European Asylum Support Office

Quantitative assessment of asylum-related migration

A survey of methodology

August 2017

SUPPORT IS OUR MISSION

Acknowledgements

This work has been prepared by Jakub Bijak, Jonathan J Forster and Jason Hilton from the ESRC Centre for Population Change, University of Southampton, and was funded by the European Asylum Support Office (EASO), under the contract EASO/2015/290. The authors would like to thank Christoph Curchod, Andres Delgado, Louise Waldenström, Johanna Delac, Haidar Al-Talibi, and other colleagues from SEM and Migrationsverket for sharing their insights into the Swiss and Swedish models.

Disclaimer

The opinions expressed in the report are those of the authors and do not necessarily reflect the views of the European Asylum Support Office. EASO does not guarantee the accuracy of the data included in this study. No person acting on EASO's behalf may be held responsible for the use which may be made of the information contained therein.

European Asylum Support Office

Quantitative assessment of asylum-related migration

A survey of methodology

AUGUST 2017

SUPPORT IS OUR MISSION

Europe Direct is a service to help you find answers to your questions about the European Union.

Freephone number (*):

00 800 6 7 8 9 10 11

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

More information on the European Union is available on the internet (http://europa.eu).

Luxembourg: Publications Office of the European Union, 2016

PDF ISBN 978-92-9494-607-2 doi:10.2847/71794 BZ-01-17-586-EN-C

Print ISBN 978-92-9494-608-9 doi:10.2847/642161 BZ-01-17-586-EN-N

Printed in Belgium

PRINTED ON ELEMENTAL CHLORINE-FREE BLEACHED PAPER (ECF)

Contents

Tables and Figures

Quantitative assessment of asylum-related migration: A survey of methodology

Report Highlights

- Asylum-related migration is an analytical category with rather elusive boundaries, which does not necessarily map well onto established definitions of migration. In this report, it is understood as international forced displacement, caused by persecution, armed conflict, violence, or violations of human rights. As such, this is one of the most uncertain types of migration flows.
- There are many explanation of migration processes, including forced migration, but they lack an overarching framework. Here, we endorse utilising a pragmatic approach for modelling purposes, which concentrates on individual push, pull and intervening factors operating across a range of levels of analysis: from individuals, to groups, to whole countries and societies.
- The majority of data on asylum applications offer similar estimates, and the key advantage of the EASO data is their timeliness. Since 2008, European data have been harmonised, and are thus of higher quality and better comparable than before. Data on drivers of asylum are of varying quality, and their use in predictions may be problematic, except for crucial policy variables.
- Quantitative studies of asylum typically rely on aggregate-level statistical or econometric models of varying complexity. There are trade-offs between explanatory panel regressions and predictive time series models. On the other hand, examples of the best national practice include some form of expert opinion. There is a knowledge gap in terms of learning from ex-post model validations.
- For practical use, we recommend building an early warning model based on EASO data series, supplemented by synthetic asylum policy indices, expert opinion on individual asylum-related flows, and stakeholder views on what migration levels should trigger a warning. The model should be ideally expressed in terms of changes of asylum-related migration levels.
- In the short-term, work efforts need to be concentrated on specification of the model based on user needs. Longer-term modelling endeavours can explore interactive simulations and the use of 'big data' from social media and other sources, to enhance the early warning capacity of the models. Clear communication of model aims, uncertainty and limitations, is crucial for its success.
- Given the importance of the challenges related to asylum in Europe, a successful development of an early warning system is a priority, and would change the perspective in policy and decision-making at the EU level from reactive to pro-active. At the same time, contingency planning would become naturally embedded at the heart of the modelling process.

Executive Summary

The report presents an overview and typology of existing empirical models of asylum-related migration, followed by an analysis of empirical data available for modelling purposes, and by an evaluation of models prepared in specific countries and by international organisations. Its key aims are to identify, review and assess the various quantitative models for explaining and predicting asylum-related migration, and to aid the development of a bespoke European asylum model in the longer term, as a part of the wider work programme of the European Asylum Support Office (EASO).

In this study, asylum-related migration is defined as international flows of people who are – or who claim to be – forcibly displaced due to persecution, armed conflict, violence, or violations of human rights. This definition excludes internally displaced persons (IDPs), who seek protection within their countries of nationality, as well as people who are forced to migrate by other extreme factors, for example famines, environmental crises, or development projects, except when these factors are related to forced displacement defined above. The category thus encompasses, but is not limited to, refugees, asylum seekers, people with temporary protected status, or others in refugee-like situations, and is independent from the duration of stay in the destination country (**Section 1**).

An overarching theoretical framework for analysing asylum-related migration does not exist despite the vast literature on the factors and drivers behind the forced population flows. Moreover, there is a disconnection between the explanations offered for asylum-related migration and for migration flows generally, and the corresponding bodies of literature rarely intersect. An analysis of push and pull factors of migration, together with individual-level and other intervening variables, even though reductionist, remains a pragmatic choice for operationalising the conceptual framework to guide the subsequent analysis. The specific character of asylum-related flows is related to the presence of extreme push factors behind the migration, which render the resulting processes less predictable than other migration types. Uncertainty and chance are thus central features of the asylum-related decisions and moves, what is reflected in the proposed analytical framework (**Section 2**).

The majority of data on asylum applications in Europe ultimately derives from national state sources, and so major providers of data (UNHCR, Eurostat, EASO, and national statistics offices or migration ministries) tend to give estimates that are extremely similar. EASO data has the advantage of providing monthly information in a timely fashion, about two weeks after the reference period. Coherent definitions and harmonisation due to new European legislation mean that data from 2008 onwards is of higher quality and better comparable across countries. Before this date, definitional problems and missing data is in evidence. A trade-off between the length of the data series needed to make good generalisations and data quality must thus be considered (**Sections 3.1 and 3.2**). Data on drivers including economic factors, conflict and human rights are available from a number of sources and vary in quality. In many cases, the inherent unpredictability of many of these factors makes it difficult to include them as covariates for prediction. Information on policy 'pull' factors is more accessible; the IMPALA database provides comparable and detailed indicators on relevant asylum policies and processes (**Sections 3.3 and 3.4**).

The existing academic approaches for modelling and predicting asylum-related migration have been reviewed and assessed according to six criteria: adequate specification, detailed resolution, justifiable assumptions, predictive potential, uncertainty assessment, and model generalisability. (**Section 4.1**). In statistical terms, almost all models of asylum-related migration can be seen as instances of generalised linear models (GLM) – generalisations of simple linear regression models. Three groups of models are analysed: simple regression models, panel regression models, and more complex approaches – time series, structural and log-linear models. The available models differ with regard to their complexity, yet all of them describe macro-level asylum processes, typically for countries or groups of countries, with usually yearly time resolution. There are trade-offs between the models from different groups, with the panel models being more comprehensive in terms of drivers and assumptions, and the more complex models better suited for prediction. Important features of different models involve political and policy variables,

for which there exist encouraging examples for modelling and forecasting of synthetic indices of policy or conflict (**Section 4.2**).

However, the national practice in predicting asylum and preparing contingency plans – even though varying widely across Europe with respect to the levels of complexity – typically involves significant expert and contextual knowledge. The existing approaches do not rely on advanced statistical techniques but are more flexible and timely and thus potentially better suited to match the needs of their users (**Section 4.3**). The existing models are usually not formally assessed *ex post*, based on new data, although some national models get updated up to several times a year. This shows the potential for building a comprehensive forecasting framework, which would incorporate expert opinion and allow the predictions and models to be dynamically adjusted based on the results of periodic validation (**Section 4.4**).

For practical purposes, we recommend building an early warning model based on EASO data series on weekly or monthly changes in the level of asylum applications, supplemented by asylum policy indicators. Other important variables, such as armed conflict intensity measures for different regions of the world, could also be used, alongside the expert opinion on the envisaged changes of individual asylum-related flows, as well as stakeholder views on the warning thresholds – what levels of migration should trigger a model reassessment or a policy action. The asylum data used in the model would be origin-and-destination specific, the expert knowledge and conflict measures would be related to countries of origin, and the policy indices to the European countries of destination. A two-stage model structure would reflect the distinction between the push and pull factors. The various elements would be combined together by the means of Bayesian statistics (**Sections 5.1 and 5.2**).

The key knowledge gaps are related to asylum policy developments across Europe, with synthetic policy indices ideally updated in real time – existing projects, such as the IMPALA database, offer an excellent blueprint here. Other challenges are related to expert knowledge elicitation, reliant on a pool of experts on the countries of origin, and to specifying the warning thresholds for the models, ideally in collaboration with the stakeholders. One of the main issues here are related to the trade-offs between the costs of generating false alarms by the early warning system and the costs of its inaction in the face of real challenges (**Section 5.3**). In this report, we provide a simplified illustration of how the modelling framework may operate, by using the examples of Syria and Iran (**Section 5.4**).

In terms of recommendations, we propose that the work programme concentrates on specification of early warning models tailored to user needs. Formal migration models in general can help with explaining the data, but not with prediction, which is difficult in the light of the many unpredictable push and pull factors. Longer-term modelling endeavours can also explore interactive simulations, where qualitative and other survey information could form additional input. From an academic perspective, there are also knowledge gaps with respect to people who remain immobile during large-scale crises and impacts of asylum on sending countries. Another area worth exploring is related to the use of 'big data' from social media and other sources to enhance the early warning capacity of the models. Key prerequisites of success of any formal modelling include an appropriate communication of the model aims, results, uncertainty and limitations to stakeholders (**Section 6**).

1. Introduction and Background

This study is intended to contribute to a wider work programme of the European Asylum Support Office (EASO), aimed at investigating the possibility of building a quantitative model of asylum-related migration in Europe. Within this programme, the key objectives of this report are threefold: we aim to identify, review and assess the various quantitative models for explaining and predicting asylum-related migration, with the view of aiding the development of a bespoke asylum model in the longer term. In particular, the report covers the following three topic areas:

1. A comprehensive overview and typology of existing empirical models of asylum-related migration;
2. An analysis of empirical data available for modelling purposes and gaps in the existing information;
3. A critical evaluation of models prepared in specific countries and by international organisations.

Within this framework, we present a feasibility study concerning the possibility of building a Europe-wide quantitative models of refugee and other asylum-related migration flows, accompanied by a discussion of the opportunities and limitations of different modelling approaches. The review also includes a set of recommendations and guidelines on the construction and interpretation of these models, and on communication of their results and limits to European policy users.

According to the 1951 United Nations (UN) Convention and the 1967 Protocol Relating to the Status of Refugees, the term *refugee* is very precisely defined in legal terms as someone, who:

> "owing to well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group or political opinion, is outside the country of his nationality and is unable or, owing to such fear, is unwilling to avail himself of the protection of that country; or who, not having a nationality and being outside the country of his former habitual residence as a result of such events, is unable or, owing to such fear, is unwilling to return to it." (UNHCR 1951/1967; Art. 1 A (2))

The focus of this report, however, is on *asylum-related migration*, and is kept deliberately broad in order to encompass a variety of flows of people seeking different forms of humanitarian protection outside their respective countries of nationality. The key distinct groups here include refugees and asylum seekers – people who are seeking refugee status or another form of protection in a country other than their own; which includes foreigners seeking or being granted temporary protected status, or others in refugee-like situations, admitted on humanitarian grounds (UN 1998: 34).

For the purpose of this study, asylum-related migration has therefore to jointly meet two criteria: first, it needs to be **international** in nature, and second, it has to be – or claimed to be – related to **forced displacement**, defined as forced migration due to persecution, armed conflict, violence, or violations of human rights[1]. This definition thus excludes internally displaced persons (IDPs), who seek protection within their countries of nationality, as well as people who are forced to migrate by other extreme factors, for example famines, environmental crises, or development projects[2], except when these factors are related to forced displacement defined above.

As this definition needs to be related to the operational requirements of EASO, it does not distinguish between asylum seekers who have fled conflict or persecution, or those who only claim to have done so while seeking protection. Similarly, it does not distinguish between the channels of migration, and includes people subject to human smuggling or victims of human trafficking as appropriate, insofar they meet the main criteria listed above. Figure 1 presents a conceptual mapping of the various terms and definitions used in this study.

---

[1] Compare, for example, the following definition of the World Bank: "Forced displacement refers to the situation of persons who are forced to leave or flee their homes due to conflict, violence and human rights violations" (World Bank; http://www.worldbank.org/en/programs/forceddisplacement, as of 1 June 2016).

[2] See, for example, the following definition, attributed to the International Association for the Study of Forced Migration (IASFM), which defines forced migration as "a general term that refers to the movements of refugees and internally displaced people (those displaced by conflicts) as well as people displaced by natural or environmental disasters, chemical or nuclear disasters, famine, or development projects" (Forced Migration Online; http://www.forcedmigration.org/about/whatisfm, as of 1 June 2016).

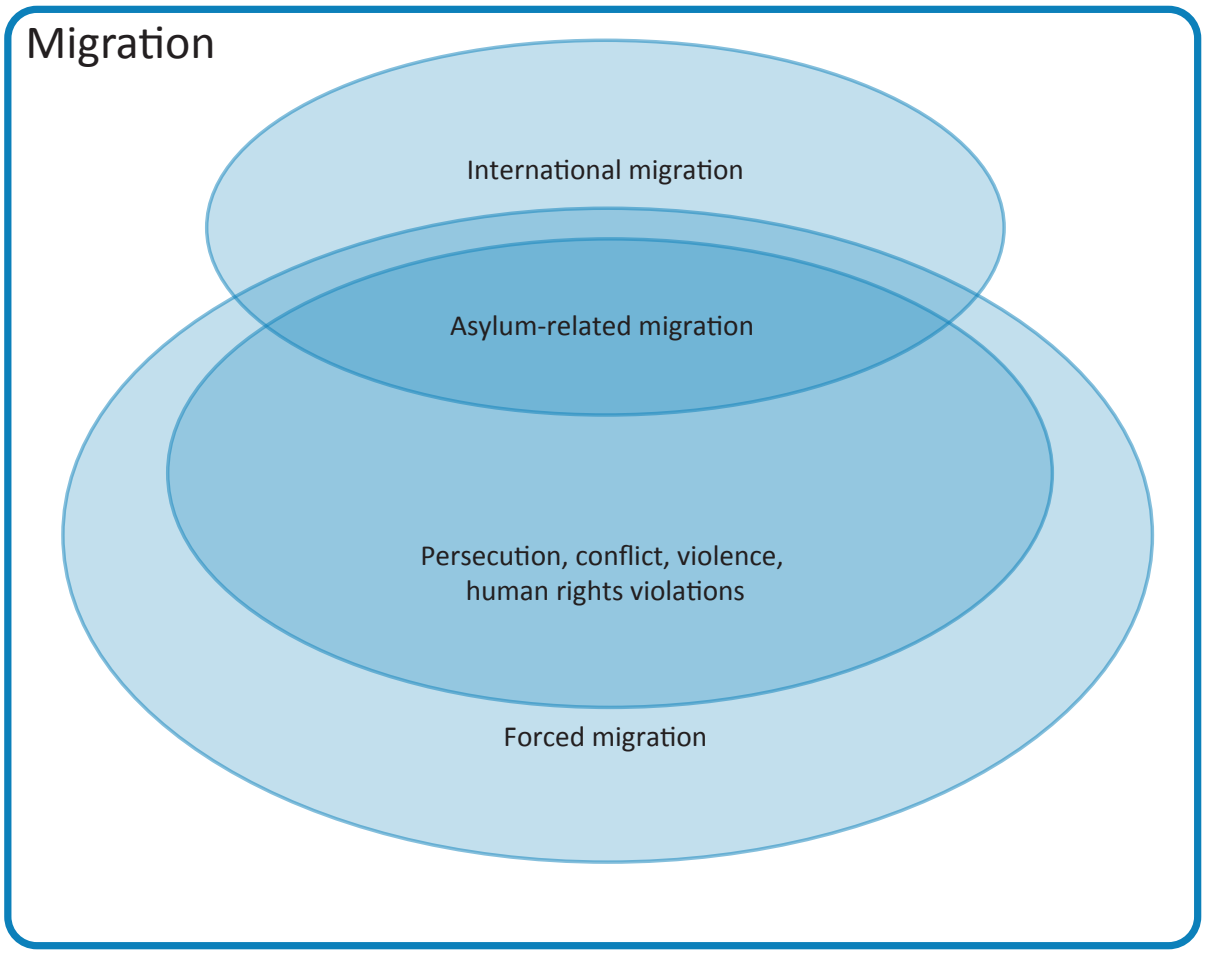

Figure 1. Conceptualising asylum-related migration within a broader area of migration studies

Additionally, given the asymmetry in resource implications, unless noted otherwise the focus in this study is at **inflows of asylum-related migrants**, with subsequent migrations being outside the scope of the current enquiry. Finally, it is worth noting that asylum-related flows are not related to any specific time criterion for the actual or intended duration of stay, which apply to other groups of migrants (UN 1998)[3].

The structure of the remainder of this report is as follows. In Section 2, a conceptual framework for a quantitative analysis of asylum-related migration is introduced, building on the earlier study on migration factors and drivers (Maastricht University & IOM Berlin 2016). In Section 3 the various data sources on asylum-related migration and a range of its potential explanatory variables are discussed and evaluated, with focus on macro- and mezzo-level factors and drivers. Section 4 is devoted to evaluation of formal approaches aimed at modelling and prediction of asylum-related migration. A subsequent synthesis of the information from a range of sources and modelling practices, presented in Section 5, enables a feasibility study to be carried out related to constructing a bespoke EASO model of asylum flows. The report concludes with a set of practical recommendations related to short- and longer-term perspectives for model-building, which are discussed in Section 6 alongside the current gaps in data, and the limitation of the model uses. The report is accompanied by a glossary of key terms, an inventory of data available for modelling purposes (Appendix A), as well as a detailed assessment of various data and models (Appendix B).

[3] For a more general discussion of issues related to defining migration in Europe, see Poulain et al. (2006) or Raymer et al. (2013).

2. State of the art: The conceptual framework for analysing asylum-related migration

This section is devoted to sketching a conceptual framework for quantitative studies of asylum-related migration. Its aims are, first, to critically assess the state of the art in the conceptual analysis of asylum-related processes, and second, to propose a framework for operationalising the various push and pull factors, as well as intervening obstacles, which will guide the evaluation of individual data sources and models presented in the subsequent sections of this report.

What remains beyond doubt is that the key characteristic feature of migration – and asylum-related migration in particular – is its very high volatility and uncertainty, resulting from the susceptibility of migration flows to a broad range of factors, drivers, and shock events (Bongaarts and Bulatao 2000; Bijak 2010). This high volatility leads to the very limited predictability of migration flows overall (Pijpers 2008; Bijak 2010), with horizons of prediction of a few years at most for forecasts made on an annual basis (Bijak and Wiśniowski 2010). The flows of refugees and asylum seekers, by their very nature, are the most volatile of all migration flows, yet can have extremely high impact, not only on the countries of origin but also on the receiving ones (Disney et al. 2015).

Despite a long history of attempts to generalise the observed migration patterns, dating back to Ravenstein (1885), a theoretical framework of migration studies remains patchy. The main criticisms of the existing approaches are that they are fragmented, entrenched in the individual disciplines of scientific enquiry, and lacking an all-encompassing view of migration processes (Arango 2000), as well as predictive capacity (Kupiszewski & Bijak 2013). Besides, specifically in the context of asylum, forced migration rarely features in theoretical enquiries on migration flows (FitzGerald 2015).

With notable exceptions, such as Zolberg (1989) or a few economic theories related to the provision of asylum and choices among asylum seekers (for a review, see Suriyakumaran & Tamura 2016), attempting to theorise about forced migration often boils down to namechecking armed conflict and violence among the key drivers of population flows. As observed by FitzGerald (2015), specialised areas of research, such as "refugee studies" or "forced migration studies" rarely intersect with the mainstream of international migration enquiries, especially in the theoretical domain. Given this disconnection, FitzGerald calls for more work on the transferability of statements on migration in general to the asylum processes, and on the interrelations between asylum and other types of migration, such as economic factors or social networks.

Where theories of asylum-related migration exist, they can be difficult to operationalise in a formal model without losing the depth of the original thought. For example, Zolberg's (1989) rich narrative brings together a comprehensive analysis of the dynamics of world's political and economic systems, with political theories of the state and social order. His framework can be formalised with a range of proxy variables but this approach would not do justice to the complexity of the theoretical model. Similar problems are faced in the economic domain: complex theoretical explanations of asylum processes are usually simplified when it comes to the modelling (Suriyakumaran & Tamura 2016).

Still, such a reductionist approach, whereby broad theories are distilled into a suite of several explanatory variables, is at least a pragmatic choice. In this context, a convenient overarching structure for analysing any migration flows – including asylum-related ones – is the classical 'push-pull' theoretical model of Everett Lee (1966). Briefly, in this framework, migration is a result of interplay between four types of factors: those 'pushing' migrants from their origin countries or regions; those 'pulling' them to the areas of destinations; the 'intervening factors' facilitating or obstructing migration; and finally the individual characteristics of migrants. The framework on its own is not predictive – not least because it would require predicting all the factors working at different levels – but it can nevertheless help with framing the modelling endeavours.

As remarked by Öberg (1996), the push factors differ in their severity and importance. Those labelled as 'hard' include overt violence, the presence of armed conflict, or natural disasters, and lead to very strong and immediate pressure to migrate. On the other hand, the ones referred to as 'soft' – such as the

economic factors – are associated more with conscientious choices related to migration. As argued by Öberg (1996), the 'soft' factors are associated with more stable long-term migration trends, whereas the 'hard' ones are very hard to predict, if at all. Besides, the dominant factors shape the composition of the migrant groups: more self-selected when 'soft' factors are chiefly at play, and much more indiscriminate when the 'hard' ones dominate.

Of course, sometimes it can be very difficult to tease out the various reasons for migration, even for individual migrants or households. In an influential article, King (2002) has argued that some of the dichotomies present in migration studies, such as the one of *voluntary* versus *forced migration*, are largely obsolete given the complex nature of contemporary migration flows. Instead, it may be more helpful to look at the "a complex continuum of coercion and free-will in migration decisions" (King 2002, p. 92). For example, armed conflict may pose a very direct and immediate threat to people's lives, or may result in prolonged economic collapse and loss of livelihood of people affected.

Similar conclusions were reached in a slightly different, yet related context of environmentally-driven migration flows (Government Office for Science 2011): as the various migration factors and drivers are intrinsically linked, it is difficult to isolate the 'pure' environmental effects. Importantly, the Government Office for Science (2011) pointed out to the often-overlooked problems related to *immobility* of those referred to as 'trapped populations' – vulnerable people who are unable to move in the face of losses of livelihood or even more immediate dangers. This links to one key gap in the literature mentioned by Arango (2000), who argued that migration scholars focus too much on those who move, at the expense of those who *do not*, and the drivers behind lack of mobility.

The different drivers of asylum-related migration have been recently reviewed by Maastricht University and IOM Berlin (2016), where a detailed survey of these factors is presented. A graphical synthesis of the analytical framework, as seen through the lens of the push-pull approach is offered in Figure 2, with the addition of several additional factors identified by Böcker and Havinga (1998). In particular, the roles of 'image' and 'reputation' of the countries of destination – in other words, perceptions of, rather than actual political, social or economic situation – can be crucial in shaping the decisions of the asylum seekers. In the age of mass Internet access and social media, this aspect has become very important – and very dynamic, adding to the already high volatility of asylum processes. Besides, crucially given the uncertain character migration, Böcker and Havinga (1998) mention *chance* as an inherent element associated with the decision processes in asylum-related migration, in addition to the higher-level uncertainties of the various societal processes and drivers.

Hence, as shown in Figure 2, the general framework for analysing asylum-related migration flows encompasses various push and pull factors, operating at the most general (macro) levels, mediated by the decisions at the individual (micro) level, and a range of cross-level intervening factors. The operationalisation of these factors by using specific variables and data sources is discussed next.

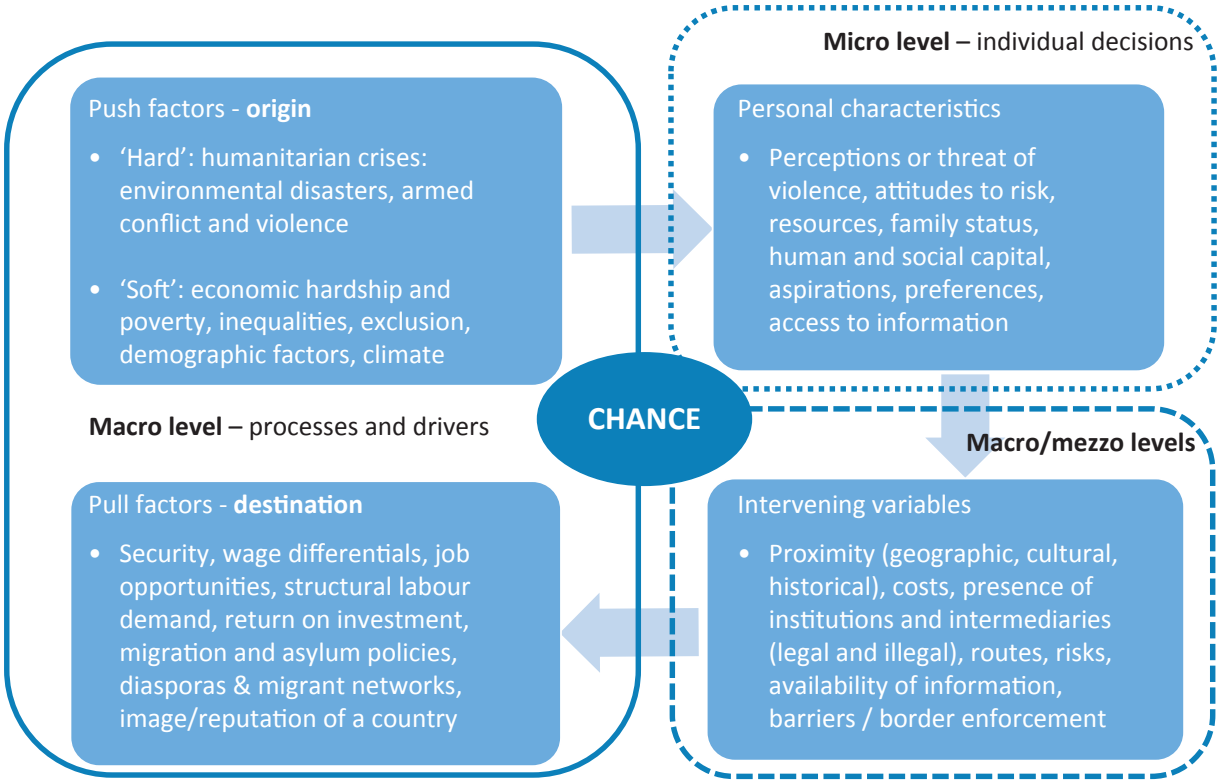

Figure 2. Conceptual framework for modelling asylum-related migration

Source: Böcker & Havinga (1998); Maastricht University and IOM Berlin (2016); framework after Lee (1966) and Öberg (1996)

3. Review and assessment of the existing sources of data on asylum-related migration

3.1 Assessment framework and criteria

Any potential predictive model of asylum requires appropriate data. In order to facilitate the systematic assessment of the available data sources related to asylum flows, this section sets out consistent criteria that specifically relate to the intended goal of modelling and predicting asylum. Measuring asylum-relating migration is often difficult, not least because practical and definitional issues may arise in the collection and analysis of data (Bakewell 1999). Furthermore, the asylum system is hugely political, and many groups with competing interests attempt to attach their own narrative and interpretations to the 'ground truth' (Crisp 1999; Bakewell 1999).

Due to the specific focus of this project on asylum flows into Europe and asylum applications specifically, it is possible to sidestep some of the more severe difficulties associated with data on forced migrants in developing countries (see Bakewell 1999, UNHCR 2002). European countries are much better able to devote resources to enumerating arrivals, and the formalised nature of European labour markets together with the attractiveness of refugee status in European countries mean that those fleeing conflict are likely to be absorbed relatively quickly into the asylum application system (Bakewell 1999). The legal nature of the process of applications for asylum also means that the state has direct information about numbers, both of applicants and grants of status, which means that asylum applicants are in some cases better measureable than other forms of migration (Disney et al. 2015).

Despite this, there remain issues to consider when examining present and historical asylum data. Firstly, domestic political concerns, geopolitics, and national reputational issues may lead to secrecy or selective reporting of national statistics (Crisp 1999). For instance, during the Cold War, Western states viewed large asylum flows as propaganda victories over the communist bloc. Conversely, domestic opinion is often hostile to asylum seekers, and so this may incentivise states to focus on their ability to return failed claimants (ibid).

Incentives for individual asylum seekers may also lead to problems with asylum data. Those who are unsuccessful with a first attempt at gaining protected status may attempt to re-apply elsewhere, leading to the possibility of double counting in asylum statistics, particularly if applicants adopt false identities (Singleton 2016). The EURODAC database[4], which allows the sharing of information (including fingerprints) about applicants, facilitates the re-identification of asylum seekers and thus reduces the incentives to take this route.

Furthermore, differences in administrative procedures across 28 member states may lead to additional problems. The length of the application process may differ over the continent, so that migrants who begin the application process at the same time in different states may show up in the statistics in different periods, which may be problematic when trying to relate outflows to inflows in over time (Singleton 2016). Definitions of data reported may also differ between sources; some may relate number of cases rather than number of individuals, while others describe numbers of first-time applicants while excluding repeat applicants counted by other states (UNHCR 2002).

Finally, given that the aim of this report is to make recommendations regarding a model to aid decision-making and planning among European member states, consideration of whether potential data sources meet operational requirements is also vital. In particular, it is important that data arrive quickly enough for model outputs to be relevant and actionable, and that the frequency of the dataset is granular enough to allow responsive interventions to changes in the asylum system.

[4] Details at http://ec.europa.eu/dgs/home-affairs/what-we-do/policies/asylum/identification-of-applicants/index_en.htm

Potential data sources are thus judged against the following criteria:

1. **Frequency –** How frequently are the data updated? Equivalently, what is the time period between consecutive measurements?
2. **Definitions –** To what extent do the data fit with the required definition of asylum-related migration, as discussed above?
3. **Coverage –** What time periods and spatial area or areas does the data describe, and how does this relate to what is required by stakeholders?
4. **Accuracy –** How far can the data be relied upon to describe the true state of reality with respect to asylum-related migration? What biases may the data have, and how uncertain are the data estimates?
5. **Timeliness –** How soon after the period which the data describe are the statistics released?
6. **Quality assurance processes –** What processes are in place to ensure that the data meet predefined quality standards and that possible errors are caught prior to publication?

Clear side-by-side comparison of the various data-sources is facilitated by the use of a traffic-lights scoring system. Each data series is given a rating for each criterion against a three-point colour scale, where 'red' indicates that the data are problematic with respect to that measure; "amber" indicates that some problems exist; while "green" indicates that the variables in question meet standards needed for the task at hand. Table B1 in Appendix B gives specific descriptions of each point on the scale for every dimension of assessment. For an application of a similar evaluation framework, see Disney et al. (2015). Links to all data sources described are given in Appendix A.

3.2 Data on asylum-related migration

Based on the criteria developed in the previous sub-section, this section contains a review of existing data on asylum-related migration flows. Applying for asylum is a legal process, and so the main source of direct information about numbers of applications and numbers granted status is from the EU member states. As a result, data on asylum-related migration to European countries might be expected to be replicated between different series, as the ultimate source is likely to be the same (Singleton, 2016). However, this is not always the case, as Mouzourakis (2014) describes in relation to discrepancies between EUROSTAT data, EASO data, Frontex data, and that provided by individual member states. Recent moves by Eurostat and EASO to better harmonise definitions and reporting is likely to have alleviated some of these problems for more recent periods, however.

The most extensively used source of data regarding asylum and refugees is that provided by the **UNHCR**. In its online database, the organisation provides data from 1951 onwards on stocks of refugees, asylum seekers, internally displaced persons, returnees and stateless individuals, together with other 'persons of concern'. These data are comprehensive in terms of its temporal and spatial coverage, detailing both sources and destinations of asylum-related migration globally. However, Crisp (1999) suggests that while there were significant improvements in quality after the end of the Cold War, prior to this time, a combination of political pressures and lack of harmonisation across administrative areas (country 'desks') lead to substantial problems with data quality. A desire to overstate numbers in order to maximise resources may also have been a factor in past crises, particularly as often not all promised supplies would arrive on time. However, more recently, UNHCR data collection and quality assurance procedures have improved dramatically (ibid); its website provides an Emergency Information Management Toolkit aimed at providing field operatives with guidance and tools on exactly what to collect (http://data.unhcr.org/imtoolkit/).

The UNHCR also provides details of asylum applications to industrialised nations from the early 1980s. These data are more difficult to access than the stock data described above, being located across statistical annexes of reports from several different years[5]. The information is disaggregated across both sending and receiving countries, although a cross tabulation showing specific flows is only available for aggregated multi-year periods, rather than for all individual years. For developed nations, the data are based on reports

[5] Specifically, the UNHCR Statistical Yearbook (2001, 2008-2014), the Asylum Trends report (2005-07) (as detailed in Hatton and Moloney (2015)), and for 2015, the Global Trends report.

from receiving governments. More detailed information is provided on specific emergency situations, such as the 2015 European migration crisis, but use of such information for continuous monitoring on an EU-wide basis is not plausible.

As would be expected, individual **national statistical agencies** or government agencies responsible for migration also publish data on asylum applications and recognition independently. For instance, the German Federal Agency for Migration and Refugees provides data on applications and first applications from 1995 onwards, while the UK Home Office provides similar data from 1979. The main problem with these data sources are inconsistencies in coverage and definitions between data collected by different agencies. Although recent moves towards harmonisation within the European Union have improved matters somewhat, historical data may still exhibit differences in definitions and reporting, as a review in the UNHCR Statistical Yearbook 2001 makes clear (UNHCR, 2002; Table VI.2). The need to consider the European system holistically is evident from the discussion in Section 2, and from the review of drivers conducted by the Maastricht University & IOM Berlin (2016), so data sources which are consistent would be preferred.

These differences between national sources have lessened with the introduction in 2007 of European regulations creating obligations on the part of member states to collect statistics on asylum in accordance with specific definitions. **Eurostat** collates these harmonised data from all member states, including not only applications, but also status decisions, numbers of unaccompanied minors and resettlements. Furthermore, since 2008, details of the age and sex composition of the applicant flow are also available. Eurostat data are only available on a consistent and comparable basis from 2008 onwards, with data before this period missing for some countries and country/period combinations. However for many of the primary recipient countries, the data are present for a substantial period prior to 2008, and metadata is provide detailing possible issues. Eurostat does provide disaggregation of applications by month, allowing seasonal patterns to be identified. To ensure that data meet the required standards, quality assurance procedures are in place for submissions to Eurostat, including the use of reporting templates to ensure consistency and checks for internal consistency[6].

EASO's Early-warning and Preparedness System (EPS) data on applications are derived from the same sources as the Eurostat data, but its focus on operational support to member states means that the turnaround on the data is much quicker than for Eurostat, being provided only 15 days after the reference period (versus several months at least for Eurostat data) (EASO, 2016). However, this faster turnaround means that the data do not undergo the same validation processes and are not signed off as official statistics, and thus may differ from revised data published later. However, EASO provides detailed guidance to member states regarding definitions, illustrated by practical examples, and the incentives are on member states to reduce duplication by getting the data right first time, so differences between EPS data and equivalent Eurostat data might be expected to be limited. Some additional series not provided by Eurostat are included in EASO's collection efforts. In particular, recent moves to begin collecting data on "applications made" (as opposed to the currently reported "applications lodged") may provide a more comparable indication of the timing of arrivals to the country in question, as the former is less affected by differences in administrative procedure in different countries.

Figure 3 displays the time series of application data from UNHCR, Eurostat, the UK Home Office and the German Federal Agency for Migration and Refugees for a selection of top recipient countries. As can be seen, for the most part, the data overlap almost exactly, with some moderate differences reflecting revisions and/or definitional differences.

[6] The annexes to the metadata for the asylum data series provide details of these procedures: http://ec.europa.eu/eurostat/cache/metadata/en/migr_asyapp_esms.htm

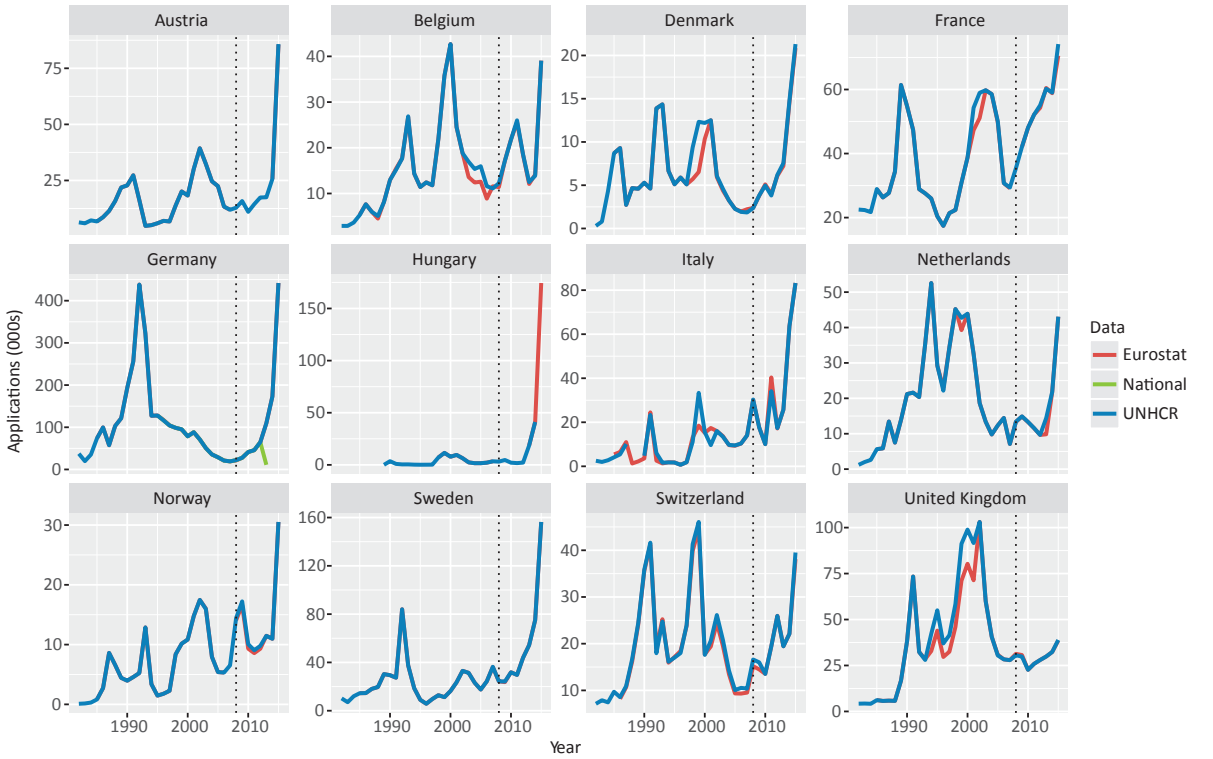

Figure 3. First-time asylum applications to 12 European states

Source: Eurostat, UNHCR and, for the UK and Germany, national reporting bodies. Note that often the data coincide exactly and thus only the top-most line can be observed. A vertical line at 2008 indicates the date from which Eurostat data are harmonised

Frontex (the European Agency for the Management of Operational Cooperation at the External Borders of the Member States of the European Union) collects data on detected crossings at the EU's external borders. Of course, not all border crossing are detected, and furthermore migrants may make multiple attempts to cross a border and may cross external EU borders more than once during their journey (Singleton, 2016). Thus, while Frontex data provide well-defined and timely operational data on attempted crossings, it can only approximate the flows or crossings themselves. It is also possible that some borders are intrinsically more difficult to police than others, so that the proportion of crossings detected differs according to border type and intensity of policing operations, making interpretation difficult.

Table 1. Evaluation of the existing data on asylum-related flows and their main drivers

Variable	Source	Frequency	Definitions	Coverage	Accuracy	Timeliness	Quality Assurance	Class
a. Asylum-related migration								
Refugee Stocks	UNHCR[1]							
Asylum Applications	National Agencies							
	EUROSTAT							
	EASO[2]							
	UNHCR							
Border Crossings	Frontex[3]							
b. Drivers of asylum-related migration								
Conflict	UNCDP Battle Deaths Dataset							
	PRIO Armed Conflict Database							
Human Rights	Political Terror Scale							
	Freedom House							
Economy	World Bank							
Distance	CEPII							

Key to assessment: **[Green]** Good aspects; **[Amber]** Average aspects; **[Red]** Problematic aspects. For category descriptors, see Table B1 in Appendix B.

Notes: 1. UNHCR receives low ratings for accuracy mainly due to concerns about historical data;
2. EASO data are rated 'amber' for coverage due to the relative short period for which data are available;
3. Frontex data are considered poor as a measure of total border crossings, as it can only capture *detected* crossings.

Sources: See table and Appendix B

Overall assessments of potential sources of input data are given in Table 1, with evaluations made against the criteria previously set out. Given the high degree of similarity between the various sources of application data, the smaller delay between the reference period and data delivery suggests that EASO data is the preferable source of information for the dependent variable in any attempt to produce actionable predictions of asylum-related migration. This timeliness comes at the cost of a drop in the level of quality controls applied to the data but it is not expected that this will be significant.

The extent of historical data is also at issue. Using longer time series provides a greater pool of data for analysis and thus potentially better inference but this needs to be traded off against the potentially lower data quality. Data from before 2008 are available from UNHCR and Eurostat, although the latter is lacking more information than the former, but definitions are not harmonised between sources. Frontex data, while necessarily imperfect, provide a complementary sources, as it is an important as a leading indicator of short-term flows and about routes taken to access application procedures.

3.3 Data on drivers: push and pull factors, and intervening obstacles

Drawing on the work in the Section 2 conceptualising the processes of asylum-related migration, this section identifies metrics and/or proxies for the drivers of flows into Europe and assesses their reliability and utility as covariates.

Economic variables are often used to proxy for the attractiveness of host countries; more prosperous countries are more likely to be considered attractive as potential destinations for asylum seekers. The World Bank provides a comprehensive and widely used repository of economic indicators. However, some care must be taken when including economic covariates of this type. Thought must be given to which of these indicators are most relevant for those seeking to claim asylum, and also one must consider whether top-line indicators such as GDP reflect either the actual experience of those in the country or the information potential asylum seekers and refugees have about destination countries (c.f. Haldane 2016). Unemployment rates and quantiles of the real wage distribution might provide more information about such targets. Economic metrics are of course measured only with uncertainty and are notoriously difficult to predict into the future.

CEPII, a French research centre for international economics, provides a range of information on the **distance** between states. As well as geographic distance, other less physical indicators such as linguistic similarities and 'gravity' indicators based on trade links are also provided. However, CEPII does not appear to be currently maintained – this is not a problem for distance and linguistic ties, but the latest 'gravity' indicators date from 2006. Relatedly, the size of diasporas already at a destination country can be considered a potential pull factor for migratory flows (Beine, 2011). The OECD Database on Immigrants in OECD Countries (DIOC) provides detailed information on migrant populations in industrialised nations for 2000/01, 2005/06 and 2010/11, including breakdowns by age and education, based on census data and labour force surveys. Although these are not provided by individual years, and only recent data are available, migrant stocks presumably change more slowly than flows, and so such data may still be useful as covariates.

Measures of **human rights** can be considered good indicators of push factors in particular states. Freedom House provides yearly data from 1972 on both civil liberties and political rights for all countries. These two metrics are rated on a 7-point scale by a team of in-house experts, using a broad range of sources. The extremes of the scale reflect ratings of 'Free' and 'Not Free'. The focus is placed on realised rather than theoretical rights, so that the final score reflects not just what the law is but what is actually practiced. The Political Terror Scale adopts a similar approach by synthesising the information found in reports from Amnesty International, the US State Departments and Human Rights Watch, giving each report a quantitative rating out of 5. The difficulties of providing consistent ratings across different cultural and legal contexts and gathering information from such a wide range of countries notwithstanding, these datasets each appear to provide the best proxies that can be reasonably expected for measuring abstract concepts such as these. Hatton and Moloney (2015) found that both the civil liberties component of the Freedom House index and the Political Terror Scale were significant in their explanatory regressions.

Conflict is also obviously a major driver of asylum flows. The Peace Research Institute Oslo (PRIO) and the Uppsala Conflict Data Program (UCDP) provide the Armed Conflict Dataset, which details every post-World-War-II conflict resulting in more than 25 conflict-related deaths together with information on the intensity and type of conflict. These numbers are based on summary estimates of conflict deaths by experts, and so their accuracy is questionable, but the database is relatively comprehensive and has large amounts of historical data. UCDP also produces estimates on the number of battle deaths in conflicts since 1989 using reports from news media and from Non-Governmental Organisations on the ground. These latter estimates are conservative, designed to provide a lower bound on the total number of deaths in a particular conflict, and thus are certainly an underestimate of total deaths. Given that the ability to confirm deaths may differ in unknown ways between different conflicts, the uncertainty in comparisons of deaths across conflicts must be borne in mind.

Summaries of the evaluations of the described variables are given in Table 1, section b. Although the driver variables described above are likely to show correlations with asylum flows, utilising them in predictive modelling remains difficult, due to the problems involved in predicting such variables into the future. In some cases, lagged variables may prove useful in one-step-ahead predictions, particularly if we consider that asylum seekers will not necessary have up-to-date knowledge of current conditions in destination countries (Neumayer 2004). For other variables that change slowly (if at all), such as migrant stock and geographic and linguistic distance, this will not be as much of a problem.

3.4 Meta-information on asylum policies across Europe

Sources and meta-information of information on institutional, policy and legal arrangements relating to asylum are also potentially crucial drivers of asylum flows[7]. As Suriyakumaran & Tamura (2016) describe in their review, economic models of asylum suggest that asylum provision can be considered a public good. Free-riding by unscrupulous states is thus possible by setting tougher standards for incoming claimants than do their neighbours, and thus allowing them to bear the costs of supporting the refugees while benefiting from the externalities of knowing that asylum-seekers have gained sanctuary. Comparable measures of policy strictness are thus important in this regard. Collating information across member states is a significant task, so it will be necessary to rely on existing work in this area.

The current developments in the area of policy analysis most notably include the International Migration Policy and Law Analysis (IMPALA) Database[8] – a result of a project aiming to assemble a comparative database of various migration policy and law indicators in 20 OECD countries. One of the areas of interest is explicitly linked to humanitarian and asylum migration. Beine et al. (2015) present the tentative findings from IMPALA for 1999–2008 for nine countries (Australia, France, Germany, Luxembourg, the Netherlands, Spain, Switzerland, United Kingdom, and the US). The indicators considered for the humanitarian and asylum domain include country-specific numbers of legal 'tracks' under which a person can be admitted (between 6 and 43 in the countries under study in 2008), as well as a number of questions related to possible regulatory instruments (Beine et al. 2015: 9), which are subsequently used in creating synthetic measures (indices) of policy 'stringency'.

The Asylum Information Database provides detailed comparisons of 20 major European countries across three policy areas: Asylum Procedure; Reception Conditions; and Asylum Detention. In each area, countries are rated on a large number of measures on everything to availability of interpreters for asylum interviews, to the availability of appeal procedures, to whether family members are allowed to visit detention centres. While the database is comprehensive and up-to-date, reports only go back to 2013 and thus obtaining time-varying data may be hard. Synthesis along the lines of that conducted by the IMPALA team would be required in order to utilise such information for modelling.

[7] For a critical evaluation of asylum policies in the developed countries, see Gibney and Hansen (2004).

[8] See http://www.impaladatabase.org/, as of 5 July 2016.

An alternate approach is to rely on less direct proxies for the stringency of asylum policy. The 'recognition rate' is sometimes used in the literature, being the proportion of positive decisions in any particular time period (e.g. Neumayer 2004). Such a rate is easily calculable from data provided from Eurostat or EASO and can be lagged to account for endogeneity between policy changes and applications. However, this approach is not without problems; very stringent rules might be assumed to deter migrants who do not have genuine claims, meaning countries with tough policies may end up with *higher* recognition rates than those with loose conditions, but lower volumes. Thus recognition rates may not truly capture the effect of policy as expected.

Synthetic policy indices provide possible proxies for the comparative attractiveness of asylum-granting countries. One potential consideration is that policy changes are often predictable in advance, due to their basis in legislative processes which take time to reach conclusions, and the fact that changes in policies may come into effect only after a delay. Linking a consistent coding of asylum practices with information about planned future changes in legislation and policy might be possible within existing mechanisms such as EASO's Information and Documentation System.

4. Review and assessment of the existing methods for modelling asylum-related migration

4.1 Model evaluation framework and criteria

Throughout this section, the existing quantitative approaches for modelling and predicting asylum-related migration are reviewed and assessed, with the aim to provide guidance for the development of a bespoke EASO model. We begin by setting out the process for evaluating the main approaches available in the literature and other independent sources, looking at a range of factors that may be important for practical uses. Thus, for evaluating the models, the following six criteria are used:

1. **Specification** – how well are the individual models specified in relation to the state of the art in quantitative social sciences?
2. **Resolution** – how detailed is the spatial and temporal resolution of the model?
3. **Assumptions** – how detailed are the assumptions on the drivers of asylum-related processes, and how many aspects are taken into account?
4. **Predictiveness** – are the models intended to be predictive or chiefly explanatory, yet with some predictive potential?
5. **Uncertainty** – is the predictive uncertainty assessed in the models in a formal way, or at least discussed in the narrative?
6. **Generality** – how transferable are the models to other situations, countries, or periods?

As in Section 3, for each criterion, the particular model classes are rated by using traffic-lights codes, ranging from 'green' where the particular aspects of the models are well-designed and conform to the best modelling practice, through 'amber' for average, yet still acceptable quality, to 'red' for the presence of problematic aspects. As the aim of the assessment is to evaluate the available modelling approaches rather than individual models, the assessments have been aggregated for all models in a particular class. In borderline cases, mixed ratings (e.g. 'amber/green') have been used to indicate that some models in a given model class have demonstrated potential to go beyond what is the dominant approach, or that there are differences between the models in a given class.

The quality descriptors for the three assessment levels for each of the six criteria are presented in Table B2 in the Appendix. Additionally, Table B3 in the Appendix contains more specific notes on individual models or suites of models prepared by the same authors or teams, including the listing of variables that have been used in model construction. The variables are grouped into six categories, encompassing the various drivers of asylum-related migration conceptualised in Figure 2. These categories include: policy and politics, conflict, culture, demography and migration, economy, and other (encompassing geography and infrastructure). Within each category, individual variables used in particular studies have been mapped onto a few selected broader indicators, to help identify synergies and conceptual overlap between the models. Thus, for example, visa requirements and points-based systems would be classified together under the 'Policy: immigration overall' header, while the asylum recognition rates, rejection rates and other variables measuring the restrictiveness of the asylum process would be grouped into the 'Policy: asylum-related'.

Finally, there are two additional aspects that are no less important: data quality and transparency. They have been left out from the formal assessment process presented in this section, since data quality as such has been extensively discussed in Section 3, and transparency should be a *sine qua non* condition for the publication of any piece of scholarly research. This is the case for a vast majority of models examined in this study, and any exceptions to that rule are specifically discussed.

4.2 Statistical and econometric methods and models

There exist several methodological reviews of modelling and forecasting methods for overall migration flows (Keilman 2008; Bijak 2010, 2012; Disney et al. 2015). On the other hand, in the academic literature, it has long been argued that asylum seekers and refugees should be modelled and forecast separately from other migration types (de Beer 2008). Comprehensive studies on the *Analysis and forecasting of international migration by major groups*, carried out for Eurostat in the 1990s and early 2000s (Salt & Singleton 1995; van der Gaag & van Wissen 1999; Hilderink et al. 2002) provide – among other insights – a conceptual and empirical background for such models.

In particular, at the theoretical level, Salt and Singleton (1995) identified the "political restructuring" in the third countries as a macro-level process driving asylum seekers into Europe. However, at the empirical level, the predictive potential of asylum trends was found to be largely limited to 'now-casting' of the very short-term trends (van der Gaag & van Wissen 1999). The updated analysis (Hilderink et al. 2002) did not identify specific spatial patterns of flows with regard to destination countries, but in some cases did so for broad groupings of countries of origin.

A useful overarching framework for analysing the different effects for all migration flows, including asylum-related, is offered by well-known and established statistical and econometric methods, falling under the umbrella of generalised linear models (GLM) – various generalisations of simple linear regression models. In migration studies, Willekens and Baydar (1986) proposed an operationalisation of this approach, distinguishing the general effects related to regions of origin, destination, their interactions, and socio-economic and political environment. An example of application is given by Willekens and Baydar (1986: 207), who proposed the following model:

$$m_{ijt} = N_t w_{it} p_{ijt}. \quad (1)$$

In (1), the variable under study is m_{ijt}, migration from country i to country j in year t, while N_t denotes the overall migration level in that year; w_{it} is the *generation effect*, the probability that a migration starts in i; and p_{ijt} is the *distribution effect*, the probability that migrations from i end up in j. In this framework, all elements of (1) can be additionally explained by using socio-economic and other covariates (Willekens & Baydar 1986). Other recent examples of the use of the GLM approach are gravity models for the whole world, where the key driving forces behind migrations are the population sizes of countries (Cohen 2012).

The underlying theoretical framework behind constructing and analysing GLMs is related to the push and pull factors (Lee 1966), conceptualised in Section 2. The background level of forced migration, N_t, is related to the overall political stability worldwide. The push factors or migration triggers – armed conflict, persecution – are related to the generation effects, w_{it}. Then, the pull factors attracting asylum seekers and refugees to particular destinations – policies, legal frameworks, welfare provision, access to certain rights – are linked to the distribution effects, p_{ijt}. Each of these variables can be modelled with appropriate explanatory factors. The overall forced migration levels N_t can be approximated by some random (stochastic) processes, for example autoregressive ones, justified by the self-perpetuating nature of many migration flows, including asylum-related.

In practice, the dominant approaches to quantitative modelling of asylum-related flows present in the academic literature apply some forms of GLMs. The available models differ with regard to their complexity and to the attention paid to particular drivers, factors and the underpinning mechanisms of asylum processes but they share some general-level commonalities. First, all existing models describe macro-level asylum processes. Second, typically the unit of analysis is asylum-related migration into a particular country or a group of countries, with a notable exception of sub-national flows in Czaika (2009). Third, the temporal resolution is usually yearly, again with a few exceptions: monthly in Stanley (1987) and Jennings (2010), and weekly in Schellman and Stewart (2007). The detailed specification of selected quantitative models of asylum-related migration reviewed in this report is presented in Table B 3, in Appendix B, which summarises the relevant drivers and factors of migration taken into account in different models.

Among the formal models of asylum-related migration, three broad groups can be distinguished. Firstly, there are **simple regression models** (univariate or multivariate), estimated for individual countries, usually

with at most a few explanatory factors. Examples include Stanley (1987), Böcker and Havinga (1998), simple asylum-related models presented in a wider study of van der Gaag & van Wissen (1999), or the automatic asylum forecasts of Trading Economics (n.d.). These models do not have a high predictive potential, and their applicability is usually limited to the individual task at hand. In simple models, the uncertainty of predictions is usually not analysed in a formal way. The exact model specification behind the Trading Economics forecasts is unknown – judging by the outcomes, it looks like a polynomial trend, but the model is not documented and is not transparent. From the point of view of modelling asylum-related migration, such models are of limited use.

The second group of methods encompasses **panel regression models**, whereby panels of country-level indicators are modelled across time. Examples of models from this group include Moore & Shellman (2004); Neumayer (2004, 2005a, 2005b); Hatton (2004, 2009); Czaika (2009); Keogh (2013); Czaika & Hobolth (2014), and Hatton & Moloney (2015). The models in this category are typically explanatory, and have been designed with the aim of assessing the various drivers of asylum-related variables. On the one hand, the universe of these models collectively covers a very broad group of various macro-level trends and drivers of migration from different domains (see Table B3 in Appendix B). On the other hand, the predictive properties of such models are limited, and any predictive applications would need to involve predicting the drivers first, which can carry substantial uncertainty on its own (Bijak 2010). The uncertainty of estimates is usually not reported, beyond significance tests for impacts of particular drivers on asylum processes.

Finally, the third group covers the most complex approaches – **time series, structural and log-linear models**. This group is represented by formal time series analysis, either standalone (Schmeidl 1997; Jennings 2010; and Toshkov 2014), or embedded within larger-scale structural model with economic elements (Schellman & Stewart 2007), as well as log-linear models for counts (van Wissen & Jennissen 2008; Vink & Meijerink 2003). Additionally, the studies by de Beer (2008) and van Wissen and Jennissen (2008) present decomposition of overall asylum-related flows into distribution and respectively substitution or distribution effects, akin to the framework presented in (1). Most of the models in this group have clear predictive aims, and sometimes even present quantifiable uncertainty assessment of the estimates (here, a good recent example is Toshkov 2014). The range of explanatory variables in such models is typically limited, though, and they most reliably include past (lagged) asylum or wider migration flows. Only Vink and Meijerink (2003) rely on a completely theory-free model with country-specific and time-specific effects, as well as their interactions.

A detailed assessment of the different classes of models with respect to the criteria introduced in Section 4.1 is presented in Table 2. From the analysis of the properties of various model groups, it becomes apparent that simple regression models are too crude to quantifiably explain or predict asylum-related migration. As for the models from the second and third groups, there are obvious trade-offs between them. While the panel models are typically more comprehensive in terms of drivers, their predictive capacity remains only a potential possibility. On the other hand, the more complex models, involving time series or more advanced log-linear structures, are better suited for prediction, but at the expense of being able to identify the key drivers of asylum-related processes.

Table 2. Evaluation of the existing methods for modelling and predicting asylum-related flows

Model type	Examples	Specification	Resolution	Assumptions	Predictiveness	Uncertainty	Generality	Overall
Simple regression macro-models	Böcker & Havinga (1998); Stanley (1987); Trading Economics (n.d.); van der Gaag & van Wissen (1999)*							
Panel regression macro-models	Czaika (2009); Czaika & Hobolth (2014); Hatton (2004, 2009); Hatton & Moloney (2015); Keogh (2013); Moore & Shellman (2004); Neumayer (2004, 2005a, 2005b)							
Time series, structural and log-linear models	de Beer (2008)*; Jennings (2010); Schellman & Stewart (2007); Schmeidl (1997); Toshkov (2014); van Wissen & Jennissen (2008); Vink & Meijerink (2003)							

* The assessment only takes into account models of asylum, and not models for other types of migration presented in the respective publications.
Key to assessment: **[Green]** Good aspects; **[Amber]** Average aspects; **[Red]** Problematic aspects. For category descriptors, see Table B2 in Appendix B.
Sources: See table and Appendix B

As mentioned before, the GLM framework is one of the most universal and flexible approaches for analysing and predicting asylum-related migration. In the most general form, the GLM family encompassing a vast majority of the existing approaches to modelling asylum flows from country i to country j in time t, y_{ijt}, can be operationalised as:

(2), $$f(y_{ijt}) = c + \sum_{k=0}^{K} B^k\left(\boldsymbol{\beta}_{it}'\mathbf{x}_{it} + \boldsymbol{\beta}_{jt}'\mathbf{x}_{jt} + \boldsymbol{\beta}_{ijt}'\mathbf{x}_{ijt}\right) + \sum_{l=1}^{L} B^l\left(\varphi_i y_{it} + \varphi_j y_{jt} + \varphi_{ij} y_{ijt}\right) + \varepsilon_{ijt},$$

where f is a link function (for example, logarithmic or logistic), c is the constant, B is the backshift operator ($B^m(x_t) = x_{t-m}$), $\mathbf{x}$ denote vectors of explanatory variables, $\boldsymbol{\beta}$ their respective regression coefficients, φ the autoregression parameters, and ε_{ijt} are (some) error terms. This framework is flexible enough to encompass panel regression models, with or without lagged explanatory variables, panel and other time series, as well as log-linear models for counts (with all values of $\mathbf{x}$ set to $\mathbf{1}$, and with no lags).

In terms of drivers $\mathbf{x}$, across all models, the most prevalent explanatory variables are related to migration and asylum policies, politics and governance, trends in asylum and migration from the past, as well as other demographic variables, armed conflict, and economic differentials, chiefly related to income as measured by the GDP per capita. On the other hand, cultural, spatial or infrastructural factors have been examined only in a handful of studies (see Table B3 in Appendix B). In most cases, the data on the drivers are of acceptable quality and come from reputable sources, such as established databases of international organisations (UN, UNHCR, Eurostat, World Bank).

Among explanatory variables, especially noteworthy are attempts to create synthetic indices, in the context of asylum policy (e.g. Theilemann 2004 and Czaika 2005 – standalone pieces of work on policy indices; as well as Hatton 2004 and Hatton & Moloney 2015) and armed conflict intensity (e.g. Stanley 1987; Moore & Shellman 2004; Neumayer 2004, 2005a, 2005b; Hatton 2009; Hatton & Moloney 2015). Importantly, the role of other countries' policies has also been acknowledged in several studies, both from the point of view of analysing their impact explicitly (e.g. Czaika & Hobolth, 2014; Hatton & Moloney 2015), or implicitly, through distribution or substitution effects (de Beer 2008; van Wissen and Jennissen 2008).

In this context, an important aspect of the underlying policy analysis is related to the 'burden sharing' related to asylum between different destination countries. This points to differences in policy norms related to asylum and humanitarian protection across Europe (Theilemann 2003; 2004), which have converged somewhat in the recent years, although not to the point of becoming irrelevant (Toshkov & de Haan 2013). This implies that the policies of other countries within the EU, as well as possibly of the third countries, need to be taken into account in any formal modelling of asylum-related migration even for individual countries. In particular, the variables from the IMPALA database discussed in Section 3 have a great potential in providing quantifiable input to formal models of asylum-related migration, subject to annual time granularity.

A separate question related to models based on the various push and pull factors of migration is: how to predict the individual factors and drivers? The forecasting of demographic and macroeconomic predictors constitute very large topics in their own right, and as such remain beyond the scope of this report. Here, the relevant review studies include: Jennissen (2004) on explanatory economic models of migration; Suriyakumaran and Tamura (2016) on economic drivers of asylum; Bijak (2010), with not very successful attempts to use economic variables in migration forecasting; Keilman (1990) and Alho and Specer (2005), with a comprehensive overview of a range of statistical methods for population forecasting; Bijak (2012) and Disney et al. (2015) with recent overviews of migration forecasting methods. However, in the context of asylum-related flows, political and policy variables – in particular, the presence of armed conflict and policy developments – require at least as much if not more attention than structural macroeconomic or demographic ones.

Some good examples of models for various political and especially conflict-related push factors of migration are available from the political science literature. Examples include multinomial logit models for transitions between different levels of conflict (Hegre et al. 2013) at a country level, or spatio-temporal autoregressive

sub-national Bayesian models with binary response for conflict presence (Weidmann & Ward 2010)[9], both formally acknowledging the predictive uncertainty. The obvious caveat is that the uncertainty of prediction from such models can be very high, but there is a literature consensus on the presence of four key factors leading to conflict: (1) social, economic or political grievances; (2) competition between various groups; (3) readiness to use violence; and (4) lack of non-violent alternatives (Regehr 2014: 16).

The policy-related pull factors can follow the general literature on the role of policies as drivers of migration (Zolberg 1989), including specific examples related to asylum-related migration (Neumayer 2004; Hatton & Moloney 2015). Of course, general caveats about the weak predictive power of migration theories (Arango 2000) remain in force: one key issue is that the drivers of asylum may be even less predictable than asylum as such (Bijak 2010). Besides, with most of the models (with a notable exception of Schellman and Stewart 2007), timeliness remains an issue – for the data may not become available until after the desired prediction horizon. This suggests that such models, which rely exclusively on formal relationships between asylum-related migrations, are not the best choice for operational and planning purposes. With that in mind, the examples of national practice of trying to predict asylum flows in selected European countries are evaluated next.

4.3 National practice in predicting asylum-related flows in Europe

The national practice in trying to predict the inflows of asylum seekers varies between the member states of the European Union (EU). Most of the approaches used to predict future asylum-related flows, and thus demand for the relevant state services, are based on some analysis of the quantitative data coupled with expert knowledge. In this section, several of the approaches developed by the relevant national authorities in selected European countries are reviewed, and are supplemented by more detailed meta-information on the specification of forecasting models used in two countries that are at the forefront of methodological developments: Sweden and Switzerland.

The information reported in this section largely comes from a publicly available *Ad-Hoc Query on Forecasting and Contingency Planning Arrangements for International Protection Applicants* of the European Migration Network (2014). This information is current as of 19 December 2014, and covers responses from 15 EU countries, eight of which reported having some mechanisms for predicting asylum flows in place, and whether relevant contingency plans are in place. The details of the responses of individual countries are summarised in Table 3, together with the relevant meta-information concerning Switzerland, based on the information obtained from the State Secretariat for Migration (SEM, Staatssekretariat für Migration). Similarly, the entry for Sweden is updated based on the meta-information provided by the Swedish Migration Agency (Migrationsverket).

The approaches nine countries presented in Table 3 rely on internally-developed predictions and early warnings, which usually combine quantitative data series on asylum trends with qualitative insights from experts, border intelligence, media, as well as different types of contextual information about the countries of origin and routes. The approaches are not too complex in statistical terms – sometimes being outright descriptive or relying on simple extrapolation of trends – but their main strength lies in the potential richness of different types of auxiliary information that may be difficult to operationalise and quantify in a timely manner. The forecast horizons are usually short – typically up to a year – and are usually updated several times a year. There is not much methodological overlap between the national practice and academic models reviewed in Section 4.2. The countries listed in Table 3 typically either have some contingency plans related to asylum – of a varying degree of complexity – or are in the process of developing and enhancing them. A detailed description of the approaches used in Switzerland and in Sweden is provided in Box 1 and Box 2, respectively.

[9] This is a very advanced, very well specified model, which has been applied to the 1990s war in Bosnia and Herzegovina, although the robustness of its findings is difficult to judge, as the paper does not take into account the body of evidence assembled at the UN International Criminal Tribunal for the former Yugoslavia (Tabeau 2009), relying instead on crude indicators of conflict presence or absence in particular municipalities.

Table 3. National practice in asylum-related forecasting and links with contingency planning

Country	Type of approach	Quantitative data sources	Other sources of information	Contingency planning
Estonia	Risk assessment and data analysis	Various series of statistical data	Border intelligence, information exchange with other countries	Under development
Germany	'Entry forecasts' up to a year ahead	Monthly arrivals of asylum seekers	Experts, countries of origin, EASO COI* portal, government reports, NGOs, media	In place
Ireland	Informal forecasting (extrapolations), trend monitoring	Refugee and asylum flows in Ireland and nearby countries, UNHCR, EASO	Existent refugee communities	Basic
Netherlands	A rolling forecast, updated monthly	National data on refugees and asylum seekers	Under development: to include a wide range of contextual information	In place
Norway	Four forecasts a year, forecasting monthly arrivals	Eurostat, EASO, IOM / IGC**, Frontex, other countries, asylum applications	Media, Frontex, EASO COI*, diplomatic missions	Advanced
Poland	Under development	International and national and data on refugees and asylum flows; UNHCR, EASO	Under development: to include a wide range of contextual information	Basic
Sweden***	Four-five forecasts a year: quantitative predictions with qualitative insights	International and national data; UNHCR, EASO, other countries of asylum, Frontex	Media, NGOs, EASO, Lifos: national COI*, diplomatic missions	Advanced
Switzerland***	One-year ahead expert-based asylum predictions, including two alternative variants	Past trends, quantification of expert opinion, border crossings	Expert opinion on individual countries of origin, migration drivers and routes	Advanced
United Kingdom	Asylum intake reports, regularly updated, with annual forecasts by nationality	Home Office asylum data, tracing the intake reports and provide early warnings	Via involvement in EASO activities	In place

* COI – Country of Origin Information database; ** IGC – Intergovernmental Consultations on Migration, Asylum and Refugees (IOM Geneva)

*** For Sweden and Switzerland, see the detailed descriptions in Box 1 and Box 2, respectively

Sources: European Migration Network (2014); for Switzerland: SEM; for Sweden also Migrationsverket

Box 1. Case study: Switzerland

Model design

The Swiss State Secretariat for Migration (SEM) has been producing year-ahead prognoses of the likely number of asylum applicants to Switzerland for more than a decade, making the SEM process one of the most established in Europe. The yearly prognosis is produced in January and updated in July once the migration 'season' is underway. The SEM prognosis can be considered an expert-based model; formal statistical methods are not used to produce the central forecast. Instead, detailed in-house knowledge is brought to bear to assess how asylum flows into Switzerland may change relative to the previous year.

More particularly, the Swiss approach starts by considering the push factors in sending countries. Country experts from the SEM are surveyed to determine whether they believe flows to Europe as a whole will increase or decrease versus the previous year, and to what extent. As well as giving a range of increase or decrease on a 7 point scale, experts are also asked to justify their assessment.

This assessment of pull factors is complemented by information about changes along the various migratory routes into Europe, incorporating both quantitative information about border crossings and modifications in border controls. An extremely wide range of sources are used to gather this information, including information from partner agencies and countries, all of which are continuously monitored throughout the year. The whole prognostic process is illustrated in Figure 4.

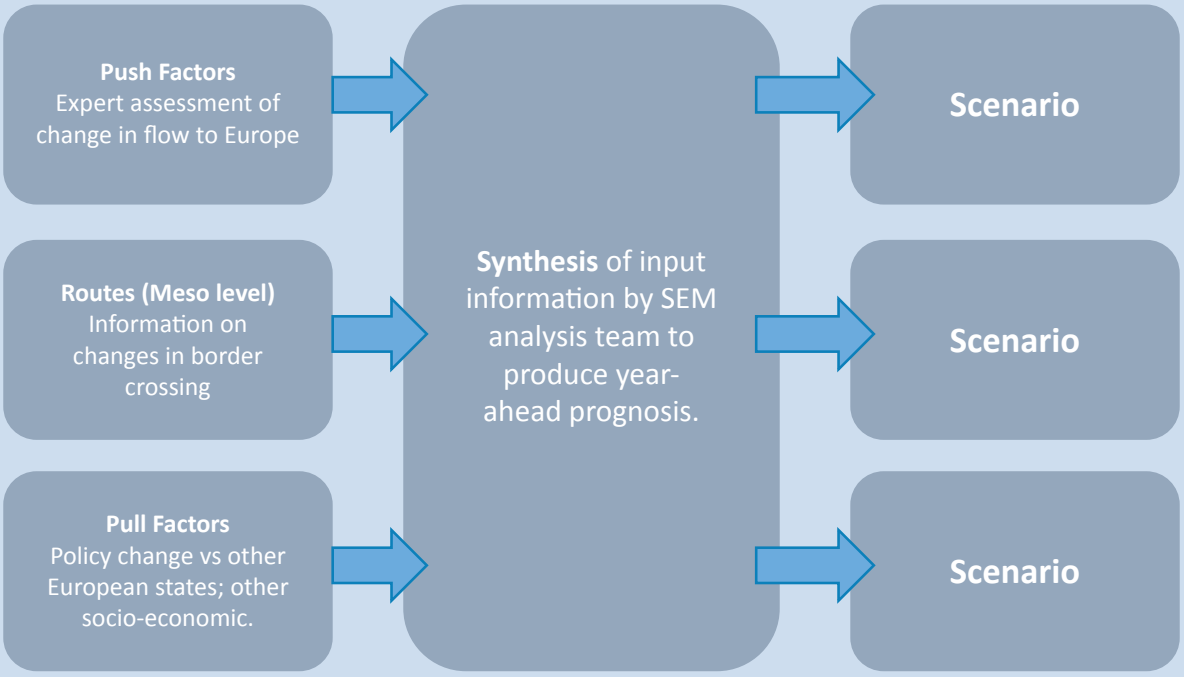

Figure 4. Diagrammatic representation of the Swiss State Secretariat prognosis for asylum flows

Source: State Secretariat for Migration (SEM)

Finally, pull factors in both Switzerland and other European destination countries also play an important role in the SEM's thinking. As well as social and economic factors, asylum policy and the welfare arrangements for applicants are also crucial in the final reckoning.

This information on the various factors determining change in the asylum applicants is synthesised by the SEM analysis team to arrive at their final prognosis. Although production of the model does not involve statistical techniques, an indication of the extent of uncertainty surrounding the year-ahead estimates of application levels is provided in two ways. First, the prognosis describes different possible scenarios, each describing particular changes in determining factors that may lead to different outcomes in terms of asylum outcomes. Each outcome is given a specific (subjective) probability of being realised, based on the opinion of the SEM analysts. Secondly, a "plus or minus" range around each scenario estimate is provided, delimiting a range of approximately 20 % of the central value.

Box 1 (cont.)

Model assessment

The prognosis of the SEM migration analysis team is used by the Secretariat and by the Swiss Cantons for planning purposes, indicating it is trusted in government. In terms of validation against the number of asylum applicants who apply in Switzerland, the prognoses have generally performed well. Exceptions have largely been in years of extreme changes in the European migration system. Specifically, in the aftermath of the Arab Spring in 2011, asylum applicants to Switzerland were understandably underestimated, although this assessment was corrected in the mid-year revisions. Similarly, the migration crisis in 2015 also led to underestimations. In contrast, in 2013, the prognosis was an over-estimation in the number of applications, largely due to a large than expected effect of the disparity between Swiss and other European asylum policy.

Despite its success, there are some problems with advocating such approaches for a Europe-wide system of asylum forecasting. Firstly, although the SEM is admirably clear about its conceptualisation of the migratory system and the determining factors in change in asylum applicants, due to the reliance on the specific expertise and judgment of the analysis team, it is not clear that this could be reproduced by other bodies. Secondly, the difficulties of applying a judgment-based model rather than a statistical/mathematical model are magnified when more than one country needs to be considered – particularly with regard to interactions between asylum policies.

However, the strengths and sustained success of the Swiss model certainly suggest some areas from which we can learn. In particular, the method of surveying experts for opinions on future changes in flows from sources countries is worth pursuing, and could be formalised further as a process of Bayesian elicitation in order to obtain probability distributions over future flow changes.

Source: State Secretariat for Migration (SEM)

The use of expert judgment has also featured in several quantitative academic studies of migration, including predictions of immigration to several European countries (Bijak and Wiśniowski 2010), of environment-related migration to the United Kingdom (Abel et al. 2013), and of possible migration to and from Scotland expected after the 2014 independence referendum (Wiśniowski et al. 2014), as well as in the estimates of intra-European migration flows (Wiśniowski et al. 2013). In most cases, this information was formally elicited from a heterogeneous group of experts within the framework of quasi-Delphi techniques, and was formally combined with quantitative data by the means of Bayesian statistical inference (see Bijak & Bryant 2016 for a review). Expert judgment has also been incorporated – albeit implicitly – in Bayesian predictive models of migration for the whole world, prepared for the United Nations by Azose and Raftery (2015). Additionally, some purely expert-based studies have followed a futures approach, with comprehensive narratives about the possible scenarios of environment-related migration (Government Office for Science 2011); cross-border flows in the EU (Ariely et al. 2011); or migration in the Horn of Africa (RMMS–IMI 2012).

Box 2. Case study: Sweden

Model design

In recent years, the Swedish Migration Agency (Migrationsverket) has developed a comprehensive national migration-related database (Milos), and, with focus on asylum, an advanced mechanism for preparing forecasts of asylum applications. The forecasts are utilised by a number of stakeholders, from internal operations to the Ministry of Justice, and are prepared five times a year: four quarterly and one annual. The forecasts are broken down by weeks, allowing for seasonal variations, with a horizon of one to two years, and are used for planning, policy evaluation and resource allocation.

The current model is based on a mix of qualitative and quantitative information. The quantitative data come from the Milos database and Statistics Sweden, supplemented by the information from European agencies (EASO, Eurostat, Frontex, Eurodac), UNHCR, and other important countries of asylum, with information sharing arrangements in place between the Nordic countries. The qualitative information, in turn, is collated from a wide range of sources, including national database on the countries of origin (Lifos), contact points at diplomatic missions overseas, media or NGOs. This information concerns various push and pull factors of migration flows, an assessment of possible migration routes, and – crucially – of asylum policies and their anticipated changes across Europe, as well as at the EU level. Of particular significance are the temporarily-reinstated border controls in the Schengen area.

The qualitative insights are subsequently quantified by using expert judgment through an application of *morphological analysis*, to produce the forecasts, which are then updated on a quarterly basis. The method, dating back to Zwicky (1948, 1969), relies on a systematic approach to coherently assessing the structure of multi-dimensional complex problems. The process is summarised in Figure 5.

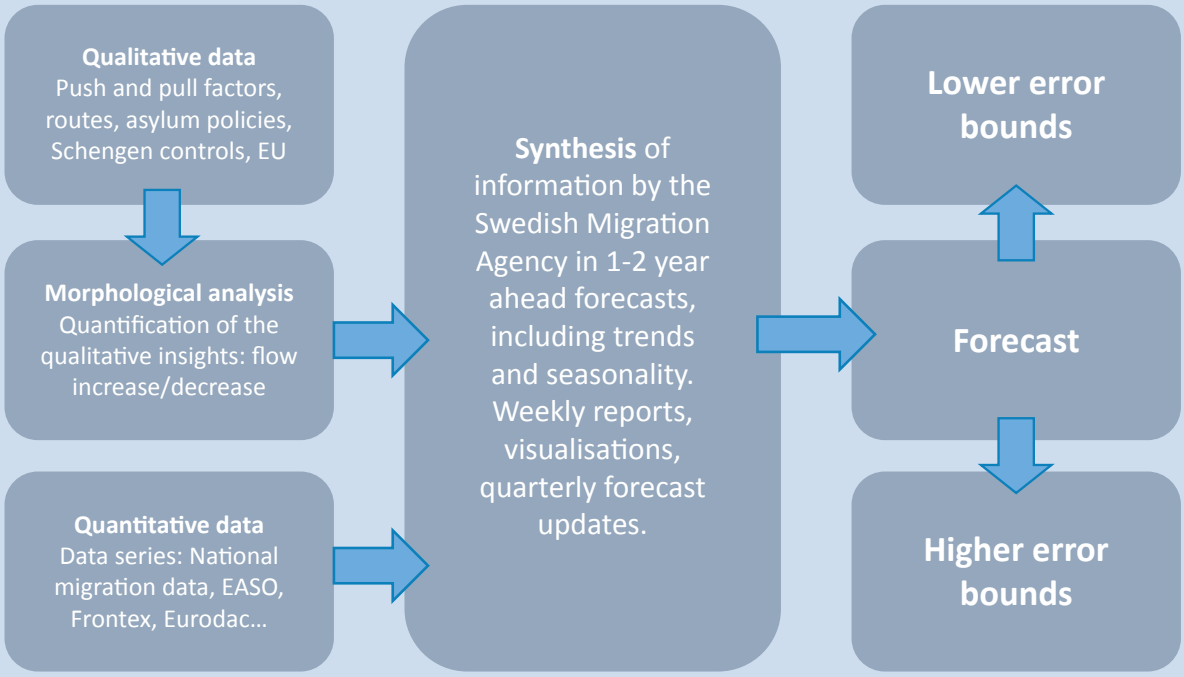

Figure 5. Diagramatic representation of the Swedish forecasts for asylum flows

Source: Swedish Migration Agency (Migrationsverket)

Based on the main forecast, the error bounds are subsequently derived, aiming to cover with 90 % probability the number of asylum applications in the next forecast period.

The Swedish Migration Agency is also developing a comprehensive system for predicting flows and modelling the circulation of migrants with Sweden, with the aim of optimising resource allocation and increasing the preparedness in case of sudden changes in the volume of flows. This experimental system will combine extrapolative models (time series, exponential smoothing, and possibly some machine learning techniques) with queuing theory and a statistical model for the spatial distribution of flows. Through building a suite of interconnected formal models, the agency aims to further improve the predictive capacity of the existing approaches.

Box 2 (cont.)

Model assessment

The forecasts prepared at the Swedish Migration Agency are routinely monitored against the real-life developments, and have so far proved to be relatively accurate *ex post*, especially as for such a volatile phenomenon as asylum-related migration. The main drivers of this accuracy are a comprehensive set of quantitative and qualitative input information, a high level of detail of the background analysis behind the forecasts, and periodic updating of predictions. The forecasting and monitoring tools also enable analysis of the asylum flows at a very high level of detail, for example by individual nationalities, routes or channels of entry into Sweden, etc.

The numbers of applications being monitored and forecast are slightly biased due to including multiple entries for the same persons, for example when they change 'track', switching from other types of migration to asylum, or in the case of circulation of asylum seekers between different EU countries. Children born to asylum seekers and refugees in Sweden are also included in the data, even though technically they cannot be classified as 'new arrivals'. These issues are recognised, and the users of the forecasts need to be aware of these shortcomings.

One obvious limitation of the Swedish approach is that it is very resource intensive, both in terms of human effort, as well as the required data infrastructure. Still, the impact of the forecasting work done at the agency is not to be underestimated: by providing accurate predictions, and by contributing to national contingency planning, it can generate efficiencies throughout the Swedish asylum system.

Source: Swedish Migration Agency (Migrationsverket)

The review of national practice in asylum forecasting has revealed four important features of the methods used. First, even among those countries that carry out forecasts of asylum-related migration, there is a wide disparity in terms of complexity and sophistication of the methods used. However, the best available models and contingency plans – such as those prepared in Sweden or Switzerland, but also Norway or the United Kingdom – offer quite comprehensive tools to prepare resources for prospective asylum flows. As such, there is much scope for harmonisation of the various approaches across the EU and for learning from best practices. Second, most of the predictions rely on expert opinion but its level of formalisation varies from country to country. Third, the approaches usually take into account a wide range of contextual information about asylum processes. Fourth, with a few exceptions, the examples of national practice in predicting migration are typically more flexible and timely than academic studies, which are usually based on secondary data published with a time lag. These features, especially the formal inclusion of expert judgment and of the detailed intelligence about the context, as well as timeliness, are worth retaining in the eventual construction of the bespoke EASO model of asylum flows.

4.4 Notes on the validation of models

An important question pertaining to all the models and approaches reviewed in this section is how they and their performance were – or could have been – validated by using real data. There are two main approaches to validation: *ex ante* and *ex post*. In the ex-ante approach, the *expected* properties of the model, such as estimation errors, goodness of model fit, etc., are evaluated on the basis of the same dataset on which the model is estimated. In the ex-post approach, the model predictions are compared with the actual observations that go beyond the data sample, which allows for calculating different measures of error, based on the differences between the predictions and observations. Here, usually some summary measures are calculated, such as mean percentage errors, mean absolute errors, or (root) mean square errors, based on averaged across all out-of-sample points, on the basis of which the validation is made (cf. Bijak 2010: 161).

For probabilistic methods and models, which describe their predictions not in terms of single values, but rather whole probability distributions of the possible future outcomes (Alho & Spencer 2005; Bijak et al. 2015), there are a few additional considerations. Firstly, an open question is which characteristics of the predictive distributions should be used for calculating the errors: typically these are mean or median values from the distributions, but other points can be also used, depending on the actual needs of the forecast users (*idem*). Secondly, the predictive distributions themselves need to be well calibrated, in that the distributions obtained from the modelling should match the empirical frequencies of the actual errors

(Gneiting et al. 2007). In other words, in well-calibrated models, the errors ex ante and ex post should match: for example, 50-per cent predictive intervals should cover the true values ex post around half of the time, and so on.

Among the academic models reviewed, only some ex-ante model validation has been usually done, mostly through determining how the various models fit the data at hand. In hardly any case an ex-post analysis of errors was attempted (with very rare exceptions, such as the conflict model of Weidmann & Ward 2010). Even though conducting an ex-post validation exercise for a wide range of models and for different data series is beyond the scope of this study, Disney et al. (2015) have undertaken such evaluation for a range of migration-related data series for the United Kingdom, including the number of asylum applications based on the Home Office data. Several autoregressive (AR) and autoregressive moving average (ARMA) models have been estimated based on the data for 1984 – 2003, and the resulting forecasts for the subsequent 10 years, 2004–2013, were compared to the actual observations from that period, yielding the results shown in Table 4.

Table 4. Results of the validation exercise of Disney et al. (2015) for the forecasts of UK asylum seekers, for 2004–2013

Model	Mean Percentage Error	Mean Absolute Error	Root Mean Square Error	Calibration: 50-percent interval	Calibration: 80-percent interval	Summary of errors	Interval calibration
Classical (frequentist) time series extrapolation models							
Random walk	–110%	25,293	25,590	45%	100%	Very high	Some issues
General AR(1)	–49%	11,352	12,185	82%	100%	High	Too conservative
ARMA(1,1)	–13%	4,116	4,867	100%	100%	Medium	Extremely wide
AR(1) *differences*	–70%	15,950	16,286	100%	100%	Very high	Extremely wide
AR(1) *de-trended*	–1,446%	328,091	391,522	0%	0%	Very high	Extremely tight
Bayesian extrapolation models							
Random walk	445%	99,096	113,061	0%	40%	Very high	Too optimistic
General AR(1)	–98%	22,339	22,598	0%	100%	Very high	Miscalibrated
ARMA(1,1)	–85%	19,281	19,585	40%	100%	Very high	Some issues
AR(1) *differences*	–330%	73,577	82,831	0%	100%	Very high	Miscalibrated
AR(1) *de-trended*	–368%	82,425	90,115	0%	100%	Very high	Miscalibrated

Source: Adapted from Disney et al. (2015: Table B5, p. 60)

The interpretation of the above results provided by Disney et al. (2015) confirms the nature of asylum-related flows as one of the worst predictable migration processes. Still, in the context of practical applications of asylum modelling, validation can be more than a one-off exercise to confirm the suitability of a particular model. There is a gap with respect to how much is learned from especially ex-post validation exercises, and how useful such approaches could potentially be. A more comprehensive approach would have the validation exercise built into the forecasting as a part of a wider framework – regularly updated as new data become available, as in several examples of national practice discussed in Section 4.3. There exists advanced statistical theory that formalises such a framework – the *prequential* (*pr*obabilistic and *sequential*)

forecasting (Dawid 1984). For practical applications, however, if the framework was to be adopted across Europe, a simpler pragmatic solution may be preferable, which would formalise what is already common practice in some countries, and augment it with expert opinion, within a joint statistical framework. The feasibility of such a modelling approach is discussed in the next section.

5. Towards a bespoke EASO model of asylum-related migration: A feasibility study

5.1 Synthesising the possible modelling frameworks

As a first step in assessing the feasibility of various modelling approaches for asylum-related migration flows, in this section we present the communalities and desirable features of the various models and methods identified earlier. In particular, the aim of this section is to synthesise the meta-information on the available data and models, and to analyse the potential areas of overlap or synergy between the modelling approaches. This assessment is subsequently used to guide the assessment of feasibility of the various approaches, in the context of their appropriateness for the development of a bespoke EASO model of asylum flows.

As mentioned, the existing quantitative models of asylum-related migration are macro-level, with a varying time resolution (from weekly to yearly). They encompass explanatory approaches (panel models), predictive ones (time-series models), and expert-based endeavours (best examples of the national practice). In the synthesis, we recommend retaining some elements of all three types of methods while bearing in mind their practical limitations. To start with, the prediction of all individual drivers of asylum-related migration is not a feasible task, as it is likely to introduce even more predictive uncertainty to the model than the asylum-related flows alone (see also Bijak 2010). Still, some information on especially policy-related drivers is crucial.

In the light of very weak predictability of forced migration, the proposed approach needs to take a form of an **early warning system** (Clark 1989; Schellman & Stewart 2007; Regehr 2014; Disney et al. 2015), rather than making futile attempts at predicting longer-term asylum trends. Given the high level of uncertainty, the use of Bayesian statistical methodology is especially recommended. The Bayesian approach updates the *prior* beliefs about the phenomena of interest with the information from data, in order to produce the *posterior* estimates of model parameters and derive the forecasts on their basis (Bijak and Bryant 2016). As such, it is well suited for the description of estimation and forecast errors in such circumstances, while also allowing for including structured expert knowledge in the models in a formal, open and transparent way (Bijak 2010; Azose and Raftery 2015).

There are several important arguments for including expert opinion. First, when the data are sparse or time series are short, they can be formally supplemented by expert judgment. More generally, in the Bayesian approach, there are important trade-offs between the prior information and the data: the more data are available, the less influential the prior assumptions are, and the less weight they have in the final estimates – and vice versa (see e.g. Gelman et al. 2014 for a detailed discussion). Secondly, given the high uncertainty of migration flows, expert information can provide insights that are not included in the data sample. In that way, expert-based prior information, for example on low predictability of migration flows or on similarities between various types of flows, can lead to better calibrated assessment of forecast uncertainty and help the users avoid the pitfalls of spurious accuracy of prediction (Bijak 2010; Azose and Raftery 2015). In the proposed model, the role of experts would be also to provide better timeliness and agility of model amendments beyond what is permitted by the available data, as done in the Swiss and Swedish models (Section 4.3).

Hence, in general terms, a macro-level early warning model could be based on the **time series of EASO data**, supplemented by synthetic asylum **policy indices** for a range of countries and regions, in order to capture some of the distribution or substitution effects in asylum processes. Other important variables, such as armed conflict intensity measures for different regions of the world, could also be used. Another key input would be **expert opinion** on the envisaged changes of individual asylum-related flows within a pre-defined horizon, for example one year ahead, as well as stakeholder views on the **warning thresholds** – what levels of asylum-related migration should trigger a reassessment of the model or a policy action.

The asylum data used in the model would be origin-and-destination specific, the expert knowledge and conflict intensity measures would be related to countries of origin, and the policy indices to the European countries of destination. A **two-stage model structure** would reflect that, possibly through adopting a GLM approach in the spirit of Willekens and Baydar (1986), mirroring the conceptual distinction between the push and pull factors (see Figure 2, Section 2), and possibly modelling the intervening factors through origin-destination interaction terms. The various elements would be combined via **Bayesian statistics**, and inferences would be drawn on the basis of an overall model, taking the warning thresholds mentioned above as a basis for decisions.

The key ingredients of the proposed approach are summarised in Figure 6. An example of the model design and structure is described next.

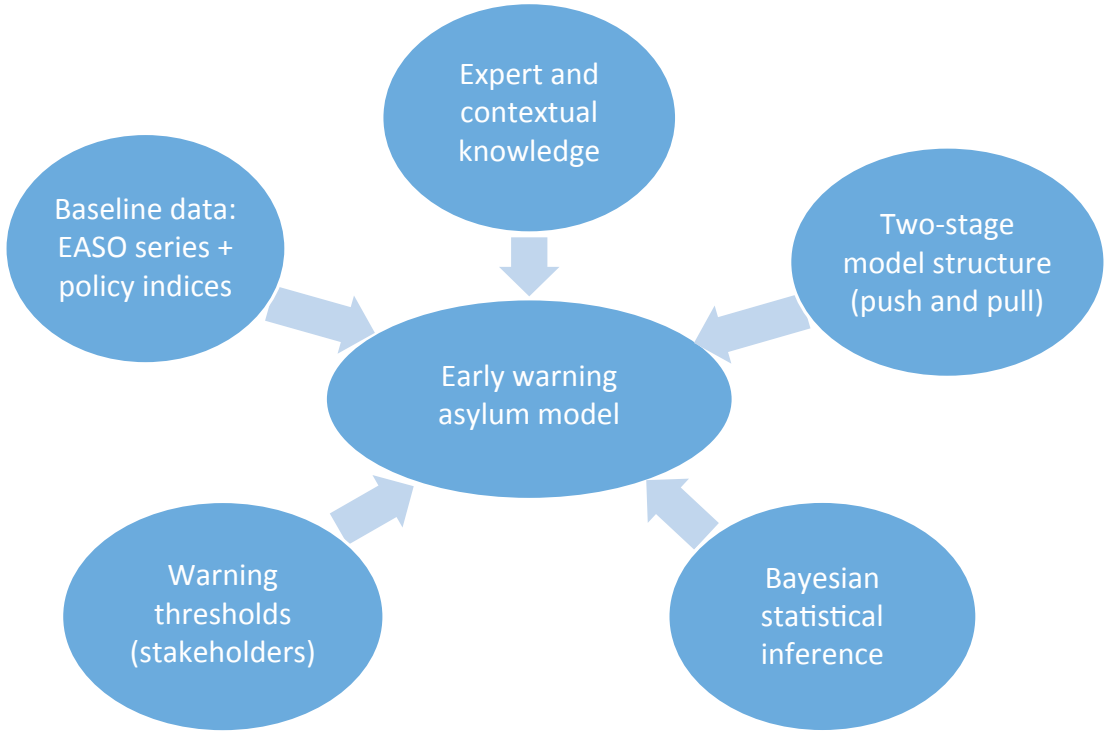

Figure 6. Elements of the proposed modelling approach

5.2 Design of the bespoke EASO model structure

Based on the knowledge of the asylum processes and the state of the art in formal statistical modelling, in this section we propose a class of models that may be potentially used for designing the EASO model at the later stages of the research programme.

Let us start with the dependent variable: given very high volatility of asylum-related outflows, it is recommended to model weekly or monthly **changes** in the level of asylum applications. The changes will likely exhibit more stable properties than the flow levels (see also Disney et al. 2015). This approach requires timely and good quality baseline (current) data and up-to-date information on policies across different countries, as well as on the new developments in terms of push factors, migration routes, smuggling channels, etc. – all of which would be very difficult to operationalise in with quantitative indicators but could be approximated through expert knowledge.

One of the key purposes of the model will be to provide early warning signals through constant monitoring of asylum-related flows. The question underlying the modelling process would be: to what extent are the asylum processes stable, disregarding the impact of events identifiable at the time of model estimation? If there is evidence that the processes become unstable this should lead to a policy action as well as to a re-evaluation of the model. Hence, for the time resolution, the asylum data would need to be as detailed as possible: ideally weekly, or at least monthly. In our view, the time horizon of the model should be limited to one year ahead, first, this is a period typically related to budget and planning, and second, given the high volatility of asylum-related migration, it may not be possible to create reasonable forecasts in a longer perspective.

Convenient methods for answering the question above are offered by the statistical control theory, in particular the cumulative sum (*'cusum'*) approaches (Page 1954, 1961). In this approach, the indicator of choice (in our case, asylum levels) is traced by cumulatively adding the subsequent observations over time, checking whether such cumulative sums (cusums) exceed certain thresholds, and making decisions on that basis. Several formalisations of the cusum model exist; for example, Harrison and Veerapen (1994) discuss a three-element decision space, with process being either 'in control' under a certain *acceptance threshold*, μ_0, or 'out of control' above some *rejection threshold*, $\mu_1 \geq \mu_0$, or requiring further monitoring for the values between μ_0 and μ_1. The thresholds can be determined based on the past performance of the indicator, or alternatively example by using some formal model. Denoting the process value at time t by y_t, the cusum C_t can be defined recursively, through the following equation, starting from $C_0 = 0$ (Harrison and Veerapen 1994: 30):

(3) $$C_t = C_{t-1} + y_t - (\mu_1 + \mu_0)/2 \text{ for } t > 0.$$

A stylised illustration of the approach is illustrated in Figure 7, where the process triggers a warning by crossing a rejection threshold in the 20th observation, shown by red vertical lines on the process graph (left panel) and the corresponding cusum chart (right panel). As stipulated already by Page (1954), a downward path of a cusum indicates satisfactory performance of the indicator, and an upward path denotes an unsatisfactory performance.

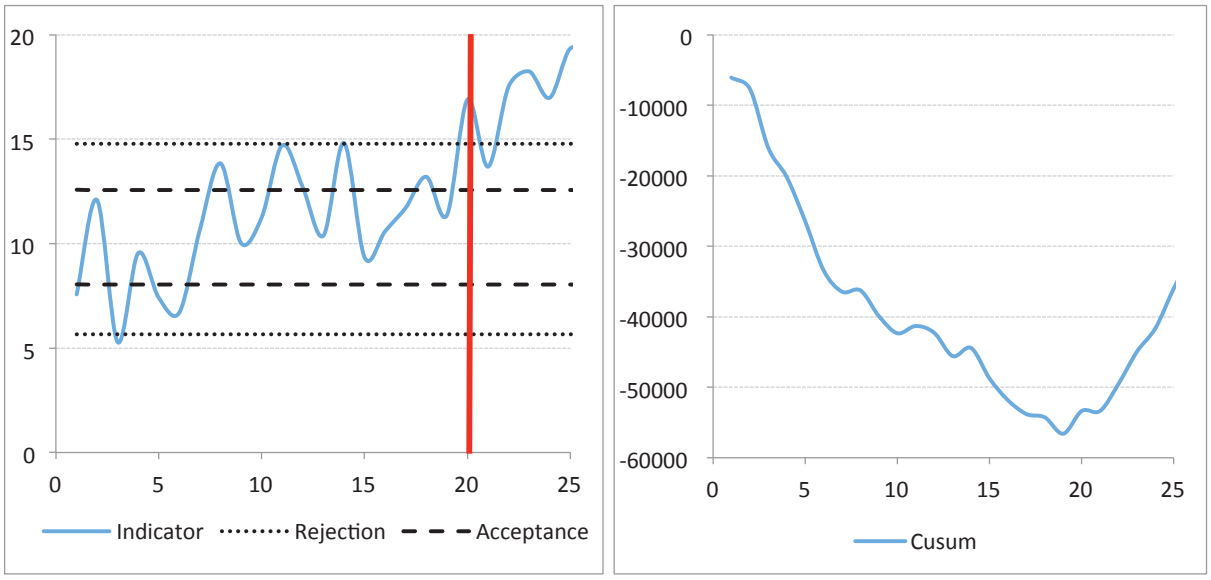

Figure 7. A stylised cusum analysis – the process (left) and the corresponding cusums (right)

There are different formulations of the statistical control problem, including those involving Bayesian inference (West 1986; see also a recent review in Tartakovsky and Moustakides 2010), and also natural links with the statistical decision theory (Harrison & Veerapen 1994). One important premise of the application of the method is an appropriate setting of the thresholds, which rely on good calibration of the underlying model and on appropriate trade-offs between 'false alarms' (*Type 1 errors*) – issuing warnings based on

observations that ultimately prove to be following the trend, and 'complacency' (*Type 2 errors*) – not generating warnings when there are genuine shifts in the asylum trends. As remarked by the original creator of the cusum method, ES Page, these errors can be measured by the expected length of time the process spends respectively below or above the threshold before the methods triggers a warning which leads to a possible action (Page 1954: 101).

To obtain the one-year-ahead forecasts, which we suggest should be monitored using the above techniques, we propose that for each sending country experts are asked to synthesise the relevant push-factor information (together with their assessment of likely changes) into an estimate of the change in applications N for the year ahead. Use of formal elicitation methods (e.g. O'Hagan 1998) allows uncertainty in this estimation to be accounted for, and thus obtaining a distribution over the possible extent of change, which can be assumed to be normal:

$$\Delta N_{it} \sim N(n_{it}, \sigma_{it}) \tag{4}$$

The subscripts *i* and *t* here refer to country of origin and year respectively. A normal distribution is assumed for illustrative purposes but in practice other distributions might be used. Given these changes and their implications for levels of future applications, the overall process can be modelled as exhibiting the following structure:

$$Y_{ijm} \sim \text{Poisson}(\mu_{ijm}) \tag{5}$$

$$\mu_{ijm} = N_{it}\pi_{ijt}S_{m(t)}.$$

In (5), N_{it} represents the implied elicited outflow from country *i* in year *t*, and $S_{m(t)}$ describes the distribution of this flow across months *m* of the year, easily transferrable to weekly or even daily data. The estimated dyad-specific effects π_{ijt} describe how flows are split across receiving countries *j*, and incorporates both policy indicators for recipient country *j*, and historical and geographical information describing the tendency for refugees from country i to apply for asylum in country *j*. As the model is probabilistic in nature, its outcome will consist of full probability distributions of the future values of N_{it} and Y_{ijt}. Based on those distributions, the various point estimates can be derived, such as predictions of the central tendency (mean, median) and of other contingencies (for example, based on quantiles), measures of dispersion (standard errors, inter-quantile ranges), and so on.

The matrix $\mathbf{\Pi}$ incorporating the dyadic effects π_{ijt} allows for the capturing of dependencies between flows. For example, if a recipient country *j* tightened its asylum policy toward country *i*, this would manifest itself in a reduction in the respective π_{ijt} term. Given that these terms describe the distribution over recipient countries of the expected inflows N_{it} from country *i*, this will result in a redirection of some of the flow from *i* to *j* to other countries, in proportion to their attraction to migrants from the sending country (represented by the other terms $\pi_{i.t}$).

Several extensions can be proposed to this basic model. Firstly, the Poisson assumption can be relaxed to allow variances to be over-dispersed with respect to the mean count estimate. This is likely to give a more realistic account of the uncertainties inherent in the model. Secondly, in order to account for the importance to the eventual destination of the choice of routes into Europe, we would suggest separate $\mathbf{\Pi}$ matrices and seasonal distribution effects $S_{m(t)}$ be estimated for each route. Then, migrant flows can be probabilistically distributed across routes according to past trends, Frontex data, and expert opinion.

Monitoring the model on a weekly or monthly basis means that parameters can be updated using Bayes' rule as more information becomes available, thus allowing the predictions themselves to be altered as new data come in. This applies equally to several elements of the model: the prevalence of different routes into Europe can be updated as further Frontex information becomes available; and the seasonal and flow levels may also be updated in light of new EASO data. Similarly, other auxiliary information about the asylum-related migration process might be incorporated into the model, either through periodically asking experts to re-evaluate their assessments, or through more formal arrangements. For instance, noisy data may be available on numbers of migrants present in a border country such as Greece or Italy but who have yet to make an application for asylum (perhaps because they prefer to apply in other member states). Such

data could inform expert assessments of change in flows, or it might be used to formally update elicited assessments based on past relationships in the data, or it might be utilised to inform about route choice, as with Frontex data.

In terms of the workflow, the process could be summarised in the following stages. At first, the necessary information needs to be assembled, in the form of quantitative data, expert knowledge and stakeholder input as to the sensitivity of the warning thresholds. On that basis, the model is estimated and is updated every time new data on asylum become available. At this stage, the statistical control mechanisms are also triggered: if the process is found to be still 'in control' (below the threshold) or if it requires monitoring, no action is taken until the successive data become available, when the situation is being reassessed. If, however, at any stage the process is deemed 'out of control' (above the threshold), this triggers a policy warning to the stakeholders and a re-evaluation of the whole model, including a re-assessment of the situation by the experts. The control is then re-set, and the evaluation cycle resumes as described above. The particular elements of the workflow are graphically summarised in Figure 8.

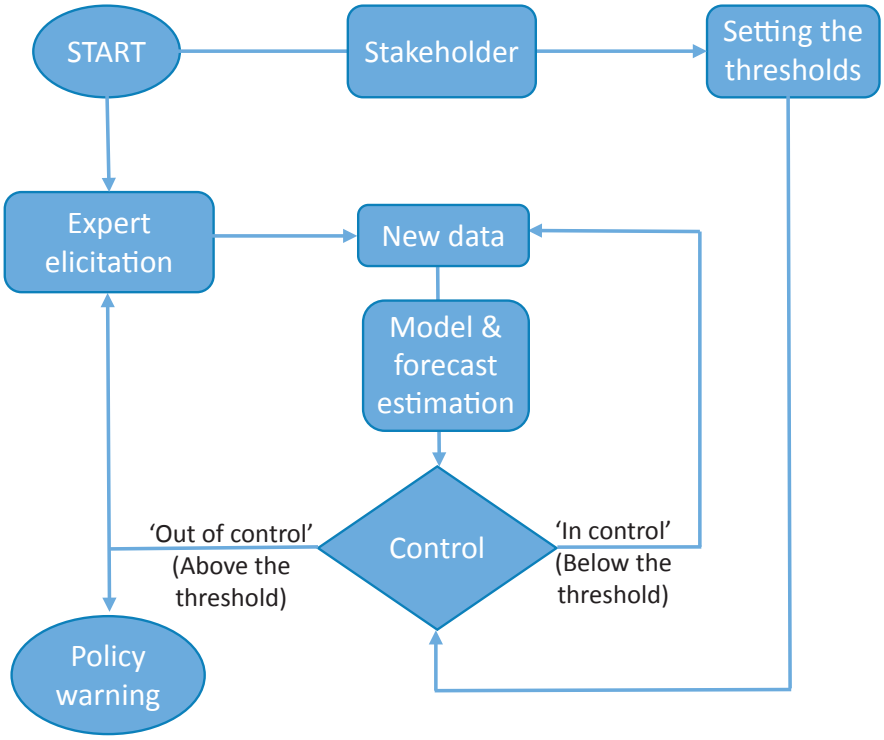

Figure 8. An algorithmic representation of the workflow in the early warning asylum model

5.3 Identifying the data/information gaps and limitations

In this section, we look at the specific data and information gaps for the particular models proposed above, and make suggestions on the ways of filling these gaps, through additional analytical work and data collection. We also look into possible resource implications of such endeavours.

As mentioned, the key quantitative variable in the model would be the EASO estimates of asylum flows at the highest available frequency. In terms of further model development, it would also enable EASO to update the model very quickly, making use of the synergy with data collection efforts that are already in place. One question that arises is whether the **outcome measure** for asylum applications should be applications made or applications lodged. As described in the data section, the 'applications made' metric is less likely to be affected by differences in administrative procedures between countries, and so should give a better indication of the timing of inflows. However, this metric has only begun to be collected consistently in recent periods, so there may not be enough data to inform the various elements of the model. However, as more data become available, using the 'applications made' measure may become more practicable, and it may also be possible to model the relationship between the two measures.

As for other quantifiable data, a key modelling priority is to keep abreast of the **asylum policy developments** across Europe and in other important countries of destination. This implies the need to create and update synthetic policy indices: an excellent starting point is offered by the IMPALA database, but for the purposes of the EASO model it would need to be extended in terms of geography, possibly at the expense of some of the detail. Then, in order to be useful in the modelling process, the indices would need to be regularly monitored and updated. All the related tasks could be performed in-house by EASO, by having a dedicated ca. 10 % of a researcher's time.

A more complex task has to do with **expert knowledge elicitation**. This would require assembling a pool of reliable experts on country-specific issues with respect to origin regions of asylum seekers, whose expertise could be called upon at a relatively short notice. These constraints imply that there ideally should be a restricted-access online survey, pre-tested as a part of further lots of the EASO research programme to ensure as unambiguous questions as possible. Ideally, a wide range of areas of expertise should be considered, and the expert answers would enter into the model in a quantified form. In principle, it might be beneficial to allow for multi-round elicitation, such as in a Delphi framework, but practical considerations are likely to limit the time available for the elicitation exercise. Through a survey, the expert knowledge could be elicited quickly and efficiently. Examples of migration-related formal elicitation frameworks are given in Bijak and Wiśniowski (2010), Abel et al. (2013), and Wiśniowski et al. (2013, 2014).

Another important task related to the modelling process described above consists in specifying the **warning thresholds** for the models. Ideally, the thresholds would need to be specified through a separate elicitation process, this time from policy stakeholders. One of the main aims of the elicitation exercise would be to establish trade-offs between the costs of 'false alarms' and the costs of 'complacency'. These trade-offs could be subsequently formalised by using statistical decision analysis, to translate them into probability levels for warning thresholds (see Alho & Spencer 2005: Chapter 12.4, and Bijak 2010: Chapter 11, for further discussion and examples). For example, warnings could be triggered by 'once in two years' events (1 % probability threshold for weekly data), 'once in a decade' events (ca. 0.2 %), or at other levels. As the specification of thresholds would be a one-off task, or at least would not require routine or periodic updating, we recommend convening a meeting with the representatives of the member states and other interested parties, where the issues related to thresholds could be discussed and agreed upon. This is especially crucial, as warnings may have important financial implications, so there needs to be a consensus amongst the model users, at which probability level the warnings should be triggered.

Given that the recommended models are macro-level early warning mechanisms, there is no pressing need for carrying out additional survey-based studies amongst the asylum seekers and refugees. While the information that could potentially be gathered is important for a variety of operational reasons – for example to identify priority areas of service provision – it is unlikely to give new, game-changing insights into the dynamics of the asylum processes under study. In that respect, longer-term model development goals may include monitoring of social media, as well as other approaches designed to harvest information

from various 'big data' sources (some suggestions are provided in Section 6). However, at the first phase of model development this is something that does not need to be done independently and could be subsumed within the broader task of expert elicitation.

5.4 A simple illustration of the framework

To demonstrate the workings, information requirements, envisaged outputs, and the resulting application of the proposed approach for decision making for single countries, we present the workings of a simple version of a prototype control model using data from the Swedish Migrationverket (migration agency) on flows into Sweden from Syria and Iran. Taking the differences in log-counts, or equivalently, the log ratios of successive observations, we monitor how flow growth evolves over time for these two sources of applications. In reality, elicitation exercises would allow meaningful boundaries to be set determining when the flows are in and out of control (Harrison and Veerapen, 1994). Lacking this information, the current exercise uses a standardised version of the cusum measure given in Equation 6 below, and takes the decision boundary beyond which a loss of control is flagged as lying five standard errors above zero (Page 1954; Scrucca 2004).

(6) $$S_t = \max(S_{t-1} + x_t - k, 0)$$

This is similar to the cusum measure described above but acts to reset the cusum to zero each time a new minimum value is reached. The value k determines what size of shift is picked up by the cusum monitoring process. In this case, one standard error is used. Figure 9 below displays the evolution of this measure for both Iran and Syria. Each chart contains two measures, one above the x axis examining positive shifts in log-ratios (that is, increases in rate of growth of the number of asylum seekers), and the other displaying negative shifts, plotted below the axis. The red areas of the plot mark boundaries of the region considered out of control. When put into practice, monthly data would be used, and the latest observations would be model-generated forecasts, but for this example, actual, annual data are used to aid illustration.

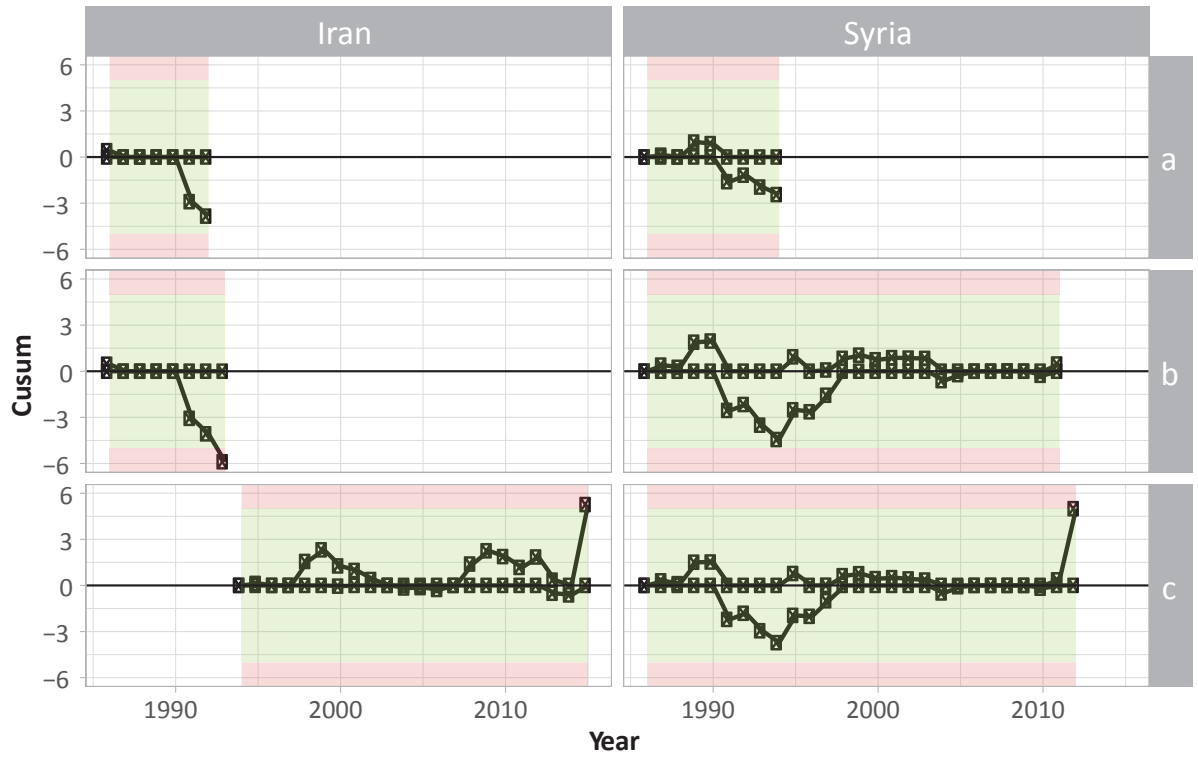

Figure 9. Cusum control plots for Syria and Iran

Examining Figure 9, we can observe how a monitoring process would work in practice. Starting with Iran (left column), we see that initial the process is in control (panel a, top row), despite some decreases in

rate. Looking at the panel b (second row), we see that in 1994, a significant downward shift in the rate is detected and the model is recalibrated. In panel c, the process remains in control until a sharp increase triggers a policy warning in 2015.

Moving to Syria (right-hand column), we see that in the initial observations (panel a), we observe some slight downward shifts but not enough to trigger a recalibration. The process remains under control throughout the later 1990s and 2000s (panel b), but the effects of the civil war are picked up by the monitoring process, and trigger a warning in 2012 (panel c).

The above description gives examples of how model forecasts could be used to trigger policy warnings when significant shifts are detected in asylum flows. A final model would take into account uncertainty in predictions and flows and allow calibration of warning thresholds according to the preferences of policy makers.

6. Conclusions and recommendations

There are a few key conclusions from the data and methods review conducted in this study. Firstly, in the context of asylum-related migration prediction is difficult in the light of the presence of 'hard', unpredictable push and pull factors, such as armed conflict or political change (Öberg 1996). As is evident from the review, most formal models of migration can help explain the patterns seen in the data, but on their own do not have much predictive power. This confirms the findings of other studies related to modelling and forecasting of migration flows (Bijak 2010; Raymer et al. 2013).

For that reason, in terms of recommendations, we propose that the future EASO work programme concentrates on specification of early warning models tailored to the specific needs of user at the EU level. We propose a framework based on the tenets of statistical control theory, tailored to the specific needs of the model users. In particular, different features can be assessed for whether they remain 'in control': observations, for monitoring purposes; forecasts, for pro-active contingency planning; or model errors, for assessing, whether the model needs respecification. Depending on the specific needs of EASO, and on the available resources, the models can be estimated and evaluated at different levels, from all-European aggregates, to individual flows of asylum-related migration.

One of the key limitations of the proposed modelling work is that they cannot be seen as predictive in the formal sense. The practical implication of the low predictability of asylum-related migration is that the policy decisions will be always inevitably be made under the conditions of uncertainty. Even though there are techniques for supporting decision-making in uncertain circumstances, for example statistical decision analysis (see Alho & Spencer 2005; Bijak 2010), the fundamental uncertainty remains a crucial issue, which the policymakers need to be fully aware of. Hence, key prerequisites of success of any formal modelling approach include an appropriate communication of the model aims, results, uncertainty and limitations to stakeholders, in an accessible, yet at the same time honest and transparent way. The traffic-lights approach and risk management matrices, as used in the migration context, for example by Disney et al. (2015), and adopted for the evaluation of data and methods in this study, can provide a very useful framework in that respect.

There are some natural extensions of the proposed approach that can be accommodated in the next stages of the EASO research programme. One area that is worth exploring in that context is related to 'big data' from social media, electronic communications, and other sources, to enhance the early warning capacity of the models. As is well documented in the accounts of the asylum seekers' journeys (Kingsley 2016), new technologies are widely used by people seeking asylum to navigate through the challenges and obstacles along the way. The analytical potential of these sources remains so far unused but it can be potentially harnessed for the benefit of the whole European asylum system, provided that appropriate ethical and privacy safeguards are put in place, to protect the rights of individuals while allowing greater and timely insights into population-level phenomena.

Besides the practical recommendations listed above, there are many other aspects of asylum-related migration that would benefit from further and more detailed research, which goes well beyond the EASO work programme but should be of substantial academic interest. One aspect is related to the flip-side of asylum-related migration, which is immobility even in the face of dramatic events, armed conflict or persecution. Similarly to the findings of the Government Office for Science (2011) in the context of environment-related migration, it can be argued that the most vulnerable populations – those who do not have enough resources to move – remain 'trapped' in very adverse conditions. Migration and mobility are important strategies through which people are able to adapt to difficult circumstances, so it becomes crucial to examine the potential ways in which the EU member states and other international bodies can reach out and help the populations who are unable to move. Another aspect worth exploring is the impact of asylum-related migration and its consequences on countries of origin, which has been discussed only in a handful of studies (Koser and Van Hear 2004).

From the methodological point of view, there are also longer-term opportunities for policy-relevant model development, for example through the design of interactive simulations, such as agent-based models,

whereby the behaviour of individual actors – in this case, refugees, as well as policy-making institutions – is modelled *in silico*, within computer programmes (Gilbert & Troitzsch 2005). In the context of environment-related migration, an example of such a model has been proposed by Smith et al. (2008) with the framework discussed in policy terms by Ginnetti (2015), and a review of the current state of the art in this area has been offered by Klabunde and Willekens (2016). One particularly appealing feature of such policies is the testing of policy alternatives through simulations (examples may include the possible policy developments suggested by Gibney and Hansen 2004). Qualitative and other survey information on the mechanisms of decision making of the asylum seekers as fully autonomous agents (e.g. Robinson & Segrott 2002; Kingsley 2016) could constitute important input into such models. Still, in terms of the work effort required, these approaches would go well beyond the EASO research programme but are worth considering as a longer-term investment that could be co-funded and possibly co-produced with academic stakeholders.

At this stage, however, given the importance of the challenges related to asylum in Europe, the development of an early warning system is a priority. Just by doing that, the perspective in policy and decision-making at the EU level, including at EASO, would change from reactive to pro-active and contingency planning would become naturally embedded at the heart of the modelling, the role of which would shift from an one-off product to a constantly-updated **process**. We consider this to be the main appealing feature of the proposed approach to modelling asylum, which remains honest and realistic about the limitations of any formal models in the context of asylum-related migration.

The six key recommendations made throughout this report are summarised in Box 3.

Box 3. Summary of recommendations

Recommendation 1. The formal approach pursued as a part of the EASO research programme should be an early warning system, modelling *changes* in asylum flows based on internal EASO data.

Recommendation 2. The proposed way of operationalising this approach is through the statistical control theory, and by applying Bayesian methods for describing uncertainty in the model.

Recommendation 3. The main inputs into the model should include expert knowledge on countries of origin, possibly augmented by formal conflict intensity indices for countries of origin and policy variables for destinations, as well as stakeholder views on the sensitivity of the system.

Recommendation 4. The application of the approach should be iterative, and the model should be periodically revisited and reassessed, especially every time a change in asylum trends is suspected.

Recommendation 5. Communication is key: the users of the model should be made aware of the aim of the modelling process and limitations of the results, especially in relation to the uncertainty.

Recommendation 6. From the modelling point of view, further work should concentrate on harnessing the power of 'big data', for example from social media and other electronic communication means, while bearing in mind the associated ethical and privacy concerns.

Glossary

Asylum-related migration: In this report, migration that is defined as international in nature, and being – or being claimed to be – related to forced displacement, defined as forced migration due to persecution, armed conflict, violence, or violations of human rights (page 6)

Bayesian approach: A statistical methodology based on *prior* beliefs about the phenomena of interest, updated with the information from data, in order to produce the *posterior* estimates of model parameters and derive the forecasts on their basis (page 32; more in Bijak & Bryant 2016)

Cusum: Cumulative sum – an indicator used in statistical quality control models and **early warning systems**, defined through cumulating the values of the underlying variable over time (pages 33–34)

Delphi technique: A multi-stage technique for **expert elicitation**, whereby results of the elicitation from one stage of the exercise are communicated anonymously back to the experts, who are then given the option of adjusting their views on that basis (page 27)

Early warning system: A statistical modelling framework designed to signal a need for undertaking action if some control indicators (such as **cusums**) exceed warning thresholds, which can be specified by taking into account the **type 1 and type 2 errors** related to making incorrect decisions (page 32)

Expert elicitation: An exercise aimed at gauging the views of a experts on certain aspects of the model, such as the prior distributions for parameters in the **Bayesian approach**, or target values for certain processes or indicators; carried out formally within a pre-defined framework (page 42)

Generalised linear models (GLM): Generalisations of simple linear regression models, where the error terms are allowed to have different probability distributions than normal. In this report, we consider models that enable distinguishing various effects: generation, distribution, etc. (page 20)

Log-linear models: A class of linear regression models with log-transformed response variables. In this report, these are mainly log-linear models for counts in multidimensional contingency tables, with explanatory variables including various individual effects and interactions (pages 21 and 50)

Panel regression: A type of regression models, whereby groups (panels) of the same variables or indicators are modelled across time for different objects, such as countries (page 21)

Push and pull factors: In Lee's (1966) framework, these are drivers 'pushing' migrants from their origins and 'pulling' them to the destinations, mediated by the 'intervening factors' and individual migrant characteristics. Öberg (1996) additionally distinguished 'hard' and 'soft' factors (pages 8–9)

Probabilistic methods: Methods and models of forecasting, which describe their predictions not in terms of single values, but whole probability distributions of the possible future outcomes (page 30)

Synthetic indices: Artificial variables created by combining the information on different dimensions of the same phenomenon, for example stringency of the migration or asylum policy, or intensity of armed conflict (pages 18 and 23)

Time series analysis: A class of **probabilistic methods** of prediction, whereby the forecasts depend, among other things, on past values of the variables being predicted (page 21)

Traffic-lights approach: A method of evaluating various aspects of quality of the data or models, assigning them the 'green' (good), 'amber' (moderate) or 'red' (poor) ratings (pages 12 and 19)

Type 1 and type 2 errors: Errors in making incorrect decisions based on signals in data – either 'false alarms' (Type 1 errors), related to spurious warnings in data, or 'complacency' (Type 2 errors) – not generating warnings when the indications in the data are genuine (page 34)

Bibliography

Abel G, Bijak J, Findlay A, McCollum D and Wiśniowski A (2013) Forecasting environmental migration to the UK: an exploration using Bayesian models. *Population and Environment*, 35(2): 183–203.

Alho J and Spencer B (2005) *Statistical demography and forecasting*. New York: Springer.

Arango J (2000) Explaining migration: a critical view. *International Social Science Journal*, 52 (165): 283–296.

Ariely G, Warnes R, Bijak J and Landesmann R, with contribution by Y Poria (2011) Futures of Borders: A Forward Study of European Border Checks until 2030. Report for EU/Frontex. Liron Systems and University of Southampton.

Azose J and Raftery AE (2015) Bayesian Probabilistic Projection of International Migration. Demography, 52(5): 1627–1650.

Bakewell O (1999) Can we ever rely on refugee statistics? *Radical Statistics*, 72, art. 1. Accessible via: www.radstats.org.uk/no072/article1.htm

Beine M, Boucher A, Burgoon B, Crock M, Gest J, Hiscox M, McGovern P, Rapoport H, Schaper J, and Thielemann E (2015) Comparing Immigration Policies: An Overview from the IMPALA Database. *International Migration Review*, Online First, DOI: 10.1111/imre.12169.

Bijak J (2010) *Forecasting International Migration in Europe: A Bayesian View.* Dordrecht: Springer.

Bijak J (2012) Migration Assumptions in the UK National Population Projections: Methodology Review. Report for the Office for National Statistics. Southampton: University of Southampton.

Bijak J, Alberts I, Alho J, Bryant J, Buettner T, Falkingham J, Forster JJ, Gerland P, Keilman N, King T, O'Hagan A, Onorante L, Owens D, Raftery AE, Ševčíková H and Smith PWF (2015) Probabilistic population forecasts for informed decision making. Letter to the Editor. *Journal of Official Statistics*, 31(4), 537–544.

Bijak, J, & Bryant, J (2016) Bayesian demography 250 years after Bayes. Population Studies, 70(1), 1-19. DOI:10.1080/00324728.2015.1122826

Bijak J and Wiśniowski A (2010) Bayesian forecasting of immigration to selected European countries by using expert knowledge. *Journal of the Royal Statistical Society A*, 173 (4): 775–796.

Böcker A and Havinga T (1998) Asylum migration to the European Union: Patterns of origin and destination. Luxembourg: European Commission.

Bongaarts J and Bulatao R (eds) (2000) *Beyond six billion: Forecasting the world's population*. Washington DC: National Academy Press.

Clark L (1989) *Early Warning of Refugee Flows*. Washington DC: Refugee Policy Group.

Cohen JE (2012) Projection of net migration using a gravity model. Tenth. Coordination Meeting on International Migration, Population Division, United Nations, New York, 9–10 February 2012.

Crisp, J. (1999). Who has counted the refugees? UNHCR and the politics of numbers. New Issues in Refugee Resarch No. 12, Geneva: UNHCR

Czaika M (2005) A Refugee Burden Index: Methodology and its application. *Migration Letters*, 2(2): 101–125.

Czaika M (2009) *The Political Economy of Refugee Migration and Foreign Aid*. Basingstoke: Palgrave Macmillan.

Czaika M and Hobolth M (2014) Deflection into irregularity? The (un)intended effects of restrictive asylum and visa policies. DEMIG project paper 15. Oxford: International Migration Institute.

Dawid AP (1984) Present Position and Potential Developments: Some Personal Views: Statistical Theory: The Prequential Approach (with Discussion), *Journal of the Royal Statistical Society A*, 147(2): 278-292.

de Beer J (2008) Forecasting international migration: time series projections vs argument-based forecasts. In: J Raymer and F Willekens (eds.), *Estimation of international migration in Europe: issues, models and assessment*. Chichester: Wiley, pp. 283–306.

Disney GN, Wiśniowski A, Forster JJ, Smith PWF and Bijak J (2015) Evaluation of existing migration forecasting methods and models. Report to the Migration Advisory Committee. Southampton: CPC. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/467405 /Migration_Forecasting_report.pdf (as of 1 June 2016).

European Asylum Support Office (2016) Guide to EPS Data Collection. Unpublished Draft Report, European Asylum Support Office (May 2016 Version).

European Migration Network (2014) Ad-Hoc Query on Forecasting and Contingency Planning Arrangements for International Protection Applicants. https://www.udi.no/globalassets/global/ european-migration-network_i/ad-hoc-queries/ad-hoc-query-forecasting-and-contingency-planning-arrangements-international-protection-applicants-closed.pdf (as of 1 June 2016)

FitzGerald DS (2015) The sociology of international migration. In: CB Brettell and JF Hollifield (eds.), *Migration Theory, Talking across Disciplines*, 3rd ed. Routledge: New York, pp. 115–147.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A and Rubin DB (2014) *Bayesian Data Analysis: Third Edition*. Boca Raton, FL: CRC Press / Chapman and Hall.

Gibney MJ and Hansen R (2004) Asylum Policy in the West: Past Trends, Future Possibilities. In: GJ Borjas and J Crisp (eds.) *Poverty, International Migration and Asylum*. Basingstoke: Palgrave Macmillan, pp. 70–96.

Gilbert N and Troitzsch KG (2005) *Simulation for the social scientist*. Milton Keynes: Open University.

Ginnetti J (2015) Modelling displacement. *Forced Migration Review*, 49: 10–11.

Gneiting T, Balabdaoui F and Raftery AE (2007) Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society B*, 69(2): 243–268.

Government Office for Science (2011) Migration and Global Environmental Change. Future Challenges and Opportunities. Final Project Report. London: Government Office for Science. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/287717/11-1116-migration-and-global-environmental-change.pdf (as of 1 June 2016)

Haldane, A. (2016) Whose Recovery? Speech given in Port Talbot, Wales, 30/06/16. Bank of England. http://www.bankofengland.co.uk/publications/Documents/speeches/2016/speech916.pdf (as of 1st August 2016)

Harrison PJ and Veerapen PP (1994) A Bayesian decision approach to model monitoring and cusums. *Journal of Forecasting*, 13(1): 29–36.

Hatton TJ (2004) Seeking asylum in Europe. *Economic Policy*, 19(1): 5–62.

Hatton TJ (2009) The Rise and Fall of Asylum: What Happened and Why? *The Economic Journal*, 119: F183–F213

Hatton TJ and Moloney J (2015) Applications for Asylum in the Developed World: Modelling Asylum Claims by Origin and Destination. CEPR Discussion Paper DP10678. Available at SSRN: http://ssrn.com/abstract=2624633

Hegre H, Karlsen J, Nygård HM, Strand H and Urdal H (2013) Predicting Armed Conflict, 2010–2050. *International Studies Quarterly*, 57(2): 250–270.

Hilderink H, van der Gaag N, van Wissen L, Jennissen R, Román A, Salt J, Clarke J and Pinkerton C (2002) Analysis and Forecasting of International Migration by Major Groups (Part III). Eurostat Working Paper 3/2002/E/nº17. Eurostat, Luxembourg.

Jennings W (2010) Bureaucratic Performance and Control in British Politics: Asylum Policy 1994–2007. *The British Journal of Politics & International Relations*, 12(4): 539–568.

Jennissen RPW (2004) *Macro-economic Determinants of International Migration in Europe*. Amsterdam: Dutch University Press.

Keilman N (1990) *Uncertainty in national population forecasting: Issues, backgrounds, analyses, recommendations*. Amsterdam: Swets & Zeitlinger.

Keogh G (2013) Modelling Asylum Migration Pull-Force Factors in the EU-15. *The Economic and Social Review*, 44(3): 371–399.

King R (2002) Towards a New Map of European Migration. *International Journal of Population Geography*, 8(2): 89–106.

Kingsley P (2016) *The New Odyssey: The Story of Europe's Refugee Crisis*. London: Guardian / Faber & Faber.

Klabunde A and Willekens FJ (2016) Decision-making in agent-based models of migration: State of the art and challenges. *European Journal of Population*, 32(1): 73–97.

Koser K and Van Hear N (2004) Asylum Migration: Implications for Countries of Origin. In: GJ Borjas and J Crisp (eds.) *Poverty, International Migration and Asylum*. Basingstoke: Palgrave Macmillan, pp. 121–142.

Kupiszewski M and Bijak J (2013) The use of international migration theories in migration forecasting – A practical approach. In: M Kupiszewski (ed.) *International migration and the future of populations and labour in Europe*. The Springer Series on Demographic Methods and Population Analysis Vol. 32, Dordrecht: Springer, pp. 35–55.

Lee ES (1966) A Theory of Migration. *Demography*, 3(1): 47–57.

Maastricht University and IOM Berlin (2016) Literature review of significant pull/push factors for determining asylum-related migration. Report for the European Asylum Support Office. Maastricht: Maastricht University – Maastricht Graduate School of Governance/UNU-MERIT & Berlin: IOM Global Migration Data Analysis Centre.

Moore WH and Shellman SM (2004) Fear of Persecution. Forced migration 1952–1995. *Journal of Conflict Resolution*, 40(5): 723–745.

Mouzourakis M (2014) 'Wrong number?' The Use and Misuse of Asylum Data in the European Union. CEPS Paper No. 69, Centre for Europe and Policy Studies, Brussels.

Neumayer E (2004) Asylum Destination Choice: What Makes Some Western European Countries More Attractive than Others? *European Union Politics*, 5(2): 155–180.

Neumayer E (2005a) Bogus Refugees? The Determinants of Asylum Migration to Western Europe. *International Studies Quarterly*, 49: 389–409.

Neumayer E (2005b) Asylum Recognition Rates in Western Europe – Their Determinants, Variation and Lack of Convergence. *Journal of Conflict Resolution*, 49(1): 43–66.

Öberg S (1996) Spatial and Economic Factors in Future South-North Migration. In: W. Lutz (ed.) *The Future Population of the World: What Can We Assume Today?* London: Earthscan, pp. 336–357.

O'Hagan, A. (1998). Eliciting Expert Beliefs in Substantial Practical Applications. *Journal of the Royal Statistical Society*, *Series D*, 47(1), 129–136.

Page ES (1954) Continuous inspection schemes. *Biometrika*, 41(1–2): 100–114.

Page ES (1961) Cumulative Sum Charts. *Technometrics*, 3(1): 1–9.

Pijpers R (2008) Problematising the 'orderly' aesthetic assumptions of forecasts of East–West migration in the European Union. *Environment and Planning A*, 40 (1): 174–188.

Poulain M, Perrin N and Singleton A (eds) (2006) *THESIM: Towards Harmonised European Statistics on International Migration*. Louvain-la-Neuve: Presses universitaires de Louvain.

Raymer J, Wiśniowski A, Forster JJ, Smith PWF and Bijak J (2013) Integrated Modeling of European Migration. *Journal of the American Statistical Association* 108 (503), 801–819.

Regehr E (2014) Armed Conflict: Trends and Drivers. Background paper for the workshop "Civil society views on next generation peacebuilding and conflict prevention policy and programming issues and responses", Ottawa, 14.03.2011.

RMMS–IMI (2012) Global Migration Futures: Using scenarios to explore future migration in the Horn of Africa & Yemen. Project report, November 2012. Nairobi and Oxford: Regional Mixed Migration Secretariat and International Migration Institute.

Robinson V and Segrott J (2002) *Understanding the Decision-Making of Asylum Seekers*. Home Office Research Study 243. London: Home Office.

Salt J and Singleton A (1995) Analysis and forecasting of international migration by major groups. Report to the Working Party on Demographic Projections. Eurostat, Luxembourg.

Schellman SM and Stewart BM (2007) Predicting Risk Factors Associated with Forced Migration: An Early Warning Model of Haitian Flight. *Civil Wars*, 9(2): 174–199.

Schmeidl S (1997) Exploring the Causes of Forced Migration: A Pooled Time-Series Analysis, 1971–1990. *Social Science Quarterly*, 78(2): 284–308.

Scrucca, L. (2004). qcc: an R package for quality control charting and statistical process control. *R News* 4/1, 11-17.

Singleton A (2016) Migration and Asylum Data for Policy-Making in the European Union – The Problem with Numbers. CEPS Paper No. 89, Centre for Europe and Policy Studies, Brussels.

Smith Ch, Kniveton D, Wood S, Black R (2008) Predictive modelling. *Forced Migration Review*, 31: 58–59.

Stanley WD (1987) Economic Migrants or Refugees from Violence? A Time-Series Analysis of Salvadoran Migration to the United States. *Latin American Research Review*, 22(1): 132–154.

Suriyakumaran A & Tamura Y (2016) Asylum Provision: A Review of Economic Theories. *International Migration* (early view); DOI: 10.1111/imig.12228.

Tabeau E (ed.) (2009) *Conflict in Numbers: Casualties of the 1990s Wars in the Former Yugoslavia (1991–1999)*. Testimonies, vol. 33. Belgrade: Helsinki Committee for Human Rights in Serbia.

Tartakovsky AG and Moustakides GV (2010) State-of-the-Art in Bayesian Changepoint Detection. *Sequential Analysis: Design Methods and Applications*, 29(2): 125-145.

Theilemann ER (2003) Between Interests and Norms: Explaining Burden-Sharing in the European Union. *Journal of Refugee Studies*, 16 (3): 253–273.

Theilemann ER (2004) Why Asylum Policy Harmonisation Undermines Refugee Burden-Sharing, *European Journal of Migration and Law*, 6: 47–64.

Toshkov DD (2014) The dynamic relationship between asylum applications and asylum recognition rates. *European Union Politics*, 15(2): 192–214.

Toshkov D, de Haan L (2013) The Europeanization of Asylum Policy: An assessment of the EU impact on asylum applications and recognitions rates. *Journal of European Public Policy*, 20(5): 661-683.

Trading Economics (n.d.) Asylum Applications, Forecast 2016–2020. Automatic online forecasting tool, http://www.tradingeconomics.com/forecast/asylum-applications (Accessed on 1 July 2016).

UNHCR (1951/1967) Text of the 1951 Convention Relating to the Status of Refugees; Text of the 1967 Protocol Relating to the Status of Refugees; Resolution 2198 (XXI) adopted by the United Nations General Assembly with an Introductory Note by the Office of the United Nations High Commissioner for Refugees. Geneva: UNHCR.

UNHCR (2002) Statistical Yearbook 2001: Refugees, Asylum-seekers and Other Persons of Concern -

Trends in Displacement, Protection and Solutions, United Nations High Commission for Refugees, Geneva.

United Nations (1998) *Recommendations on Statistics of International Migration*. Series M, No. 58, Rev. 1. New York: UN Statistics Division.

van der Gaag N and van Wissen L (1999) Analysis and Forecasting of International Migration by Major Groups (Part II). Eurostat Working Paper 3/1999/E/nº9. Eurostat, Luxembourg.

van Wissen L and Jenissen R (2008) A simple method for inferring substitution and generation from gross flows: Asylum seekers in Europe. In: J Raymer and F Willekens (eds.), *Estimation of inter-national migration in Europe: issues, models and assessment*. Chichester: Wiley, pp. 235–251.

Vink M and Meijerink F (2003) Asylum Applications and Recognition Rates in EU Member States 1982–2001: A Quantitative Analysis, *Journal of Refugee Studies*, 16(3): 297–315.

Weidmann NB and Ward MD (2010) Predicting Conflict in Space and Time. *Journal of Conflict Resolution*, 54 (6): 883–901.

West M (1986) Bayesian Model Monitoring. *Journal of the Royal Statistical Society B*, 48(1): 70–78.

Willekens F and Baydar N (1986) Forecasting place-to-place migration with generalized linear models. In: R Woods and PH Rees (eds.), *Population structures and models*. Allen & Unwin, London: 203–244.

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

Wiśniowski A, Bijak J and Shang HL (2014) Forecasting Scottish migration in the context of the 2014 constitutional change debate. *Population, Space and Place*, 20 (5), 455–464.

Zolberg AR (1989) The Next Waves: Migration Theory for a Changing World. *International Migration Review*, 23: 403–430.

Zwicky F (1948) Morphological astronomy. *Observatory*, 68(845): 121–143.

Zwicky F (1969) *Discovery, invention, research – through the morphological approach*. Toronto: Macmillan.

Appendix A. Inventory of selected data and meta-data sources

- Asylum – key series of interest
  - **EASO**
  - UNHCR (http://www.unhcr.org/uk/figures-at-a-glance.html)
- National Statistical Institutes and ministries of interior (or equivalent authorities)
  - Germany: Bundesamt für Migration und Flüchtlinge (https://www.bamf.de)
  - The Netherlands: Statistics Netherlands (https://www.cbs.nl)
  - Sweden: Migrationsverket (http://www.migrationsverket.se)
  - Switzerland: Staatssekretariat für Migration (https://www.sem.admin.ch)
  - United Kingdom: National Statistics on asylum (https://www.gov.uk/government/publications/immigration-statistics-april-to-june-2015/asylum)
- Other EU and international agencies
  - Eurostat (http://ec.europa.eu/eurostat/web/asylum-and-managed-migration/data/-database)
  - Frontex (Selected publications at http://frontex.europa.eu/publications/)
  - Database on Immigrants in OECD countries (http://www.oecd.org/els/mig/databaseonimmigrantsinoecdcountriesdioc.htm)
- Data on drivers of asylum-related flows
  - World Bank (http://data.worldbank.org/)
  - Uppsala Conflict Data Program (http://www.pcr.uu.se/research/ucdp/datasets/)
  - CEPII (http://www.cepii.fr)
  - Freedom House (https://freedomhouse.org/)
  - Political Terror Scale (http://www.politicalterrorscale.org/)
- Meta-data and research projects
  - Asylum Information Database (AIDA, http://www.asylumineurope.org/)
  - International Migration Policy And Law Analysis project database (IMPALA, http://www.impaladatabase.org)

Appendix B. Detailed assessment of data and methods

Table B 1. Framework for evaluating the data on asylum-related migration – quality descriptors

Dimension	Good aspects	Average aspects	Problematic aspects
Frequency	Daily, weekly, or fortnightly series	Monthly, *n*-monthly, or yearly series	Occasional or one-off studies
Definitions	Following international standards	Roughly corresponding to established concepts	Not defined or definitions imprecise
Coverage	Full population of interest / extensive historical series.	Well-defined regions or subpopulations / shorter times series.	Unclear, unsystematic, opportunity
Accuracy	Good-quality comprehensive register	Other register or good-quality survey	Poor-quality survey or unknown
Timeliness	Promptly after the reference period	Reasonable timeliness, without long delays	Infrequent production, one-off collections, long delays
Quality assurance	In place, rigorous, and well documented	Some, documentation may be patchy	Unclear or non-existent quality processes

Table B 2. Framework for evaluating the existing models and methods – quality descriptors

Dimension	Good aspects	Average aspects	Problematic aspects
Specification	Appropriate method, well-designed model	Some issues, but method generally fine	Design unclear or not fit for purpose, specification errors
Resolution	Detailed spatial and/or temporal resolution	Broad spatial and temporal resolution	Case studies: single area and single time period
Assumptions	Coherent, logical, well explained, many aspects	Implicit, broad, some aspects considered	Missing, incoherent, not justified, single aspects
Predictiveness	Predictive model by design (e.g. time series)	Explanatory model, some pred. potential	Descriptive model; with no predictive potential
Uncertainty	Explicitly addressed & formally modelled	Discussed in the narrative / limitations	Not acknowledged explicitly, dismissed or ignored
Generality	Model transferable to other problems	Model suited for a given class of problems	Model tailored only for a very specific problem

Table B 3. Specification of selected quantitative models of asylum-related migration (For sources, see table)

Source	Independent variable groups >	Policy and politics							Conflict		Culture			Demography and migration						Economy						Other		Notes
	Variables > / Model specification	Policy: immigration overall	Policy: Asylum-related	Governance variables	Rights and liberties	Welfare state	Policy index	Other political variables	Conflict intensity / index	Other conflict variables	Colonial ties	Language ties	Cultural ties, tourism	Population size / density	Age structure	Migrant / refugee stocks	Urbanisation variables	Ethnic composition	Past asylum / migration	Income and differentials	Unemployment rates	Consumption levels	Inflation	Trade relations	Poverty and inequalities	Infrastructure	Proximity / neighbourhood	
A. Simple regression models																												
Böcker & Havinga (1998)	Multivariate regression of a single-period index of interdependence of origins and destinations										x	x	x											x			x	Significance unknown; weak explanatory power
Stanley (1987)	A regression model of Guatemala-US migration, mediated by en route factors in Mexico			x					x	x									x									Explanatory model with a predictive potential, based on monthly data
Trading Economics (n.d.)	Simple extrapolation model of asylum as one of the many macro-level variables																		x									Not documented; model specification unknown - presumably polynomial trend
van der Gaag & van Wissen (1999)	Simple regression models linking asylum with total migration (contemporaneous & lagged)																		x									Weak predictive potential; inconclusive - depend on countries of origin / destination
B. Panel regression models																												
Czaika (2009)	Probit model of conflict occurrence; OLS and quantile regression of population change			x													x	x							x	x		Population change as proxy for conflict-induced out-migration; explanatory models; sub-national
Czaika & Hobolth (2014)	Panel models of asylum deterrence and resulting irregular migrant flows	x	x	x										x		x				x							x	Several specifications tested; explanatory models
Hatton (2004, 2009)	Panel models of the numbers of asylum claims by countries of origin and destination	x	x	x	x		x		x						x	x				x	x							Explanatory models, with a comprehensive set of variables. Timeframe: 1981-1999
Hatton & Moloney (2015)	Panel models of the numbers of asylum claims by countries of origin and destination	x	x		x	x	x		x	x										x	x							As above, for the period 1997-2012
Keogh (2013)	A parsimonious panel model with random effects for countries, time and geography		x													x				x								Explanatory model aimed at estimating the effects; exploratory analysis with loess trends
Moore & Shellman (2004)	A panel model aimed to estimate the long-term determinants of forced migration flows			x				x	x	x									x	x								Long-term perspective, since implementing of the 1951 UN Refugee convention
Neumeyer (2004, 2005a, 2005b)	A suite of panel models of asylum, destination choice and recognition rates		x	x	x	x		x	x	x	x	x	x	x	x	x	x		x	x	x						x	Explanatory models, with a very comprehensive set of predictors

Table B 3. (cont.)

Source	Independent variable groups >	Policy and politics							Conflict		Culture			Demography and migration						Economy						Other		Notes
	Variables >	Policy: immigration overall	Policy: Asylum-related	Governance variables	Rights and liberties	Welfare state	Policy index	Other political variables	Conflict intensity / index	Other conflict variables	Colonial ties	Language ties	Cultural ties, tourism	Population size / density	Age structure	Migrant / refugee stocks	Urbanisation variables	Ethnic composition	Past asylum / migration	Income and differentials	Unemployment rates	Consumption levels	Inflation	Trade relations	Poverty and inequalities	Infrastructure	Proximity / neighbourhood	
C. Time series, structural and log-linear models																												
de Beer (2008)	Framework for modelling asylum flows as one of the sub-categories of overall migration																		x									Atheoretical model, separating the generation and substitution effects; no forecast per se
Jennings (2010)	Time series (Box-Tiao) model of "bureaucratic control" of asylum in the UK	x	x	x				x											x									Predictive model, monthly data, 1994-2007
Schellman & Stewart (2007)	A structural model of (forced) migration Haiti-USA (error correction; vector autoregression)			x				x		x									x	x			x					Predictive model, intended as an early warning mechanism, with weekly data
Schmeidl (1997)	Pooled time-series analysis of annual refugee stocks, based on panel data				x			x		x				x					x			x					x	Predictive model, including interactions between selected variables
Toshkov (2014)	A suite of AR models (from pooled to multi-level structures) of applications and recognition rates		x	x															x	x	x							Solid econometric analysis; tests of Granger causality and cointegration; confidence ranges
Van Wissen & Jennissen (2008)	Log-linear framework for decomposing flows into generation and distribution effects																		x									The model aims at separating the two effects; based on origin x destination x time tables
Vink & Meijerink (2003)	A log-linear model of asylum and recognition rates, with country, time and interaction effects																											Explanatory model, based on annual data; the only predictors are the main / interaction effects

HOW TO OBTAIN EU PUBLICATIONS

Free publications:

- one copy:
  via EU Bookshop (http://bookshop.europa.eu);
- more than one copy or posters/maps:
  from the European Union's representations (http://ec.europa.eu/represent_en.htm);
  from the delegations in non-EU countries (http://eeas.europa.eu/delegations/index_en.htm);
  by contacting the Europe Direct service (http://europa.eu/europedirect/index_en.htm) or
  calling 00 800 6 7 8 9 10 11 (freephone number from anywhere in the EU) (*).

  (*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

Priced publications:

- via EU Bookshop (http://bookshop.europa.eu).

BZ-01-17-586-EN-N

Publications Office