 |

LP - Lowest proportion, HP - Highest proportion, Cert - Certainty

Source: Delphi survey

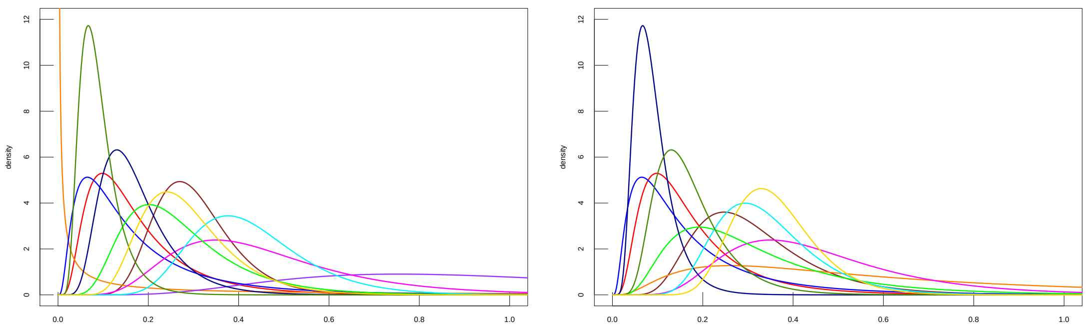


Figure 13: Expert answers transformed to densities for duration overcount, 3m vs. 6m, Rounds (left) 1 & 2 (right)

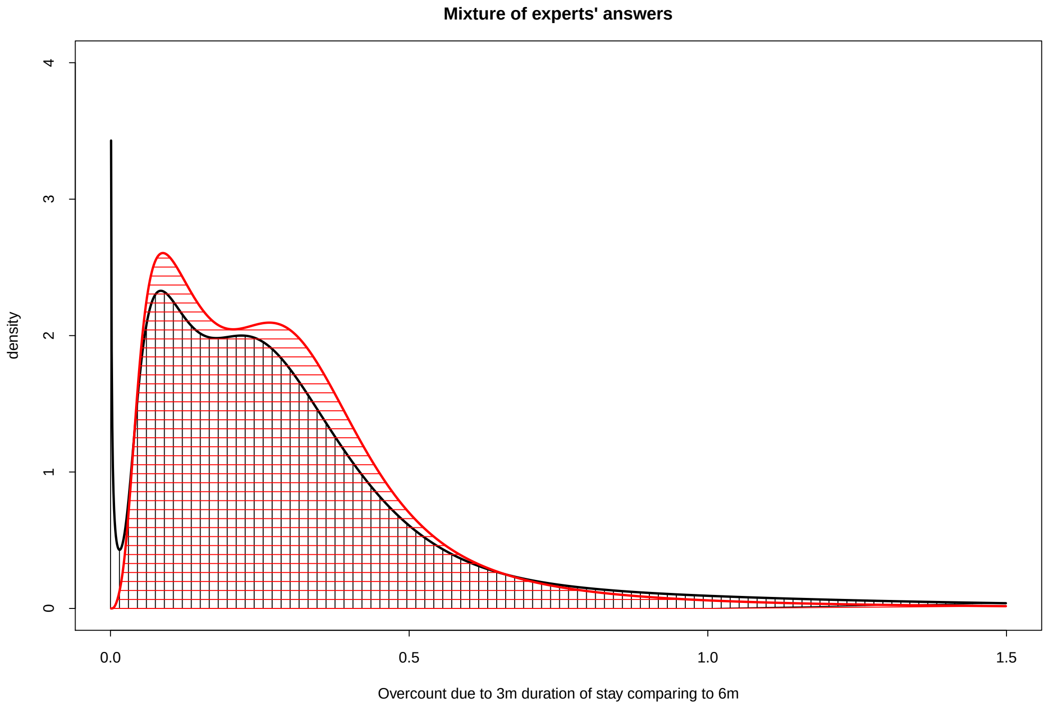


Figure 14: Mixture prior densities for duration overcount, 3m vs. 6m, Rounds 1 (vertical) & 2 (horizontal)

on the log-scale, was distributed normally with mean zero and precision $\tau$. Given the $\pm 5\%$ deviation from the true level of migration and two probabilities of such an event provided by the experts $P_i$, $i = 1, 2$, it followed that

$$P_i = \Phi(\log(1.05)\sqrt{\tau_i}) - \Phi(\log(0.95)\sqrt{\tau_i}). \quad (25)$$

Using the approximation $\log(1.05) \approx -\log(0.95) \approx 0.05$, we simplified the above equation into the following

$$P_i = 2\Phi(0.05\sqrt{\tau_i}) - 1. \quad (26)$$

Then the precision $\tau_i$ was computed as

$$\tau_i = 400 \left[\Phi^{-1}\left(\frac{P_i + 1}{2}\right)\right]^2, \; i = 1, 2. \quad (27)$$

For expert specific distribution of $\tau_i$ a gamma $\mathcal{G}(a, r)^3$ density was assumed. We could find the parameters $a$ and $r$ by solving set of equations

$$\begin{cases} F_g^{-1}(P_1, a, r) = q_1 \\ F_g^{-1}(P_2, a, r) = q_2. \end{cases} \quad (28)$$

This was achieved, similarly as in Section 3.2.1, by minimising to zero the expression

$$\min_{a,r} \left\{ \sum_{i=1}^{2} \left(F_g^{-1}(P_i, a, r) - q_i\right)^2 \right\}, \quad (29)$$

where

$$q_1 = \frac{(1-c)P_1}{1 + P_1 - P_2},$$
$$q_2 = \frac{(1-c)(1-P_2)}{1 + P_1 - P_2}$$

were proportional quantiles as given by Equation 14. Again, $c$ represents expert's confidence. For the cases where experts provided 0% or 100% probabilities, the formula cannot be used because it has no unique solution. To overcome this, these types of answers were transformed by replacing 0% with 0.01% and 100% with 99.99%.

As a starting point values for the optimising algorithm a log-normal approximation with parameters $\mu$ and $\sigma$ was used. They were calculated as

$$\sigma = \frac{\log(\tau_2) - \log(\tau_1)}{\Phi^{-1}(1 - q_2) - \Phi^{-1}(q_1)}, \quad (30)$$

$$\mu = \log(\tau_2) - \sigma\Phi^{-1}(1 - q_2). \quad (31)$$

Then, the expected value and the variance of the approximating log-normal density were computed as follows

$$E(\tau) = \exp(\mu + \sigma^2/2)$$
$$\mathrm{Var}(\tau) = (\exp(\sigma^2 - 1)\exp(2\mu + \sigma^2).$$

Finally, in order to find the starting point values for the minimisation algorithm, we solved the basic equations $E(\tau) = a/r$ and $\mathrm{Var}(\tau) = a/r^2$ for $a$ and $r$.

Table 8: Experts answers concerning accuracy of emigration measurement

| Resp | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Round 1 | | | | | | | | | | | |
| LP | 0.8 | 0.8 | 0.1 | 0.001 | 0.5 | 0.9 | 0.7 | 0.5 | 0 | 0 | 0.6 |
| HP | 0.95 | 0.9 | 0.2 | 0.1 | 0.7 | 0.95 | 0.8 | 1 | 0 | 0.1 | 0.9 |
| Cert | 0.9 | 0.75 | 0.75 | 0.9 | 0.4 | 0.5 | 0.5 | 0.9 | 0.95 | 0.95 | 0.9 |
| Round 2 | | | | | | | | | | | |
| LP | 0.8 | 0.8 | 0.1 | NA | 0.5 | 0.9 | 0.6 | 0.8 | 0.8 | 0 | 0.7 |
| HP | 0.95 | 0.9 | 0.2 | NA | 0.7 | 0.95 | 0.9 | 1 | 0.95 | 0.2 | 0.9 |
| Cert | 0.75 | 0.75 | 0.75 | NA | 0.4 | 0.5 | 0.5 | 0.9 | 0.75 | 0.9 | 0.75 |

LP - Lowest probability, HP - Highest probability, Cert - Certainty

Source: Delphi survey

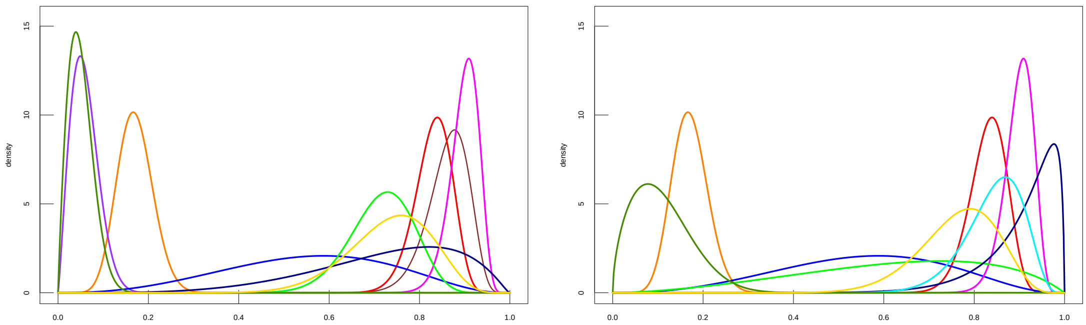

Figure 15: Expert answers transformed to densities for accuracy of emigration measurement, Rounds 1 & 2

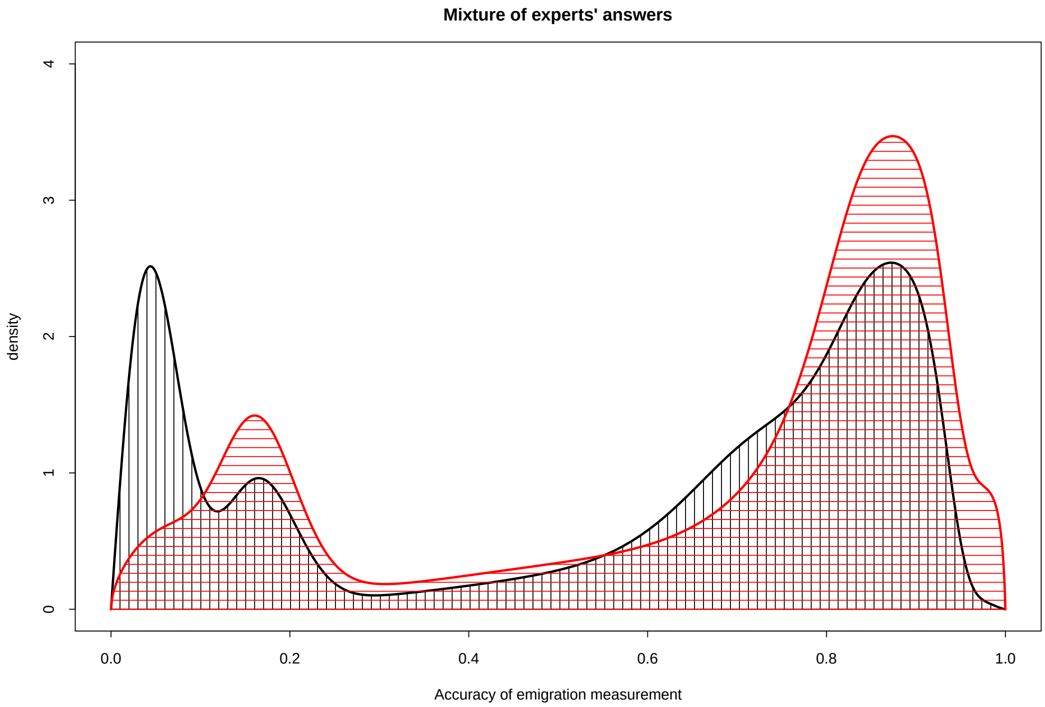


Figure 16: Mixture prior densities for accuracy of emigration measurement, Rounds 1 (vertical) & 2 (horizontal)

Table 9: Experts answers concerning accuracy of immigration measurement

| Resp | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Round 1 | | | | | | | | | | | |
| LP | 0.9 | 0.9 | 0.2 | 0.001 | 0.6 | 0.9 | 0.8 | 0.65 | 0.5 | 0 | 0.8 |
| HP | 1 | 0.95 | 0.4 | 0.1 | 0.8 | 0.95 | 0.9 | 1 | 0.6 | 0.25 | 0.95 |
| Cert | 0.9 | 0.9 | 0.05 | 0.9 | 0.5 | 0.5 | 0.5 | 0.9 | 0.75 | 0.95 | 0.9 |
| Round 2 | | | | | | | | | | | |
| LP | 0.9 | 0.9 | 0.2 | NA | 0.6 | 0.9 | 0.7 | 0.85 | 0.8 | 0.2 | 0.8 |
| HP | 1 | 0.95 | 0.4 | NA | 0.8 | 0.95 | 0.95 | 1 | 0.95 | 0.5 | 1 |
| Cert | 0.9 | 0.9 | 0.25 | NA | 0.5 | 0.5 | 0.5 | 0.9 | 0.75 | 0.9 | 0.75 |

LP - Lowest proportion, HP - Highest proportion, Cert - Certainty

Source: Delphi survey

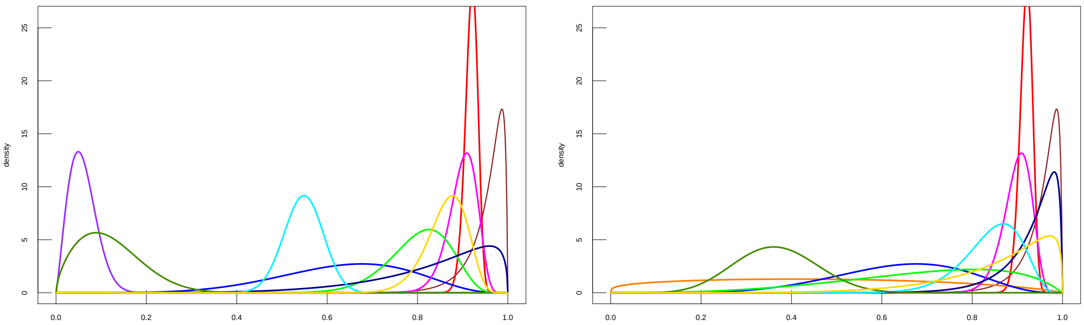

Figure 17: Expert answers transformed to densities for accuracy of immigration measurement, Rounds 1 & 2

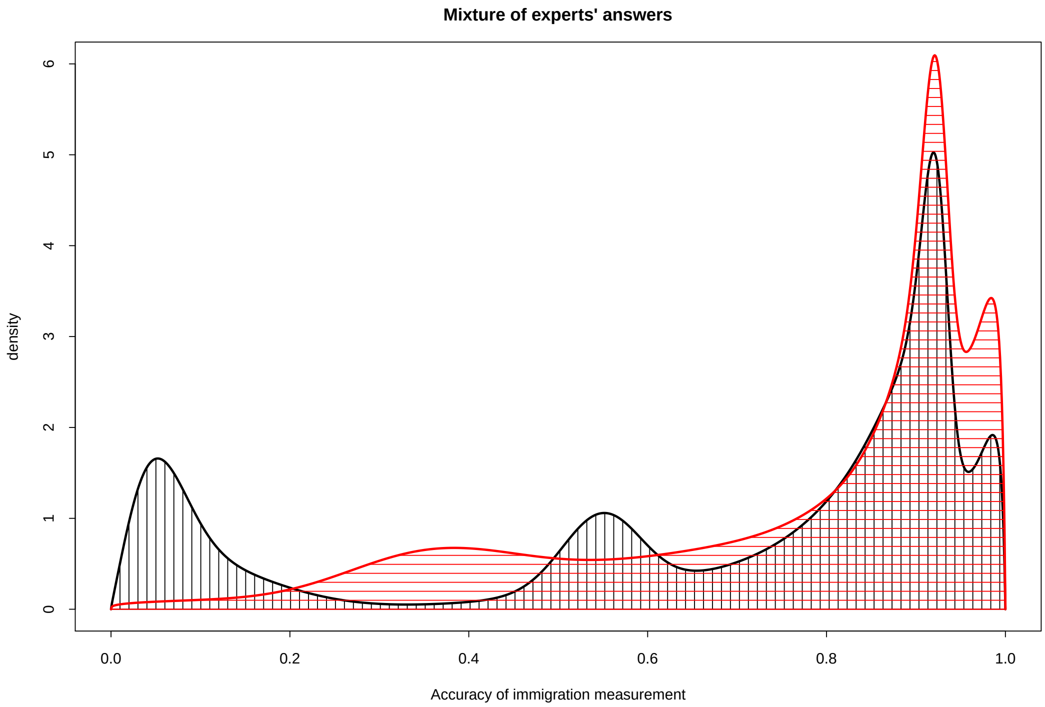


Figure 18: Mixture prior densities for accuracy of immigration measurement, Rounds 1 (vertical) & 2 (horizontal)

### 3.4.2 Expert answers and resulting prior densities

In Table 8 and Table 9 the expert answers in Round 1 and 2 to the question about accuracy of measuring emigration and immigration, respectively, are presented. We observe that in both first and second round the answers were diversified. About a third of all experts suggested that the measurement of both emigration and immigration is rather poor (they provided low probabilities), while the rest of experts stated that the data collection systems are accurate (they provided probabilities higher than 0.5). This heterogeneity could stem from the different backgrounds and experiences with various data collection systems in Europe and, as a result, different assessment of their accuracy.

Comparing to the first round, five respondents left their answers unchanged. The only major change in answer concerned respondent 9, who, in the first round, provided hardly interpretable numbers. The graphical representation of expert answers is presented in Figures 15 and 17. These individual representations were then used to produce the mixture prior densities for precision of the error terms. They are shown in Figures 16 and 18 for emigration immigration, respectively. We observe that the Round 2 prior mixture for the emigration implies slightly better evaluation of the accuracy comparing to Round 1 prior (it has more mass closer to 1 and less closer to 0). For immigration prior (Figure 18), we observe that a mode between 0 and 0.1 in the first round disappeared in the second round. This resulted from Respondent 4 dropping out of the study.

Although experts perceived the measurement of the immigration to be more accurate that of the emigration, their opinions were far from unanimous. Moreover, one of the experts, having seen the results of the first round, reduced the level of confidence in the second round.

# 4 Summary

In situations where data are inconsistent and weak, the inclusion of expert judgements are essential for improving the estimation and reflecting the uncertainty. In the IMEM model, we seek to provide the best possible estimates and measures of uncertainty based on available data, covariate information and expert judgements. These three pieces of information are integrated into a single model for estimating harmonised sets of migration flows between 31 countries in the EU and EFTA from 2002 to 2008. In this paper, we have described our methodology for obtaining expert information on migration data to supplement reported flows and covariate information in the IMEM model. Expressing knowledge and judgements in statistical terms and translating expert judgements into prior distributions is not an easy task and requires lots of care.

After two rounds of the Delphi survey, we found that experts often disagreed on the various measurement aspects of migration. The feedback from the first round did not lead to any significant changes in the opinions. However, we did not aim at convergence, as this could lead to an artificial reduction of uncertainty. Moreover, due to the heterogeneity of expert judgements expressed in the survey, the results are an important assessment of the quality of the data collection systems across Europe.

[3] Parameterisation of the gamma distribution is such that the expected value is $a/r$ and variance is $a/r^2$.

## Acknowledgements

This research is part of the Integrated Modelling of European Migration (IMEM) project funded by the New Opportunities for Research Funding Agency Co-operation in Europe (NORFACE). The authors would like to thank Rob van der Erf, Guy J. Abel and the twelve experts

- Corrado Bonifazi, Institute of research on population and social policies, Rome, Italy,
- Harri Cruijsen, Democast, the Netherlands
- Frank Heins, Institute of research on population and social policies, Rome, Italy,
- Michael Jandl, International Centre for Migration Policy Development, Vienna, Austria,
- John Kelly, Canada,
- Ewa Kępińska, Centre of Migration Research, Warsaw University, Poland,
- Marek Kupiszewski, Central European Forum for Migration Research, Warsaw, Poland,
- Giampaolo Lanzieri, Eurostat, Luxembourg,
- Joao Peixoto, SOCIUS, School of Economics and Management, Technical University of Lisbon , Portugal,
- Michel Poulain, Université Catholique de Louvain, Belgium,
- Dorota Kupiszewska, Central European Forum for Migration Research, Warsaw, Poland,
- Nicolas Perrin, Belgian Department of Immigration, Belgium,

for participating in this endeavour.

## References

[1] Bijak J, Wiśniowski A (2010) Bayesian forecasting of immigration to selected European countries by using expert knowledge. Journal of the Royal Statistical Society, Series A, 173.

[2] Jandl, M., C. Hollomey; and A. Stepien (2007): Migration and Irregular Work in Austria. Results of a Delphi-Study. International Migration Papers 90. International Labour Office; International Centre for Migration Policy Development. Geneva: ILO.

[3] Johnson, NL (1949) Systems of freqtuency curves generated by methods of translation. Biometrica 36, 149-76.

[4] Poulain M, Perrin N and Singleton A (eds.) (2006) THESIM: Towards harmonised European statistics on international migration. Louvain-la-Neuve: UCL Presses Universitaires de Louvain.

[5] Raymer J, Forster JJ, Smith PWF, Bijak J, Wiśniowski A and Abel GJ (2011) Integrated Modelling of European Migration: Background, specification and results, Paper prepared for the IMEM Workshop. Southampton, United Kingdom, 25-27 May 2011.

[6] Rowe G. and G. Wright (1999): The Delphi technique as a forecasting tool: issues and analysis. International Journal of Forecasting, 15 (4): 353-375.

[7] Rowe G. and G. Wright (2001): Experts Opinions in forecasting: the Role of the Delphi technique In: J.S. Armstrong (ed.): Principles of Forecasting: A Handbook of Researchers and Practitioners. Kluwer Academic Publications, Boston: 125-144.

[8] 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.

[9] United Nations (2008): Principles and Recommendations for Population and Housing Censuses, Revision 2.

[10] Wiśniowski, A. and J. Bijak (2009): Elicitation of expert knowledge for migration forecasts using a Delphi survey, CEFMR Working Paper, 2/2009.

# A Delphi questionnaire Round 1

NORFACE

Exit this survey

UNITED NATIONS DEFINITION OF INTERNATIONAL MIGRATION (1)

Long-term migrant
A person who moves to a country other than that of his or her usual residence for a period of at least a year (12 months), so that the country of residence effectively becomes his or her new country of usual residence. From the perspective of the country of departure, the person will be a long-term emigrant and from that of the country of arrival, the person will be a long-term immigrant.

The following point should be noted.

• In theory, the UN definition includes undocumented (“illegal”) migrants. In practice, the migration statistics in most countries do not cover undocumented migrants. When we refer to the UN definition as the benchmark, we do however include undocumented migrants.

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

IMMIGRATION AND EMIGRATION FLOWS

In this survey, immigration and emigration refer to total flows (two numbers per country).

Prev Next

NORFACE

Exit this survey

Section A

UNDERCOUNT IN MIGRATION BETWEEN THE EU/EFTA COUNTRIES

As stated in the introduction, we use the UN definition of a long-term migrant as benchmark.
(If, at any stage, you would like have another look at the UN definition please scroll down to the end of the questionnaire to go back to the previous page).

Consider a European country with a good population register e.g. Sweden or Finland that has fully adopted the UN definition.

Because migrants do not always have sufficient incentives to report their moves to the relevant authorities, migration statistics are often lower than the true total level. For immigrants this difference is thought to be smaller than for emigrants.

The following questions are restricted to intra-EU / EFTA migration.

**1a) By how many per cent do you expect that EMIGRATION flows are undercounted in the published statistics, as compared to the true total level of emigration? Please provide a range in percentages.**

Lowest percentage

Highest percentage

**1b) Approximately, how certain are you that the true undercount will lie within the range that you provided above?**

**I am about … % certain. Please tick the appropriate circle**

○ 50% ○ 75% ○ 90% ○ 95%

○ Other percentage (please specify)

**2a) By how many per cent do you expect that IMMIGRATION flows are undercounted in the published statistics, as compared to the true total level of immigration? Please provide a range in percentages.**

Lowest percentage

Highest percentage

**2b) Approximately, how certain are you that the true undercount will lie within the range that you provided above?**

**I am about … % certain. Please tick the appropriate circle**

○ 50% ○ 75% ○ 90% ○ 95%

○ Other percentage (please specify)

**3) If you have comments or arguments related to your answers above, please state them here.**

Section B

THE DEFINITION OF MIGRATION, INCLUDING DURATION OF STAY

The UN definition is based on duration of stay of at least 12 months. Few countries in the EU / EFTA actually use the 12 month criterion. Other countries use criteria equal to three or six months. In some cases, all moves into or out of the countries are directly registered, irrespective of duration. Still other countries only count permanent moves as migration.

The following questions are related to durations of stay of at least 12 months, six months, and three months. Note that the issue of undercount, asked for in the previous questions, does not play a role here. The questions are restricted to intra-EU/EFTA migration, as before.

Question 4

Consider a European country that uses a 12 month criterion. Now imagine that the six month criterion is used instead. With this new criterion, more persons are considered migrants compared to the previous criterion.

**4a) By how many per cent do you expect that the level of migration with the SIX MONTH criterion is higher than with the 12 MONTH criterion? Please provide a range in percentages.**

Lowest percentage

Highest percentage

**4b) Approximately, how certain are you that the true value will lie within the range that you provided above?**

**I am about … % certain. Please tick the appropriate circle**

○ 50% ○ 75% ○ 90% ○ 95%

○ Other percentage (please specify)

Question 5

Imagine now that the six month criterion is changed into a three month criterion. The result is that even more persons are now considered migrants.

**5a) By how many per cent do you expect that the level of migration with the THREE MONTH criterion is higher than with the SIX MONTH criterion? Please provide a range in percentages.**

Lowest percentage

Highest percentage

**5b) Approximately, how certain are you that the true value will lie within the range that you provided above?**

**I am about … % certain. Please tick the appropriate circle**

- 50%
- 75%
- 90%
- 95%
- Other percentage (please specify)

**6) If you have comments or arguments related to your answers above, please state them here.**

Section C
ACCURACY OF MIGRATION MEASUREMENT

The following questions are about random fluctuations in measured migration. We distinguish between emigration and immigration. Again, we restrict ourselves to intra-EU/EFTA migration. We also restrict ourselves to countries with population registers.

Consider a European country with a population register in which there is no systematic bias in the measurement of migration. In this case, we may expect random factors, for instance administrative errors in the processing of the data, to affect the level of migration that is actually measured.

**7) For EMIGRATION, how probable do you think it is that the published statistics are within an interval from minus 5% to plus 5% compared to the true total level of emigration? (If it helps think of how often the annual published statistics are within this interval during a period of 100 years). Please provide a range in percentages.**

Lowest percentage

Highest percentage

**8) Approximately, how certain are you that the true value will lie within the range that you provided above?**

**I am about … % certain. Please tick the appropriate circle**

- 50%
- 75%
- 90%
- 95%
- Other percentage (please specify)

**9) For IMMIGRATION, how probable do you think it is that the published statistics are within an interval from minus 5% to plus 5% compared to the true total level of immigration, as determined by the UN definition of migration? Please provide a range in percentages.**

Lowest percentage

Highest percentage

**10) Approximately, how certain are you that the true value will lie within the range that you provided above?**

**I am about … % certain. Please tick the appropriate circle**

50% 75% 90% 95%

Other percentage (please specify)

**11) If you have comments or arguments related to your answers above, please state them here.**

Section D

UNDERCOUNT IN MIGRATION FROM/TO COUNTRIES OUTSIDE THE EU/EFTA

Thus far, the questionnaire has focused on migration between EU/EFTA countries. Now we want to ask for your opinions on the undercount of migration from or to countries outside the EU / EFTA. These flows are distinct from the intra-EU / EFTA flows in that the migrants are often required to obtain permission to enter, remain or work. Remember to also include undocumented migrants in your answers.

**12a) By how many per cent do you expect that EMIGRATION flows to countries outside the EU / EFTA are undercounted in the published statistics, as compared to the true total level of emigration? Please provide a range in percentages.**

Lowest percentage

Highest percentage

**12b) Approximately, how certain are you that the true undercount will lie within the range that you provided above?**

**I am about … % certain. Please tick the appropriate circle**

50% 75% 90% 95%

Other percentage (please specify)

**13a) By how many per cent do you expect that IMMIGRATION flows from countries outside the EU / EFTA are undercounted in the published statistics, as compared to the true total level of immigration? Please provide a range in percentages.**

Lowest percentage

Highest percentage

**13b) Approximately, how certain are you that the true undercount will lie within the range that you provided above?**

**I am about … % certain. Please tick the appropriate circle**

50% 75% 90% 95%

Other percentage (please specify)

**14) If you have comments or arguments related to your answers above, please state them here.**

**Final Comments**
**Thank you very much for answering these questions. Your help is very much appreciated. We would be interested in any general comments, suggestions or questions you might have. These can be written in the box below.**

**Please tick below in case you are interested in receiving a brief summary of the project findings, which we intend to prepare after the research project has been finalised.**

Please send me a brief summary of the project findings

Prev Done