Nico Keilman, Georgios Aristotelous

Expert opinion on migration data

Deliverable 6.1

QuantMig has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 870299.

History of changes

Version	Date	Changes
1.0	12 January 2021	Issued for Consortium Review
1.1	15 February 2021	First version submitted as official deliverable to the EC

Suggested citation

Keilman N, Aristotelous G (2021) Expert Opinion on Migration Data. QuantMig Project Deliverable D6.1. Southampton: University of Southampton.

Dissemination level

PU Public

Acknowledgements

Comments made by Arkadiusz Wiśniowski, Emilio Zagheni, Emanuele Del Fava, Constantinos Melachrinos, Hania Zlotnik, and Giampaolo Lanzieri are gratefully acknowledged. This document reflects the authors' views. The Research Executive Agency of the European Commission are not responsible for any use that may be made of the information it contains.

Cover photo: iStockphoto.com/Guenter Guni

Elicitation of expert opinions

QuantMig Deliverable 6.1 Part I *

Nico Keilman †

*QuantMig is funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 870299.

†University of Oslo, Norway. Email: nico.keilman@econ.uio.no

Contents

1 Introduction

The aim of the QuantMig project is to produce comprehensive, multi-perspective, and robust quantitative migration scenarios to support European policy makers when they prepare and evaluate policy on international migration to Europe, and between the countries. The QuantMig-scenarios will be based on a set of statistical estimates of international migration flows and their drivers. The scenarios rely on *estimates*, not on *observed* statistics (e.g. from statistical agencies), because there are several problems connected to these data, such as inconsistencies in data availability, quality, and collection mechanisms. Statistical agencies in some countries publish statistics on international migration annually, and they are of good quality. Other agencies provide us with crude estimates derived from a comparison of population stocks from two censuses, with inter-censal counts of births and deaths. Population registration systems result in more accurate statistics for migration flows than migration surveys. Registration of emigrants is generally thought to be less accurate than that of immigrants, but the undercount of emigration flows is larger in some countries than in others. Therefore, available information on migration flows across Europe needs to be reconciled.

This report documents some of our work in connection with Work Package 6 (WP6) of QuantMig. The project proposal of QuantMig describes the aim of WP6 as "… (t)o develop a method for estimating European migration flows based on the available data, with uncertainty assessment, and to apply it to creating a custom-made, harmonised dataset based on reconciling secondary data from different sources, augmented by using expert opinion …". WP6 is one of the 12 work packages of QuantMig. Figure 1 shows how it is situated in the whole project. It uses migration data and other information from several sources as input, whereas its results, i.e. the estimates of (true but unknown) European migration flows, form the input to a set of other work packages that deal with the building of various scenarios for future migration in the region.

WP6 consists of several tasks. One of these (Task 6.1) concerns eliciting meta-information about systems for data collection in Europe, such as measurement accuracy and under-

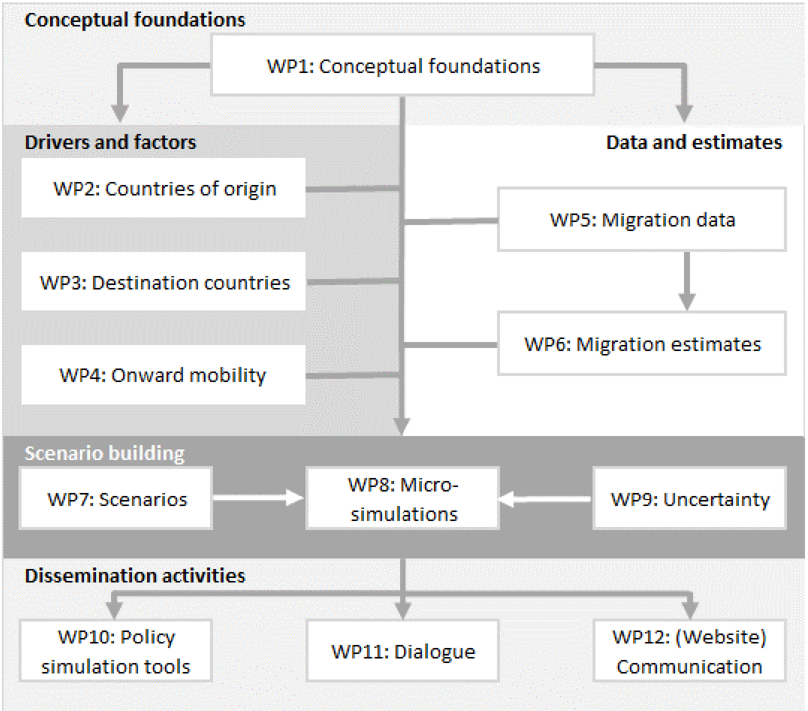


Figure 1: Logical structure and key relationships of QuantMig

count, etc., from experts on European migration data. Other tasks for WP6 involve assessing the quality of information about migration flows, migrant stocks, and a range of covariates, constructing a model for estimating migration flows, and dissemination of the results.

The aim of this report is to document work done for Task 6.1. The task was to answer the following research question: how well do publicly available statistics on international migration for European countries, more in particular statistics on international migration from national statistical agencies, reflect the true migration flows? We have attempted to answer this question by eliciting expert opinions on measurement accuracy and undercount. During the summer and autumn of 2020, we carried out a two-round Delphi survey among 15 migration experts, preceded by a pilot survey with eight participants. The online questionnaire included questions on measurement accuracy of migration data from population registration systems and from migrant surveys, on possible undercount of immigration and emigration flows, and on the assumed impact of the Covid-19 pandemic on European migration flows.

Our work on accuracy and undercount builds on the experience and lessons learned from the project "Integrated Modelling of European Migration" (IMEM) (Raymer et al. (2013); Wiśniowski et al. (2016)). The estimation involves Bayesian modelling (Bijak and Bryant (2016)), which provides a coherent description of uncertainty at different levels, and integrates data from different national sources. Since the parameters of the model cannot be identified from the data alone, we elicited them from our 15 domain experts in the form of probability distributions (Wiśniowski et al. (2013)).

The estimates will be calculated for 2009–2018, extending the IMEM dataset by another decade, and will be broken down by sex, age in five-year groups, broad region of origin/destination outside Europe, and nationality group (EU/non-EU).

Table 1 gives the time line of the survey. Eight migration experts took part in the pilot survey, and 16 responded to the Delphi survey. We acknowledge gratefully the help of the 24 experts. All of them contributed in their own personal capacity to the project. While the responses were anonymous, we wish to list, with thanks, the experts' names: Guy Abel, Jakub Bijak, Corrado Bonifazi, Jon Forster, Anne Goujon, Karen Haandrikman, Frank Heins, Bela Hovy, Giampaolo Lanzieri, Wolfgang Lutz, Marie McAuliffe, Beata Nowok, Nicolas Perrin, Joao Peixoto, Michel Poulain, James Raymer, Philip Rees, Luule Sakkeus, Nikola Sander, Ann Singleton, Peter Smith, Frans Willekens, Arkadiusz Wiśniowski, Hania Zlotnik.

This report consists of two parts. Part I, written by Nico Keilman, focuses on elicitation of expert opinions. Part II, of which Georgios Aristotelous is the author, documents how expert opinions on undercount and accuracy were transformed into statistical distributions. The latter distributions form part of the input for the model for estimating true migration flows.

In Part I, Section 2 summarizes our approach and gives a qualitative description of the model used for estimating migration flows. Next, in Section 3 we describe the way we

13 May 2020	Ten experts invited to take part in the pilot survey
28 May – 3 June 2020	Pilot survey opened to eight experts
28 May – 15 June 2020	Responses received from eight experts
18 – 19 June 2020	24 new experts invited to take part in the Delphi survey
8 July 2020	1st round opened to 16 experts
8 July – 3 August 2020	Responses received from 16 experts
26 October 2020	2nd round opened to 15 experts
27 October – 16 November 2020	Responses received from 15 experts

Table 1: Timeline for Task 6.1

selected our experts for the Delphi survey and for the pilot survey. The questionnaire for the pilot survey is presented in Section 4. Section 5 discusses the responses and feedback we received from the experts in the first round of the Delphi survey. Translation of the responses into to probability distributions (see Part II) revealed that some experts had given invalid answers to one or more questions. Hence, the second Delphi round (Section 6) included improved explanations and wording of the questions. Indeed, the results of the second round, also described in Part II, turned out to be more consistent. Section 7 summarizes our work on the Delphi survey and concludes. In Part II, after the introductory Section 1, Section 2 describes the methods used to translate experts' answers into prior probability distributions for the parameters. Next, Section 3 presents the resulting probability distributions, for both rounds of the Delphi questionnaire. Section 4 concludes with some discussion points about lessons learned from the elicitation process.

2 Approach

2.1 Definition of migration

The survey aimed at eliciting expert opinions about how a specific measurement of international migration deviates from a benchmark. As the benchmark, we have adopted the United Nations definition (United Nations 1998). This definition corresponds with the definition included in EU Regulation nr. 862/2007 on Community statistics on migration and international protection, adopted by the European Parliament in 2007.

United Nations definition of international migration:

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 UN definition employs the notion of "country of usual residence". This is to be understood as the country in which the person has lived or intends to live for a period of at least 12 months, where by "lived" more precisely is to be understood as the place where this person has taken his or her daily rest (see U.N. (1998)). In practice, many countries in Europe use a de jure definition (especially countries with a population register), and a country's de jure definition may very well be different from the rule of daily rest. We assumed, however, that such differences (other than the rule of 12 months, which will be captured by a factor called "duration"; cf. Section 2.2) are not very important for international migration (as opposed to migration within one country) and that these definitions will generally result in the same country of residence. Hence, we ignored possible differences in measured migration flows caused by different definitions of country of usual residence.

Related to the previous point is the notion of undocumented migration. In theory, the UN definition includes undocumented ("illegal") migrants. However, article 9 of EU regulation 862/2007 states that "This Regulation does not cover estimates of the number of persons illegally resident in the Member States". Indeed, the migration statistics in most countries do not cover undocumented migrants. When we refer to the UN definition as the benchmark, we do not include undocumented migrants.

Also, note that in some countries, short-term labour migrants apply repeatedly for renewal of their temporary (less than twelve months) months working permits. To the extent that they do not leave the country in-between, they are long-term migrants.

The European Union introduced the above definition formally in 2007. At that time, a number of EU-member countries used it already in their official statistics on international migration, whereas other countries employed a slightly different one, for instance based on an intended or actual period of stay of six months, instead of twelve. Still other countries applied the rule that any person, who had left the country only temporarily, still was considered as a resident of the country. Only those who had left permanently were counted as emigrants.

2.2 True migration flows and reported migration flows

As stated in the introduction, the aim of the model developed in WP6 is to estimate European migration flows based on the available data, with uncertainty assessments, and to apply it to creating a custom-made, harmonised dataset based on reconciling secondary data from different sources, augmented by using expert opinion. The model updates and extends an earlier model developed in the framework of the IMEM project mentioned before. The model estimates counts for flows of international migration for 32 European countries: EU-27 plus EFTA countries (Iceland, Norway, Liechtenstein, and Switzerland), plus the United Kingdom. Collectively, we refer to them as EU+. The flows concern both inter-EU+ migration, immigration to each EU+ country from outside the region, and emigration from each country by persons who leave the region.

The model assumes that the reported flow between two countries (reported either by the sending or by the receiving country) is a function of the real (but unknown) flow, and four parameters that reflect certain features of the data collection system.

1. A duration parameter, cf. the discussion above on actual or intended duration of stay abroad.
2. An undercount parameter, reflecting the fraction of the true flow that is captured by the data collection system in a given country. 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.

3. A coverage parameter, reflecting the discrepancies between the observed data and the true flows that are not captured by the more general undercount parameters. Coverage refers to the fact that some countries do not include all types of migrants in their measurements, for instance international students or nationals. Hence, coverage errors in reported migration flows are a consequence of a data collection system that does not follow the UN definition. They differ from undercount errors, which are caused by individuals who do not report their moves, although they should.

4. A general error term, capturing any remaining difference between reported and actual flows. This term relates to random errors in migration measurement, for instance administrative errors in the processing of the data. It is different from "coverage", which refers to systematic errors (“bias”). The variance of this error term measures the accuracy of the data collection system for each country. It reflects the quality of the data and measurement mechanisms utilised to collect the data.

The four parameters listed here cannot be estimated from data alone, because the real migration flow is unknown. In case one wants to obtain such estimates, a number of additional assumptions are necessary. Our approach is to do that in a systematic way, and to involve several experts on European migration. We have organized a two-round Delphi survey among 15 experts and asked them to give their (admittedly subjective) opinions about these issues. To reduce the survey burden, we have restricted ourselves to eliciting opinions on undercount (point 2 above) and accuracy (point 4). Adoption of EU Regulation nr. 862/2007 may have resulted in less undercount and better accuracy in recent years, compared to the period covered by the IMEM-project (2002 – 2008). Hence, it was necessary to update the IMEM-estimates for undercount and accuracy. Information on duration (point 1) and coverage (point 3) is taken from the IMEM-project.

Following the approach taken in the IMEM-project, we asked each expert to give us a set of values concerning certain parameters, which we then converted into probability distributions. The totality of resulting expert opinions was subsequently combined into a single set of distributions, allowing for the introduction of yet another source of uncertainty, related to the heterogeneity of experts; see Part II.

Undercount, accuracy, and other characteristics regarding the measurement of international migration, have varied over time. We asked the experts to answer the questions as related to the situation, on average, during the period 2009 - 2019 (before Brexit) in the EU+ countries. This means, among other things, the extent to which countries have implemented the EU Regulation 862/2007.

Given the special situation in 2020 with many countries struck by the Covid19-pandemic, we took this opportunity and asked the 23 experts to give a qualitative assessment of the effects the pandemic might have on European migration. We restricted this topic to the pilot round and the first Delphi round.

2.3 Undercount and accuracy

It is clear that measurement practices differ widely among European countries. Yet it is not feasible to ask questions for each of the EU+ countries. Instead, we asked the experts to state their estimates for each parameter in terms of a range, together with an associated level of confidence or certainty. The ranges and confidence levels should reflect the expert's belief about the variability between countries, but also how certain they were about the answers.

We asked the experts to give undercount as a percentage. This percentage refers to the ratio between not counted and the real (but unknown) flow. More formally, assume that there are no errors caused by duration, coverage, or accuracy. Let $P \times 100\%$ ($P \in [0, 1]$) denote the percentage of undercount in a given situation. The underlying assumption regarding undercount is that $(1 - P)y = z$, where y is the true flow and z is the reported flow. Thus, we can interpret P as a fraction of the true flow that is not captured in the reported data.

An ideal measurement system has an undercount of zero. The larger the percentage of undercount, the worse the system performs. An undercount of 100 % applies to the extreme situation in which none of the migrants is recorded by the system.

For instance, assume one expert stated that he is about 75 per cent certain that undercount was between 20 and 60 per cent in a given situation. We have used these numbers to construct a probability distribution for this type of undercount such that chances are $100 - 75 = 25$ per cent that undercount was less than 20 per cent or more than 60 per cent. Hence, each range must be seen together with the stated certainty – they all reflect one probability distribution; see Part II for details.

For undercount, we have grouped the 32 countries into two categories: low undercount countries and high undercount countries. The grouping is based on the findings in the IMEM-project; see Wiśniowski et al. (2013).
Low undercount countries: The Netherlands, Sweden, Finland, Norway, Denmark, Germany, Iceland, Austria, Belgium, United Kingdom, Cyprus, Ireland, Italy, France, Luxembourg, Switzerland, and immigration to Spain.
High undercount countries: Bulgaria, Croatia, Estonia, Lithuania, Latvia, Poland, Slovenia, Slovakia, Romania, the Czech Republic, Greece, Hungary, Liechtenstein, Malta, Portugal, and emigration from Spain.

With reference to accuracy, we have distinguished two types of systems, namely population registration systems and migration surveys. Cyprus, Ireland, Portugal, and the United Kingdom use a survey for collecting data on international migration. With the exception of France, Greece, and Liechtenstein, the remaining EU+ countries use a register [1]. Whereas random errors may occur in both systems, a survey has an additional error source, namely sampling errors.

[1] France uses an alien register for immigration, whereas some information on immigration to Greece is available from residence permits for foreigners. Liechtenstein does not collect any information on migration

As noted above, undercount is an error caused by the individual migrant who does not report his or her move. Clearly, this is predominantly a problem in register countries. Undercount may occur when migration information is derived from a survey, but that is, in fact, the case when a survey respondent does not answer the question about migration. Coverage errors, we repeat, occur when national authorities do not follow the UN-definition and disregard certain population sub-groups. For a survey country, this means that the survey population from which the sample was drawn excludes that subgroup and hence the migration question is not asked to the persons concerned.

Both for questions related to undercount and for those on accuracy, we distinguished between nationals, i.e. persons with nationality in one of the 32 EU+ countries, and non-nationals.

3 Experts

To select the experts for the Delphi survey was a difficult task. An expert, in this connection, should be a person who is knowledgeable about data collection systems for international migration data in many, if not all, EU+ countries. At the same time, experts should have a basic understanding of quantitative aspects of migration flows. In addition, we aimed for a certain regional distribution, and for both men and women. An important factor was to achieve a certain heterogeneity among the experts, since this was thought to stabilize the estimates of the model. For a brief discussion, see Wiśniowski et al. (2013), and the references therein.

Rowe and Wright (2001) recommend between five and 20 experts for a Delphi survey. Of the 24 experts whom we invited (6 from Eastern Europe, 12 from the remaining parts of Europe, 6 from international organisations; 7 women and 17 men), 16 agreed to take part in our survey. In the end, we received useful responses for the two Delphi rounds from 15 experts. The Delphi survey was preceded by a pilot survey among eight experts, partly selected from the group of QuantMig researchers and Advisory Board members.

Following Wiśniowski et al. (2013), the expert-specific probability density functions constructed in part II got equal weights when we computed the overall density. This method is simple and robust. We could have asked the experts about a particular variable, but we did not know the real values of any of the parameters. The expertise we are eliciting is very specific – to ask related questions (e.g. about European flows) would miss the mark.

4 Pilot round

In the middle of May of 2020, we invited the eight experts to participate in the pilot round. The questionnaire was open for responses from the end of May until the middle of June, when all eight had responded.

Appendix A contains the questionnaire used in the pilot round. It is an online questionnaire, designed by means of the "Nettskjema"-package, which was developed by the University of Oslo; see https://nettskjema.no/index.html?lang=en . After an introduction, in which we explain various aspects of measuring international migration and the way we intend to elicit the participants' opinions, there are four sections with questions. Section 1 contains questions on undercount for immigration and emigration flows of nationals and non-nationals. The questions are restricted to countries with low undercount. Section 2 repeats these questions for migration flows to and from high undercount countries. The eight questions in these two sections are all of the form: "By how many per cent do you expect that emigration/immigration flows of nationals/non-nationals who enter low undercount/high undercount countries are undercounted in the published statistics of those countries, as compared to the true level of emigration/ immigration of nationals/non-national? Please provide a range in percentages between 0 and 100." After having specified the range, the experts were asked how certain they were that the undercount of the reported flow (compared to the actual flow) was within the range they had specified. The experts could select from a set of predefined percentages (50, 75, 90, or 95 per cent), but also they had the option to state a different percentage. Section 3 addresses the accuracy of data collection in countries that have a population register, or that use a survey. A register country can collect information both for immigration and for emigration flows, whereas the information for a survey country is restricted to immigration. Again, there are separate questions for national and non-nationals. Finally, Section 4 of the questionnaire has questions on the assumed short-term (the year 2021) and long-term (2021 – 2026) impact of Covid-19 on migration flows in Europe. We asked the experts to give their assessment on a nine-point scale, ranging from very much lower flows in future years, compared to the flow for an average pre-Covid-19 year, to very much higher flows.

The results of the pilot round were encouraging. Eight experts had agreed to take part in the pilot round, and we received useful information on undercount and accuracy from seven of them. One expert informed us that (s)he did not have enough knowledge of official statistical systems to have any informed opinion on these issues, but sent us responses on the Covid-19 effects on European migration.

On average, the experts used about 35 minutes to answer the questions. The answers that the experts provided about ranges (1a to 14a) and confidence (1b to 14b) were of less interest than the comments they gave. We received a number of useful suggestions for improved formulations and lay out. Some comments indicated that the pre-amble did not explain the difference between undercount and coverage clear enough. One expert proposed to weigh up experts from countries that use a survey to collect information about international migration, because such cases are quite special. This topic is addressed in Part II. Finally, there was a comment regarding questions 15 and 16: the difference between "slightly lower/higher" and "somewhat lower/higher" was unclear. We deleted the latter category in the Delphi rounds, such that seven possible answers remained (in addition to "don't know").

The other suggestions led to a number of small changes in the questionnaire for the first Delphi round. Therefore, we do not report this version here.

5 First Delphi round

In the middle of June of 2020, we invited 24 experts to participate in the Delphi survey. The 16 experts, who had agreed to do so, were given access to the online questionnaire on 8 July. On 3 August, we had received responses for most, if not all questions from 15 experts, whereas one expert only had answered questions 15 and 16 on the effect of Covid-19 on European migration flows in the future.

Overall, the responses we received were informative, although there were a few problems. One expert was not able to answer questions 13 and 14. A few experts had difficulties to interpret questions 9-14 on accuracy correctly. In most cases, this could be resolved by direct email communication, but this was not always possible. One expert was very sceptical about the Delphi approach and about formulating subjective beliefs for undercount and accuracy in terms of ranges and levels of confidence. Another comment from this expert: "My answers are based on pure guesswork. Much better would have been to present some recent statistics on each of the collective flows and asked for probability ranges around those statistics ... These answers are pure guesses, as before." Finally, some answers revealed an impossible combination of range and confidence, i.e. a range between 0 and 100 per cent, with an associated confidence of less than 100 per cent. Problems of this kind, which did not show up in the pilot round, are demonstrated more in detail in Part II. In the second Delphi round, we reported the problematic cases back to the experts. This gave them the possibility to improve their answers. In addition, we improved the wording and explanations of relevant sections in the questionnaire for the second Delphi round. Note that the results reported in this section are based on raw data, i.e. they include a few invalid answers from some experts.

Tables 2 and 3 summarize the responses from 15 experts in the form of descriptive statistics. In Part II, we show how we used these responses to construct probability distributions.

Table 2 gives results for questions 1 – 8 on undercount of reported migration flows. The table shows average values and standard deviations [2] for the reported lower and upper bounds of the ranges, as well as the confidence parameter. The averages and standard deviations are computed across the 15 experts.

	Low undercount countries				High undercount countries			
	Nationals		Non-nationals		Nationals		Non-nationals	
	Emigration	Immigration	Emigration	Immigration	Emigration	Immigration	Emigration	Immigration
	Average (%)							
Lower bound	10.9	7.3	16.2	8.4	23.5	19.1	26.5	14.9
Upper bound	35.3	24.3	44.1	23.7	61.8	55.1	59.0	42.6
Confidence	72.3	75.0	69.7	73.3	61.0	63.7	62.7	65.3
	Standard deviation (% p.)							
Lower bound	8.8	7.1	11.7	6.7	15.5	15.7	18.2	12.4
Upper bound	19.1	19.4	22.6	14.8	23.8	28.2	25.4	25.5
Confidence	12.6	16.7	17.5	15.5	13.9	15.3	14.0	15.1

Table 2: Summary statistics for responses on undercount, first round of Delphi survey. The average expert thinks that the reported emigration flow of nationals from low undercount countries is between 10.9 and 35.3 per cent lower than the real flow, and is 72.3 per cent certain that this range is correct.

Not surprisingly, the upper panel of Table 2 shows that undercount is thought to be less, on average, in countries with low undercount compared to high undercount countries: the average bounds (both upper and lower bounds) for the ranges are lower and the average levels of confidence are higher. For instance, the average range for immigration of non-nationals to countries with high undercount stretches from 14.9 to 42.6 per cent, with an average level of confidence of 65.3 per cent. Compare this with the corresponding range for countries with low undercount: from 8.4 to 23.7 per cent on average, while average confidence is 73.3 per cent. In addition, the average expert gives a wider range and is less certain about this range for emigration compared to immigration. All these results are as one could expect. Indeed, as one of the respondents commented: "As far as it is known, the measurement of immigration is usually more accurate [3] than the measurement of

[2]Since the purpose is to present descriptive statistics, we used "population" standard deviations with $N = 15$ (and not "sample" versions with $N - 1 = 14$) in the denominator.

[3]Author's comment: "accurate" in the sense of low undercount.

emigration. The same occurs with the movements of non-nationals, when compared to nationals." A narrow range (for instance low undercount, compared to high undercount, or immigration, compared to emigration) combined with high confidence leads to a more peaked probability density function (i.e. having relatively low variance) than a wide range combined with low confidence; see Part II.

Is there much variation across experts in the ranges and in the confidence levels they reported? The lower panel of Table 2 gives the standard deviation in each answer. The ranges are in percentages, and hence the standard deviations are in percentage points. The standard deviations are difficult to interpret as such, but comparing them across categories may be useful. Not surprisingly, the variation in reported upper and lower bounds of ranges is a bit larger in countries with high undercount than in low undercount countries. There is no clear pattern in variation in confidence levels, which are rather stable at roughly 13 to 17 percent points.

Table 3 summarizes expert responses about accuracy (questions 9 – 14). We asked the experts to give a range, indicating how probable it is that reported migration flows are within ±5 per cent of the true flows due to random errors only. This means that a higher accuracy parameter means a more positive view, as the answer reflects the probability that the error is within the ±5% interval (the smaller, the worse) [4]. Again, we report average values and standard deviations for the upper and lower bounds of each range, and average confidence.

On average, the ranges are wider for immigration to survey countries (61.9 – 38.6 = 23.3 percentage points for nationals, 19.4 percentage points for non-nationals) than for immigration to register countries (12.1 and 12.0 percentage points, respectively). At the same time, the experts have less confidence (around 60 per cent) in having stated correct intervals for survey countries, compared to register countries (between 66 and 72 per cent). Again, this looks reasonable, given the fact that sampling errors are an extra source of

[4] As opposed to the question on undercount, where a higher undercount parameter means a more negative view (the smaller, the better).

	Register countries				Survey countries	
	Nationals		Non-nationals		Immigration of nationals	Immigration of non-nationals
	Emigration	Immigration	Emigration	Immigration		
	Average (%)					
Lower bound	45.9	62.2	46.5	59.1	38.6	43.7
Upper bound	63.4	74.3	67.2	71.1	61.9	63.1
Confidence	68.0	72.0	66.0	69.3	60.7	59.6
	Standard deviation (% p.)					
Lower bound	35.7	33.9	32.5	33.1	25.9	16.8
Upper bound	33.5	34.4	30.6	32.2	28.2	21.4
Confidence	22.8	23.7	22.7	23.7	21.0	21.4

Table 3: Summary statistics for responses on accuracy, first round of Delphi survey. The average expert thinks that it is between 45.9 and 63.4 per cent likely that the reported emigration flow of nationals from register countries is within -5 and +5 per cent of the real flow, and is 68 per cent certain that this range is correct.

uncertainty in survey countries. The standard deviations do not exhibit any clear pattern.

Table 4 shows results for questions on the effect of Covid-19 on future migration flows (questions 15 and 16). Here we give the qualitative assessments of 16 experts. Only one expert believes that Covid-19 will lead to (slightly) larger European migration flows to 2026, compared to migration in an average year in the recent past. The majority expects much smaller flows, although in the long term, i.e. for the years 2021 – 2026, about one-third foresees very much smaller flows. As one expert commented: “The COVID-19 will have depressed international migration hugely in 2020. Assuming no effective vaccine, this will continue over the medium term. Even with a vaccine, the vaccine might only confer immunity for a limited period. So migration will continue to be depressed.”

Obviously, the responses reflect the experts’ qualitative assessments. We are not at all certain if “slightly lower” for one expert is very different from “much lower” for another expert. Moreover, in order to keep the response burden within reasonable limits, we did not explicitly request the experts to give arguments for their choices (although a few did so when they gave comments). Hence, the results in Table 4 only give a rough indication of the assumed effects of Covid-19 on future migration flows. Yet they will provide useful input to other QuantMig work packages that deal with the building of various scenarios for future migration in the region. In these scenarios, other drivers than Covid-19 will also have an effect on migration flows.

	… be very much lower.	… be much lower.	… be slightly lower.	… not change.	… be slightly higher.	Sum
Because of Covid-19, …						
			Year 2021			
… migration flows between EU+ countries will …	2	10	4	0	0	16
… immigration flows into EU+ countries will …	3	8	5	0	0	16
… emigration flows from EU+ countries will …	1	10	5	0	0	16
			Period 2021 – 2026			
… migration flows between EU+ countries will …	7	7	0	1	1	16
… immigration flows into EU+ countries will …	5	9	0	2	0	16
… emigration flows from EU+ countries will …	6	10	0	0	0	16

Table 4: Frequency table for qualitative assessment of the effect of Covid-19 on future European migration flows. Answers received from 16 experts.

We considered carefully the critical comments by some experts. We had to be very restrictive in changing the questions for the second round, because doing so would distort the Delphi approach. A few experts commented our grouping of countries into high and low undercount countries; see the list in Section 2.3. The former group is very heterogeneous, which means that it might be difficult to answer the questions for this group. One expert commented that for non-nationals (non-EU), the use of residence permits (not available for nationals) could make a huge difference to limit the undercount. This an issue that we had not considered before. Residence permits are definitely an important data source in migration statistics. However, their use is not without issues, and there is no clear one-to-one correspondence between migrants and permits [5]. Another expert mentioned that a number of Central and Eastern European countries allow "secondary" residency. This means that some emigrants do not de-register when they leave. However, since we use the UN definition of migration as the benchmark, this implies that in fact,

[5] For more details, see https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.10/2018/mtg1/Eurostat_Integration_ENG.pdf.

we focus on “primary” residency. Problems with the notion of “accuracy” were reported a few times, as mentioned in the beginning of this section.

6 Second Delphi round

Because of the problems some experts had in the first round with a correct specification of the ranges (see Part II), we improved the text that explained this notion in the questionnaire, and included automatic checks on values for upper and lower bounds (larger than zero and lower than 100 per cent) in the online questionnaire. In addition, we included for each of the 14 questions, a table with anonymous results from the first round [6]. This allowed the experts to reconsider and revise their opinions. Appendix B contains the questionnaire for the second round.

On 26 October 2020, we invited the 15 experts to respond to the second round of the Delphi survey. They received the technical note with specific comments on their own responses in the first round, if necessary; see Part II for examples. We closed the online questionnaire on 16 November 2020, when all 15 experts had responded.

All responses on ranges were inside the interval between zero and one hundred per cent. Hence the technical note served at least one of its important purposes, namely to avoid invalid responses. As one expert commented: “ ... it was very helpful to see the graphs of my answers as well as the other respondent results. In the future, you might want to integrate an interactive graph so the respondents can immediately see the results of their answers ...” Another expert: “... Maybe next time you can try and build an interface that shows the generated distribution instantly, saving the respondent the mental burden of trying to imagine the distribution that they are trying to imply.... “. Indeed, invalid answers in the first round could have been avoided in case we would have included in the survey an online tool that builds probability densities. This was not foreseen at the time when we planned this task nr. 6.1, but it is certainly a lesson learned for the future.

[6] Note that expert nr. 8708183 specified many upper and lower bounds that are not multiples of five. This expert used an explicit method for estimating undercount that builds on two tables of migration flows by country of origin and country of destination. One table contains such numbers from the perspective of the countries of origin, while a second table is based on information for the countries of destination.

Table 5 summarizes the responses for undercount. The general picture that emerges is close to that we saw earlier for the first round. Other things remaining the same, ranges on undercount are smaller for low than for high undercount countries, and for nationals compared with non-nationals. Average bounds in this second round are very similar to those in the first round. The experts have become somewhat more confident that the ranges they specified include the real undercount value, in particular for countries with high undercount.

	Low undercount countries				High undercount countries			
	Nationals		Non-nationals		Nationals		Non-nationals	
	Emigration	Immigration	Emigration	Immigration	Emigration	Immigration	Emigration	Immigration
	Average (%)							
Lower bound	11.3	6.9	14.9	6.9	23.7	20.7	27.5	14.7
Upper bound	33.9	22.8	43.7	22.4	61.5	53.9	63.2	46.3
Confidence	75.3	78.7	74.0	73.0	71.7	70.0	69.0	73.3
	Standard deviation (% p.)							
Lower bound	7.2	6.5	9.9	4.8	14.3	15.3	18.8	9.6
Upper bound	8.4	8.9	16.0	7.8	17.5	19.1	22.3	19.4
Confidence	6.4	12.0	12.4	14.0	13.6	14.6	13.7	12.3

Table 5: Summary statistics for responses on undercount, second round of Delphi survey.

When we compare standard deviations for upper and lower bounds between the two rounds, we can notice a striking difference: expert answers became more homogeneous. With only one exception (Question nr. 7; emigration of non-nationals from high undercount countries), the second round standard deviations are smaller than those in the first round, in particular for the upper bounds of the ranges. In some cases (Questions 1 – 3; emigration and immigration of nationals from and to countries with low undercount, as well as emigration of non-nationals from these countries), they are only half as large, or even less. Since the average values of the bounds have not changed much between the two rounds, we can conclude that experts who specified an extreme value for the a certain bound (as compared to the other experts) in the first round, have given answers closer to the mean value in the second round. In other words, we observe a certain regression towards the mean. Indeed, as one expert commented "...modifying the immigration answers (...) in agreement with some of the experts in the first round results".

Some experts commented our categories of countries with high and with low undercount: " ... I am not sure that the UK should fall into this group (of low undercount) ..."; " ... I found ... the grouping of the countries to be rather problematic ..."; also, one expert found it problematic to include Slovenia in the group of high undercount countries. We agree that attempts to improve the quality of statistics on international migration during the past ten years may have made the IMEM-classification a bit outdated.

Table 6 summarizes responses on accuracy. For immigration to register countries, ranges are narrower and confidence is a few percentage points higher compared to immigration to survey countries. When the interest is in nationals versus non-nationals, both ranges and confidence are very similar. Compared to the previous round, standard deviations for answers on accuracy are substantially smaller in this round, similar to what we found for undercount. Again, the experts' opinions have become more similar.

	Register countries				Survey countries	
	Nationals		Non-nationals		Immigration of nationals	Immigration of non-nationals
	Emigration	Immigration	Emigration	Immigration		
	Average (%)					
Lower bound	50.7	65.0	53.9	66.1	43.2	42.9
Upper bound	78.2	87.4	82.9	88.6	71.4	72.1
Confidence	72.5	76.8	71.4	76.4	67.9	67.9
	Standard deviation (% p.)					
Lower bound	24.1	26.5	18.3	20.5	15.5	11.5
Upper bound	20.2	15.3	11.9	8.1	17.6	12.8
Confidence	15.8	15.9	15.1	15.5	11.3	11.3

Table 6: Summary statistics for responses on accuracy, second round of Delphi survey.

The results of this second round formed the input to the probability distributions documented in Part II.

7 Summary and conclusions

This report documents the work done in connection with Task 6.1 of the QuantMig project. The task was to answer the following research question: how well do publicly available statistics on international migration for European countries reflect the real migration flows? We attempted to answer that question by eliciting meta-information about systems for migration data collection in Europe, such as measurement accuracy and undercount, etc., from experts on European migration data. This task is part of Work Package 6 (WP6) of the project. Other tasks for WP6 involve assessing the quality of information about migration flows, migrant stocks and a range of covariates, constructing a model for estimating migration flows, and dissemination of the results.

Part I describes the elicitation of expert opinions, whereas Part II explains how expert opinions were translated into probability distributions.

During the period July – November 2020, we organized a two round Delphi survey among 15 European migration experts. The questionnaire (see Appendix B) was tested by eight experts, selected from the QuantMig research team and its Advisory Board. The survey focused on two important aspects of migration flows between European countries, namely undercount and accuracy. By undercount, we mean measurement errors caused by the fact that some migrants do not report their moves to the relevant authorities, although they should do so. The consequence is that reported flows are smaller than true flows. By true flows, we mean the number of migration moves between two countries that correspond to the United Nations' definition of long-term migration. This definition states that a long-term migrant is a person who moves to a country other than that of his or her usual residence for a period of at least 12 months. By accuracy, we mean that some errors arise in reported flows purely because of randomness, for instance administrative and clerical errors, or sampling errors (in case measurement of international migration is based on a survey).

We asked the experts to give their opinions on undercount and accuracy for 32 European countries (28 EU countries including the United Kingdom, plus four EFTA countries: Iceland, Norway, Liechtenstein, and Switzerland). Together, we refer to them as EU+ countries. We asked the experts to answer the questions as related to the situation, on average, during the period 2009 - 2019 (before Brexit) in the EU+ countries. The experts had to give their opinions in terms of a range and a level of confidence (or certainty). The range reflected the amount of undercount (as a fraction of the true count) across the 32 countries, in percentages between 0 (no undercount) and 100 per cent (none of the moves are reported by the migrants). In addition, we asked the experts how certain they were that the range they specified reflected the true range. There were separate questions for emigration and immigration, for nationals (i.e. nationality of one of the EU+ countries) and non-nationals, and for countries with high or low undercount as specified by us. For each question, the ranges and levels of confidence were translated to an aggregate probability density function. This density function is used as one (of several) inputs into the model mentioned above that estimates (true) migration flows in Europe.

Our work on accuracy and undercount builds on the experience and lessons learned from the project "Integrated Modelling of European Migration" (IMEM) (Raymer et al. (2013); Wiśniowski et al. (2016)). We have extended that work by distinguishing between national and non-nationals, and updated it, by eliciting information for the period 2009 – 2019.

Given the special situation in 2020 with many countries struck by the Covid19-pandemic, we took this opportunity and asked the experts to give a qualitative assessment of the effects the pandemic might have on European migration. We restricted this topic to the pilot round and the first Delphi round.

Sections 5 (for the first Delphi round) and 6 (second Delphi round) report our findings, in the form of average ranges and average confidence levels. Probability distributions derived from expert responses are in Part II. The results from the first round showed that in a few cases, an expert had specified a range for a certain question outside the interval between zero and one hundred per cent. In addition, some of the confidence levels were not consistent with the range the expert had specified. Extensive feedback to the experts concerned between the two rounds and improved explanations in the preamble of the second round questionnaire resolved these issues. The results we summarize here only relate to the second round.

We asked questions about undercount and about accuracy. Average ranges for undercount in the second Delphi round are lowest for immigration to low undercount countries: 7 – 23 per cent, where experts were 73 per cent (for immigration of non-nationals) to 79 per cent (immigration of nationals) certain, on average. The strongest undercount, according to the average expert, occurs for emigration of non-nationals from high undercount countries: 28 – 63 per cent, with a certainty of 69 per cent. In general, average undercount is less for low undercount countries than for high undercount countries. At the same time, the ranges are wider for high undercount countries than for low undercount countries, and the ranges tend to be wider for emigration than for immigration. The average expert is more certain about ranges for countries with low than high undercount. Moreover, (s)he is more certain about immigration ranges than about emigration ranges (but the differences are only a few percentage points). All these results are as one could expect.

When comparing the first and the second round results, we noted that experts had become somewhat more confident that the ranges they specified would include the real undercount value, in particular for countries with high undercount. We analysed standard deviations across the 15 experts for the ranges' upper and lower bounds, and noticed a certain regression to the mean between the two Delphi rounds. In other words, expert answers had become more homogeneous. With only one exception (emigration of non-nationals from high undercount countries), the second round standard deviations were smaller than those in the first round, in particular for the upper bounds of the ranges. In some cases (emigration and immigration of nationals from and to countries with low undercount, as well as emigration of non-nationals from these countries), they were only half as large, or even less. Since the average values of the bounds had not changed much between the two rounds, we conclude that experts who specified an extreme value for a certain bound (as compared to the other experts) in the first round, have given answers closer to the mean value in the second round.

The main results in the second Delphi round for the accuracy of migration measurement systems are as follows. For immigration to register countries, ranges were narrower and confidence was a few percentage points higher compared to immigration to survey countries. When the interest is in nationals versus non-nationals, both ranges and confidence were very similar. Compared to the first Delphi round, standard deviations for answers on accuracy were substantially smaller in the second round, similar to what we found for undercount. Again, the experts' opinions had become more similar.

An important lesson learned from the elicitation process is that such processes may significantly benefit if they are conducted using visual and interactive tools. The great advantage of these tools is that, being visual, they allow the expert to see their density as they provide an answer, and, being interactive, they allow the expert to see how this density changes as their answer changes. As a result, such tools can avoid the possible confusion relating to the way answers translate to probability statements and densities.

Finally, we report the qualitative assessments of 16 experts concerning the impact of Covid-19 on European migration flows [7]. Only one expert believes that Covid-19 will lead to (slightly) larger European migration flows to 2026, compared to migration in an average year in the recent past. The majority expects much smaller flows, although in the long term, i.e. for the years 2021 – 20126, about one-third foresee very much smaller flows. As one expert commented: "The COVID-19 will have depressed international migration hugely in 2020. Assuming no effective vaccine, this will continue over the medium term. Even with a vaccine, the vaccine might only confer immunity for a limited period. So migration will continue to be depressed."

[7] Initially, 16 experts had agreed to participate in the Delphi survey. One expert answered questions on Covid-19 and future migration flows only, but was unable to respond to undercount and accuracy questions. Therefore, we have answers for the entire survey from 15 experts, and from one more on Covid-19 questions.

References

Bijak, J. and Bryant, J. (2016). Bayesian demography 250 years after bayes. *Population Studies*, 70(1):1–19.

Raymer, J., Wiśniowski, A., Forster, J. J., Smith, P. W. F., and Bijak, J. (2013). Integrated modeling of european migration. *Journal of the American Statistical Association*, 108(503):801–819.

Rowe, G. and Wright, G. (2001). Expert opinions in forecasting: The role of the delphi technique. In *International Series in Operations Research & Management Science*, pages 125–144. Springer US.

U.N. (1998). *Recommendations on statistics of international migration, revision 1*. Number 58 in Statistical Papers Series M. Department of Economic and Social Affairs, Statistical Division, United Nations, New York.

Wiśniowski, A., Bijak, J., Christiansen, S., Forster, J. J., Keilman, N., Raymer, J., and Smith, P. W. (2013). Utilising expert opinion to improve the measurement of international migration in europe. *Journal of Official Statistics*, 29(4):583–607.

Wiśniowski, A., Forster, J. J., Smith, P. W. F., Bijak, J., and Raymer, J. (2016). Integrated modelling of age and sex patterns of european migration. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179(4):1007–1024.

A Questionnaire pilot round

QuantMig pilot survey copy

Page 1

QuantMig

Please fill in your name (first name and surname). *

Note: questions marked by an asterisk (*) require an answer.

About QuantMig

The aim of QuantMig is to produce comprehensive, multi-perspective and robust quantitative migration scenarios to support various areas of European migration policy. A consortium of seven academic institutions take part in the project, which is headed by Jakub Bijak of the University of Southampton, Department of Social Statistics and Demography. The project receives financial support from the European Union, through its Horizon 2020 programme. See http://quantmig.eu.

Page break

Page 2

The survey

The QuantMig-scenarios will be based on a set of statistical estimates of international migration flows and their drivers. These estimates will result from a statistical model for migration count data.

The statistical model uses, as part of its input, information about migration measurement systems, elicited, through a survey, from the migration experts.

This is the pilot version of the survey.

We kindly ask you to submit the completed form **not later than Wednesday 10 June.**

IMPORTANT

Your answers will be stored only once, namely, at the time you submit the completed questionnaire. Until then, all your answers are in your browser. This means that you should not close your browser as long as you are working on the survey. Furthermore, the program that directs the questionnaire will automatically close 4 hours (240 minutes) after you have typed your last text. As long as you continue answering the survey within the window of 4 hours, this will not occur.

Page break

Page 3

Ethics

Participation is voluntary and anonymous. Only the survey administrator (Nico Keilman), and no one else, will have the possibility to link your name to your answers.

The survey is administered at the Department of Economics of the University of Oslo. It complies with ethical aspects of research as regulated in the Norwegian Research Ethics Law of 2017 – "Lov om organisering av forskningsetisk arbeid (forskningsetikkloven)"; see https://lovdata.no/dokument/NL/lov/2017-04-28-23 .

*

○ I agree to these conditions.

Page break

Page 4

We plan to acknowledge, in project reports and the like, the help of all migration experts who provided their opinions about migration measurement. Below we offer you the option to have your name included in the list, or not.

*

○ I agree to have my name included in report(s) about the results of the survey

○ I prefer that you do not include my name

Page break

Page 5

32 European countries

The questions relate to migration measurement in 32 European countries: EU-27 plus EFTA countries (Iceland, Norway, Liechtenstein, Switzerland) plus the United Kingdom. Collectively, we refer to them as EU+.

Page break

Page 6

Undercount and accuracy

This questionnaire has three main sections, with questions about undercount in two groups of countries (Sections 1 and 2), and accuracy of migration measurement (Section 3). These questions relate to the meaurement of migration.

Undercount is to be understood as the extent to which the true migration flow is captured by the data collection system in a given country. It is different from "coverage" (not addressed in this survey), which refers to the fact that some countries do not include all types of migrants in their measurements, for instance students or nationals.

Accuracy relates to random errors in migration measurement, for instance administrative errors in the processing of the data.It is different from "coverage", which refers to systematic errors ("bias").

Finally, there are two additional questions about the possible impact of Covid-19 on European migration flows

Note that the questions below relate to different types of uncertainty. When we ask you to specify how certain you are about a given degree of undercount, this refers to epistemic uncertainty, caused by a lack of knowledge. The real degree of undercount can be known in principle, if only we would gather enough data.

On the other hand, aleatory uncertainty plays a role when we ask your opinion about accuracy. Here, random factors are responsible for the uncertainty.

Page break

Page 7

Pre-Brexit

These and other characteristics regarding the measurement of international migration, have varied over time. Your answers to the questions below should relate to the situation, on average, during the period 2009 - 2019 (before Brexit) in the EU+ countries.

This means, among other things, the extent to which countries have implemented the EU Regulation 862/2007 on Community statistics on migration and international protection, adopted by the European Parliament in 2007.

Page break

Page 8

Answers in terms of a range

It is clear that measurement practices differ widely among European countries. Yet it is not feasible to ask questions for each of the EU+ countries. Instead, we invite you to state your estimates in terms of **ranges**, together with associated **levels of confidence.**

The ranges and confidence levels should reflect your belief about the variability between countries, but also your uncertainty in the answers.

To give an example, assume that you state that you are about 70 per cent certain that undercount was between 20 and 60 per cent in a given situation. We will use these numbers to construct a probability distribution for this type of undercount such that chances are 100 – 70 = 30 per cent that undercount was less than 20 per cent *or* more than 60 per cent.

Page break

Page 9

United Nations definition as benchmark for migration

In many cases, we ask you to give your opinion about how a specific measurement of international migration deviates from a benchmark.

As the benchmark, we have adopted the **United Nations definition**, which corresponds with the definition included in the EU Regulation 862/2007 mentioned earlier.

The definition is given below. On the next page, we give a few clarifying comments. If, at any stage, you would like to have another look at the UN definition, you can scroll back to this page, or scroll down to a copy of this page at the end of the questionnaire.

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.

[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.

Page break

Page 10

Notes:

- The UN definition employs the notion of "country of usual residence". This is to be understood as the country in which the person spends at least 183 nights of the year. In practice, many countries in Europe use a *de jure* definition (especially countries with a population register), and a country's *de jure* definition may very well be different from the 183 nights rule. We assume, however, that such differences are not very important for international migration (as opposed to migration within one country) and that these definitions will generally result in the same country of residence. Hence we ignore possible differences in measured migration flows caused by different definitions of country of usual residence.
- Related to the previous point is the notion of **undocumented migration**. 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 not include undocumented migrants.

Page break

Page 11

Low and high undercount countries

For undercount, we have grouped the 32 countries into two categories: low undercount countries and high undercount countries. The grouping is based on our findings in the IMEM-project (http://www.imem.cpc.ac.uk/About.aspx).

- Low undercount countries: The Netherlands, Sweden, Finland, Norway, Denmark, Germany, Iceland, Austria, Belgium, United Kingdom, Cyprus, Ireland, Italy, France, Luxembourg, Switzerland, and immigration to Spain.
- High undercount countries: Bulgaria, Croatia, Estonia, Lithuania, Latvia, Poland, Slovenia, Slovakia, Romania, the Czech Republic, Greece, Hungary, Liechtenstein, Malta, Portugal, and emigration from Spain.

Page break

Page 12

Section 1. Undercount in migration to and from "low undercount countries"

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.

When we ask you to give undercount as a percentage, this refers to the ratio between not counted and the real (but unknown) flow. An ideal measurement system has an undercount of zero. The larger the percentage you give, the worse is your assessment of the system.

The following questions are restricted to migration to and from countries defined as "low undercount countries" (The Netherlands, Sweden, Finland, Norway, Denmark, Germany, Iceland, Austria, Belgium, United Kingdom, Cyprus, Ireland, Italy, France, Luxembourg, Switzerland, and immigration to Spain).

First, we ask your opinion about immigration and emigration flows for nationals, i.e. persons with nationality in one of the 32 EU+ countries. Next, there are questions for non-nationals.

Page break

Page 13

Question 1. Emigration of nationals

1a) By how many per cent do you expect that emigration flows of nationals who leave low undercount countries are undercounted in the published statistics of those countries, as compared to the true level of emigration of nationals? Please provide a range in percentages between 0 and 100, by writing in the box below "xx to yy %".

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

- 50 %
- 75 %
- 90 %
- 95 %
- other (please state)

*

This element is only shown when the option "other (please state)" is selected in the question "1b) Approximately, how certain are you that the true undercount of nationals will lie within the range that you provided above?"

other

Page break

Page 14

Question 2. Immigration of nationals

2a) By how many per cent do you expect that immigration flows of nationals who enter low undercount countries are undercounted in the published statistics of those countries, as compared to the true level of immigration of nationals? Please provide a range in percentages between 0 and 100.

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

- 50 %
- 75 %
- 90 %
- 95 %
- other (please state)

*

This element is only shown when the option "other (please state)" is selected in the question "2b) Approximately, how certain are you that the true undercount of nationals will lie within the range that you provided above?"

other

Page break

Page 15

Question 3. Emigration of non-nationals

3a) By how many per cent do you expect that emigration flows of non-nationals who leave low undercount countries are undercounted in the published statistics of those countries, as compared to the true level of emigration of non-nationals? Please provide a range in percentages between 0 and 100.

3b) Approximately, how certain are you that the true undercount of non-nationals will lie within the range that you provided above?

- 50 %
- 75 %
- 90 %
- 95 %
- other (please state)

*

This element is only shown when the option "other (please state)" is selected in the question "3b) Approximately, how certain are you that the true undercount of non-nationals will lie within the range that you provided above?"

other

Page break

Page 16

Question 4. Immigration of non-nationals

4a) By how many per cent do you expect that immigration flows of non-nationals who enter low undercount countries are undercounted in the published statistics of those countries, as compared to the true level of immigration of non-nationals? Please provide a range in percentages between 0 and 100.

4b) Approximately, how certain are you that the true undercount of non-nationals will lie within the range that you provided above?

- 50 %
- 75 %
- 90 %
- 95 %
- other (please state)

*

This element is only shown when the option "other (please state)" is selected in the question "4b) Approximately, how certain are you that the true undercount of non-nationals will lie within the range that you provided above?"

other

Page break

Page 17

Comments

If you have comments or arguments related to your answers to questions 1 - 4, please state them here.

Page break

Page 18

Section 2. Undercount in migration to and from "high undercount countries"

In this section, we ask you to answer the same questions as before, but now the questions refer to the group of countries that we defined as "high undercount countries": Bulgaria, Croatia, Estonia, Lithuania, Latvia, Poland, Slovenia, Slovakia, Romania, the Czech Republic, Greece, Hungary, Liechtenstein, Malta, Portugal, and emigration from Spain.

First, we ask your opinion about immigration and emigration flows for nationals, i.e. persons with nationality in one of the 32 EU+ countries. Next, there are questions for non-nationals.

Question 5. Emigration of nationals

5a) By how many per cent do you expect that emigration flows of nationals who leave high undercount countries are undercounted in the published statistics of those countries, as compared to the true level of emigration of nationals? Please provide a range in percentages between 0 and 100, by writing in the box below "xx to yy %".

5b) Approximately, how certain are you that the true undercount of nationals will lie within the range that you provided above?

- 50 %
- 75 %
- 90 %
- 95 %
- other (please state)

*

This element is only shown when the option "other (please state)" is selected in the question "5b) Approximately, how certain are you that the true undercount of nationals will lie within the range that you provided above?"

other

Page break

Page 19

Question 6. Immigration of nationals

6a) By how many per cent do you expect that immigration flows of nationals who enter high undercount countries are undercounted in the published statistics of those countries, as compared to the true level of immigration of nationals? Please provide a range in percentages between 0 and 100.

6b) Approximately, how certain are you that the true undercount of nationals will lie within the range that you provided above?

○ 50 % ○ 75 % ○ 90 %

○ 95 % ○ other (please state)

*

This element is only shown when the option "other (please state)" is selected in the question "6b) Approximately, how certain are you that the true undercount of nationals will lie within the range that you provided above?"

other

Page break

Page 20

Question 7. Emigration of non-nationals

7a) By how many per cent do you expect that emigration flows of non-nationals who leave high undercount countries are undercounted in the published statistics of those countries, as compared to the true level of emigration of non-nationals? Please provide a range in percentages between 0 and 100.

7b) Approximately, how certain are you that the true undercount of non-nationals will lie within the range that you provided above?

○ 50 % ○ 75 % ○ 90 %

○ 95 % ○ other (please state)

*

This element is only shown when the option "other (please state)" is selected in the question "7b) Approximately, how certain are you that the true undercount of non-nationals will lie within the range that you provided above?"

other

Page break

Page 21

Question 8. Immigration of non-nationals

8a) By how many per cent do you expect that immigration flows of non-nationals who enter high undercount countries are undercounted in the published statistics of those countries, as compared to the true level of immigration of non-nationals? Please provide a range in percentages between 0 and 100.

8b) Approximately, how certain are you that the true undercount of non-nationals will lie within the range that you provided above?

○ 50 % ○ 75 % ○ 90 %

○ 95 % ○ other (please state)

*

This element is only shown when the option "other (please state)" is selected in the question "8b) Approximately, how certain are you that the true undercount of non-nationals will lie within the range that you provided above?"

other

Page break

Page 22

Comments

If you have comments or arguments related to your answers to questions 5 - 8, please state them here.

Page break

Page 23

Section 3. Accuracy of migration measurement

The following questions are about random fluctuations in measured emigration and immigration. We distinguish between nationals and non-nationals, and, moreover, between population registers and migration surveys as data collection systems.

Consider a hypothetical 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.

Questions 9 - 12 all relate to this hypothetical register country.

Question 9. Emigration of nationals

9a) For emigration of nationals, 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 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 give your assessment as a range in percentages between 0 and 100, by writing in the box below "xx to yy %".

9b) Approximately, how certain are you that the true value will lie within the range that you provided above?

○ 50 % ○ 75 % ○ 90 %

○ 95 % ○ other (please state)

*

This element is only shown when the option "other (please state)" is selected in the question "9b) Approximately, how certain are you that the true value will lie within the range that you provided above?"

other

Page break

Page 24

Question 10. Immigration of nationals

10a) For immigration of nationals, 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 level of immigration? Please provide a range in percentages between 0 and 100.

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

○ 50 % ○ 75 % ○ 90 %

○ 95 % ○ other (please state)

*

This element is only shown when the option "other (please state)" is selected in the question "10b) Approximately, how certain are you that the true value will lie within the range that you provided above?"

other

Page break

Page 25

Question 11. Emigration of non-nationals

11a) For emigration of non-nationals, 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 level of emigration? Please provide a range in percentages between 0 and 100.

11b) Approximately, how certain are you that the true value will lie within the range that you provided above?

○ 50 % ○ 75 % ○ 90 %

○ 95 % ○ other (please state)

*

This element is only shown when the option "other (please state)" is selected in the question "11b) Approximately, how certain are you that the true value will lie within the range that you provided above?"

other

Page break

Page 26

Question 12. Immigration of non-nationals

12a) For immigration of non-nationals, 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 level of immigration? Please provide a range in percentages between 0 and 100.

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

- ○ 50 %
- ○ 75 %
- ○ 90 %
- ○ 95 %
- ○ other (please state)

*

This element is only shown when the option "other (please state)" is selected in the question "12b) Approximately, how certain are you that the true value will lie within the range that you provided above?"

other

Page break

Page 27

Comments

If you have comments or arguments related to your answers to questions 9 - 12, please state them here.

Page break

Page 28

Now, consider a hypothetical European country that uses a survey to collect migration data, and assume that there is no systematic bias in the measurement of migration. In this case, we may expect the accuracy to be affected by, for example, sampling error. Here, we will only consider immigration.

Questions 13 and 14 relate to this hypothetical country that uses a survey.

Question 13. Immigration of nationals

13a) For immigration of nationals, 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 level of immigration, as defined by the UN definition of migration? Please provide a range in percentages between 0 and 100.

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

- ○ 50 %
- ○ 75 %
- ○ 90 %
- ○ 95 %
- ○ other (please state)

*

This element is only shown when the option "other (please state)" is selected in the question "13b) Approximately, how certain are you that the true value will lie within the range that you provided above?"

other

Page break

Page 29

Question 14. Immigration of non-nationals

14a) For immigration of non-nationals, 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 level of immigration, as defined by the UN definition of migration? Please provide a range in percentages between 0 and 100.

14b) Approximately, how certain are you that the true value will lie within the range that you provided above?

- ○ 50 %
- ○ 75 %
- ○ 90 %
- ○ 95 %
- ○ other (please state)

*

This element is only shown when the option "other (please state)" is selected in the question "14b) Approximately, how certain are you that the true value will lie within the range that you provided above?"

other

Page break

Page 30

Comments

If you have comments or arguments related to your answers to questions 13 and 14, please state them here.

Page break

Page 31

Two final questions: Covid-19 and European migration

The following two questions are related to the possible impact of Covid-19 on European migration flows. The kind of information that we hope to receive from you will be useful when developing scenarios and story lines on possible migration flows in the future, which is also an important task in the QuantMig project.

We ask you to give a qualitative assessment for the impact of Covid-19 on European migration flows, compared to the flows for an average pre-Covid-19 year. Question 15 is about the short-term effect (the year 2021), whereas question 16 is related to the average effect for the next five years (2021-2026).

Please give your answers on a nine-points scale, ranging from "very much lower" to "no impact" to "very much higher".

We distinguish between flows within the EU+ countries, immigration flows into EU+ countries, and emigration flows from the EU+ countries.

Question 15. Short-term effect of Covid-19

What is your qualitative assessment of the impact of Covid-19 on migration flows for EU+ countries in 2021, compared to the flows for an average pre-Covid-19 year? Because of Covid-19, ...

	very much lower	much lower	some-what lower	slightly lower	no change	slightly higher	some-what higher	much higher	very much higher	don't know
... migration flows within EU+ countries will be	○	○	○	○	○	○	○	○	○	○
... immigration flows into EU+ countries will be	○	○	○	○	○	○	○	○	○	○
... emigration flows from EU+ countries will be	○	○	○	○	○	○	○	○	○	○

Question 16. Long-term effect of Covid-19

What is your qualitative assessment of the impact of Covid-19 on migration flows for EU+ countries on average during the years 2021-2026, compared to the flows for an average pre-Covid-19 year? Because of Covid-19, ...

	very much lower	much lower	some-what lower	slightly lower	no change	slightly higher	some-what higher	much higher	very much higher	don't know
... migration flows within EU+ countries will be	○	○	○	○	○	○	○	○	○	○
... immigration flows into EU+ countries will be	○	○	○	○	○	○	○	○	○	○
... emigration flows from EU+ countries will be	○	○	○	○	○	○	○	○	○	○

Page break

Page 32

If you have comments or arguments related to your answers to questions 15 and 16, please state them here.

Page break

Page 33

Final comments

You may submit your answers from a browser on a computer, mobile phone, tablet, etc. From which device did you answer the survey?

- ☐ computer
- ☐ mobile phone
- ☐ tablet
- ☐ other - fill in below

other, namely

This element is only shown when the option "other - fill in below" is selected in the question "You may submit your answers from a browser on a computer, mobile phone, tablet, etc. From which device did you answer the survey?"

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.

How many minutes, approximately, did it take you to answer the questions?

Page break

Page 34

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.

[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.

B Questionnaire second Delphi round

QuantMig Delphi 2nd round copy

Page 1

Please fill in your name (first name and surname). *

Note: questions marked by an asterisk (*) require an answer.

Second round of QuantMig Delphi survey

As you may recall, the aim of QuantMig is to produce comprehensive, multi-perspective and robust quantitative migration scenarios to support various areas of European migration policy. The QuantMig-scenarios will be based on a set of statistical estimates of international migration flows and their drivers. These estimates will result from a statistical model for migration count data.

The statistical model uses, as part of its input, information about migration measurement systems, elicited, through a two-round Delphi survey, from the migration experts.

This is the second round of the Delphi survey.

This introduction repeats, to a large extent, the introductory comments of the first round.

IMPORTANT

Your answers will be stored only once, namely, at the time you submit the completed questionnaire. Until then, all your answers are in your browser. This means that you should not close your browser as long as you are working on the survey. Furthermore, the program that directs the questionnaire will automatically close 4 hours (240 minutes) after you have typed your last text. As long as you continue answering the survey within the window of 4 hours, this will not occur.

32 European countries

The questions relate to migration measurement in 32 European countries: EU-27 plus EFTA countries (Iceland, Norway, Liechtenstein, Switzerland) plus the United Kingdom. Collectively, we refer to them as EU+.

Undercount and accuracy

This questionnaire has three main sections, with the same questions as in the first round. The questions refer to undercount in two groups of countries (Sections 1 and 2), and accuracy of migration measurement (Section 3).

Undercount is to be understood as the extent to which the true migration flow is captured by the data collection system in a given country. It is different from "coverage" (not addressed in this survey), which refers to the fact that some countries do not include all types of migrants in their measurements, for instance students or nationals.

Accuracy relates to random errors in migration measurement, for instance administrative errors in the processing of the data.It is different from "coverage", which refers to systematic errors ("bias").

Pre-Brexit

Undercount, accuracy and other characteristics regarding the measurement of international migration, have varied over time. Your answers to the questions below should relate to the situation, on average, during the period 2009 - 2019 (before Brexit) in the EU+ countries.

This means, among other things, the extent to which countries have implemented the EU Regulation 862/2007 on Community statistics on migration and international protection, adopted by the European Parliament in 2007.

Answers in terms of a range

It is clear that measurement practices differ widely among European countries. Yet it is not feasible to ask questions for each of the EU+ countries. Instead, we invite you to state your estimates in terms of **ranges**, together with associated **levels of confidence.**

The ranges and confidence levels should reflect your belief about the variability between countries, but also your uncertainty in the answers.

To give an example, assume that you state that you are about 75 per cent certain that undercount was between 20 and 60 per cent in a given situation. We will use these numbers to construct a probability distribution for this type of undercount such that chances are 100 – 75 = 25 per cent that undercount was less than 20 per cent *or* more than 60 per cent.

Further explanations on the probability distributions have been provided in the report on the first round results that you received by separate email.

Page break

Page 2

United Nations definition as benchmark for migration

In many cases, we ask you to give your opinion about how a specific measurement of international migration deviates from a benchmark.

As the benchmark, we have adopted the **United Nations definition**, which corresponds with the definition included in the EU Regulation 862/2007 mentioned earlier.

The definition is given below.

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.

[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.

Notes

- **Undocumented migration**

In theory, the UN definition includes undocumented ("illegal") migrants. In practice, the migration statistics in most countries do not cover undocumented migrants. This is also in line with the EU Regulation 862/2007.

Therefore, when we refer to the UN definition as the benchmark, we do not include undocumented migrants.

- **Asylum seekers**

Asylum seekers are not included in the immigration flow for the year in which they arrive in the country

where they apply for asylum, because the 12 months criterion cannot be applied. Asylum seekers who have their application granted in a particular year are part of the immigration flow for that year. Asylum seekers whose application is not granted and hence in principle have to leave the country, are not included in the emigration flow for the year in question, because they could not establish usual residence in the country where they applied for asylum.

Therefore, when we refer to the UN definition as the benchmark, we only include those asylum seekers who have been granted asylum and in the year it was granted.

Low and high undercount countries

For undercount, we have grouped the 32 countries into two categories: low undercount countries and high undercount countries. The grouping is based on our findings in the IMEM-project (http://www.cpc.ac.uk/projects/13/#overview).

- Low undercount countries: The Netherlands, Sweden, Finland, Norway, Denmark, Germany, Iceland, Austria, Belgium, United Kingdom, Cyprus, Ireland, Italy, France, Luxembourg, Switzerland, and immigration to Spain.
- High undercount countries: Bulgaria, Croatia, Estonia, Lithuania, Latvia, Poland, Slovenia, Slovakia, Romania, the Czech Republic, Greece, Hungary, Liechtenstein, Malta, Portugal, and emigration from Spain.

Feedback from first round

Each question in this second round is accompanied by a table with the answers from all respondents of the first round, including your own answers. This will allow you to consider whether you wish to revise your answers from the first round, or not.

There is one column for each respondent. The rows give, for each range, its lower bound (r1) and its upper bound (r2), as well as the level of confidence (indicated by the letter c).

The respondents are anonymous - only their identification number is given. You find your own identification number (id) in the recent email and in the report on first round results that I attached to that email.

Revising your answer is particularly important in case the report explicitly mentioned, under the heading "Comments and clarifications", that some of your ranges were outside the interval from 0 to 100 per cent, or that they were inconsistent otherwise.

Page break

Page 3

Section 1. Undercount in migration to and from "low undercount countries"

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.

When we ask you to give undercount as a percentage, this refers to the ratio between not counted and the real (but unknown) flow. An ideal measurement system has an undercount of zero. The larger the percentage you give, the worse is your assessment of the system. An undercount of 100 % applies to the extreme situation in which none of the migrants are measured by the system.

As stated before, we disregard undocumented migrants.

The following questions are restricted to migration to and from countries defined as "low undercount countries" (The Netherlands, Sweden, Finland, Norway, Denmark, Germany, Iceland, Austria, Belgium, United Kingdom, Cyprus, Ireland, Italy, France, Luxembourg, Switzerland, and immigration to Spain).

First, we ask your opinion about immigration and emigration flows for nationals, i.e. persons with nationality in one of the 32 EU+ countries. Next, there are questions for non-nationals.

Question 1. Emigration of nationals

1a) By how many per cent do you expect that emigration flows of nationals who leave low undercount countries are undercounted in the published statistics of those countries, as compared to the true level of emigration of nationals? Please provide a range in percentages between 0 and 100. In other words, the lower end of the range (r1) is larger than 0, whereas the upper end of the range (r2) is smaller than 100 - but larger than r1.

r1 =

r2 =

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

- 50 %
- 75 %
- 90 %
- 95 %
- other (pease state)

*

This element is only shown when the option "other (pease state)" is selected in the question "1b) Approximately, how certain are you that the true undercount of nationals will lie within the range that you provided above?"

other

Answers from round 1 for this question

id	8526257	8533066	8546148	8553207	8593119	8628621	8673705	8675156
r1	0 %	0 %	10 %	5 %	10 %	20 %	10 %	25 %
r2	20 %	50 %	20 %	35 %	30 %	40 %	20 %	35 %
c	75 %	95 %	75 %	75 %	75 %	50 %	75 %	75 %
id	8681334	8696877	8698827	8700689	8707108	8708183	8709921	
r1	30 %	7 %	20 %	5 %	10 %	1 %	10 %	
r2	90 %	10 %	40 %	20 %	30 %	59 %	30 %	
c	50 %	90 %	75 %	75 %	75 %	75 %	50 %	

Page break

Question 2. Immigration of nationals

2a) By how many per cent do you expect that immigration flows of nationals who enter low undercount countries are undercounted in the published statistics of those countries, as compared to the true level of immigration of nationals? Please provide a range in percentages between 0 and 100 as before, with a lower end of the range (r1) and an upper end of the range (r2).

r1 =

r2 =

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

- 50 %
- 75 %
- 90 %
- 95 %
- other (pease state)

*

This element is only shown when the option "other (pease state)" is selected in the question "2b) Approximately, how certain are you that the true undercount of nationals will lie within the range that you provided above?"

other

Answers from round 1 for this question

id	8526257	8533066	8546148	8553207	8593119	8628621	8673705	8675156
r1	0 %	0 %	5 %	0 %	5 %	20 %	20 %	3 %
r2	5 %	25 %	10 %	35 %	10 %	30 %	30 %	5 %
c	50 %	90 %	95 %	95 %	90 %	50 %	50 %	75 %
id	8681334	8696877	8698827	8700689	8707108	8708183	8709921	
r1	20 %	0 %	10 %	5 %	5 %	7 %	10 %	
r2	50 %	5 %	30 %	20 %	10 %	79 %	20 %	
c	50 %	90 %	75 %	90 %	75 %	75 %	75 %	

Question 3. Emigration of non-nationals

3a) By how many per cent do you expect that emigration flows of non-nationals who leave low undercount countries are undercounted in the published statistics of those countries, as compared to the true level of emigration of non-nationals? Please provide a range in percentages between 0 and 100 as before, with a lower end of the range (r1) and an upper end of the range (r2).

r1 =

r2 =

3b) Approximately, how certain are you that the true undercount of non-nationals will lie within the range that you provided above?

- 50 %
- 75 %
- 90 %
- 95 %
- other (pease state)

*

This element is only shown when the option "other (pease state)" is selected in the question "3b) Approximately, how certain are you that the true undercount of non-nationals will lie within the range that you provided above?"

other

Answers from round 1 for this question

id	8526257	8533066	8546148	8553207	8593119	8628621	8673705	8675156
r1	10 %	0 %	20 %	10 %	20 %	20 %	40 %	30 %
r2	30 %	50 %	30 %	80 %	40 %	40 %	80 %	40 %
c	50 %	90 %	75 %	95 %	50 %	50 %	50 %	75 %
id	8681334	8696877	8698827	8700689	8707108	8708183	8709921	
r1	30 %	8 %	10 %	5 %	10 %	0 %	30 %	
r2	90 %	15 %	30 %	20 %	20 %	36 %	60 %	
c	50 %	90 %	75 %	95 %	75 %	75 %	50 %	

Page break

Question 4. Immigration of non-nationals

4a) By how many per cent do you expect that immigration flows of non-nationals who enter low undercount countries are undercounted in the published statistics of those countries, as compared to the true level of immigration of non-nationals? Please provide a range in percentages between 0 and 100 as before, with a lower end of the range (r1) and an upper end of the range (r2).

r1 =

r2 =

4b) Approximately, how certain are you that the true undercount of non-nationals will lie within the range that you provided above?

- 50 %
- 75 %
- 90 %
- 95 %
- other (pease state)

*

This element is only shown when the option "other (pease state)" is selected in the question "4b) Approximately, how certain are you that the true undercount of non-nationals will lie within the range that you provided above?"

other

Answers from round 1 for this question

id	8526257	8533066	8546148	8553207	8593119	8628621	8673705	8675156
r1	5 %	0 %	10 %	10 %	10 %	10 %	20 %	5 %
r2	15 %	10 %	20 %	40 %	30 %	20 %	40 %	10 %
c	50 %	90 %	90 %	90 %	75 %	75 %	50 %	75 %
id	8681334	8696877	8698827	8700689	8707108	8708183	8709921	
r1	20 %	0 %	5 %	5 %	5 %	1 %	20 %	
r2	60 %	7 %	20 %	10 %	10 %	24 %	40 %	
c	50 %	90 %	75 %	90 %	75 %	75 %	50 %	

Comments

If you have comments or arguments related to your answers to questions 1 - 4, please state them here.

Page break

Section 2. Undercount in migration to and from "high undercount countries"

In this section, we ask you to answer the same questions as before, but now the questions refer to the group of countries that we defined as "high undercount countries": Bulgaria, Croatia, Estonia, Lithuania, Latvia, Poland, Slovenia, Slovakia, Romania, the Czech Republic, Greece, Hungary, Liechtenstein, Malta, Portugal, and emigration from Spain.

First, we ask your opinion about immigration and emigration flows for nationals, i.e. persons with nationality in one of the 32 EU+ countries. Next, there are questions for non-nationals.

Question 5. Emigration of nationals

5a) By how many per cent do you expect that emigration flows of nationals who leave high undercount countries are undercounted in the published statistics of those countries, as compared to the true level of emigration of nationals? Please provide a range in percentages between 0 and 100 as before, with a lower end of the range (r1) and an upper end of the range (r2).

r1 =

r2 =

5b) Approximately, how certain are you that the true undercount of nationals will lie within the range that you provided above?

- ○ 50 %
- ○ 75 %
- ○ 90 %
- ○ 95 %
- ○ other (pease state)

*

ℹ This element is only shown when the option "other (pease state)" is selected in the question "5b) Approximately, how certain are you that the true undercount of nationals will lie within the range that you provided above?"

other

Answers from round 1 for this question

id	8526257	8533066	8546148	8553207	8593119	8628621	8673705	8675156
r1	20 %	0 %	20 %	10 %	50 %	20 %	50 %	30 %
r2	40 %	100 %	40 %	45 %	80 %	40 %	70 %	50 %
c	50 %	50 %	75 %	50 %	75 %	50 %	50 %	50 %
id	8681334	8696877	8698827	8700689	8707108	8708183	8709921	
r1	50 %	10 %	20 %	20 %	20 %	3 %	30 %	
r2	100 %	20 %	80 %	75 %	40 %	87 %	60 %	
c	50 %	75 %	50 %	90 %	75 %	75 %	50 %	

Page break

Page 8

Question 6. Immigration of nationals

6a) By how many per cent do you expect that immigration flows of nationals who enter high under-count countries are undercounted in the published statistics of those countries, as compared to the true level of immigration of nationals? Please provide a range in percentages between 0 and 100 as before, with a lower end of the range (r1) and an upper end of the range (r2).

r1 =

r2 =

6b) Approximately, how certain are you that the true undercount of nationals will lie within the range that you provided above?

- ○ 50 %
- ○ 75 %
- ○ 90 %
- ○ 95 %
- ○ other (pease state)

*

ℹ This element is only shown when the option "other (pease state)" is selected in the question "6b) Approximately, how certain are you that the true undercount of nationals will lie within the range that you provided above?"

other

Answers from round 1 for this question

id	8526257	8533066	8546148	8553207	8593119	8628621	8673705	8675156
r1	5 %	0 %	30 %	5 %	50 %	30 %	50 %	10 %
r2	10 %	100 %	50 %	45 %	80 %	40 %	70 %	20 %
c	50 %	50 %	75 %	50 %	75 %	50 %	50 %	50 %
id	8681334	8696877	8698827	8700689	8707108	8708183	8709921	
r1	30 %	0 %	20 %	20 %	10 %	7 %	20 %	
r2	90 %	15 %	80 %	75 %	30 %	82 %	40 %	
c	50 %	90 %	75 %	90 %	75 %	75 %	50 %	

Page break

Page 9

Question 7. Emigration of non-nationals

7a) By how many per cent do you expect that emigration flows of non-nationals who leave high under-count countries are undercounted in the published statistics of those countries, as compared to the true level of emigration of non-nationals? Please provide a range in percentages between 0 and 100 as before, with a lower end of the range (r1) and an upper end of the range (r2).

r1 =

r2 =

7b) Approximately, how certain are you that the true undercount of non-nationals will lie within the range that you provided above?

- ○ 50 %
- ○ 75 %
- ○ 90 %
- ○ 95 %
- ○ other (pease state)

*

ℹ This element is only shown when the option "other (pease state)" is selected in the question "7b) Approximately, how certain are you that the true undercount of non-nationals will lie within the range that you provided above?"

other

Answers from round 1 for this question

id	8526257	8533066	8546148	8553207	8593119	8628621	8673705	8675156
r1	40 %	0 %	10 %	50 %	60 %	40 %	50 %	30 %
r2	60 %	100 %	20 %	90 %	90 %	60 %	80 %	50 %
c	50 %	75 %	75 %	75 %	50 %	50 %	50 %	50 %
id	8681334	8696877	8698827	8700689	8707108	8708183	8709921	
r1	30 %	10 %	10 %	10 %	25 %	3 %	30 %	
r2	90 %	25 %	50 %	25 %	35 %	50 %	60 %	
c	50 %	75 %	50 %	90 %	75 %	75 %	50 %	

Page break

Page 10

Question 8. Immigration of non-nationals

8a) By how many per cent do you expect that immigration flows of non-nationals who enter high under-count countries are undercounted in the published statistics of those countries, as compared to the true level of immigration of non-nationals? Please provide a range in percentages between 0 and 100 as before, with a lower end of the range (r1) and an upper end of the range (r2).

r1 =

r2 =

8b) Approximately, how certain are you that the true undercount of non-nationals will lie within the range that you provided above?

- ○ 50 %
- ○ 75 %
- ○ 90 %
- ○ 95 %
- ○ other (pease state)

*

ℹ This element is only shown when the option "other (pease state)" is selected in the question "8b) Approximately, how certain are you that the true undercount of non-nationals will lie within the range that you provided above?"

other

Answers from round 1 for this question

id	8526257	8533066	8546148	8553207	8593119	8628621	8673705	8675156
r1	10 %	0 %	10 %	40 %	20 %	20 %	40 %	10 %
r2	20 %	100 %	20 %	60 %	40 %	30 %	60 %	20 %
c	50 %	75 %	90 %	75 %	75 %	50 %	50 %	50 %
id	8681334	8696877	8698827	8700689	8707108	8708183	8709921	
r1	30 %	0 %	10 %	10 %	10 %	3 %	10 %	
r2	90 %	15 %	50 %	50 %	20 %	44 %	20 %	
c	50 %	75 %	50 %	90 %	75 %	75 %	50 %	

Comments

If you have comments or arguments related to your answers to questions 5 - 8, please state them here.

Page break

Page 11

Section 3. Accuracy of migration measurement

The following questions are about random fluctuations in measured emigration and immigration. We distinguish between nationals and non-nationals, and, moreover, between population registers and migration surveys as data collection systems.

Consider a European country with a **population register** in which there is no systematic bias in the measurement of migration. In other words, there is no undercount, and the measurement system covers all persons who migrate (in the sense of the UN definition). In this case, still 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.

Questions 9 - 12 all relate to this hypothetical country.

Question 9. Emigration of nationals, REGISTER country

9a) For emigration of nationals, 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 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 give your assessment as a range in percentages between 0 and 100 as before, with a lower end of the range (r1) and an upper end of the range (r2).

r1 =

r2 =

9b) Approximately, how certain are you that the true value will lie within the range that you provided above?

- 50 %
- 75 %
- 90 %
- 95 %
- other (pease state)

*

This element is only shown when the option "other (pease state)" is selected in the question "9b) Approximately, how certain are you that the true value will lie within the range that you provided above?"

other

Answers from round 1 for this question

id	8526257	8533066	8546148	8553207	8593119	8628621	8673705	8675156
r1	50 %	60 %	90 %	80 %	50 %	10 %	80 %	90 %
r2	75 %	100 %		100 %	70 %	20 %	90 %	95 %
c	50 %	95 %	75 %	90 %	50 %	75 %	75 %	50 %
id	8681334	8696877	8698827	8700689	8707108	8708183	8709921	
r1	-25 %	90 %	40 %	10 %	50 %	8 %	5 %	
r2	25 %	100 %	80 %	30 %	75 %	18 %	10 %	
c	50 %	95 %	85 %	75 %	75 %	-	75 %	

Page break

Page 12

Question 10. Immigration of nationals, REGISTER country

10a) For immigration of nationals, 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 level of immigration? (If it helps, think of how often the annual published statistics are within this interval during a period of 100 years.) Please give your assessment as a range in percentages between 0 and 100 as before, with a lower end of the range (r1) and an upper end of the range (r2).

r1 =

r2 =

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

- 50 %
- 75 %
- 90 %
- 95 %
- other (pease state)

*

This element is only shown when the option "other (pease state)" is selected in the question "10b) Approximately, how certain are you that the true value will lie within the range that you provided above?"

other

Answers from round 1 for this question

id	8526257	8533066	8546148	8553207	8593119	8628621	8673705	8675156
r1	80 %	80 %	90 %	85 %	70 %	10 %	80 %	93 %
r2	100 %	100 %		100 %	90 %	20 %	90 %	95 %
c	75 %	95 %	95 %	90 %	75 %	75 %	75 %	50 %
id	8681334	8696877	8698827	8700689	8707108	8708183	8709921	
r1	-10 %	95 %	50 %	50 %	90 %	0 %	70 %	
r2	10 %	100 %	90 %	60 %	100 %	5 %	80 %	
c	50 %	95 %	85 %	75 %	90 %	-	50 %	

Page break

Page 13

Question 11. Emigration of non-nationals, REGISTER country

11a) For emigration of non-nationals, 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 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 give your assessment as a range in percentages between 0 and 100 as before, with a lower end of the range (r1) and an upper end of the range (r2).

r1 =

r2 =

11b) Approximately, how certain are you that the true value will lie within the range that you provided above?

- 50 %
- 75 %
- 90 %
- 95 %
- other (pease state)

*

This element is only shown when the option "other (pease state)" is selected in the question "11b) Approximately, how certain are you that the true value will lie within the range that you provided above?"

other

Answers from round 1 for this question

id	8526257	8533066	8546148	8553207	8593119	8628621	8673705	8675156
r1	40 %	60 %	40 %	65 %	30 %	20 %	80 %	80 %
r2	75 %	100 %	70 %	95 %	50 %	30 %	90 %	90 %
c	50 %	95 %	75 %	90 %	50 %	75 %	75 %	50 %
id	8681334	8696877	8698827	8700689	8707108	8708183	8709921	
r1	-25 %	85 %	50 %	80 %	80 %	8 %	5 %	
r2	25 %	95 %	80 %	90 %	90 %	18 %	10 %	
c	50 %	75 %	85 %	90 %	75 %	-	50 %	

Page break

Page 14

Question 12. Immigration of non-nationals, REGISTER country

12a) For immigration of non-nationals, 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 level of immigration? (If it helps, think of how often the annual published statistics are within this interval during a period of 100 years.) Please give your assessment as a range in percentages between 0 and 100 as before, with a lower end of the range (r1) and an upper end of the range (r2).

r1 =

r2 =

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

- 50 %
- 75 %

- 90 %
- 95 %
- other (pease state)

*

This element is only shown when the option "other (pease state)" is selected in the question "12b) Approximately, how certain are you that the true value will lie within the range that you provided above?"

other

Answers from round 1 for this question

id	8526257	8533066	8546148	8553207	8593119	8628621	8673705	8675156
r1	50 %	90 %	90 %	85 %	60 %	20 %	80 %	90 %
r2	75 %	100 %		100 %	90 %	30 %	90 %	93 %
c	50 %	95 %	90 %	90 %	75 %	75 %	75 %	50 %
id	8681334	8696877	8698827	8700689	8707108	8708183	8709921	
r1	-10 %	95 %	60 %	70 %	80 %	17 %	10 %	
r2	10 %	100 %	90 %	80 %	90 %	27 %	20 %	
c	50 %	95 %	75 %	90 %	75 %	-	50 %	

Comments

If you have comments or arguments related to your answers to questions 9 - 12, please state them here.

Page break

Page 15

Now, consider a European country that uses a **survey** to collect migration data, and assume that there is no systematic bias in the measurement of migration. In this case, we may expect the accuracy to be affected by, for example, sampling error. Here, we will only consider immigration.

Questions 13 and 14 relate to this hypothetical country that uses a survey.

Question 13. Immigration of nationals, SURVEY country

13a) For immigration of nationals, 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 level of immigration? (If it helps, think of how often the annual published statistics are within this interval during a period of 100 years.) Please give your assessment as a range in percentages between 0 and 100 as before, with a lower end of the range (r1) and an upper end of the range (r2).

r1 =

r2 =

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

- 50 %
- 75 %
- 90 %
- 95 %
- other (pease state)

*

This element is only shown when the option "other (pease state)" is selected in the question "13b) Approximately, how certain are you that the true value will lie within the range that you provided above?"

other

Answers from round 1 for this question

id	8526257	8533066	8546148	8553207	8593119	8628621	8673705	8675156
r1	50 %	50 %	20 %	40 %	40 %	20 %	30 %	60 %
r2	75 %	100 %		80 %	60 %	30 %	50 %	70 %
c	50 %	75 %	95 %	75 %	75 %	50 %	50 %	50 %
id	8681334	8696877	8698827	8700689	8707108	8708183	8709921	
r1	-15 %	75 %	30 %	60 %	80 %	0 %	-	
r2	15 %	90 %	70 %	70 %	90 %	5 %	-	
c	50 %	50 %	75 %	75 %	75 %	-	-	

Page break

Page 16

Question 14. Immigration of non-nationals, SURVEY country

14a) For immigration of non-nationals, 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 level of immigration? (If it helps, think of how often the annual published statistics are within this interval during a period of 100 years.) Please give your assessment as a range in percentages between 0 and 100 as before, with a lower end of the range (r1) and an upper end of the range (r2).

r1 =

r2 =

14b) Approximately, how certain are you that the true value will lie within the range that you provided above?

- 50 %
- 75 %
- 90 %
- 95 %
- other (pease state)

*

This element is only shown when the option "other (pease state)" is selected in the question "14b) Approximately, how certain are you that the true value will lie within the range that you provided above?"

other

Answers from round 1 for this question

id	8526257	8533066	8546148	8553207	8593119	8628621	8673705	8675156
r1	40 %	50 %	50 %	60 %	20 %	20 %	30 %	55 %
r2	60 %	100 %		80 %	40 %	30 %	50 %	70 %
c	50 %	75 %	90 %	75 %	50 %	50 %	50 %	50 %
id	8681334	8696877	8698827	8700689	8707108	8708183	8709921	
r1	50 %	60 %	30 %	60 %	70 %	17 %	-	
r2		80 %	70 %	70 %	80 %	27 %	-	
c	50 %	50 %	75 %	90 %	75 %	-	-	

Comments

If you have comments or arguments related to your answers to questions 13 and 14, please state them here.

Page break

Page 17

Final comments

Many thanks 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.

Transforming experts' answers into probability distributions

QuantMig Deliverable 6.1 Part II *

Georgios Aristotelous †

*QuantMig is funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 870299.

†University of Southampton, UK. Email: G.Aristotelous@soton.ac.uk

Contents

1 Introduction

Deliverable 6.1 (D6.1) consists of two parts, Part I and Part II. What follows below is Part II. As explained in the introduction of Part I, the aim of both parts is to document the work done for Task 6.1, as part of Work Package 6 (WP6). Part I, entitled "Elicitation of expert opinions", described what Task 6.1 is concerned with, provided an overview of the migration flow modelling framework and the parameters, described how the Delphi survey was carried out (choice of experts, rounds, questions, feedback), and provided some summary statistics of the responses (see Part I for more details).

Part II, proceeds to explain how expert opinions on undercount and accuracy were transformed into statistical distributions. Part II is structured as follows. Section 2 describes the methods used to translate experts' answers into prior probability distributions for the parameters. Then, section 3 presents these resulting probability distributions, for both rounds of the Delphi questionnaire. Lastly, section 4 concludes with a summary of the main results of the analysis and some discussion points about lessons learned from the elicitation process.

We note that many of the steps we undertook, towards performing the task of translating answers into probability distributions, follow the work of Wiśniowski et al. (2013), where the same task was performed, under the same modelling framework; references to Wiśniowski et al. (2013) are made at the relevant points.

The computational analysis conducted for the purposes of the prior elicitation process (prior parameter estimation algorithms and prior probability distribution plots) was performed in the statistical programming language `R` (R Core Team, 2020). The code was written by the author and can be provided upon request.

2 Methods

In this section we describe how experts' answers were transformed into prior probability distributions for the parameters, for both rounds of the questionnaire. Consider any one of the questions 1-14 of the questionnaire, and let X denote the unknown quantity (random variable) of interest, for which a distribution was desired to be elicited. Note that for questions 1-8, X was related to undercount, whereas for questions 9-14, X was related to accuracy (a description of what X represents in each question is provided in section 3 below). All questions of the questionnaire had a common formulation with a given respondent being asked to provide a range of values (in part (a) of the question), along with a certainty value (in part (b) of the question), for the quantity in question to lie withing that range (all questions of the questionnaire are included in the Appendix of Part I). Note that, instead of asking the experts to provide these values directly for the quantity of interest X, we asked them to provide these values for some quantity $g(X)$, taking values in $[0, 1]$, were g was a suitably chosen continuous, strictly monotone deterministic function (a description of what $g(X)$ represents in each question is provided in section 3 below, while the explicit expression of the function $g(x)$ in each question is given in the Appendix A). Asking the experts to implicitly provide information for X, through $g(X)$, as opposed to asking them for X directly, made the questions more accessible.

The experts were asked to provide the range values and the certainty value as percentages. On what follows below we transform the provided percentage values to real numbers (proportions), to be in line with the mathematical framework, and we use the notation $r_{i,1}, r_{i,2}$ to denote the range values, and c_i to denote the certainty value, provided by an expert i. For example if, for a given question, an expert i provided a range of 50% to 70% in part (a), and a certainty of 80% in part (b), then $r_{i,1} = 0.5$, $r_{i,2} = 0.7$ and $c_i = 0.8$. For an answer to be translatable to a probability distribution it was required that $0 \leq r_{i,1} < r_{i,2} \leq 1$ and $0 \leq c_i \leq 1$. Answers that did not satisfy these restrictions were not translated to probability distributions. We note that, as in Wiśniowski et al. (2013), answers of 0 and 1 were changed to 0.0001 and 0.9999, respectively, for algorithm stability reasons.

Before proceeding further we note that the probability distribution of a continuous random variable can be characterized by its cumulative distribution function, which we henceforth refer to as cdf, or by its probability density function, which we henceforth refer to as density. We considered two methods for translating experts' answers into probability distributions. The first method, referred to as method 1, follows along the lines of Wiśniowski et al. (2013) and it is a two-step procedure, where first the experts' individual answers are translated into individual densities, and second, the individual densities are combined into an aggregated density. The second method, referred to as method 2, is a one-step procedure where experts' individual answers are directly combined (without being first translated into individual densities) into an aggregated density. Sections 2.1 and 2.2, respectively, provide a description of the two methods, with technical details being included in the Appendix.

2.1 Method 1

As mentioned right above, method 1 is a two-step procedure, where we first translate experts' individual answers into individual densities and then combine the individual densities to form an aggregated density. As above, consider any one of the questions 1-14, and let X denote the unknown quantity (random variable) of interest, for which a distribution was desired to be elicited. For each expert i, $i = 1, 2, ..., n$, where n the number of experts which provided an answer such that elicitation was possible (see the beginning of section 2 for which answers were not possible to be translated to probability distributions), let θ_i, $F(x; \theta_i)$ and $f(x; \theta_i)$, respectively denote the parameter (vector), the cdf and the density of X, corresponding to expert i. Notice that the dependence of the cdf and the density of X on expert i is made explicit in the notation through θ_i. This notational choice serves for two purposes. First, it highlights that under method 1, an expert's individual answer corresponds to an individual density. Second, it reveals that these individual densities differ on the parameter θ_i, and not on the family of the distribution of X (the choice of family for the distribution of X is described below). The same notational rule is followed below with all quantities corresponding to expert i having an i subscript in their notation.

2.1.1 Transforming experts's individual answers into individual densities

In a nutshell, to translate an expert's answer into a density we performed three tasks. First, we translated the expert's answer into two points of the cdf of X. Second, we made a suitable choice for the family of the distribution of X. Third, we specified the value of the expert-specific parameter of the distribution of X, by substituting the two points on the cdf of X and finding the solution of the resulting system of equations. Below we provide more information on how each of these three tasks was performed.

To translate the expert's answer into two points of the cdf of X we worked as follows. As mentioned above (see section 2), in any given question, a given expert i, $i = 1, 2, ..., n$, provided a range of values for $g(X)$ (in part (a) of the question), say $r_{i,1}$ and $r_{i,2}$ ($0 \leq r_{i,1} < r_{i,2} \leq 1$), and a certainty/probability value (in part (b) of the question), say c_i ($0 \leq c_i \leq 1$), for $g(X)$ to lie within the interval $[r_{i,1}, r_{i,2}]$. Such an answer translates to the following probability statement about $g(X)$:

$$P(r_{i,1} < g(X) < r_{i,2}) = c_i \tag{2.1}$$

Note that the above statement directly implies that the probability for $g(X)$ to lie outside of the interval $[r_{i,1}, r_{i,2}]$, that is in $[0, r_{i,1}]$ or $[r_{i,2}, 1]$, is equal to $1-c_i$. Following Wiśniowski et al. (2013) we assigned the remaining probability of $1 - c_i$ (that not directly assigned by the expert) to be proportional to the length of the intervals $[0, r_{i,1}]$ and $[r_{i,2}, 1]$. In probability terms, this assignment is expressed as:

$$\begin{aligned} P(g(X) < r_{i,1}) &= k_i r_{i,1} \\ P(g(X) > r_{i,2}) &= k_i(1 - r_{i,2}), \end{aligned} \tag{2.2}$$

where $k_i = \frac{1-c_i}{1+r_{i,1}-r_{i,2}}$ is the proportionality constant, calculated by substituting equations (2.1) and (2.2), into equation $P(g(X) < r_{i,1}) + P(r_{i,1} < g(X) < r_{i,2}) + P(g(X) > r_{i,1}) = 1$. Utilizing the properties of the function g, the probability expressions for $g(X)$ in equation (2.2) right above, can be reexpressed in terms of X as:

$$\begin{aligned} P(X < g^{-1}(r_{i,1})) &= k_i r_{i,1} \\ P(X > g^{-1}(r_{i,2})) &= k_i(1 - r_{i,2}), \end{aligned} \tag{2.3}$$

or as

$$\begin{aligned} P(X > g^{-1}(r_{i,1})) &= k_i r_{i,1} \\ P(X < g^{-1}(r_{i,2})) &= k_i(1 - r_{i,2}), \end{aligned} \tag{2.4}$$

depending whether g is strictly increasing or strictly decreasing, respectively. In the above equations, and in what follows, g^{-1} denotes the inverse function of g. Since the cdf of X corresponding to expert i, $F(x;\theta_i)$, is such that $F(x;\theta_i) = P(X < x)$, the probability expressions in equations (2.3) and (2.4) right above, specify two points of $F(x;\theta_i)$, point $(x_{i,1}, y_{i,1})$ and point $(x_{i,2}, y_{i,2})$, where $x_{i,1} = g^{-1}(r_{i,1})$ and $x_{i,2} = g^{-1}(r_{i,2})$, and, $y_{i,1} = k_i r_{i,1}$ and $y_{i,2} = 1 - k_i(1 - r_{i,2})$, for g strictly increasing, or $y_{i,1} = 1 - k_i r_{i,1}$ and $y_{i,2} = k_i(1 - r_{i,2})$, for g strictly decreasing (the explicit form of g^{-1} in each question, required to calculate $x_{i,1}$ and $x_{i,2}$, is given in the Appendix A).

Regarding the second task, that is the task of specifying a family for the distribution of X, we followed the specification of Wiśniowski et al. (2013). Specifically, for the undercount questions (questions 1-8), we assumed a Beta(α, β) distribution for X, parametrized by the two shape parameters, α and β. In these questions, the expert-specific parameter θ_i was therefore $\theta_i = (\alpha_i, \beta_i)$, $i = 1, 2, \ldots, n$. For the accuracy questions (questions 9-14), we assumed a Gamma(ν, ρ) distribution, parametrized by the shape parameter ν and the rate parameter ρ. For these questions, the expert-specific parameter θ_i was $\theta_i = (\nu_i, \rho_i)$, $i = 1, 2, \ldots, n$. These families are naturally suited for the parameters in question and also bind well with the underlying modelling framework (see Wiśniowski et al. (2013) for more details). As mentioned above, the choice of family was common across experts and the individual densities differed only on the expert-specifc parameter θ_i.

To perform the third task, that is to specify a value for expert-specific parameter θ_i, we substituted the two points provided by the expert, $(x_{i,1}, y_{i,1})$ and $(x_{i,2}, y_{i,2})$, on the cdf of X and computed the solution of the resulting system of equations

$$\begin{aligned} F(x_{i,1};\theta_i) &= y_{i,1} \\ F(x_{i,2};\theta_i) &= y_{i,2} \end{aligned} \tag{2.5}$$

The solution of the above system was computed numerically by minimizing the function $d(\theta_i) = (F(x_{i,1};\theta_i) - y_{i,1})^2 + (F(x_{i,2};\theta_i) - y_{i,2})^2$, the (squared) Euclidean distance between the cdf of X and the two points, $(x_{i,1}, y_{i,1})$ and $(x_{i,2}, y_{i,2})$, i.e. by computing $\underset{\theta_i}{\operatorname{argmin}}\, d(\theta_i)$. This is because the solution of the above system, say $\hat{\theta}_i$, is such that $d(\hat{\theta}_i) = 0$ and therefore, since $d(\theta_i) \geq 0$, it is the case that $\hat{\theta}_i = \underset{\theta_i}{\text{argmin}}\ d(\theta_i)$ (sections B.1 and B.2 in the Appendix give the initial values used for these minimization algorithms, for question 1-8 and 9-14, respectively). A similar procedure to solve system 2.5 was also followed in Wiśniowski et al. (2013) (only that there, the minimization was conducted with respect to the inverse cdf, as opposed to the cdf). Notice that the specification of $\hat{\theta}_i$, simultaneouly specifies the density of expert i to be $f(x; \hat{\theta}_i)$, $i = 1, 2, ..., n$.

2.1.2 Combining experts' individual densities into an aggregated density

Following Wiśniowski et al. (2013), to combine the experts' individual densities, $f(x; \hat{\theta}_i)$, into a single, aggregated, prior density for X, say $\pi(x)$, we used an equally weighted-mixture density. That is, $\pi(x)$ was specified as:

$$\pi(x) = \sum_{i=1}^{n} \frac{1}{n} f(x; \hat{\theta}_i) \tag{2.6}$$

As discussed in Wiśniowski et al. (2013), such an equally-weighted opinion pool offers a simple, robust and general method for aggregating expert knowledge.

2.2 Method 2

Method 2 offers an alternative way for specifying prior distributions for the parameters. As already mentioned in the beginning of section 2, method 2, unlike method 1, directly combines experts' answers to specify a single, aggregated density, without first translating them into individual densities. The main difference between method 2 and method 1, as far as the form of the aggregated density, is that the aggregated density, under method 2, is typically smoother, compared to that under method 1. Having smooth prior densities can be an appealing feature, especially in settings such as the present one, where the observed data contain very little information for some of the parameters, meaning that their posterior densities will largely be determined by the form of their prior densities.

The procedure that specifies the aggregated density, under method 2, is almost the same as the procedure that specifies an expert's individual density under method 1. The subtle but defining difference is that, under method 2, the procedure is applied to all experts' answers/points at once (specifying the aggregated density in a single step), whereas under method 1 the procedure is applied seperetaly to each expert's answer/points (specifying each expert's individual density in a first step, with the aggregated density being specified in a subsequent step).

As in subsection 2.1, consider any one of the questions 1-14 and let X denote the quantity (random variable) of interest. Let also an expert be denoted by i, $i = 1, 2, ..., n$, where n the number of experts which provided an answer such that elicitation was possible (see the beginning of section 2 for which answers were not possible to be translated to probability distributions). We denote the parameter (vector), the cdf and the density of X as θ, $F(x; \theta)$ and $f(x; \theta)$, respectively. Notice that, unlike method 1, these are not indexed by i since, as already mentioned, under method 2, we do not have any expert-specific densities but rather a single common density.

Same as for the specification of the expert-specific densities in method 1, to specify the aggregrated density using method 2, we performed the following three tasks. First, we translated all experts' answers to points of the cdf of X, with each expert's answer contributing a pair of points, as described in the second paragraph of subsection 2.1.1. For example, in Question 1 of Round 1, where 15 experts provided complete and valid answers, there were 30 points. As in section 2.1.1, we use the notation $(x_{i,1}, y_{i,1})$ and $(x_{i,2}, y_{i,2})$ to denote the pair of points corresponding to expert i, $i = 1, 2, ..., n$. Second, we chose the family of the distribution of X. Same as for method 1 (see the third paragraph of subsection 2.1.1) we followed the specification of Wiśniowski et al. (2013), assuming $X \sim \text{Beta}(\alpha, \beta)$ for the undercount questions and $X \sim \text{Gamma}(\nu, \rho)$ for the accuracy questions. That is to say that, the common (over experts) parameter θ, of the distribution of X, was $\theta = (\alpha, \beta)$ and $\theta = (\nu, \rho)$, in questions 1-8 and 9-14, respectively. Third, we specified the value of θ by minimizing the (squared) Euclidean distance between the cdf of X and all points $(x_{i,1}, y_{i,1})$, $(x_{i,2}, y_{i,2})$, $i = 1, 2, ..., n$. That is, our chosen value for θ, say $\hat{\theta}$, was such that $\hat{\theta} = \underset{\theta}{\text{argmin}}\, d(\theta)$, where $d(\theta) = \sum_{i=1}^{n} (F(x_{i,1}; \theta) - y_{i,1})^2 + (F(x_{i,2}; \theta) - y_{i,2})^2$, which is the least-squares estimate of θ resulting from fitting the cdf of X through all the provided points (sections B.1 and B.2 in the Appendix give the initial values used for these minimization algorithms, for question 1-8 and 9-14, respectively). Note that this way of specifying the parameter (using least-squares) is the same as the one applied in the prior elicitation tools, The Elicitator (Bastin et al. (2013)), SHELF (Gosling (2017)) and MATCH (Morris et al. (2014)). What the above translate to is that the resulting prior density for X, say $\pi(x)$, was set as:

$$\pi(x) = f(x; \hat{\theta}). \tag{2.7}$$

3 Results

This section presents the resulting prior densities, for both rounds of the Delphi questionnaire, under both method 1 and method 2. First, we present the densities from Round 1 of the questionnaire (section 3.1), followed by the description of some feedback (section 3.2) which we returned to the experts, after Round 1 and before Round 2. Then we present the densities from Round 2 and make comparisons between the two rounds (section 3.3).

In questions 1-8, the parameter of interest X, represents the proportion of true flow counts that are reported by a given country. More specifically, ignoring the effect of the other measurement error parameters, under our assumed modelling framework (see Raymer et al. (2013)), we can describe the effect of the parameter X via the relationship $\mu = yX$, where μ can be thought of as representing a reported flow count and y is the corrsponding true flow count. Questions 1-8 asked the experts to provide information for $g(X) = 1 - X$, as opposed to X directly. That is, questions 1-8 asked for the proportion of true flow counts that are not reported by a given country, i.e. for the proportion of undercount associated with the reporting system of a given country. This was done to avoid causing confusion between high and low undercount questions, since for $g(X)$ high (low) values imply higher (lower) undercount, whereas for X, it is the other way around.

In question 9-14, the parameter of interest X, is the precision (inverse variance) of the random fluctuation error term associated with the measurement of migration. More precisely, assuming that there are no sources of systematic error, under our assumed modelling framework (see Raymer et al. (2013)), we can describe the effect of the parameter X via the relationship $\mu = y \exp(\varepsilon|_X)$, where μ represents a reported flow count, y is the corresponding true flow count, and $\varepsilon|_X \sim \mathrm{N}(0, 1/X)$, where $\mathrm{N}(m, v)$ denotes a Normal distribution with mean m and variance v. Asking the experts to provide information for X directly would be too involved (due to the fact that there are two levels of randomness at work here, that of $\varepsilon|_X$ and that of X itself) and so we asked the experts to provide information for $g(X) = \mathrm{E}(\mathbb{1}_A \mid X)$, where $A = \{0.95 < \exp(\varepsilon|_X) < 1.05\}$. That is, questions 9-14 asked for the expected proportion of observations with less than 5% error, i.e. for the expected proportion of times that the published statistics of a given country are within an interval of -5% to $+5\%$ compared to the true level of immigration, when no systematic errors exist.

In all the plots to follow, we plot densities for $g(X)$ and not for X. There are two main reasons for doing so. First, is that the questions asked information about $g(X)$ and not for X, and so it is more informative to present densities for the quantity in question. Second, it is visually advantageous to plot for $g(X)$, instead of X, since $g(X)$ has the bounded support $[0, 1]$, in all questions, unlike X whose support is not bounded in questions 9-14.

Before providing the results we make two notes. First, we note that all questions of the questionnaire are included in the Appendix of Part I, for reference. Second, as per the Delphi technique, we note that for the model, it is the prior distributions of Round 2 that are used as input. Therefore, although we fully report and comment the results of Round 1, from a modelling standpoint we are more concerned with the results of Round 2.

3.1 Round 1

Figures 1-14 respectively present the elicited densities for questions 1-14, from Round 1. Each figure presents two plots, one with all individual densities and the aggregated densities imposed, and another one with the aggregated densities on their own. The second plot uses a different y-axis scale to allow better visual illustration of the aggregated densities.

For the undercount questions (questions 1-8, figures 1-8), overall, there was a fair amount of heterogeneity in the opinions of respondents, although for some questions, such as questions 2 and 4 (see figures 2 and 4), the amount of heterogeneity was less than others, such as questions 5 and 6 (see figures 5 and 6). Some respondents provided answers with very high certainty while others were much less certain, leading to highly peaked densities or densities with a lot of mass on the tails, respectively (see for example figure 4). Sensibly, the experts' densities suggested that undercount is more likely to be less in countries with low undercount compared to high undercount countries, when EU+ status (EU+ national or non-EU+ national) and type of migration (immigration or emigration) is the same (see for example figures 1 and 5 or figures 2 and 6). Also as expected, experts appear to be more certain in their assessment of immigration compared to emigration, when type of undercount (low or high) and EU+ status is the same (see for example figures 1 and 2).

For the accuracy questions (questions 9-14, figures 9-14), the opinions of respondents were again quite heterogeneous, although in some questions, such as 9 and 11 (see figures 9 and 11) the experts can be loosely divided into those that suggested that accuracy was very poor and to those that suggested that accuracy was very good, putting a lot of the mass either near 0 or 1, respectively. Reasonably, the experts' densities, overall, have more mass near 1 for register countries compared to survey countries.

Regarding the comparison of the two aggregated densities, as expected, method 2 produces smoother densities compared to method 1. As far as uncertainty, neither of the two is systematically more uncertain than the other, in the sense that in some questions it is the density from method 1 that is more peaked around a value (see for example figure 13), whereas in other questions it is the other way around (see for example figure 2). What is noticable is that, for a lot the questions, the two aggregated densities are not that similar. For example, in the accuracy questions, it appears that the aggregated density from method 2 is much more susceptible to the pattern of answers of either very poor or very good accuracy (see above paragraph) and puts a lot of its mass around 0 and 1, the boundary values of the support (see figures 9-12).

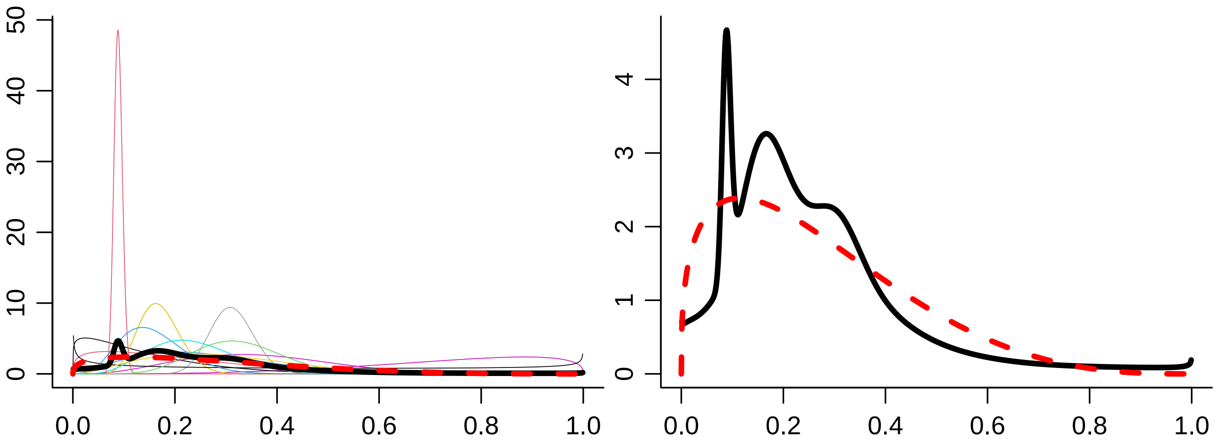


Figure 1: Experts' answers transformed to probability densities for question 1 (undercount of emigration of EU+ nationals who leave low undercount countries) of Round 1. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.

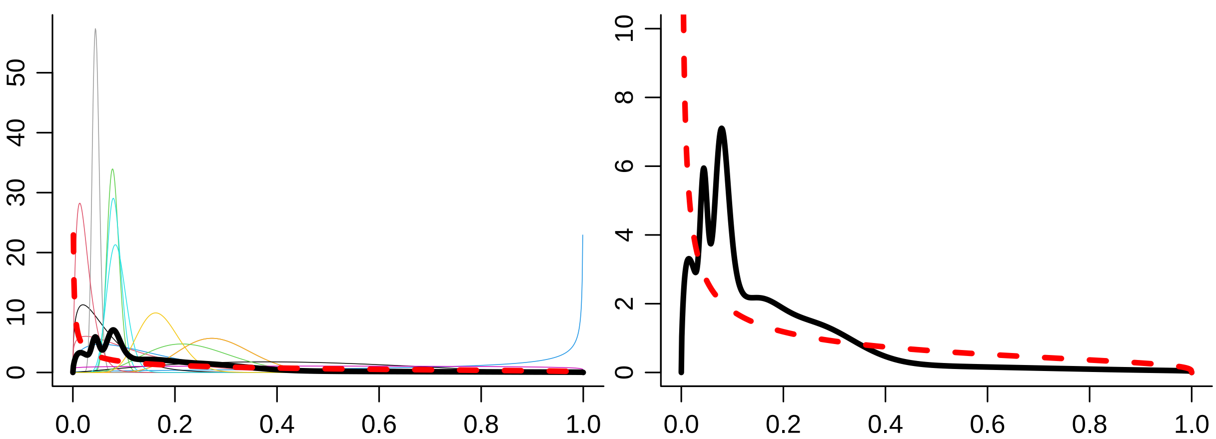


Figure 2: Experts' answers transformed to probability densities for question 2 (undercount of immigration of EU+ nationals who enter low undercount countries) of Round 1. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.

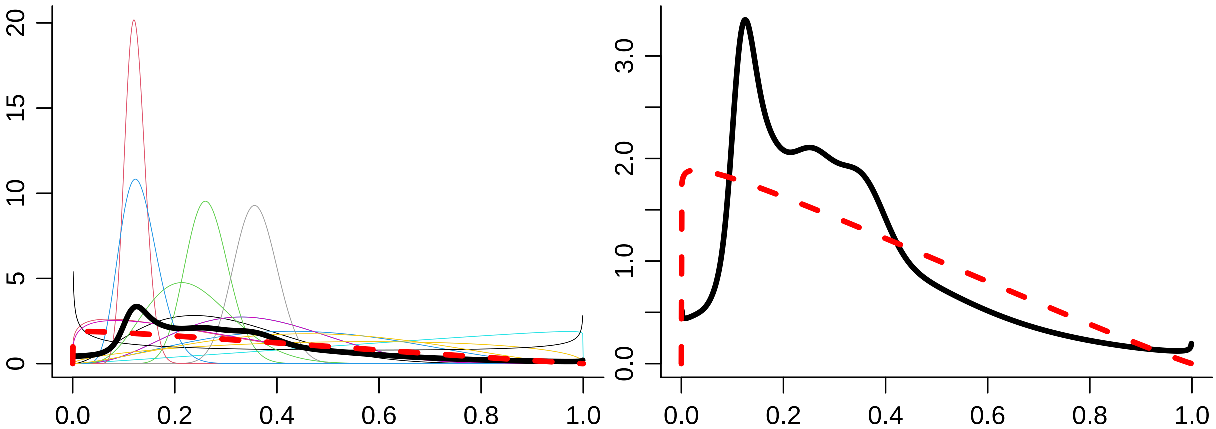


Figure 3: Experts' answers transformed to probability densities for question 3 (undercount of emigration of non-EU+ nationals who leave low undercount countries) of Round 1. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.

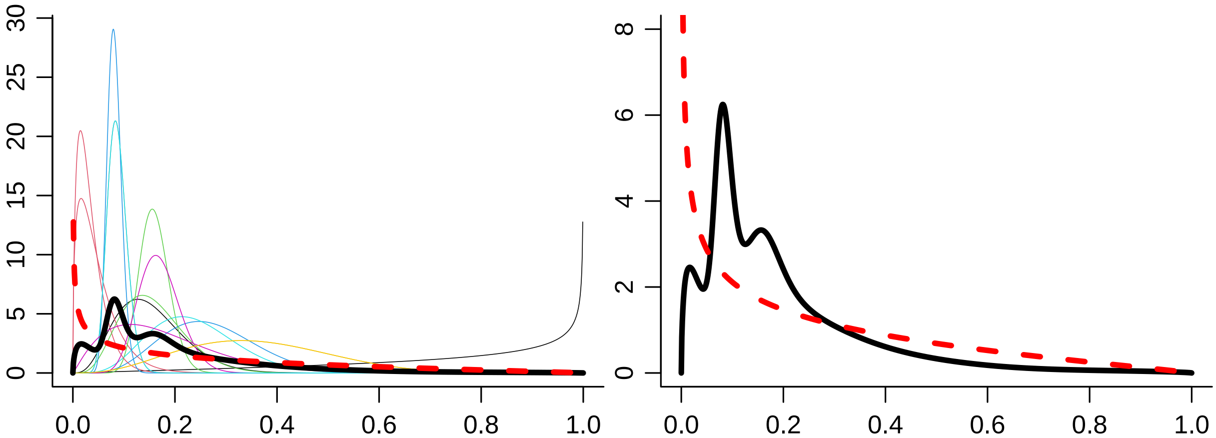


Figure 4: Experts' answers transformed to probability densities for question 4 (undercount of immigration of non-EU+ nationals who enter low undercount countries) of Round 1. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.

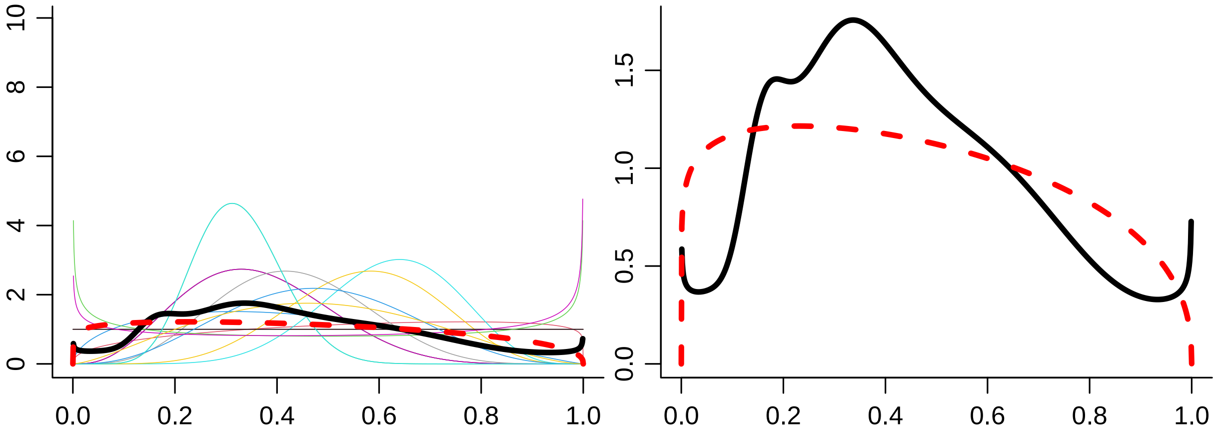


Figure 5: Experts' answers transformed to probability densities for question 5 (undercount of emigration of EU+ nationals who leave high undercount countries) of Round 1. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.

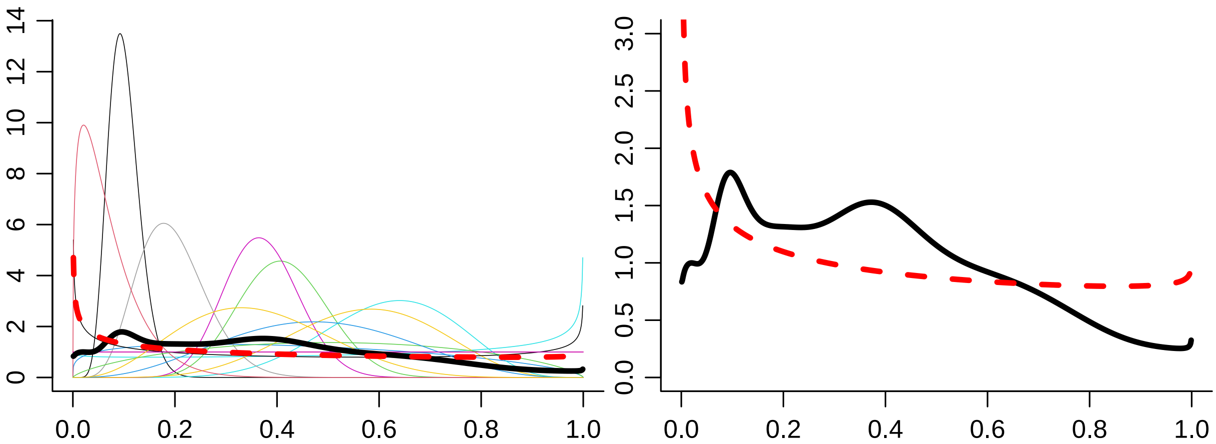


Figure 6: Experts' answers transformed to probability densities for question 6 (undercount of immigration of EU+ nationals who enter high undercount countries) of Round 1. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.

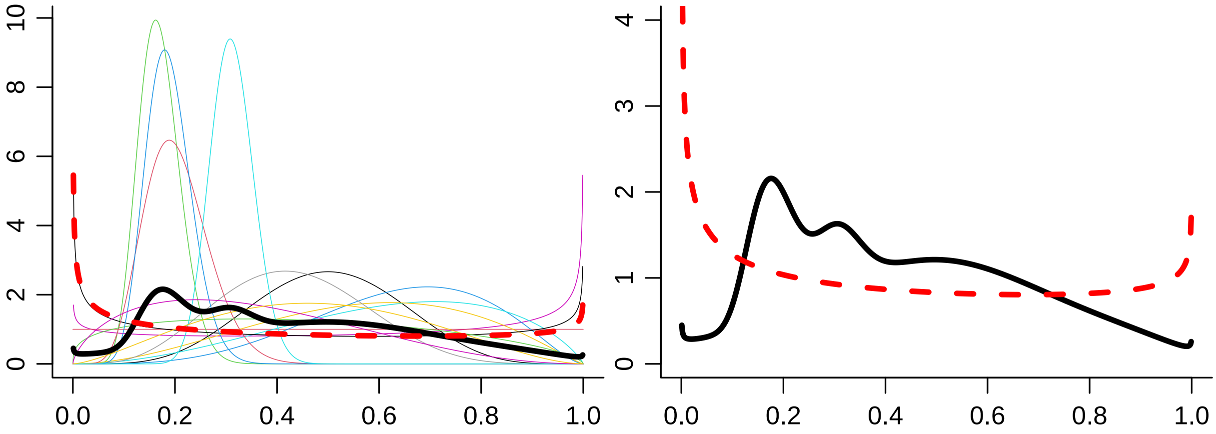


Figure 7: Experts' answers transformed to probability densities for question 7 (undercount of emigration of non-EU+ nationals who leave high undercount countries) of Round 1. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.

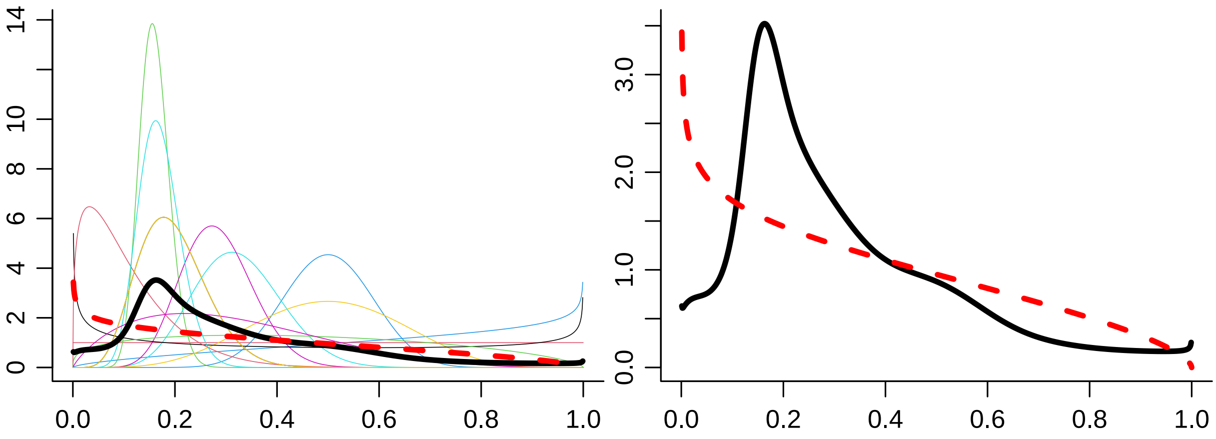


Figure 8: Experts' answers transformed to probability densities for question 8 (undercount of immigration of non-EU+ nationals who enter high undercount countries) of Round 1. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.

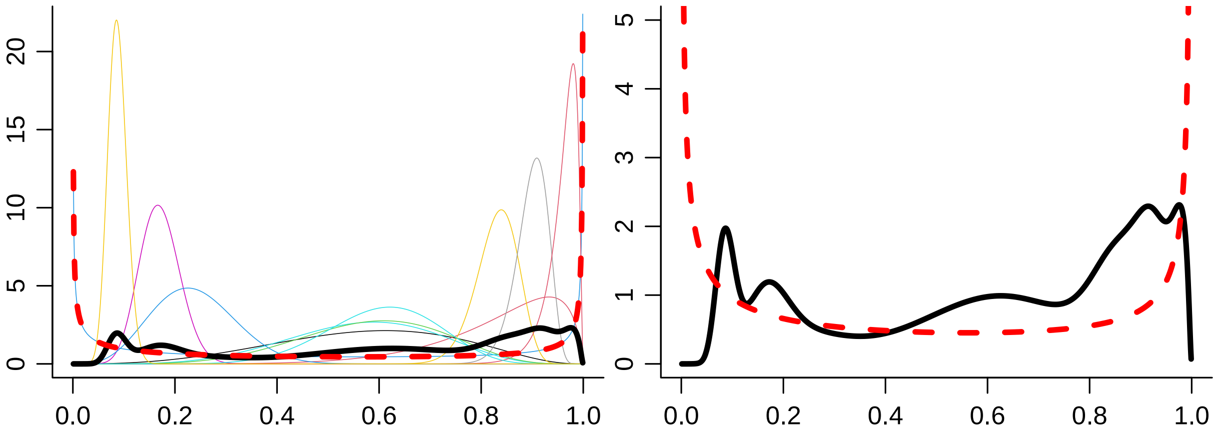


Figure 9: Experts' answers transformed to probability densities for question 9 (accuracy of emigration of EU+ nationals who leave countries recording with a register) of Round 1. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.

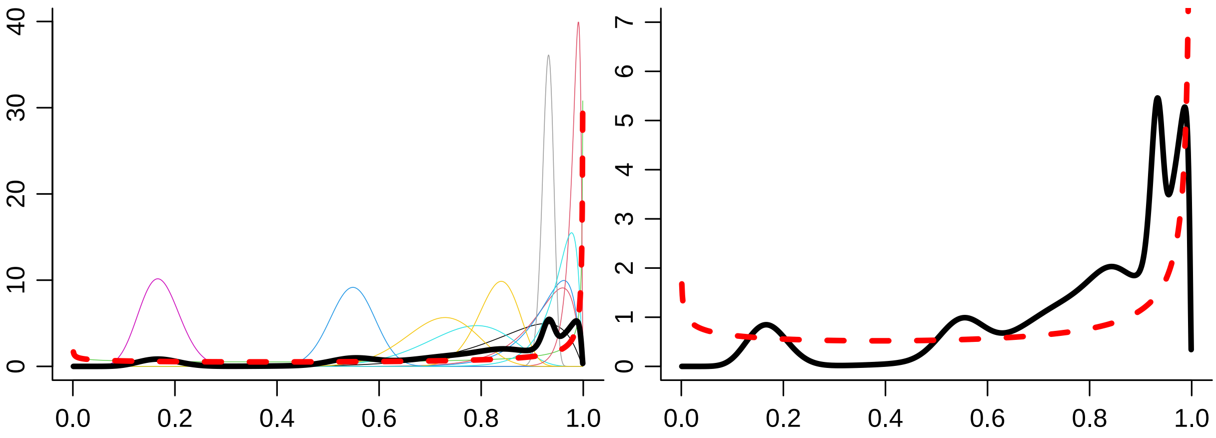


Figure 10: Experts' answers transformed to probability densities for question 10 (accuracy of immigration of EU+ nationals who enter countries recording with a register) of Round 1. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.

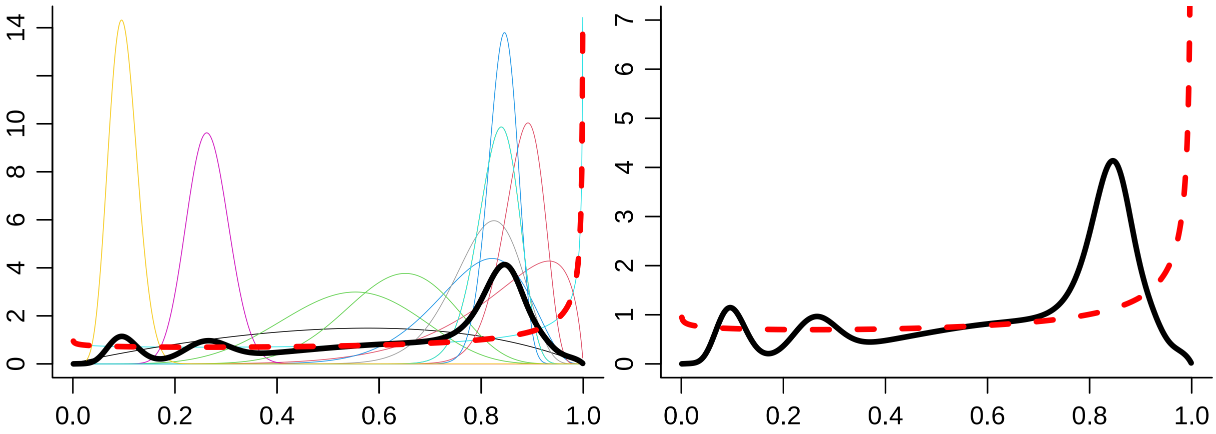


Figure 11: Experts' answers transformed to probability densities for question 11 (accuracy of emigration of non-EU+ nationals who leave countries recording with a register) of Round 1. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.

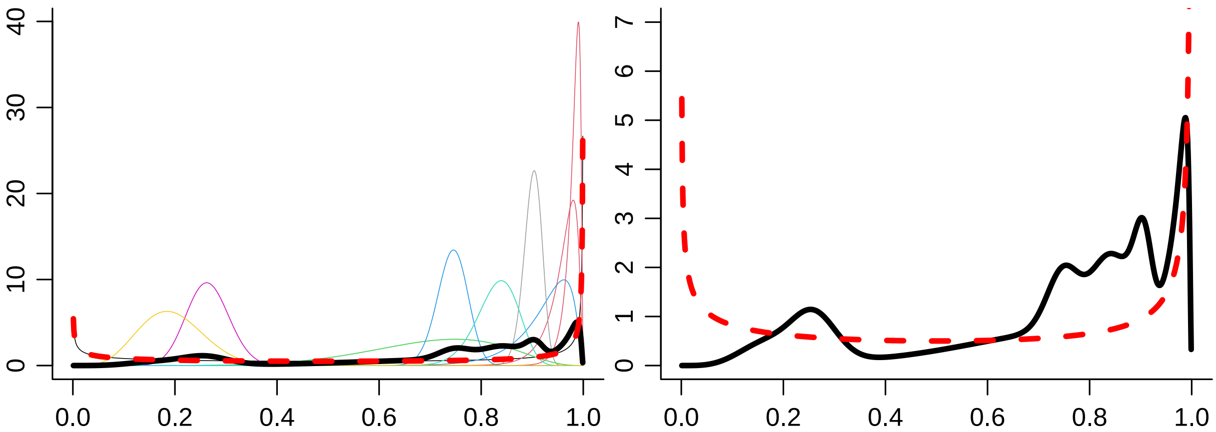


Figure 12: Experts' answers transformed to probability densities for question 12 (accuracy of immigration of non-EU+ nationals who enter countries recording with a register) of Round 1. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.

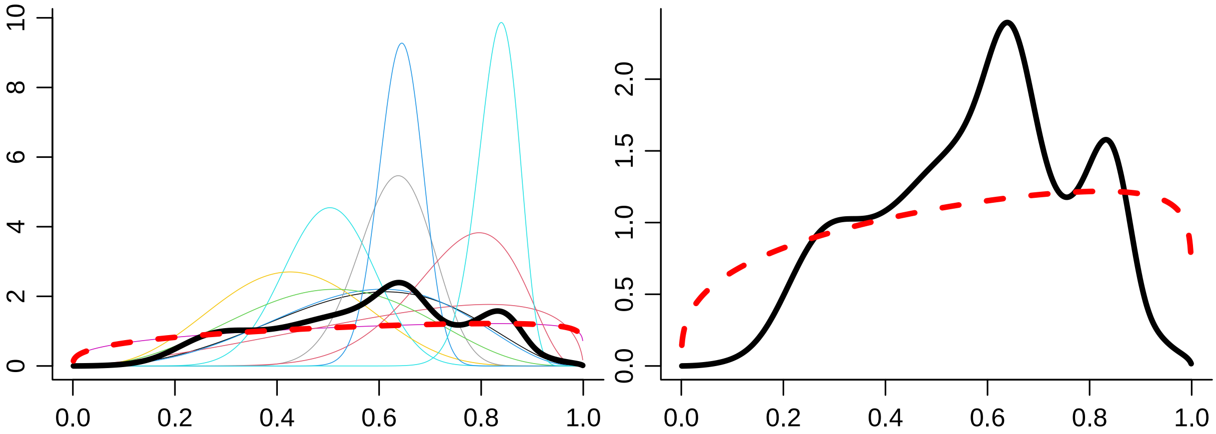


Figure 13: Experts' answers transformed to probability densities for question 13 (accuracy of immigration of EU+ nationals who enter countries recording with a survey) of Round 1. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.

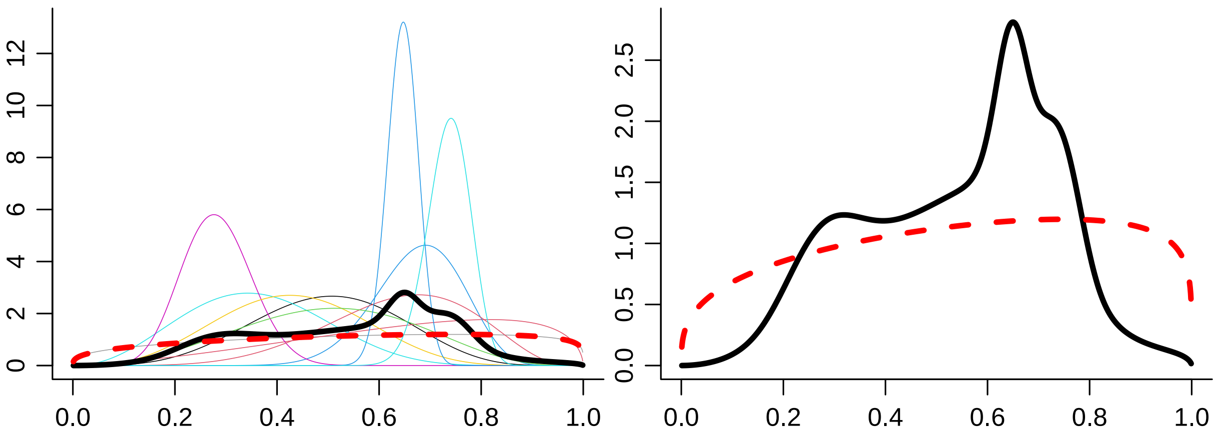


Figure 14: Experts' answers transformed to probability densities for question 14 (accuracy of immigration of non-EU+ nationals who enter countries recording with a survey) of Round 1. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.

3.2 Feedback to experts

Before starting Round 2, we provided the experts with feedback, based on their answers on Round 1. Each expert, received two documents, an expert-specific document (providing feedback specific to the expert) along with a generic document (providing generic feedback for all experts). The generic document provided an overall summary of the results of Round 1 and an explanation on how expert answers were translated to probability statements. The expert-specific document included two sets of plots, with each set containing 14 plots, one for each question. A plot of the first set presented all expert densities, with the given expert's density being highlighted, allowing the expert to compare their density with the densities of the other experts, establish a reference, and perhaps consider revising their answer for Round 2. A plot of the second set presented the expert's density on its own, serving as a visual medium, for explaining how the expert's answer was translated to a density, as well as for comments and clarifications. For completeness, an example expert-specific feedback document is attached in the Appendix C.

Something that we made sure to include in the feedback documents, was how an expert's answer translated to a probability statement. Based on some of the answers of Round 1, it was our impression that some of the experts might had not been aware that by providing a range r_1, r_2 and a certainty c, they essentially stated that the probability that the quantity in question $g(X)$ lies in the interval $[r_1, r_2]$ is equal to c, and, simultaneously, the probability that it does not lie in $[r_1, r_2]$, is equal to $1 - c$. For example, we speculate that some answers such that c was smaller than the length of the interval $[r_1, r_2]$ were given without realizing that by expressing (relatively) low certainty c for $g(X)$ to lie in $[r_1, r_2]$ one automatically expresses high certainty for $g(X)$ to lie outside of $[r_1, r_2]$. Such answers typically translated to U-shaped densities (see for example figure 7), having most of the mass near 0 and 1, which we considered to be rather implausible.

Another point that we reiterated in the documents was that for the accuracy questions, the questions asked for the sampling variability in the reporting of migration, under the assumption that there are no sources of systematic bias. We considered this reminder important to make, having speculated that perhaps some of the answers suggesting very poor accuracy in Round 1, were given by not taking into consideration the above assumption.

In the feedback documents, we also included clarifications in case an expert provided invalid answers (see section 2 for what were the requirements for an answer to be valid). In addition to that (see Part I), further explanations and online checks were included in Round 2 of the questionnaire in an attempt to avoid invalid answers.

3.3 Round 2

For Round 2, we produced the same set of plots as in Round 1. Specifically, figures 15-28 present the elicited densities for questions 1-14, respectively. To aid comparison between rounds, alongside the Round 2 plots, we included the Round 1 plots as well, using the same y-scale between rounds. For completeness, plots of the densities of only Round 2, on a different y-scale, are given in Appendix D.

For the undercount questions (questions 1-8, 15-22), similar to Round 1, the experts' densities were sensible in the sense that they put more mass on low (high) undercount values for low (high) undercount countries (see for example figures 15 and 19 or figures 2 and 20) and were more certain in their assessment of immigration compared to emigration (see for example figures 17 and 18). Compared to Round 1, the amount of heterogeneity in experts' answers was evidently lower, something that is perhaps more clearly reflected by the nature of the mixture aggregated density (aggregated density of method 1), which is typically multimodal in Round 1 and much closer to unimodal in Round 2 (see for example figures 15 and 17).

For the accuracy questions (questions 9-14, 23-28), as in Round 1, the experts' densities, as reasonably expected, had more mass near 1 for register countries compared to survey countries. Unlike Round 1, in Round 2, there was more homogeneity among experts' opinions. In particular, the pattern of experts' opinions being divided into those that suggested that accuracy was very poor and those that suggested that it was very good, was not observed in Round 2.

As far as the comparison of the aggregated densities from the two methods, they are much more similar in Round 2, compared to Round 1, reflecting the higher homogeneity among expert opinions, exhibited in Round 2. In fact, in a lot of the questions, the aggregated density of method 2 appears as a smoothed version of the aggregated density of method 1, which is what we aimed to achieve with the conception of method 2.

Based on the results of Round 2, we perceived that the feedback which we provided between the two rounds helped the experts better understand the questions. For example, it appears that, clarifying that for the accuracy questions the assumption was that no systematic bias existed, may have made some of the experts that gave answers supporting low accuracy values to revise their answers. Also, there were much fewer U-shaped densities in Round 2, compared to Round 1, suggesting that experts were more aware on how their answers translated to probability statements in Round 2. Overall, we considered that the feedback led to more informed densities in Round 2, compared to Round 1, something which was very welcome, considering that the prior distributions of Round 2 are used as input into the model.

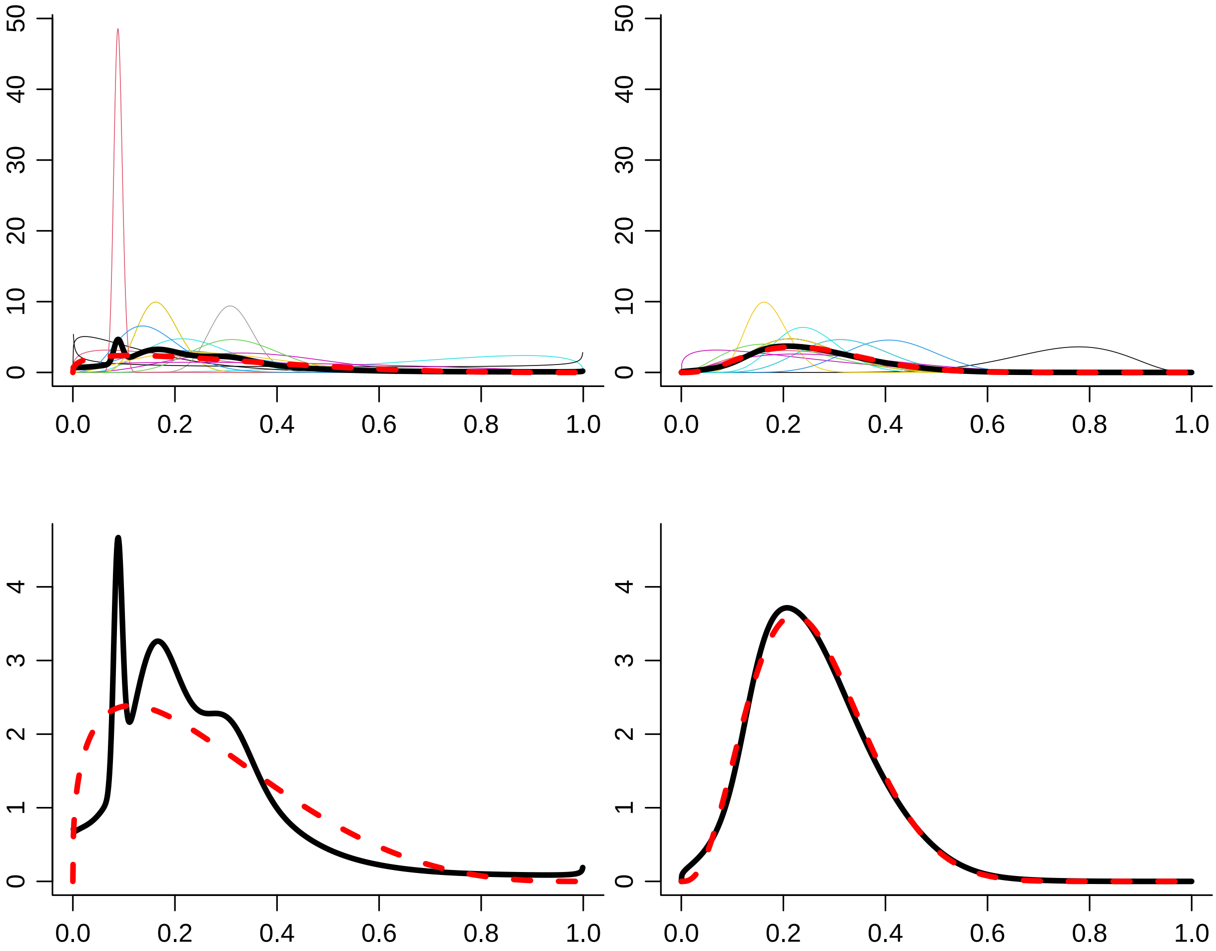


Figure 15: Experts' answers transformed to probability densities for question 1 (undercount of emigration of EU+ nationals who leave low undercount countries). Left column corresponds to Round 1 and right column corresponds to Round 2. Top row are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2) and bottom row are the two aggregated densities on their own.

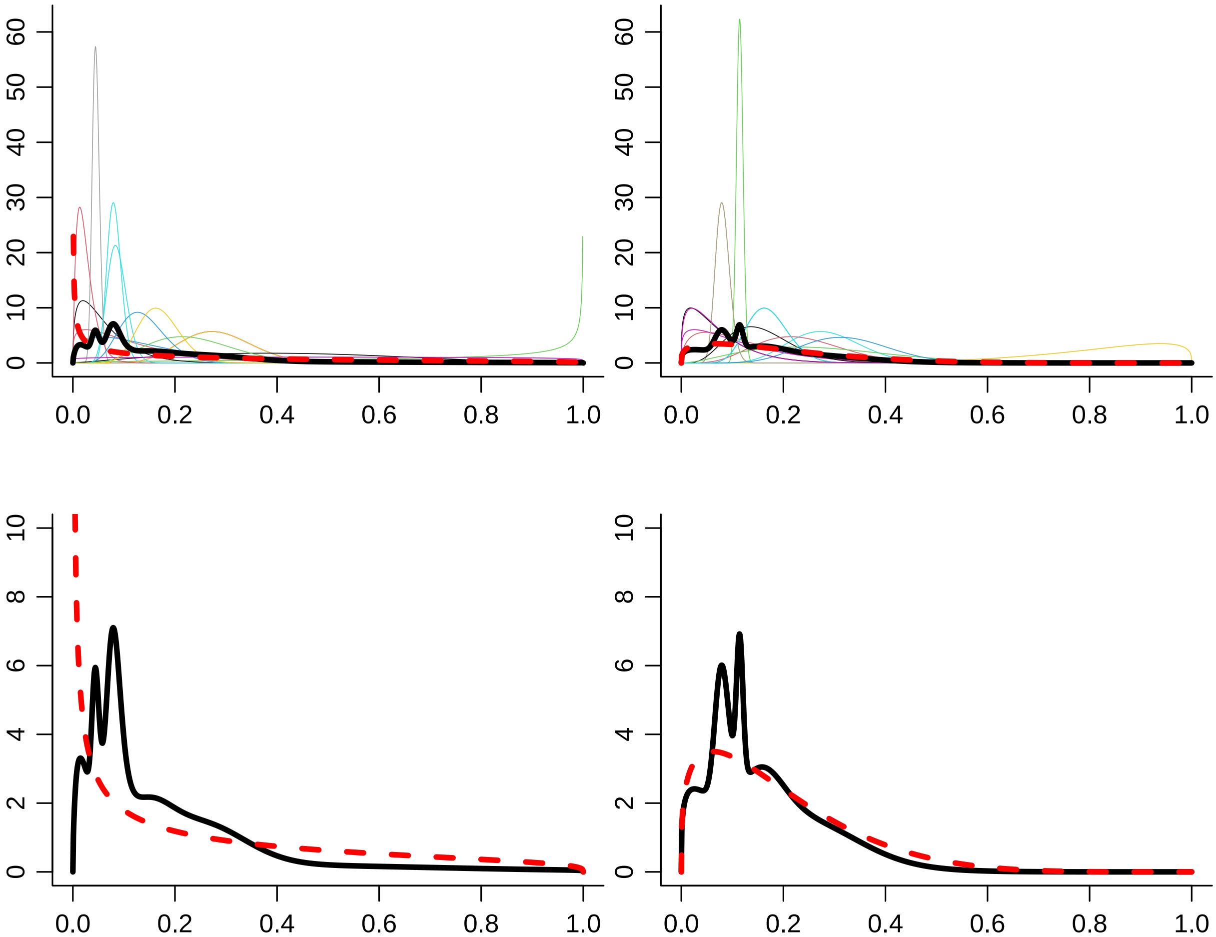


Figure 16: Experts' answers transformed to probability densities for question 2 (undercount of immigration of EU+ nationals who enter low undercount countries). Left column corresponds to Round 1 and right column corresponds to Round 2. Top row are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2) and bottom row are the two aggregated densities on their own.

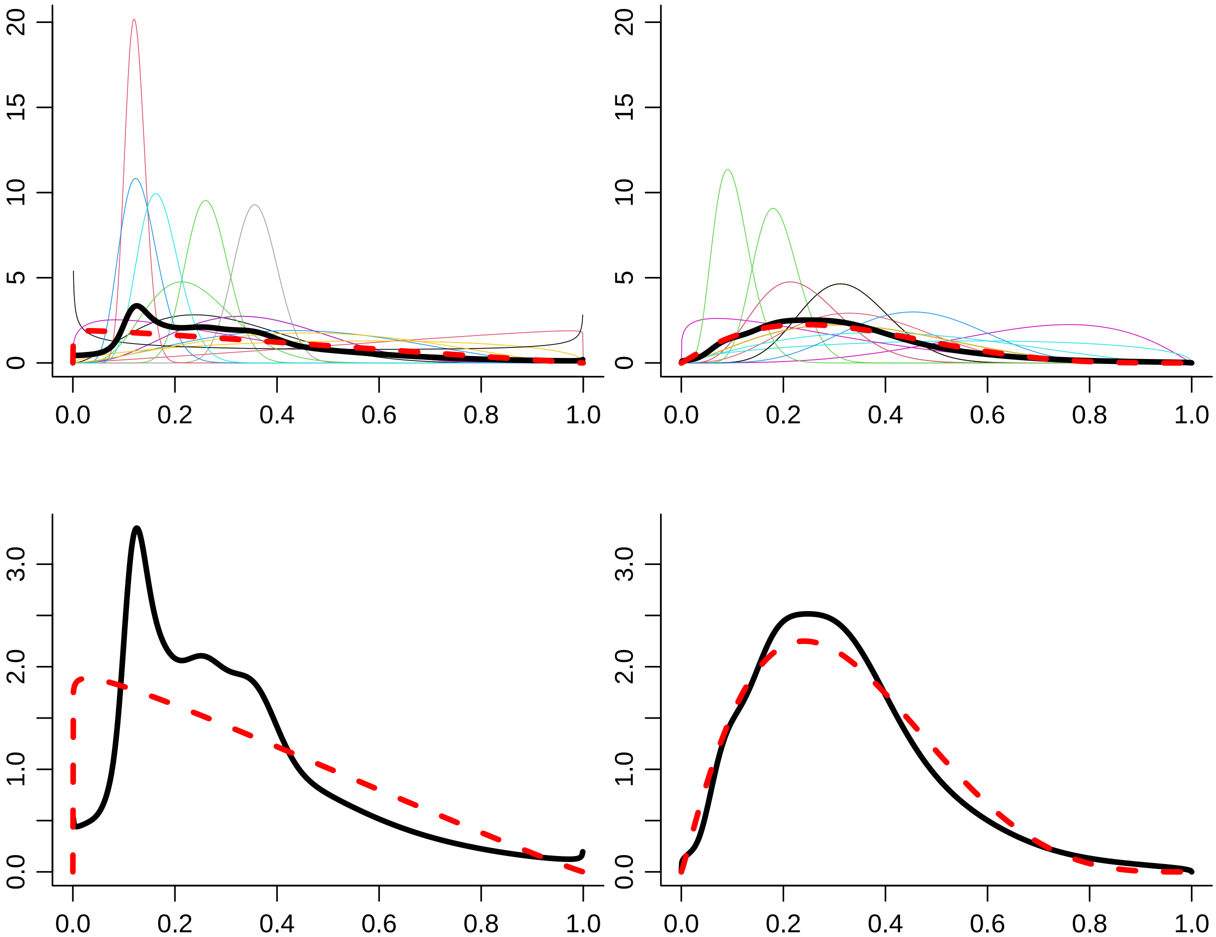


Figure 17: Experts' answers transformed to probability densities for question 3 (undercount of emigration of non-EU+ nationals who leave low undercount countries). Left column corresponds to Round 1 and right column corresponds to Round 2. Top row are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2) and bottom row are the two aggregated densities on their own.

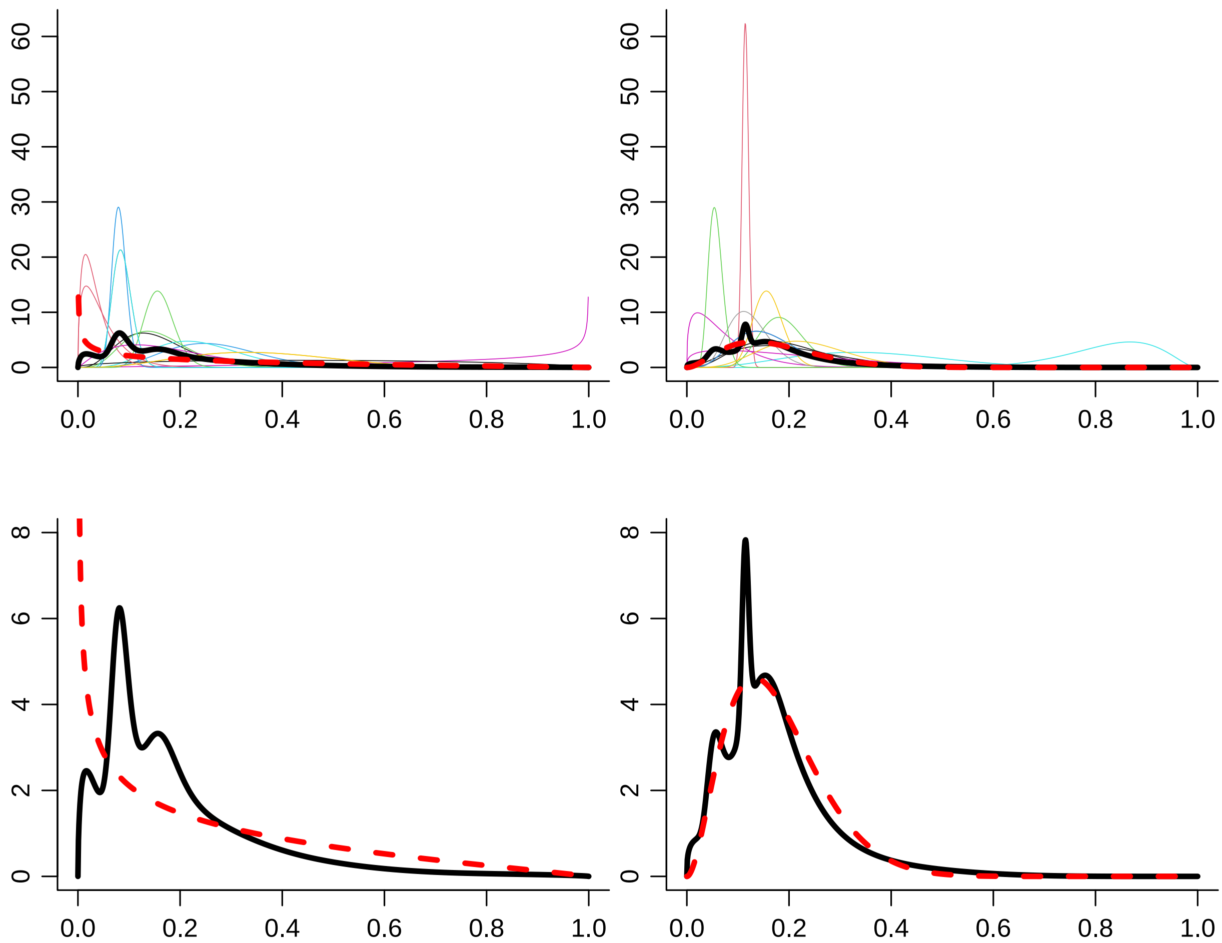


Figure 18: Experts' answers transformed to probability densities for question 4 (undercount of immigration of non-EU+ nationals who enter low undercount countries). Left column corresponds to Round 1 and right column corresponds to Round 2. Top row are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2) and bottom row are the two aggregated densities on their own.

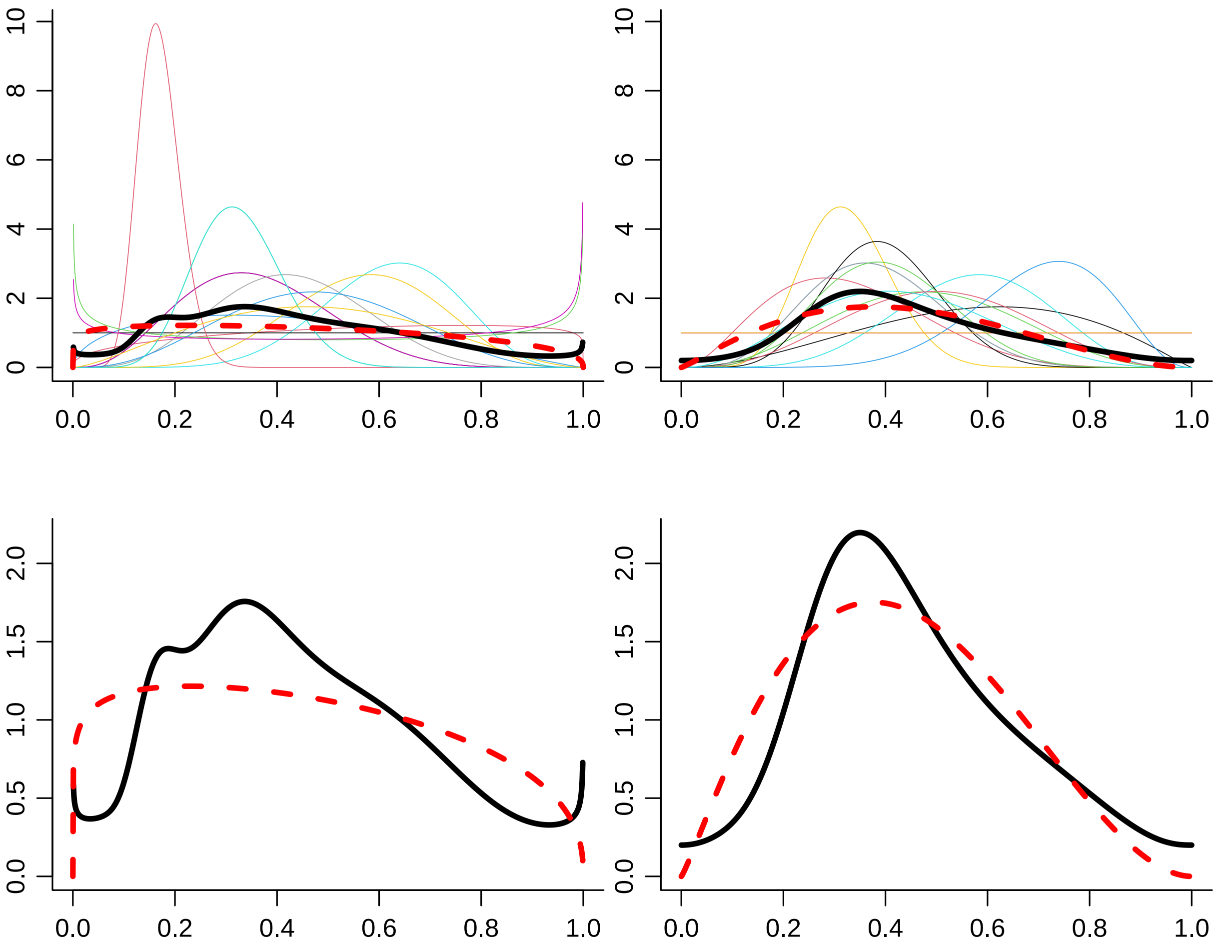


Figure 19: Experts' answers transformed to probability densities for question 5 (undercount of emigration of EU+ nationals who leave high undercount countries). Left column corresponds to Round 1 and right column corresponds to Round 2. Top row are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2) and bottom row are the two aggregated densities on their own.

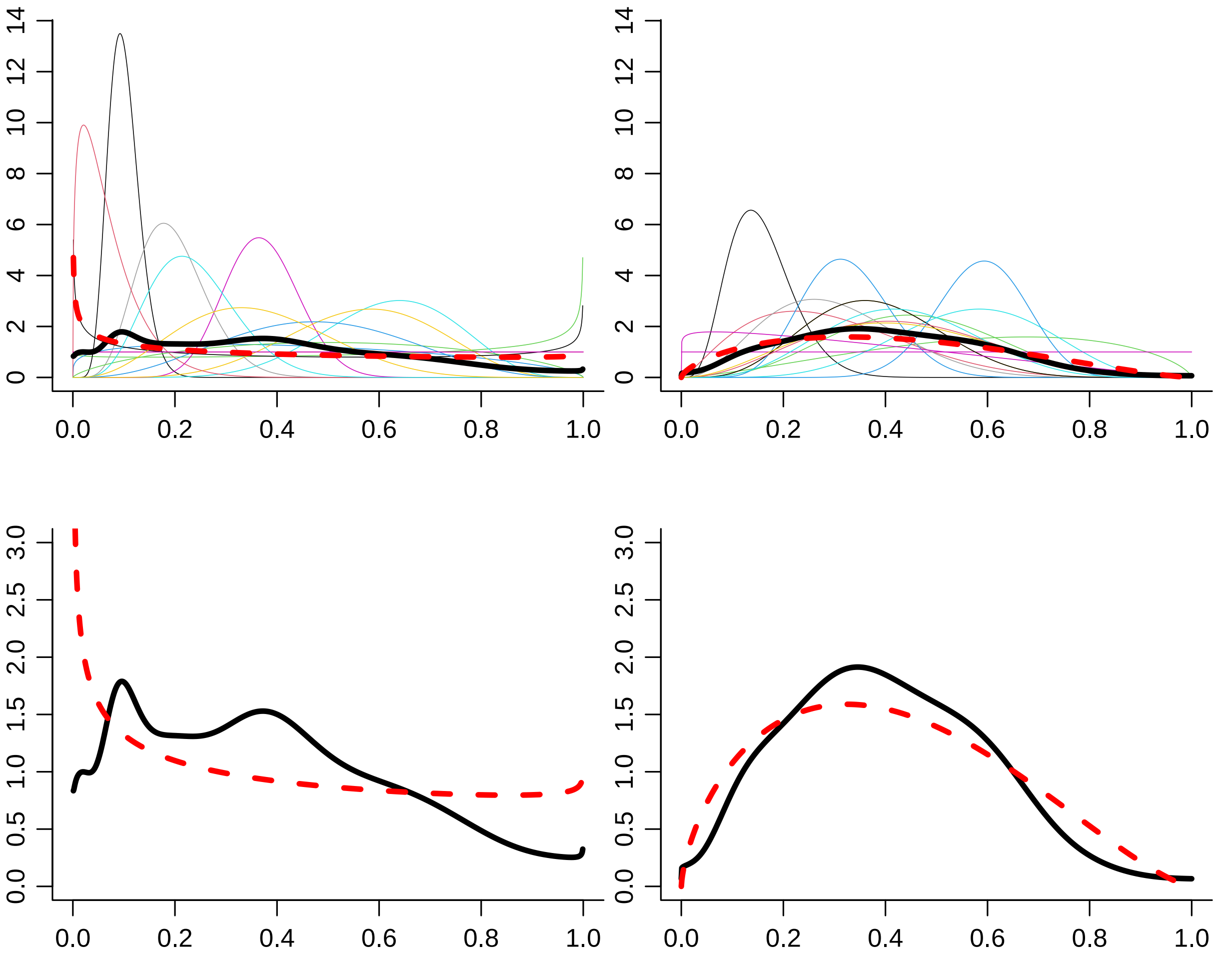


Figure 20: Experts' answers transformed to probability densities for question 6 (undercount of immigration of EU+ nationals who enter high undercount countries). Left column corresponds to Round 1 and right column corresponds to Round 2. Top row are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2) and bottom row are the two aggregated densities on their own.

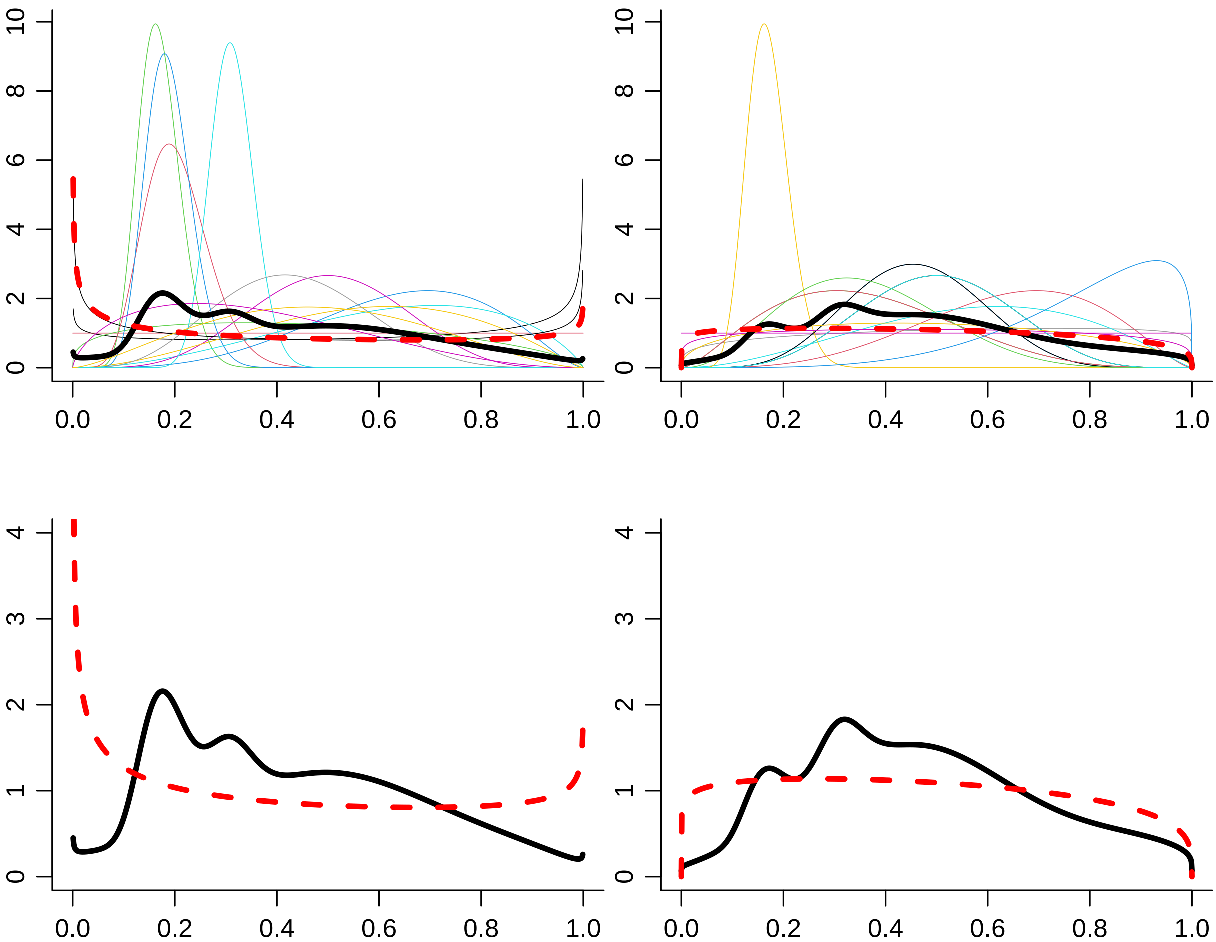


Figure 21: Experts' answers transformed to probability densities for question 7 (undercount of emigration of non-EU+ nationals who leave high undercount countries). Left column corresponds to Round 1 and right column corresponds to Round 2. Top row are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2) and bottom row are the two aggregated densities on their own.

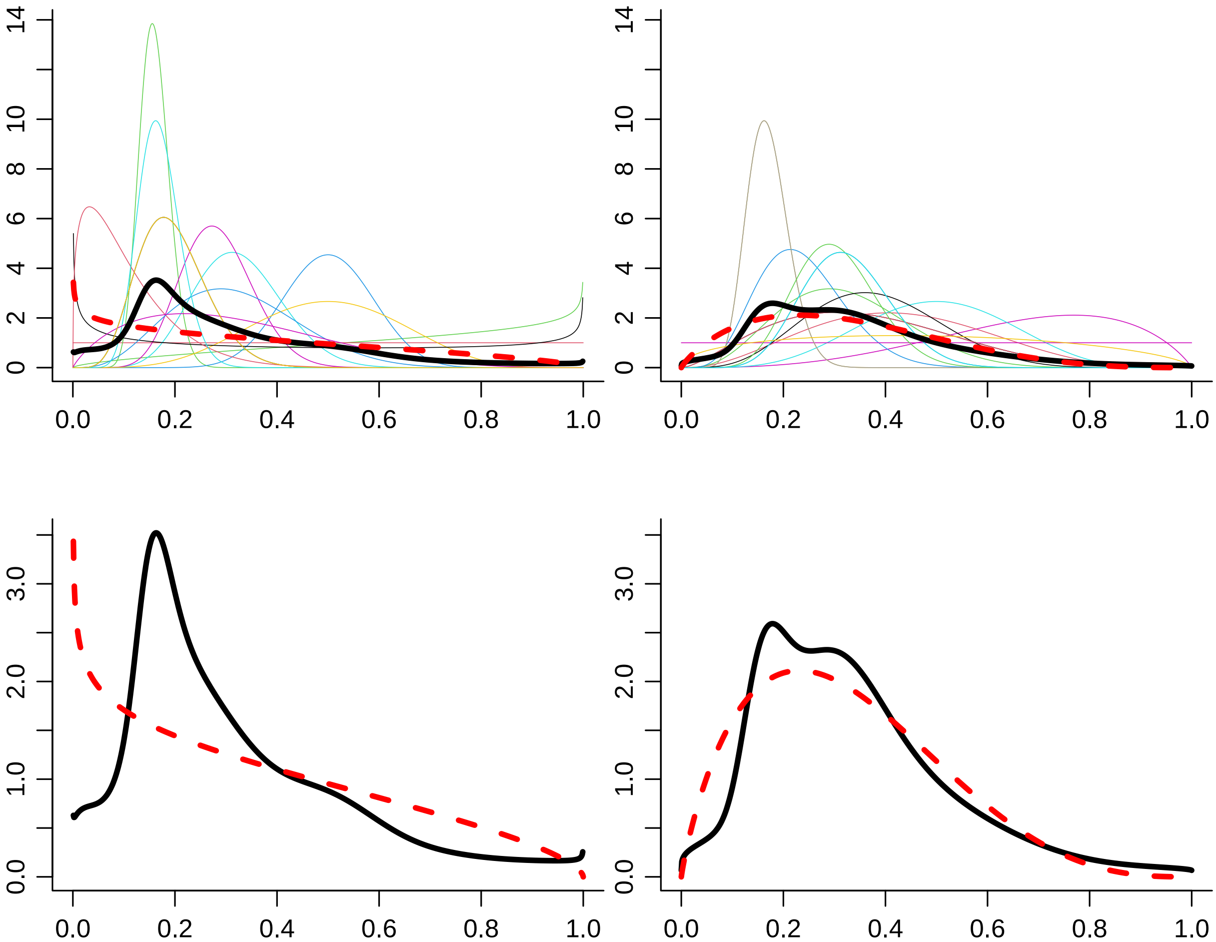


Figure 22: Experts' answers transformed to probability densities for question 8 (undercount of immigration of non-EU+ nationals who enter high undercount countries). Left column corresponds to Round 1 and right column corresponds to Round 2. Top row are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2) and bottom row are the two aggregated densities on their own.

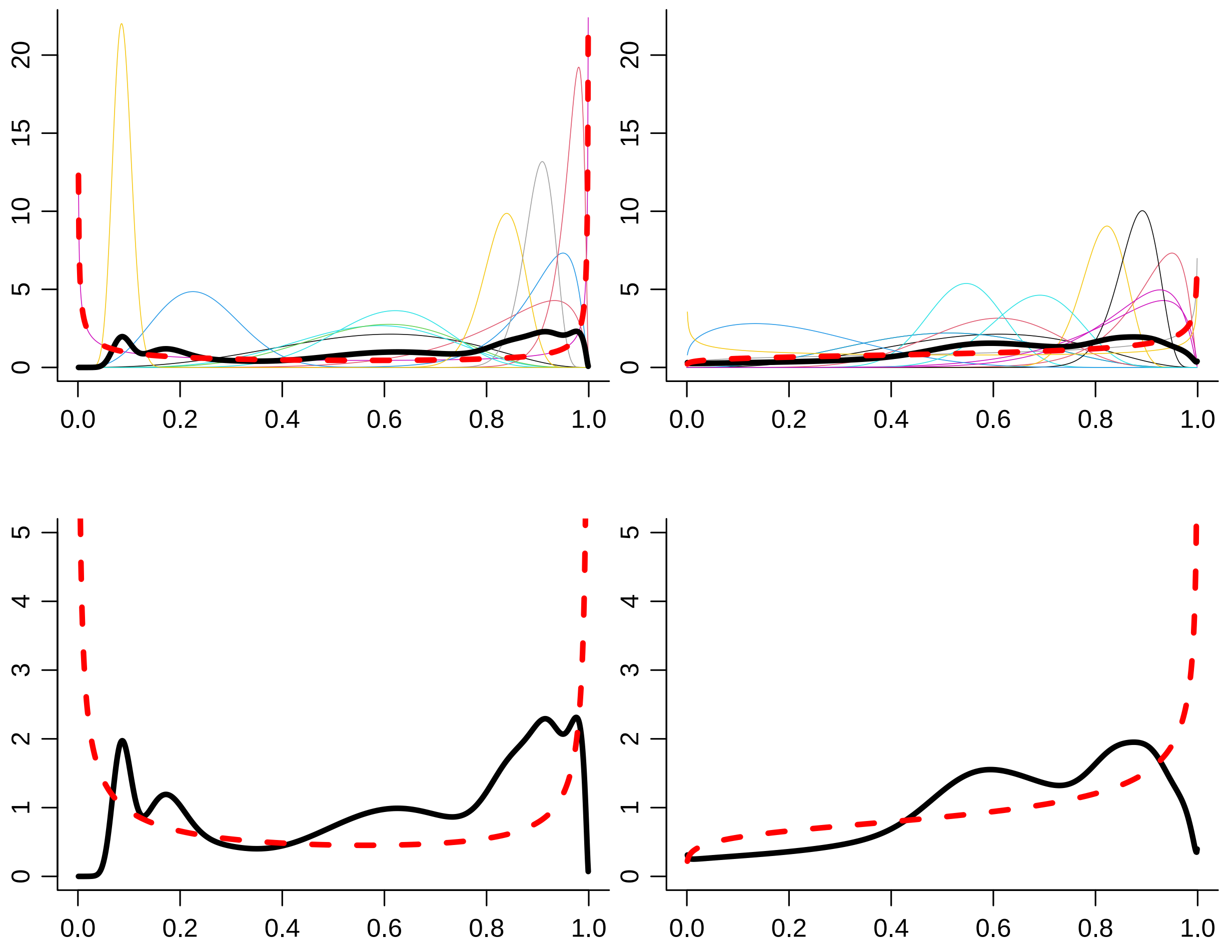


Figure 23: Experts' answers transformed to probability densities for question 9 (accuracy of emigration of EU+ nationals who leave countries recording with a register). Left column corresponds to Round 1 and right column corresponds to Round 2. Top row are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2) and bottom row are the two aggregated densities on their own.

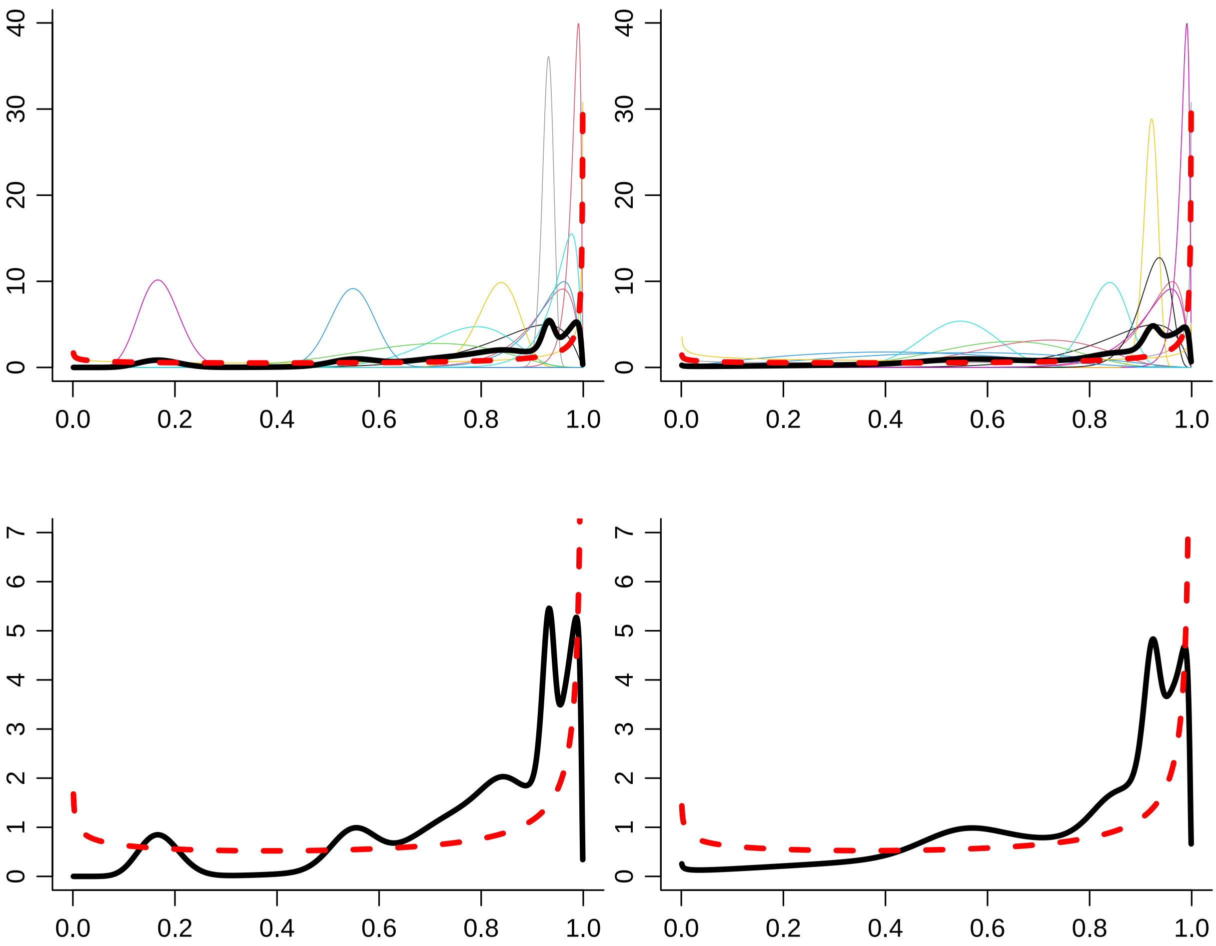


Figure 24: Experts' answers transformed to probability densities for question 10 (accuracy of immigration of EU+ nationals who enter countries recording with a register). Left column corresponds to Round 1 and right column corresponds to Round 2. Top row are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2) and bottom row are the two aggregated densities on their own.

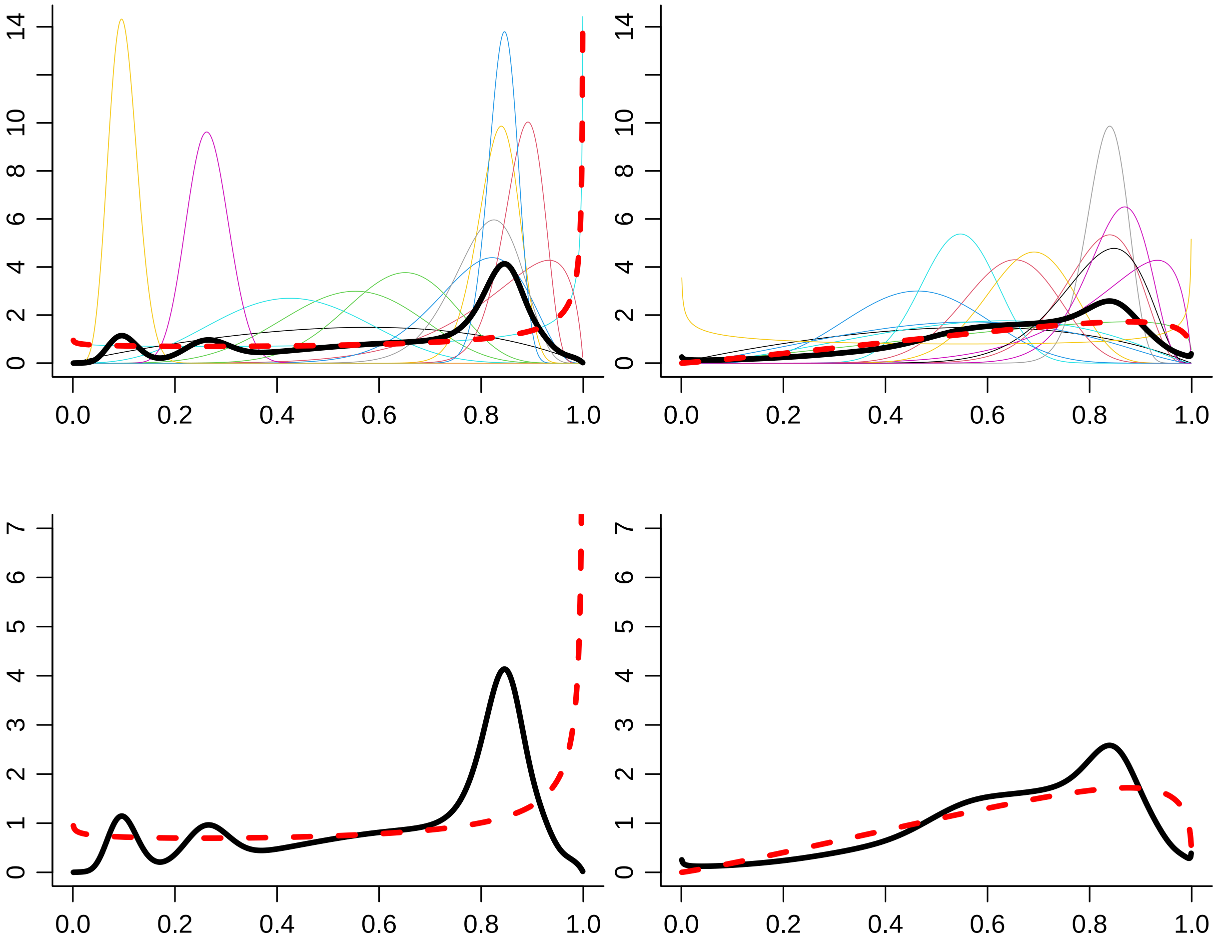


Figure 25: Experts' answers transformed to probability densities for question 11 (accuracy of emigration of non-EU+ nationals who leave countries recording with a register). Left column corresponds to Round 1 and right column corresponds to Round 2. Top row are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2) and bottom row are the two aggregated densities on their own.

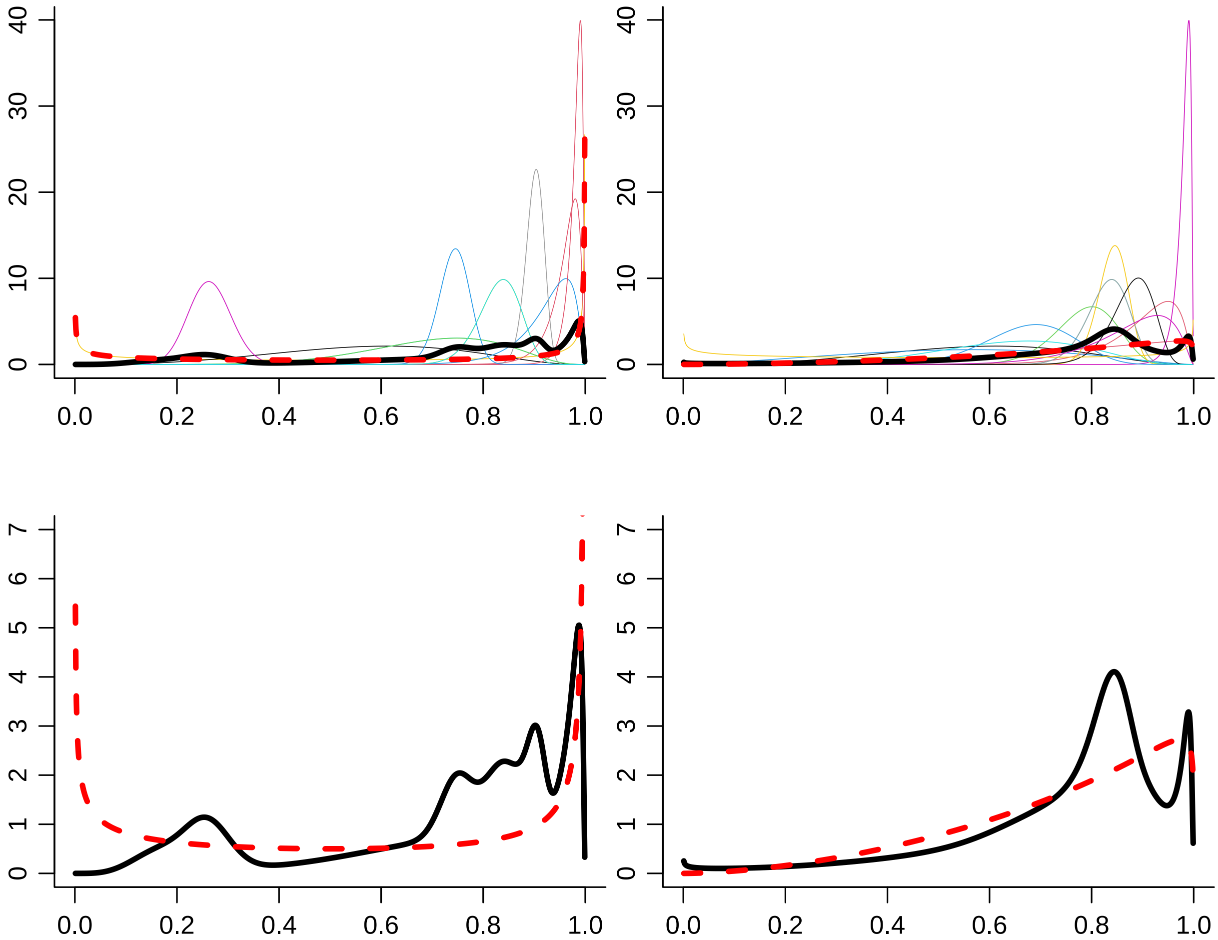


Figure 26: Experts' answers transformed to probability densities for question 12 (accuracy of immigration of non-EU+ nationals who enter countries recording with a register). Left column corresponds to Round 1 and right column corresponds to Round 2. Top row are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2) and bottom row are the two aggregated densities on their own.

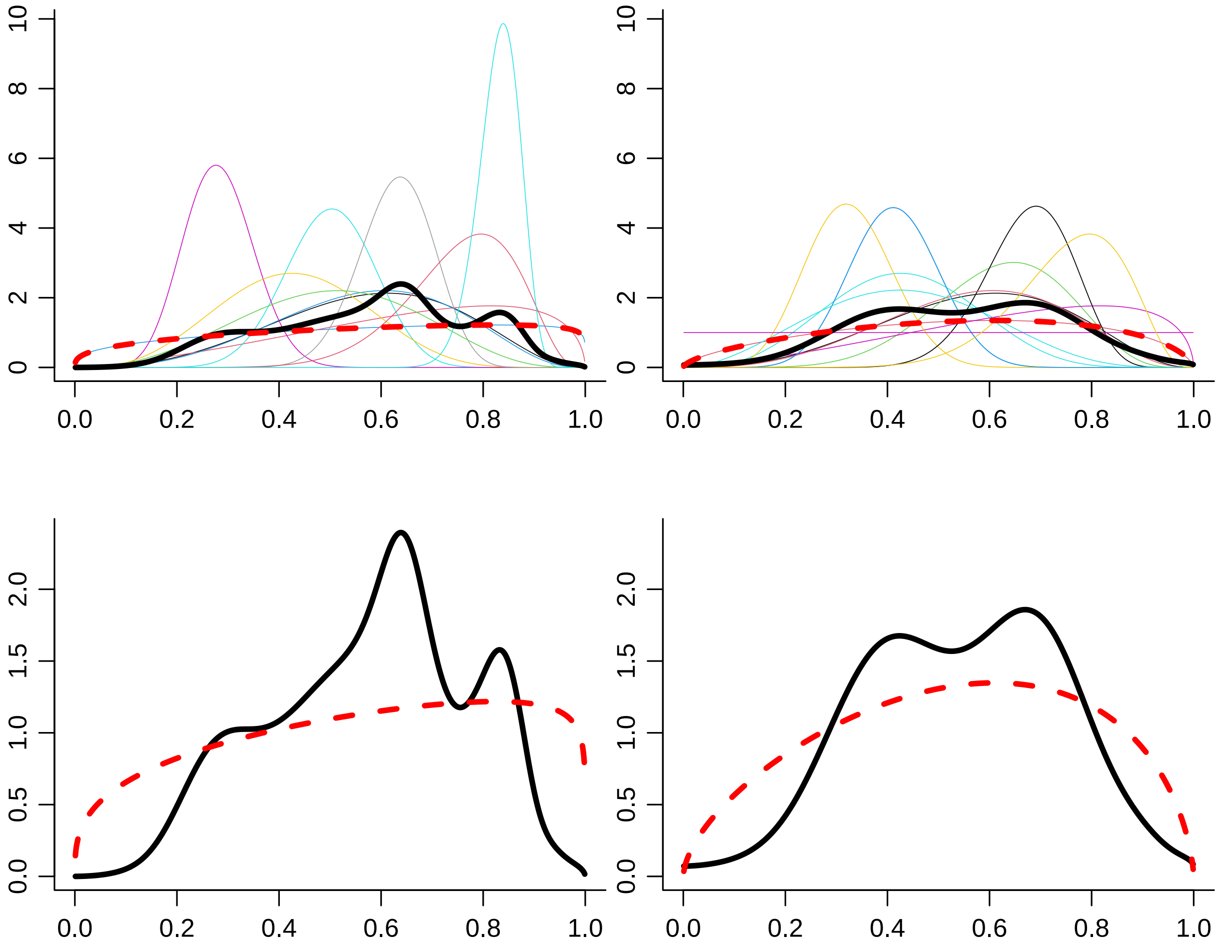


Figure 27: Experts' answers transformed to probability densities for question 13 (accuracy of immigration of EU+ nationals who enter countries recording with a survey). Left column corresponds to Round 1 and right column corresponds to Round 2. Top row are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2) and bottom row are the two aggregated densities on their own.

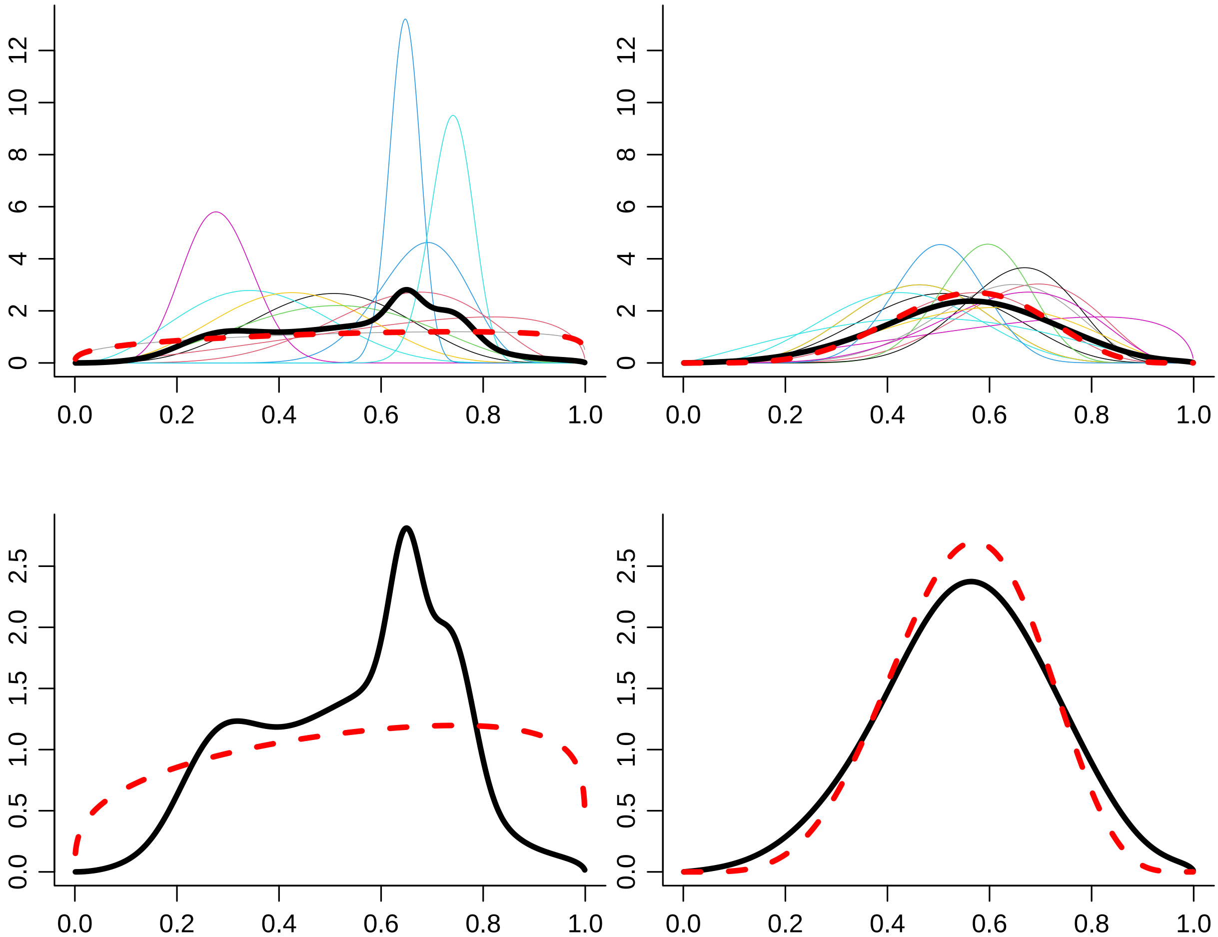


Figure 28: Experts' answers transformed to probability densities for question 14 (accuracy of immigration of non-EU+ nationals who enter countries recording with a survey). Left column corresponds to Round 1 and right column corresponds to Round 2. Top row are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2) and bottom row are the two aggregated densities on their own.

4 Discussion

Overall, the prior distributions resulting from our elicitation procedure were informative or at least weakly informative. For the undercount questions, the prior distributions were more informative for immigration, compared to emigration, and for low undercount, compared to high undercount. For example, the prior distiribution of the parameter associated with question 4 (undercount of immigration of non-EU+ nationals who enter low undercount countries; see figure 18) was the most informative from all the undecount questions, whereas the prior distiribution of the parameter associated with question 7 (undercount of emigration of non-EU+ nationals who leave high undercount countries; see figure 21) the least informative. For the accuracy questions, the prior densities for register countries had a lot of their mass near 1, as opposed to the densities for survey countries, whose mass was assigned mostly at values around 0.5. No clear distinction could be made as to whether the register or the survey densities were more informative.

In our conducted elicitation process we did not use any visual and interactive tools, such as MATCH (Morris et al. (2014)). Instead, we used the feedback documents to provide the experts with a visual illustration of their answers, between the two rounds of the Delphi questionnaire. Looking back at our conducted elicitation process, we believe that it could have significantly benefited from the use of a visual and interactive tool. For example, we believe that some of the implausible densities we encountered, such as the U-shaped densities or some very highly peaked densities, would not be encountered if the experts could see their densities while providing their answers. In addition, the use of a visual and interactive elicitation tool would also help with avoiding invalid answers, since an expert would instantly see (and be informed by the tool) that their answer can not be translated to a density. Although the Dephi structure of two rounds, along with the feedback we provided between the rounds, helped us alleviate a lot of these issues, we came to the realization that such issues might have been avoided altogether had we used a visual elicitation tool.

Our suggestion is that similar future studies should be conducted using visual and interactive elicitation tools. The great advantage of these tools is that, being visual, they allow the expert to see their density as they provide an answer, and, being interactive, they allow the expert to see how this density changes as their answer changes. As a result, such tools can avoid the possible confusion relating to how answers translate to probability statements and densities.

References

Bastin, L., Cornford, D., Jones, R., Heuvelink, G. B., Pebesma, E., Stasch, C., Nativi, S., Mazzetti, P., and Williams, M. (2013). Managing uncertainty in integrated environmental modelling: The UncertWeb framework. *Environmental Modelling & Software*, 39:116–134.

Gosling, J. P. (2017). SHELF: The sheffield elicitation framework. In *Elicitation*, pages 61–93. Springer International Publishing.

Morris, D. E., Oakley, J. E., and Crowe, J. A. (2014). A web-based tool for eliciting probability distributions from experts. *Environmental Modelling & Software*, 52:1–4.

R Core Team (2020). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Raymer, J., Wiśniowski, A., Forster, J. J., Smith, P. W. F., and Bijak, J. (2013). Integrated modeling of european migration. *Journal of the American Statistical Association*, 108(503):801–819.

Wiśniowski, A., Bijak, J., Christiansen, S., Forster, J. J., Keilman, N., Raymer, J., and Smith, P. W. (2013). Utilising expert opinion to improve the measurement of international migration in europe. *Journal of Official Statistics*, 29(4):583–607.

A Expressions for g and g^{-1}

As in the main document, for a given question, we will use X to denote the quantity of interest and $g(X)$ to denote the quantity for which the question asked information for. Recall that the assumption is that $g(X)$ takes values in $[0, 1]$ and g is a continuous, strictly monotone, deterministic function. The assumptions of continuity and strict monotinicity ensure the existence of g^{-1}, the inverse function of g, a function which is needed to translate expert answers to points on the cdf of X (see section 2.1.1). Note also that, just like X, $g(X)$ is a random variable and that, for clarity reasons, whenever we refer to g as a deterministic function of x, I will use the notation $g(x)$, making a distinction from the notation $g(X)$.

A.1 Questions 1-8 (undercount questions)

The parameter of interest X in questions 1-8 represents the proportion of true flow counts that are reported by a given country. More specifically, ignoring the effect of the other measurement error parameters, under our assumed modelling framework (see Raymer et al. (2013)), we can describe the effect of the parameter X via the relationship $\mu = yX$, where μ can be thought of as representing a reported flow count and y is the corresponding true flow count. As mentioned in the main text the choice of family for the distribution of X was Beta(α, β). Questions 1-8 asked the experts to provide information for $g(X) = 1 - X$, as opposed to X directly. That is, questions 1-8 asked for the proportion of true flow counts that are not reported by a given country, i.e. for the proportion of undercount associated with the reporting system of a given country. This was done to avoid causing confusion between high and low undercount questions, since for $g(X)$ high (low) values imply higher (lower) undercount, whereas for X, it is the other way around.

It is easy to see that $g(X) = 1-X$ takes values in $[0, 1]$ and that the function $g(x) = 1-x$, is continuous and strictly decreasing and, therefore, invertible. It is also straighforward to see that the inverse of g, g^{-1}, is given by the expression $g^{-1}(r) = 1 - r$. Equation A.1

below collects the expressions for g and g^{-1} for questions 1-8:

$$\begin{aligned} g(x) &= 1 - x \\ g^{-1}(r) &= 1 - r \end{aligned} \tag{A.1}$$

A.2 Questions 9-14 (accuracy questions)

The parameter of interest X in questions 9-14 is the precision (inverse variance) of the random fluctuation error term associated with the measurement of migration. More precisely, assuming that there are no sources of systemaric error, under our assumed modelling framework (see Raymer et al. (2013)), we can describe the effect of the parameter X via the relationship $\mu = y \exp(\varepsilon|_X)$, where μ represents a reported flow count, y is the corresponding true flow count, and $\varepsilon|_X \sim \mathrm{N}(0, 1/X)$, where $\mathrm{N}(m, v)$ denotes a Normal distribution with mean m and variance v. As mentioned in the main text the choice of family for the distribution of X was $\mathrm{Gamma}(\nu, \rho)$. Asking the experts to provide information for X directly would be too involved (due to the fact that there are two levels of randomness at work here, that of $\varepsilon|_X$ and that of X itself) and so we asked the experts to provide information for $g(X) = \mathrm{E}(\mathbb{1}_A \mid X)$, where $A = \{0.95 < \exp(\varepsilon|_X) < 1.05\}$. That is, questions 9-14 asked for the expected proportion of observations with less than 5% error, i.e. for the expected proportion of times that the published statistics of a given country are within an interval of -5% to $+5\%$ compared to the true level of immigration, when no systematic errors exist.

Mathematically, $g(X) = \mathrm{E}(\mathbb{1}_A \mid X)$ is the conditional expectation of $\mathbb{1}_A$ given X, a random variable (as a function of the random variable X), and it is easy to see how it takes values in $[0, 1]$. It is also easy to see that, as a function of x, $g(x) = \mathrm{E}(\mathbb{1}_A \mid X = x)$ is strictly increasing, since the higher the value of the precision x, the higher the expected proportion of observations with less than 5% error. Now, $g(x) = \mathrm{E}(\mathbb{1}_A \mid X = x) = P(A \mid X = x) = P(0.95 < \exp(\varepsilon|_X) < 1.05 \mid X = x) = \Phi(\log(1.05)\sqrt{x}) - \Phi(\log(0.95)\sqrt{x})$, where Φ is the cdf of the standard Normal distribution $\mathrm{N}(0, 1)$. The last equality reveals that $g(x)$ is also continuous, being the difference of continuous functions. Therefore, $g(x)$ is invertible. To find the inverse of g, g^{-1}, we must solve $g(x) = r$ with respect to x. Starting from the last of the above equalities for $g(x)$, that of $g(x) = \Phi(\log(1.05)\sqrt{x}) - \Phi(\log(0.95)\sqrt{x})$, and using the approximation $\log(1.05) \approx -\log(0.95) \approx 0.05$ we solve $g(x) = r$ with respect to x and reveal that $x = 400(\Phi^{-1}(\frac{r+1}{2}))^2$, where Φ^{-1} is the inverse of Φ. Hence, the expression for g^{-1} is $g^{-1}(r) = 400(\Phi^{-1}(\frac{r+1}{2}))^2$. Equation A.1 below collects the expressions for g and g^{-1} for questions 9-14:

$$\begin{aligned} g(x) &= \mathrm{E}(\mathbb{1}_A \mid X = x) \\ g^{-1}(r) &= 400(\Phi^{-1}(\frac{r+1}{2}))^2 \end{aligned} \tag{A.2}$$

B Starting values for the minimization algorithms

As in the main document, consider any one of the questions 1-14 and let X denote the quantity of interest. Let also an expert be denoted by i, $i = 1, 2, ..., n$, where n the number of experts providing a valid answer to the given question.

B.1 Questions 1-8 (undercount questions)

In questions 1-8, the choice of family for the distribution of X was Beta(α, β). As explained in the main text, for the purposes of method 1 (see section 2.1) the minimization algorithm is run to specify the values of expert-specific parameters $\theta_i = (\alpha_i, \beta_i)$, $i = 1, 2, ..., n$, whereas for the purposes of method 2 (see section 2.2) it is run to specify the value of a common (over experts) parameter $\theta = (\alpha, \beta)$. For all runs, under both methods, we set the starting values, say (α_0, β_0), as $(\alpha_0, \beta_0) = (1, 1)$. Note that, under these values, the Beta distribution reduces to a uniform distribution on $[0, 1]$, denoted as U$[0, 1]$, i.e. Beta$(1, 1) \equiv$ U$[0, 1]$. These starting values were very robust and worked well for all the runs of the minimization algorithm.

B.2 Questions 9-14 (accuracy questions)

In questions 9-14, the choice of family for the distribution of X was Gamma(ν, ρ). As mentioned in the main text, in method 1 (see section 2.1) the minimization algorithm is run to specify the values of expert-specific parameters $\theta_i = (\nu_i, \rho_i)$, $i = 1, 2, ..., n$, whereas in method 2 (see section 2.2) it is run to specify the value of a common (over experts) parameter $\theta = (\nu, \rho)$. Under both methods, to specify reasonable starting values for the minimization algorithm, we followed Wiśniowski et al. (2013) and approximated the Gamma distribution of X with a Log-normal distribution. We use the notation LogN(κ, σ^2) to denote a Log-normal distribution parameterized by κ and σ^2 so that if $Y \sim$ LogN(κ, σ^2) then $\log(Y) \sim$ N(κ, σ^2). The motivation behind the Log-normal approximation is that the parameters of a Log-normal distribution, unlike the parameters of a Gamma distribution, can be analytically calculated, given two points on the associated cdf. These calculated values, of the parameters of the LogNormal, can then be used to set reasonable starting values for the run of the minimization algorithm aimed to specify the parameters of the Gamma. Below we describe how this is done, first for the purposes of method 1, and subsequently for the purposes of method 2.

In method 1 the minimization algorithm is run to specify the values of the expert-specific parameters $\theta_i = (\nu_i, \rho_i)$, $i = 1, 2, ..., n$. Same as the main text, let $(x_{i,1}, y_{i,1})$ and $(x_{i,2}, y_{i,2})$ be the two points of the cdf of X, corresponding to expert i, where $x_{i,1} = g^{-1}(r_{i,1})$, $x_{i,2} = g^{-1}(r_{i,2})$, $y_{i,1} = k_i r_{i,1}$ and $y_{i,2} = 1 - k_i(1 - r_{i,2})$, and where $r_{i,1}, r_{i,2}$ are the range values and c_i the certainty value, provided by i, $i = 1, 2, \ldots, n$. Under the assumption that $X \sim \text{LogN}(\kappa_i, \sigma_i^2)$, given $(x_{i,1}, y_{i,1})$ and $(x_{i,2}, y_{i,2})$, we analytically solve the system of equations

$$\begin{aligned} P(X < x_{i,1}) &= y_{i,1} \\ P(X < x_{i,2}) &= y_{i,2}, \end{aligned} \tag{B.1}$$

and calculate σ_i and κ_i to be $\sigma_i = \frac{\log(x_{i,2}) - \log(x_{i,1})}{\Phi^{-1}(y_{i,2}) - \Phi^{-1}(y_{i,1})}$ and $\kappa_i = \log(x_{i,1}) - \sigma_i \Phi^{-1}(y_{i,1})$. Having σ_i and κ_i, we calculate the mean m_i and variance v_i of the approximating LogNormal distribution as $m_i = \exp(\kappa_i + \sigma_i^2/2)$ and $v_i = (\exp(\sigma_i^2) - 1)\exp(2\kappa_i + \sigma_i^2)$. Finally, using the calculated m_i and v_i, we conduct a method of moments (MOM) estimation to calculate parameters for the Gamma distribution, which we set as the starting values, $(\nu_{i,0}, \rho_{i,0})$, for the minimization algorithm of $\theta_i = (\nu_i, \rho_i)$. These are given by $\nu_{i,0} = m_i^2/v_i$ and $\rho_{0,i} = m_i/v_i$, $i = 1, 2, \ldots, n$.

In method 2 the minimization algorithm aims to specify the value of a common (over experts) parameter $\theta = (\nu, \rho)$. As above, let $(x_{i,1}, y_{i,1})$ and $(x_{i,2}, y_{i,2})$ be the two points of the cdf of X, corresponding to expert i, where $x_{i,1} = g^{-1}(r_{i,1})$, $x_{i,2} = g^{-1}(r_{i,2})$, $y_{i,1} = k_i r_{i,1}$ and $y_{i,2} = 1 - k_i(1 - r_{i,2})$, and where $r_{i,1}, r_{i,2}$ are the range values and c_i the certainty value, provided by i, $i = 1, 2, \ldots, n$. The intention is to work as in method 1 and use a pair of points of the cdf of X and analytically calculate the parameters of the LogNormal. However, method 2, unlike method 1, involves all $2n$ points provided by experts, and not just an expert-specific pair of points. To overcome this, we construct an artificial pair of points, representing all experts, by taking a sort of average over all points. Specifically, we consider the pair of points $(\bar{x}_1, \bar{y}_1)$ and $(\bar{x}_2, \bar{y}_2)$, where $\bar{x}_1 = \frac{1}{n}\sum_{i=1}^{n} x_{i,1}$, $\bar{y}_1 = \frac{1}{n}\sum_{i=1}^{n} y_{i,1}$,

$\bar{x}_2 = \frac{1}{n}\sum_{i=1}^{n} x_{i,2}$ and $\bar{y}_2 = \frac{1}{n}\sum_{i=1}^{n} y_{i,2}$. We then proceed as in method 1. Specifically, under the assumption that $X \sim \text{LogN}(\kappa, \sigma^2)$, we solve the system of equations

$$\begin{aligned} P(X < \bar{x}_1) &= \bar{y}_1 \\ P(X < \bar{x}_2) &= \bar{y}_2, \end{aligned} \tag{B.2}$$

to find $\sigma = \frac{\log(\bar{x}_2) - \log(\bar{x}_1)}{\Phi^{-1}(\bar{y}_2) - \Phi^{-1}(\bar{y}_1)}$ and $\kappa = \log(\bar{x}_1) - \sigma\Phi^{-1}(\bar{x}_1)$. We then calculate the mean m and variance v of the approximating LogNormal distribution as $m = \exp(\kappa + \sigma^2/2)$ and $v = (\exp(\sigma^2) - 1)\exp(2\kappa + \sigma^2)$. Lastly, we conduct a MOM estimation to calculate parameters for the Gamma distribution, which we set as the starting values, (ν_0, ρ_0), for the minimization algorithm of $\theta = (\nu, \rho)$. These are given by $\nu_0 = m^2/v$ and $\rho_0 = m/v$.

C Example expert-specific feedback document

Delphi round 1: respondent-specific feedback for respondent id=8533066

Document description

This document provides plots and feedback, for the answers to questions 1-14 of round 1 of the Delphi questionnaire. Section 1 repeats the generic (not respondent-specific) feedback that was provided in the generic feedback document. Sections 2 and 3 provide plots and feedback that are specific to **respondent id=8533066**. Recall that questions 1-8 regarded undercount parameters, questions 9-14 accuracy parameters, and note that 16 participants took part in the questionnaire.

Section 1

Overall, there was a fair amount of heterogeneity in the opinions of respondents for the undercount parameters, although for some questions (e.g. question 2 and 4) the amount of heterogeneity was less than others (e.g. question 5 and 6). Some respondents provided answers with very high certainty while others were much less certain.

The opinions of respondents regarding accuracy, were again quite diverse, although in some questions (e.g. questions 9 and 11) the respondents were loosely divided into those that suggested that accuracy is very poor and to those that suggested that accuracy is very good. An important reminder is that for the accuracy questions, we assume that there is **no systematic bias in the measurement of migration**, i.e. we assume that there is no bias due to undercount, coverage, duration criteria or any other factor, and we are asking for the accuracy related only to sampling variability. That is, we assume that the only source of variability is sampling variability.

Something that we should had made clearer before the first round, was **how the answers are translated to probability statements**. Consider any one of the question 1-14 and suppose that the quantity in question is denoted by x (for questions 1-8 x denotes undercount whereas for questions 9-14 x denotes accuracy). Say that a respondent provides a range in (a) of r_1 to r_2 and provides certainty in (b) of c. What these translate to, is to the respondent saying that the probability that x lies within the provided range $[r_1, r_2]$, is equal to c. Notice though, that the former probability statement automatically implies that the probability that x lies outside of the provided range $[r_1, r_2]$ must be equal to $1 - c$. For example, if a respondent provides a range in (a) of 10% to 30% (i.e. $r_1 = 0.1$ and $r_2 = 0.3$) and provides certainty in (b) of 75% (i.e. $c = 0.75$), then this answer translates to the respondent saying that the probability that x lies within the provided range $[0.1, 0.3]$ is equal to 0.75, and, simultaneously, that the probability that x lies outside the provided range $[0.1, 0.3]$ (i.e. that x lies within $[0, 0.1]$ or $[0.3, 1]$) is equal to $1 - 0.75 = 0.25$. It has to be noted that these two statements are equivalent, and that giving one as an answer also implies the other.

Based on some of the answers of the questionnaire we believe that some of the answers with small certainty/probability in (b) were perhaps given without fully realizing that low certainty/probability for x to lie within $[r_1, r_2]$ automatically implies high certainty/probability for x to lie outside of $[r_1, r_2]$. Similarly, some of the answers with very high certainty/probability in (b) left very small certainty/probability for x to lie outside of $[r_1, r_2]$.

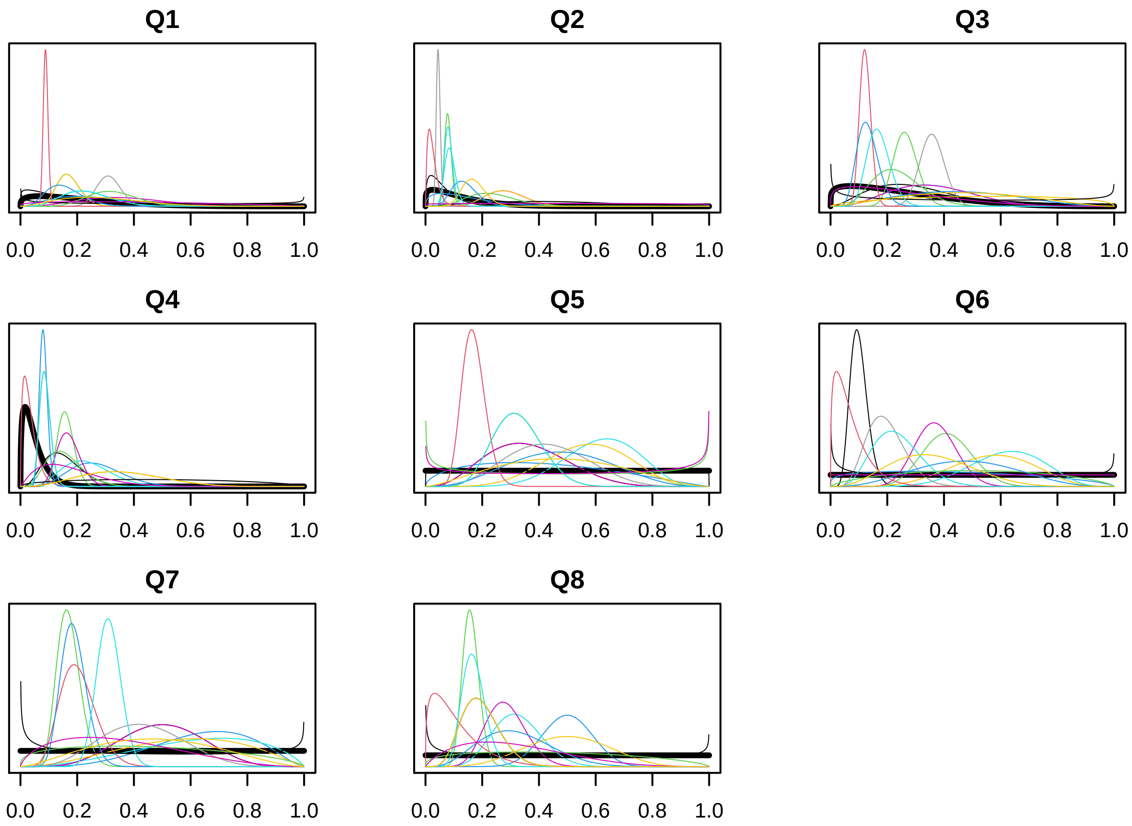
Other notable issues were answers of negative ranges, answers providing a single number instead of a range and answers not providing certainty assessments. Such answers were impossible to be translated to probability statements.

Section 2: Density curves of your answers (thick black curves) alongside the other respondents' density curves

Section 2 consists of 14 plots, one for each question. Each of the 14 plots presents your answer, alongside the answers of the other respondents, all translated to probability distributions, or more precisely to probability density curves. Your curve is highlighted as the thick black curve. The intention of these plots is to allow you to compare your answer with the answers provided by the other respondents, establish a reference, and perhaps consider revising your answer in round 2.

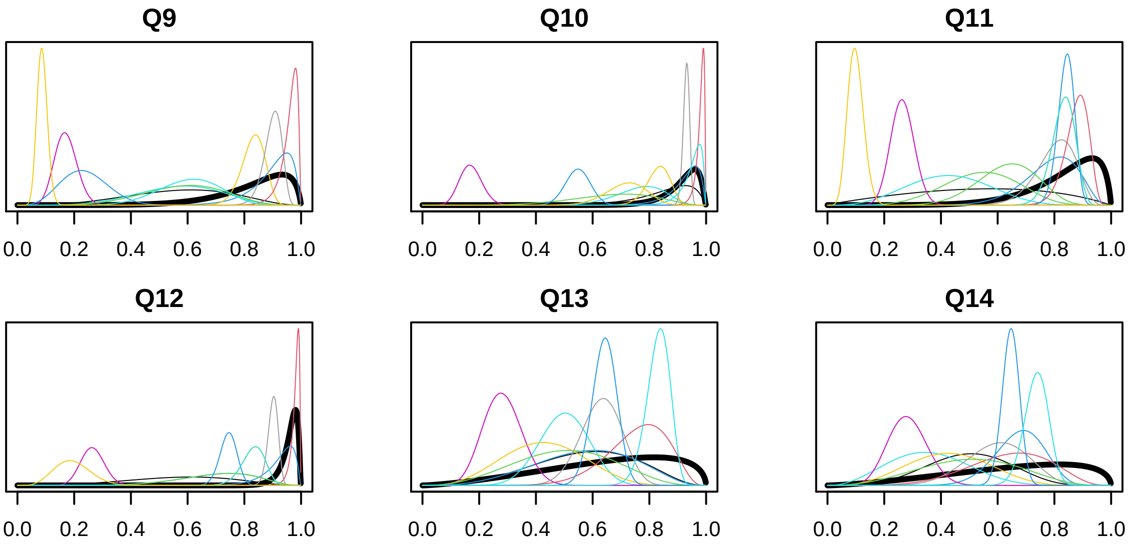
Plots: undercount questions

For reference:
Q1: country=sending, migrants=EU+, undercount=low
Q2: country=receiving, migrants=EU+, undercount=low
Q3: country=sending, migrants=non-EU+, undercount=low
Q4: country=receiving, migrants=non-EU+, undercount=low
Q5: country=sending, migrants=EU+, undercount=high
Q6: country=receiving, migrants=EU+, undercount=high
Q7: country=sending, migrants=non-EU+, undercount=high
Q8: country=receiving, migrants=non-EU+, undercount=high



Plots: accuracy questions

For reference:
Q9: country=sending, migrants=EU+, recording=register
Q10: country=receiving, migrants=EU+, recording=register
Q11: country=sending, migrants=non-EU+, recording=register
Q12: country=receiving, migrants=non-EU+, recording=register
Q13: country=receiving, migrants=EU+, recording=survey
Q14: country=receiving, migrants=non-EU+, recording=survey

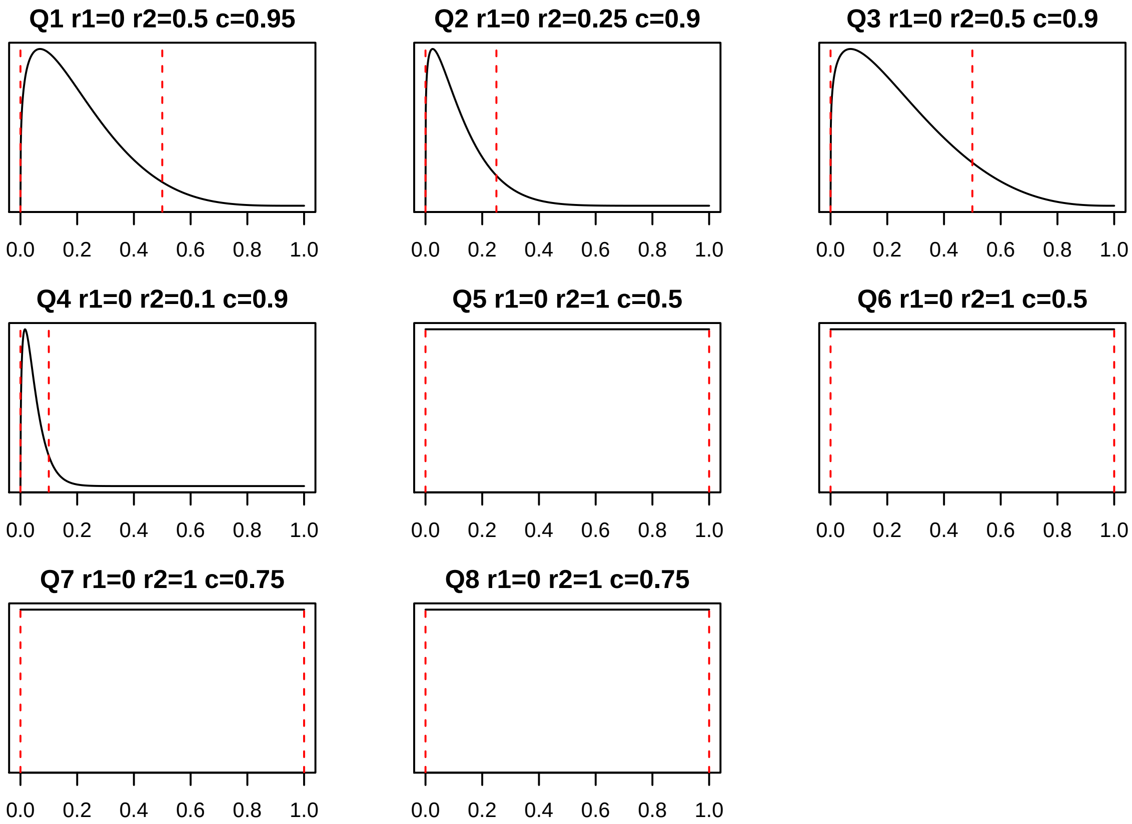


Section 3: Density curves of your responses on their own

Section 3, just like Section 2, consists of 14 plots, one for each question. Here, the plots present your answer (again translated to a density curve), only this time on its own and not alongside the answers of other respondents. Please note that the y-axis (vertical) scale is different in these plots than the plots of Section 2. The different scale helps to better illustrate your densities. The plots are followed by feedback that is specific to you. The intention of these plots is to explain how your answers were translated to density curves and to provide a visual medium for comments and clarifications.

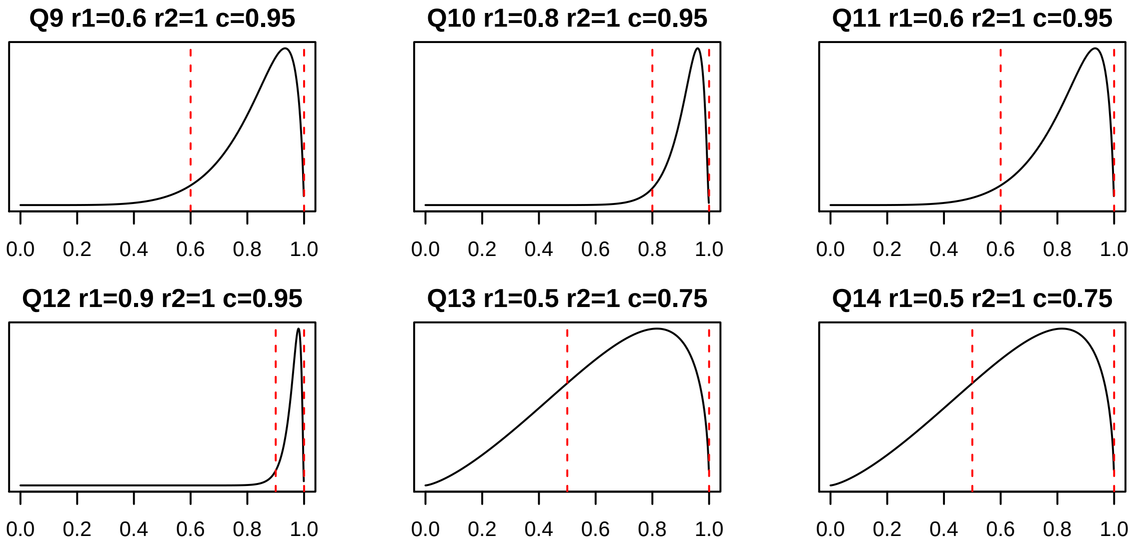
Plots: undercount questions

For reference:
Q1: country=sending, migrants=EU+, undercount=low
Q2: country=receiving, migrants=EU+, undercount=low
Q3: country=sending, migrants=non-EU+, undercount=low
Q4: country=receiving, migrants=non-EU+, undercount=low
Q5: country=sending, migrants=EU+, undercount=high
Q6: country=receiving, migrants=EU+, undercount=high
Q7: country=sending, migrants=non-EU+, undercount=high
Q8: country=receiving, migrants=non-EU+, undercount=high



Plots: accuracy questions

For reference:
Q9: country=sending, migrants=EU+, recording=register
Q10: country=receiving, migrants=EU+, recording=register
Q11: country=sending, migrants=non-EU+, recording=register
Q12: country=receiving, migrants=non-EU+, recording=register
Q13: country=receiving, migrants=EU+, recording=survey
Q14: country=receiving, migrants=non-EU+, recording=survey



Feedback specific to you

Translating the answers to density curves To clearly illustrate how your answers were translated to density curves, we will use question 1; the procedure was the same for all other questions (simply change the values of r_1, r_2 and c to the values that you provided in any given question). In question 1, you provided a range in (a) of 0% to 50%, i.e. $r_1 = 0$ and $r_2 = 0.5$, and provided certainty in (b) of 95%, i.e. $c = 0.95$. This translated to you saying that the probability that x (x is an undercount proportion in question 1) lies within the provided range $[0, 0.5]$ is equal to 0.95, and, simultaneously, that the probability that x lies outside the provided range $[0, 0.5]$ (i.e. that x lies within $[0.5, 1]$) is equal to $1 - 0.95 = 0.05$. Then, a suitable density curve was chosen so that the area under the curve, within $[0, 0.5]$ (your provided range), was equal to 0.95 (your provided certainty), and, automatically, the area under the curve, outside of $[0, 0.5]$ (i.e. within $[0.5, 1]$), was $1 - 0.95 = 0.05$. These properties of the curve can be visually appreciated by looking at the relevant plot above (the two vertical red lines on the plot represent the two range values).

Comments and clarifications To complement the generic feedback (see Section 1) here are some comments and clarifications that are specific to you:

1. In all of questions 5, 6, 7 and 8 you provided range values in (a) of 0% to 100% (i.e. $r_1 = 0$ and $r_2 = 1$). That is, you provided as range values the range of all possible values of the quantity in question. This could only be translated to you saying that all values of the quantity are equally likely, irrespective of the certainty value provided in (b). To see how this is the case consider question 5; the argument is the same for questions 6, 7 and 8. In question 5, you provided a certainty in (b) of 50% (i.e. $c = 0.5$). This translated to you saying that the probability that x (x is an undercount proportion in question 5) lies within the provided range $[0, 1]$ is equal to 0.5. Simultaneously though, this means that there must be probability equal to $1 - 0.5 = 0.5$, for x to lie outside of the provided range $[0, 1]$. However, this is not possible, since the quantity in question is a proportion, thus it can not lie outside of $[0, 1]$.

As mentioned in the generic feedback (see Section 1), a probability equal to c, for x to lie within the provided range $[r_1, r_2]$, implies a probability equal to $1 - c$, for x to lie outside of this range.

We would really appreciate if you could take these feedback into consideration before providing your answers in round 2. **Thank you**!

D Plots of Round 2 densities (without the Round 1 densities)

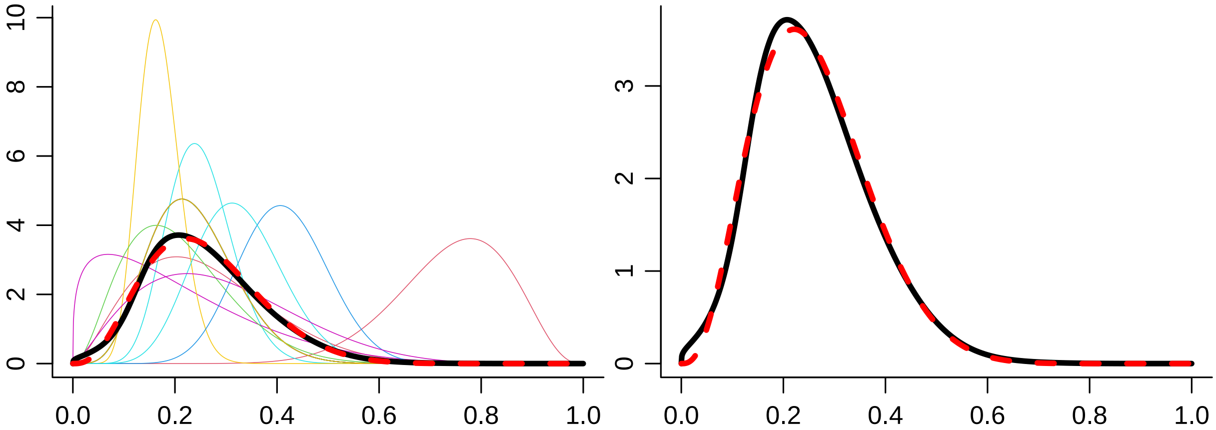


Figure 29: Experts' answers transformed to probability densities for question 1 (undercount of emigration of EU+ nationals who leave low undercount countries) of Round 2. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.

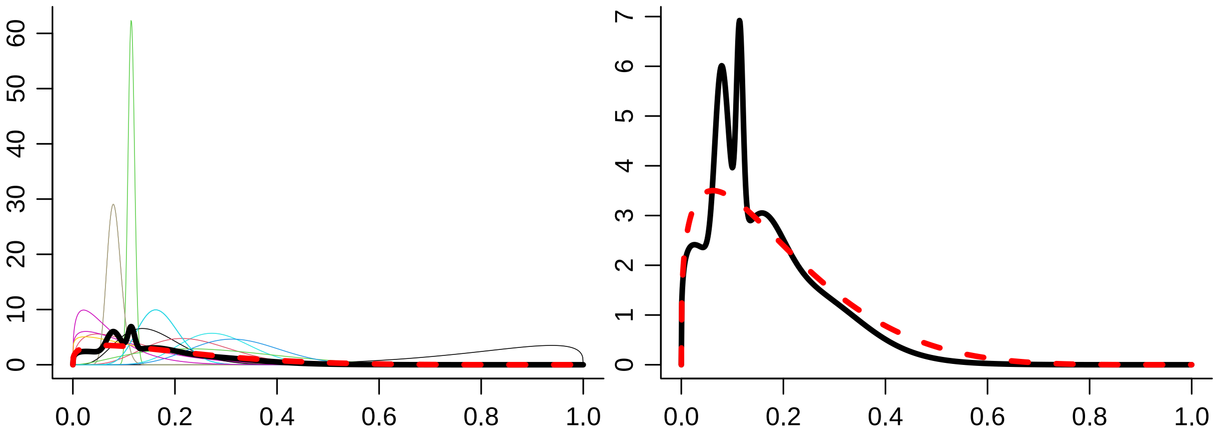


Figure 30: Experts' answers transformed to probability densities for question 2 (undercount of immigration of EU+ nationals who enter low undercount countries) of Round 2. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.

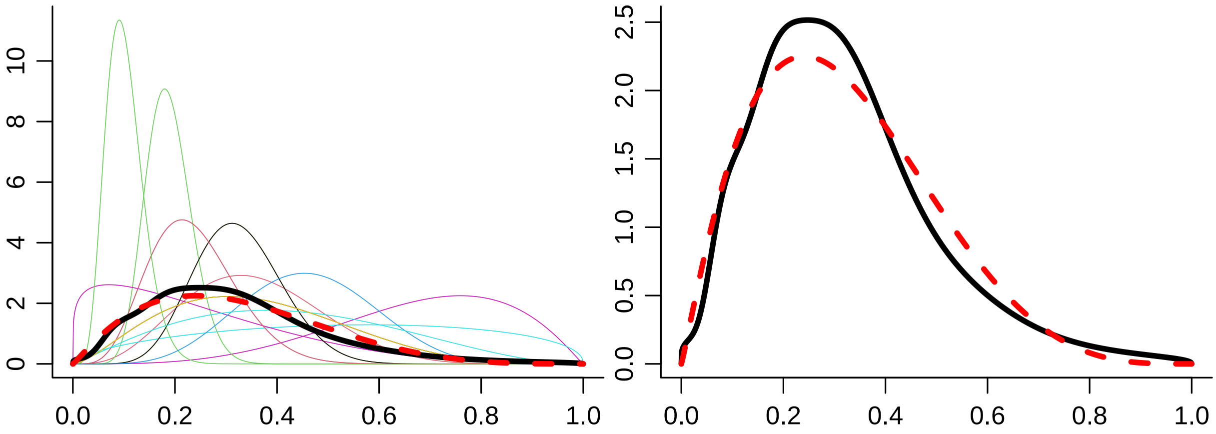


Figure 31: Experts' answers transformed to probability densities for question 3 (undercount of emigration of non-EU+ nationals who leave low undercount countries) of Round 2. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.

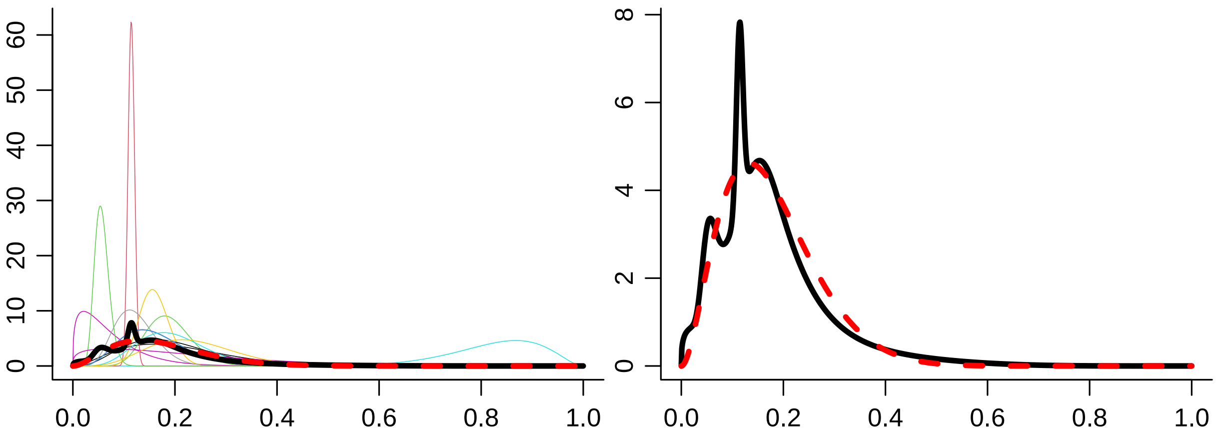


Figure 32: Experts' answers transformed to probability densities for question 4 (undercount of immigration of non-EU+ nationals who enter low undercount countries) of Round 2. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.

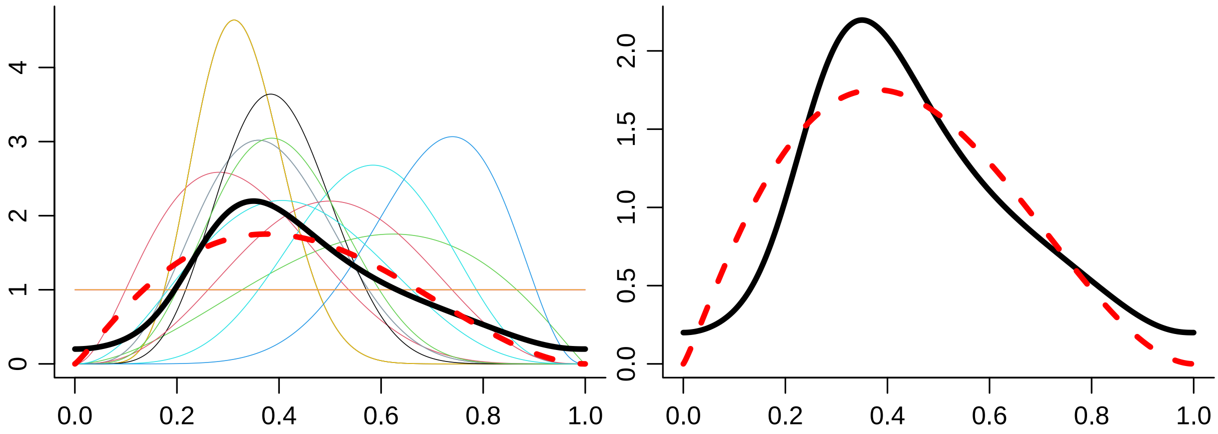


Figure 33: Experts' answers transformed to probability densities for question 5 (undercount of emigration of EU+ nationals who leave high undercount countries) of Round 2. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.

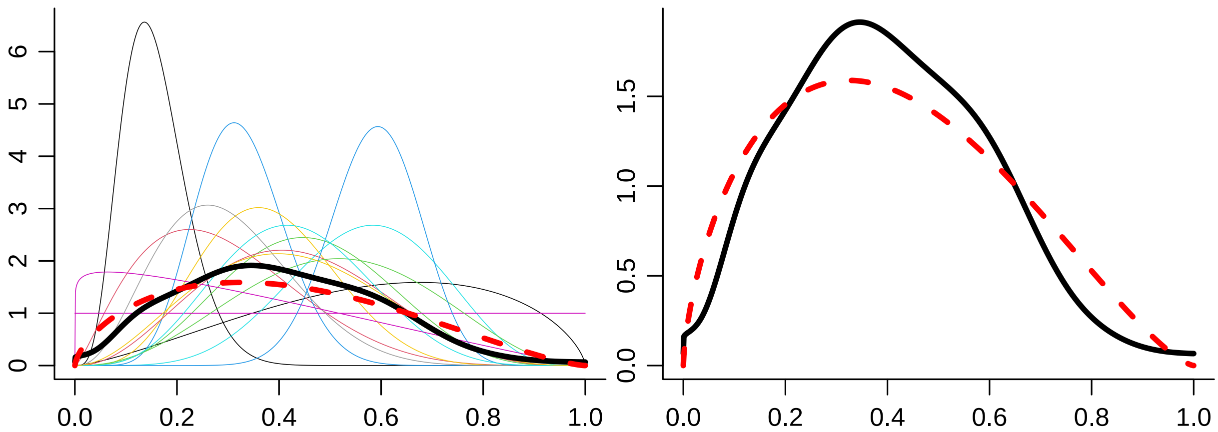


Figure 34: Experts' answers transformed to probability densities for question 6 (undercount of immigration of EU+ nationals who enter high undercount countries) of Round 2. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.

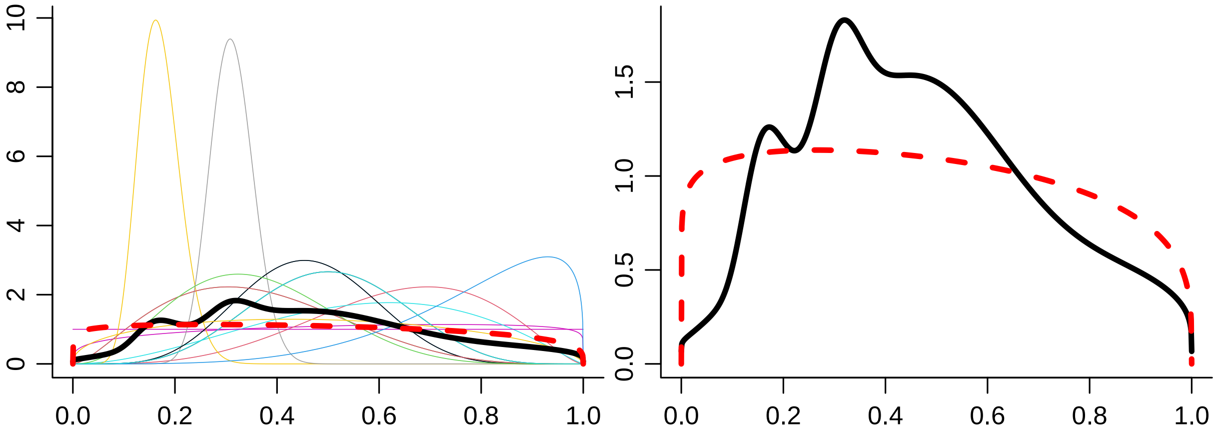


Figure 35: Experts' answers transformed to probability densities for question 7 (undercount of emigration of non-EU+ nationals who leave high undercount countries) of Round 2. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.

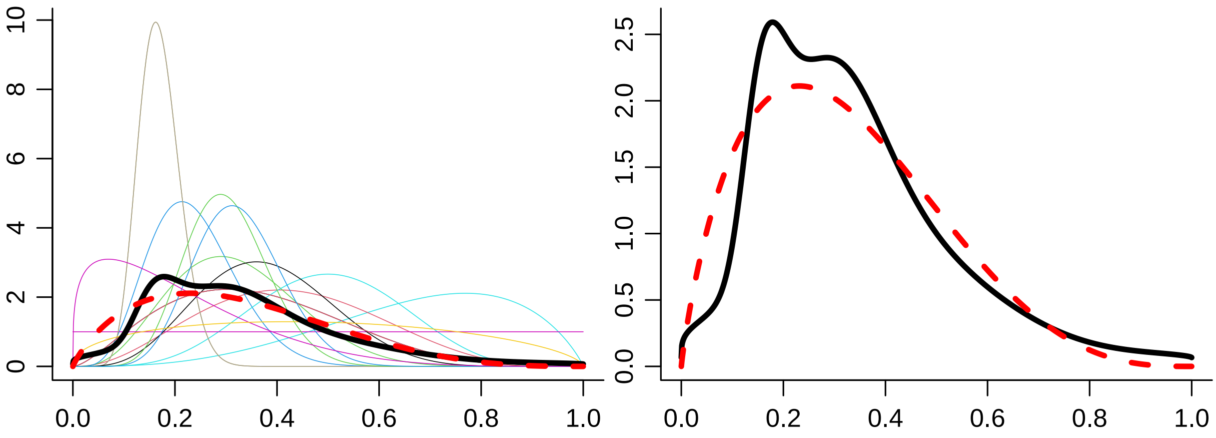


Figure 36: Experts' answers transformed to probability densities for question 8 (undercount of immigration of non-EU+ nationals who enter high undercount countries) of Round 2. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.

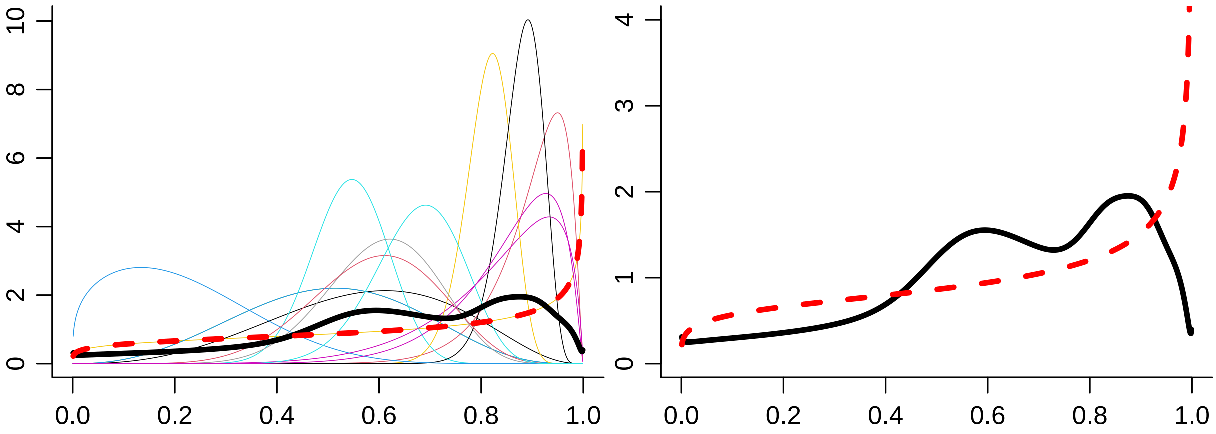


Figure 37: Experts' answers transformed to probability densities for question 9 (accuracy of emigration of EU+ nationals who leave countries recording with a register) of Round 2. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.

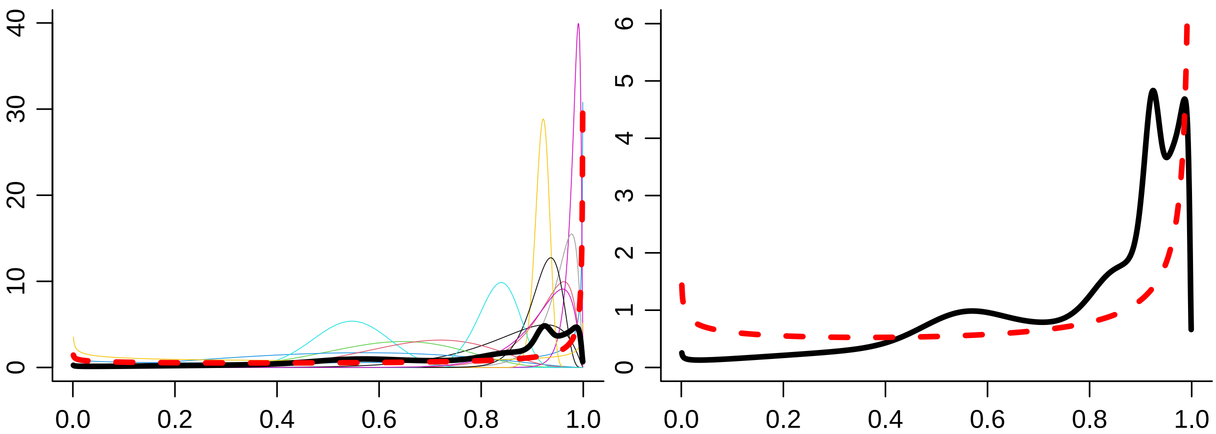


Figure 38: Experts' answers transformed to probability densities for question 10 (accuracy of immigration of EU+ nationals who enter countries recording with a register) of Round 2. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.

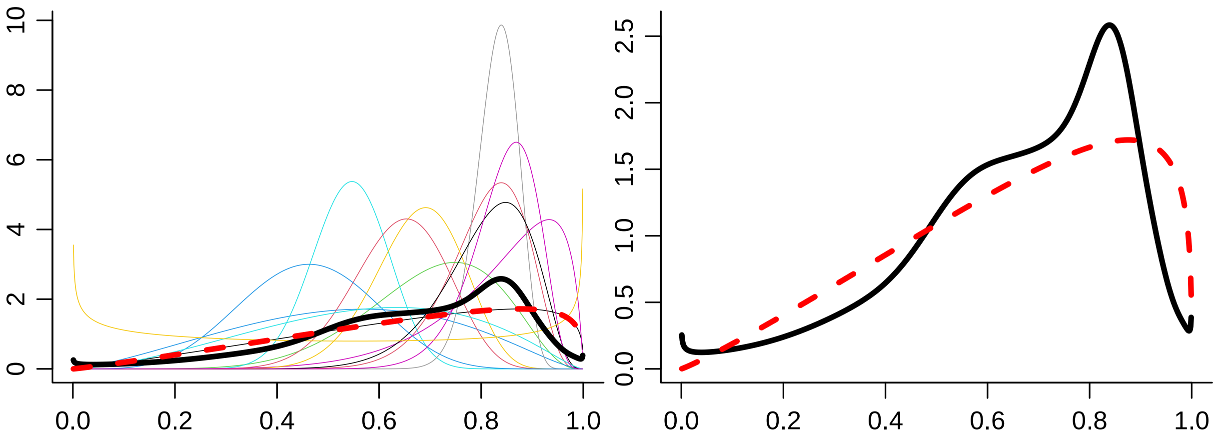


Figure 39: Experts' answers transformed to probability densities for question 11 (accuracy of emigration of non-EU+ nationals who leave countries recording with a register) of Round 2. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.

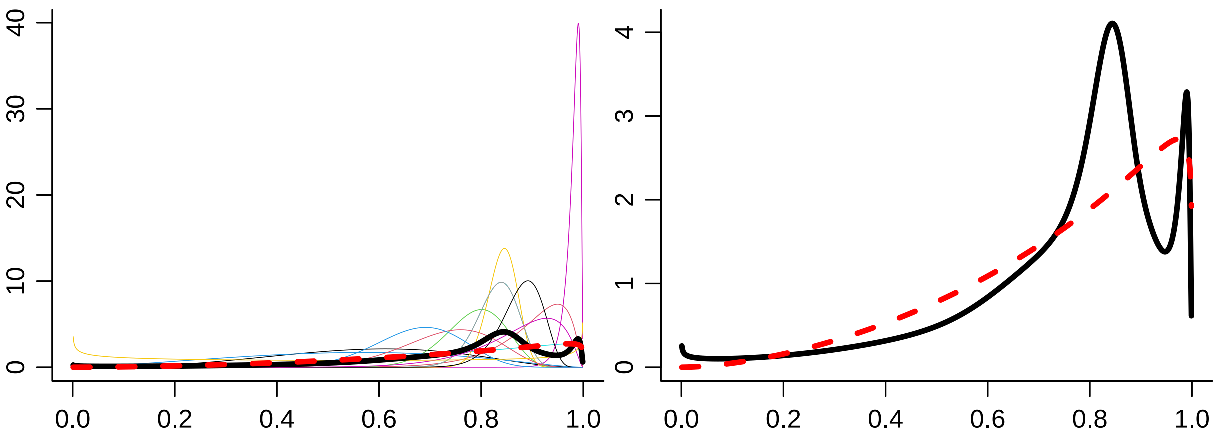


Figure 40: Experts' answers transformed to probability densities for question 12 (accuracy of immigration of non-EU+ nationals who enter countries recording with a register) of Round 2. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.

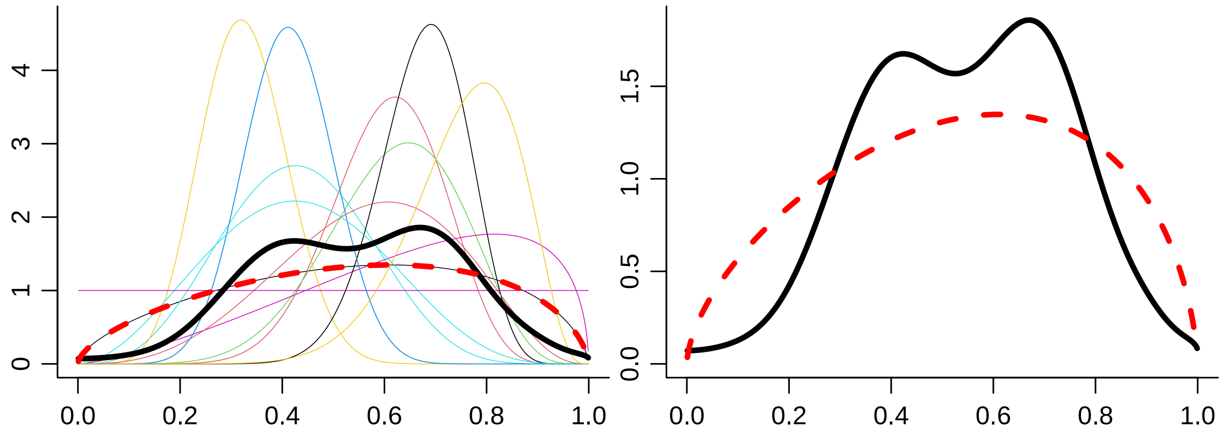


Figure 41: Experts' answers transformed to probability densities for question 13 (accuracy of immigration of EU+ nationals who enter countries recording with a survey) of Round 2. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.

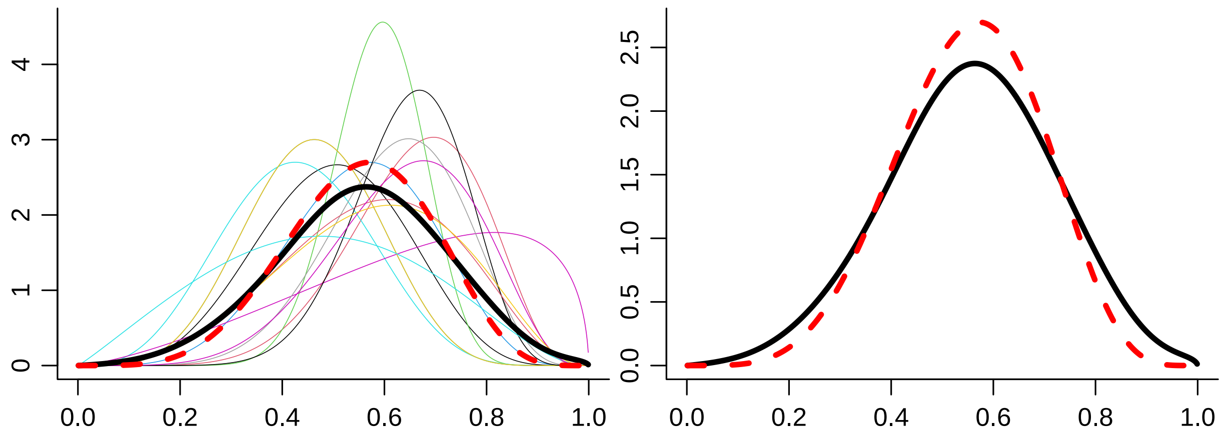


Figure 42: Experts' answers transformed to probability densities for question 14 (accuracy of immigration of non-EU+ nationals who enter countries recording with a survey) of Round 2. Left plot are the experts' individual densities (thin, coloured curves) with the two aggregated densities imposed (the thick, solid, black curve corresponds to aggregation via method 1 and the thick, dashed, red curve corresponds to aggregation via method 2). Right plot are the two aggregated densities on their own.