Evaluation of existing migration forecasting methods and models

Report for the Migration Advisory Committee

George Disney, Arkadiusz Wiśniowski, Jonathan J. Forster, Peter W.F. Smith and Jakub Bijak

ESRC Centre for Population Change, University of Southampton.

Southampton, 10 October 2015

Acknowledgments

This work has been prepared for and was funded by the Migration Advisory Committee (MAC), under the Home Office Science contract HOS/14/040. We would like to thank Allan Findlay and the members of the MAC and the MAC Secretariat for their comments on the drafts. All the interpretations in this report are those of the authors only.

Contents

Tables and Figures

Evaluation of existing migration forecasting methods and models

Key Messages

- Migration is a global phenomenon, and the United Kingdom (UK) is an important country of destination, as well as of origin, for many migrants. In recent years, migration has become an important topic in the UK policy debate. Having accurate knowledge of actual and predicted migration flows can be very useful with regards to the planning and implementation of new policy tools and instruments.

- However, there are many social, economic and political drivers which can impact migration flows, making forecasting migration an extremely difficult task. In particular, migration is very susceptible to shock events which are, by their very nature, hard to predict, such as economic cycles, military conflict and policy changes. Changes in migration flows can be subject to extreme short-term fluctuations, thereby making migration forecasts prone to very high levels of error.

- The inherent uncertainty about future migration flows is further compounded by the intrinsic errors in the data. Data sources can differ in the coverage of specific migrant groups, the accuracy of measurement methodologies, and even in the definition of migration itself.

- In the past, migration forecasts have been attempted using a wide array of methods, with the central focus frequently on one or more of the following: extrapolation of the past data, the opinion of experts in the field, and the inclusion of additional explanatory information, such as economic data and demographic characteristics. No method is considered to be universally superior, and the applicability of each method depends on the particular definition of migration under scrutiny, as well as the features of the data, such as the length of the series and the stability of trends.

- The aim of the empirical part of this study was to assess the degree of uncertainty in migration forecasting models. It was done by comparing the results of various models for different migration flows against the trends observed in the past. All the models examined in this report produced considerable uncertainty when tested against past data. For example, the more reliable models for migration from the new EU member states predicted that there was a 50 per cent chance that the average annual immigration between 2004 and 2013 would range from 100 to 200 thousand people; in reality, this proved to be just above 150,000.

- Some models produced smaller errors and consequently were deemed to have performed better than others with regards to forecasting migration. Notably, the more successful forecasts were observed when using the more stable data series, such as the migration of UK nationals, which are less susceptible to unpredictable shocks or policy changes.

- From the analysis, there is no particular model that can be considered as conclusively superior. Instead, it is recommended that any future analysis utilise a three-step approach to migration forecasting. Namely, future analysis should assess the nature of the migration flow being forecast, evaluate the available data, and design a bespoke forecasting model for the given situation. Instead of trying to do the impossible and design the 'best possible' migration forecasting method, further work in this area should focus on the ways of translating uncertain forecasts into policy advice and decisions.

- It is imperative that all migration forecasts emphasise the uncertainty involved in the predictions. This is necessary to transparently acknowledge that migration cannot be forecasted without substantial error, whilst also providing an account for the possible size of these errors. Different ways of showing the range of errors are possible, by the means of probabilities for various ranges of possible outcomes.

- Since the probability of a single forecast being correct is extremely low, it is vital that the uncertainty around migration forecasts is made explicit to decision-makers and the general public. Emphasising the uncertainty also allows decision-makers to correctly represent the fact that migration can be affected by a wide range of events, including 'shocks', all of which need to be taken into account as, although they are quite unlikely, their potential impact on migratory flows could be large.

Executive Summary

Forecasting migration flows is an extremely difficult task, characterised by very high levels of error, the highest amongst the three components of demographic change (fertility, mortality and migration). There are many social, economic, political or even environmental drivers which can impact migration flows: from unemployment, job prospects and wage differentials, to social networks and various institutions. In particular, migration is very susceptible to events which are difficult to predict in terms of timing as well as scale of impact, such as fluctuations in the economic cycle, the incidence of armed conflict, and changes in policies or political circumstances.

There exists no perfect migration theory that can be used for forecasting purposes. Explanations put forward for complex migration processes are often entrenched in particular disciplines of social sciences, such as economics, sociology or human geography, so that they tend to focus predominantly on one fragment of the picture and examine a specific group of migration drivers (economic, social, etc.) at the expense of others. Even if credible theoretical explanations of past migration flows do exist, their tenets tend to be difficult to extrapolate into the future.

This report focuses on three sources of uncertainty in migration forecasts: uncertainty about the future, errors and differences in the data, and uncertainty related to relying on a single particular forecasting model amongst the many possibilities available (**Section 2**). None of these sources of uncertainty can be eliminated completely, but it is crucial that they are acknowledged in the forecasting process in a transparent way. Uncertainty about data can be quantified, albeit with caution, but the uncertainty surrounding the nature and size of future shocks can be assessed only to a very limited extent. As such, any decision-making based on migration forecasts is particularly susceptible to error from unforeseen future events.

There exist several different sources of data on migration flows into and out of the United Kingdom, but they differ with respect to four key attributes: the definitions of migration they use; the particular migrant groups they cover; whether there is under- or over-reporting of migration; and how accurate their measurements are (**Section 3**). Even the main source of data used to measure migration in the UK, the International Passenger Survey (IPS), has several weaknesses. The IPS is a sample survey, so disaggregations of the data by countries of origin or destination of migrants can have high margins of error resulting from sampling of respondents. There can also be some bias in the numbers related to the way the data are collected, with the initial focus mainly on the largest airports and Channel crossings having caused problems after the 2004 enlargement of the European Union. Additionally, the long-term IPS estimates are based on the questions about the intended (rather than actual) length of stay in the UK or abroad.

The existing forecasting methods usually involve single deterministic scenarios, which follow some pre-defined assumptions, or probabilistic models. The former do not attempt to measure uncertainty in any way, while the latter seek to assess the chance of various future migration trajectories, and measure them with probabilities of occurrence. Most of the official migration projections prepared by the statistical offices of other countries or international organisations remain deterministic. Their experience so far is mixed – while there is a considerable variation in the accuracy of the assumptions between different countries, it can be largely attributed mainly to chance. Moreover, deterministic scenarios have not become more accurate over time. On the other hand, some countries (such as The Netherlands and New Zealand) have moved towards probabilistic methods, and began to openly acknowledge uncertainty in the predictions. The United Nations Population Division is expected to follow suit with their prototype model for migration.

The probabilistic forecasting models are usually based on past data, expert opinion, the inclusion of additional variables (in econometric models), or some combination of the three. The best model choice for a particular task depends on the characteristics of the data, such as the length of the series of observations (how many years or quarters of data there are), and whether the trends in data are relatively stable over time or are constantly changing (**Section 4**).

In the empirical part of the analysis, various models have been compared based on their forecast errors, and on how accurately they measure the uncertainty of the forecast. The results were obtained for 'pretend forecasts' prepared by using data only up to certain years in the past. In particular, this exercise aimed to mimic the forecasting process as if it were being undertaken just before the two major 'shocks' to the migration patterns observed in the previous decade: firstly, the enlargement of the European Union in 2004; and, secondly, the economic crisis in 2009. For the outcomes, several measures of error have been calculated. The forecasts have been also assessed according to how well they describe the predicted uncertainty in comparison with the real data (**Section 5**).

As might be expected, only when the underlying data series were relatively stable, such as for migration of UK nationals, were some models able to produce relatively small errors – typically deviating from the actual observations by no more than 15 per cent on average. In other situations, the applicability of various methods was either limited, when the errors were larger, or completely inappropriate, for example when the forecasts differed from the actual observations by more than 50 per cent and the actual observations remained far outside the prediction intervals. In particular, no model was able to predict migration well if the underlying data series were short, or in the presence of shocks (structural breaks), such as the enlargement of the European Union. Still, even in such cases, some models performed better than others; models that did not assume stability of trends, when none was to be expected, at least described the forecast uncertainty more accurately, and hence more honestly.

In terms of recommendations for the future, rather than suggesting any particular model or approach for all circumstances, we recommend following a three-stage process to guide the choice of the forecasting methodology for a given task (**Section 6**):

- **Stage 1.** Develop a thorough understanding of the migration flow being forecast, with a focus on whether it is stable or highly susceptible to external political or economic shocks or policy interventions. For example, asylum flows, generated by war and conflict in other parts of the world, can be expected to be less stable than return flows of UK nationals.
- **Stage 2.** The available data need to be assessed with their relative strengths and weaknesses taken into consideration, such as the length of data series or presence of structural breaks in the data. Forecasts based on short series are typically characterised by higher uncertainty, and so are predictions of such migration flows, which were subject to shocks in the past.
- **Stage 3.** An appropriate modelling approach needs to be selected given the characteristics of the migration flow in question and the available data. In particular, a data series with non-stable characteristics should not be forecasted by using models which assume stability (in technical terms, stationarity) of the process, and vice versa. Short data series may ideally require additional expert input concerning the future migration flows.

Following the process outlined above cannot guarantee that the resulting forecasts will exhibit no or only small errors. Still, it would help safeguard against making forecasts that are either unjustifiably too precise or just too uncertain to be useful, and thus protect from making radically incorrect policy decisions on the basis of such predictions. It is unrealistic to expect that there will be no uncertainty: even in good forecasts, errors are inevitable, and the longer the forecast horizon, the higher the errors. In many cases, forecast errors become too high for the forecasts to be useful beyond the horizon of five to ten years into the future.

One key recommendation is that any migration forecasts should come with explicit statements of uncertainty, ideally expressed in terms of probabilities for various ranges of possible migration outcomes. Instead of trying to do the impossible and design the 'best possible' migration forecasting method, further work in this area should focus on translating uncertain forecasts into decisions, creating early warning systems, and providing risk management strategies. As a caveat, the forecasters should not offer methods producing too certain predictions, as they will most likely fail, but neither should the decision-makers expect or require them. Bringing together potential impacts of migration policy interventions (migration caps, visa regulations) with the uncertainty of these impacts can also help policy makers make prudent and more robust decisions, for example related to controlling or influencing specific migration flows.

1. Introduction

The main aim of this report is to contribute to the evidence base in migration forecasting by providing a systematic and up to date methodological overview of academic literature and of official statistical practice in migration modelling and forecasting. This is further strengthened by the empirical comparison of ex-post performance of various methods at the later stage of the project. The report also provides policy and research recommendations related to the usefulness of various forecasting approaches for policy-makers, with a focus on the role of uncertainty and its appropriate communication to the users. Throughout this report, the terms 'forecast' and 'prediction' are used interchangeably to denote quantitative statements about future migration; whereas the term 'projections' is reserved mainly for the results of deterministic calculations of future population size and structure under a set of specific assumptions (for a discussion, see Keilman 1990 and Bijak 2010).

The report is structured as follows: after this Introduction, various layers of uncertainty in migration forecasting are introduced (Section 2), followed by an overview of the available data for the UK migration (Section 3). We then review the existing migration forecasting methods and evaluate them in the context of applying them to the available data for the UK (Section 4). Subsequently, we present the results of the ex-post analysis of selected models, followed by a summary evaluation of their performance (Section 5). The study is concluded by summarising the key findings, sketching a research framework for future work on evaluating migration forecasting methods, and presenting specific policy recommendations (Section 6). The report is supplemented by a glossary of the key terms, while the detailed results of the empirical forecasting exercises and some further technical details are reported in Appendices A–C.

2. Uncertainty and Migration

A vital consideration in forecasting migration is how to incorporate uncertainty into the estimates. The concept of uncertainty refers to the indeterminism or randomness of the phenomena under study (Bijak 2010). This section outlines the three main broad sources of uncertainty to be considered in this project drawing on the work of Willekens (1994) and Kupiszewski (2002).

The first one is the inherent uncertainty about future events. Some level of error in migration forecasting is always inevitable, as any inference about the future is made under uncertainty (Bijak 2010). There is uncertainty in all three components of demographic change – fertility, mortality and migration. Hence, a key question for producers of population forecasts is how to deal with this uncertainty in a statistically consistent manner so that the result is both informative to the users and feasible for the producing agencies (Lutz and Goldstein 2004).

Trends of migration over time tend to be the most volatile demographic process (NRC 2000). Unpredictable and shock events such as political crises, wars, economic downturns and environmental catastrophes can have a significant and varied impact on the level and characteristics of international migration flows. Compounding this uncertainty is the change over time in the relationships between the origin and destination countries, as well as the fact that events affecting migration are also complex and largely unpredictable with respect to the countries involved, timing, and magnitude. In addition, migration often leads to the establishment of networks of migrants from shared origins, which perpetuates the process further through facilitating further migration. This further increases the complexity and uncertainty of migration as a process.

The second source of uncertainty under consideration is associated with migration data themselves. To illustrate this, let us consider the three components of population change – births, deaths and migration. In comparison to births and deaths, migration is the most uncertain component. Vital events, by their very nature, are straightforward to define and measure, with birth registration a universal human right (UN 1966), for example. As such, there is a relatively high level of certainty in the measurement of births or deaths.

The same cannot be said for migration, which, to begin with, is difficult to define. Sources of migration data from different countries are often based on differing definitions (cf. Raymer et al. 2013). A further source of uncertainty is that the available data are often inaccurate, inconsistent and incomplete. Migration into and out of the UK is no exception. The precise size of international migration flows are difficult to measure; data collection systems used to record migrants often produce biased and inaccurate estimates, which need correcting (Disney 2014; Wiśniowski 2013).

The third source of uncertainty comes from the forecasting models themselves. Applications of different models to the same data can produce different forecasts, including different assessments of the uncertainty of the predictions. There is no perfect model, and choosing which model to apply is a matter of judgement, therefore justification is required. If the forecasts from various competing models are combined using formal criteria, additional uncertainty about the model is introduced (cf. Bijak and Wiśniowski 2010).

Consequently, it is clear, that in any forecast of migration there are multiple considerations of uncertainty that firstly need to be fully understood and then taken into account in an empirical analysis. Experts play a key role in developing forecasts, but their task depends on the chosen approach (Lutz and Goldstein 2004). For example, the role of the expert could be limited to choosing the forecast model and selecting the underlying sources of data, or providing expert judgement that is explicitly incorporated as a parameter in the model.

3. Data Audit and Assessment

As stated above, international migration is hard to define and measure. The available data are often inconsistent and are not designed with the purpose of monitoring migration. Hence, one of the main sources of uncertainty in any forecast of immigration comes from the data as such in terms of how they are collected, processed and disseminated (Kupiszewska and Nowok 2008).

With this in mind, it is vital to have a framework that one can use to understand the extent and nature of the uncertainty in migration data. The main sources of publicly available data are outlined and assessed in Table 1 below to help aid our understanding of this uncertainty.

Each of the sources of data are assessed in relation to the 'true flow' (cf. Raymer et al. 2013 and Wiśniowski et al. 2013). 'True flow' is the unknown number that is being estimated. It represents the number that one would obtain if one was able to monitor a given definition of immigration perfectly, without bias and undercount and with complete coverage of the population. A true flow for the purpose of this data audit uses the UN (1998) definition of long-term international migration:

> *"A person who moves to a country other than that of his or her usual residence for a period of at least a year (12 months), so that the country of destination effectively becomes his or her new country of usual residence. From the perspective of the country of departure the person will be a long-term emigrant and from that of the country of arrival the person will be a long-term immigrant" (UN 1998: page 18).*

Similarly, short-term migration is defined by length of stay between three and 12 months (idem). In general, various definitions can be used to represent the 'true flow' depending on its purpose. The data collected to measure such a flow should then aim at reflecting it as closely as possible.

The quality of each source of data can be assessed in relation to the 'true flow' according to the UN definition, by using the following analytical categories (Raymer et al. 2013; Disney 2014):

(i) **Definition** – how closely do the data match the UN definition of international migration?
(ii) **Coverage** – theoretically what portion of the total immigration flow does the data set cover?
(iii) **Bias** – is there any systematic bias as a result of the way the data are collected?
(iv) **Accuracy** – with regard to its intended purpose, how accurate are the data?

A summary table of the data audit and assessment is detailed below (Table 1). A traffic light system is used to indicate how close a match, for each given assessment criteria, each source of data are to the true flow. Green indicates a close match to true flow and red indicates that there is a large distortion leading to a large level of uncertainty or bias in the data; with orange indicating a medium distortion of the true flow and resulting moderate bias and uncertainty. Mode of data collection, data availability, availability of migration characteristics such as citizenship, country of birth, and country of previous residence, as well as an indicator as to whether the data describe migration stocks or flows are all detailed in the table.

In Figure 1, we present the migration data as measured by various sources. The observed increase in total immigration as measured by the International Passenger Survey (IPS) since the late 1990s is a result of an increase of non-British migrants, and, especially after the enlargement of the European Union (EU) in 2004, migration from the EU-8 new member states. Alongside the IPS trends, its augmented version, the Long-Term International Migration (LTIM), is presented, additionally including asylum seekers, migration to and from the Republic of Ireland, as well as a correction for an estimated number of people who change their migration intentions. A relatively smaller increase is observed for the IPS emigration, with a notable switch to emigration of non-British being larger than emigration of British nationals. Short-Term International Migration (STIM) remains on rather stable levels during the 2004-2012 period for which the data are available. We also observe that between 2002 and 2014 the number of non-EU students registered at the Higher Education Statistics Authority (HESA) more than doubled, while the number of students from the EU grew only by 50 per cent in the same period.

Table 1: Data Audit and Assessment

Source	Data Characteristics							Data Assessment			
	Data Collection	Years Available	Citizen-ship	Country of Birth	Previous residence	Flow/ Stock	Availability	Definition	Coverage	Bias	Accuracy
IPS Long-term migration (ONS) – Immigration and Emigration	Sample Survey Data collected at ports and airports. 2990 immigrant interviews.	1975-2013 (dependent on disaggregation)	Y	Y	Y	Flow	Publicly available	Question designed with UN Definition in mind	All formal migration Possible under-sampling of regional airports	Survey is intentions based. Potential bias in declaration of intention re. duration of stay	Sampling variability, especially for disaggregated flows (sample weighting is only done for the aggregate flows)
IPS Short-term migration (ONS) – Immigration and Emigration	Sample Survey	2004-2013	Y	N	Y	Flow	Publicly available	Several definitions used (1 month, 3 months, 3 months – UN standard) Lack of consistency with the long-term migration data: actual, rather than intended stay	See above	Events measured ex-post, only in one way: upon departure for short-term immigrants, upon arrival for short-term emigrants	See above
LFS – Immigration	Household Sample Survey. Two sets of variables: residence 12 months ago & national identity, citizenship questions	1992–2014 (quarterly LFS); 12 month transition question asked since Spring 2000 & since 2008 a 3 month transition	Y	Y	Y	Stock (also transition variable available)	Publicly available	Mainly stock data (although there is the question on residence 12 months ago, which effectively refers to migration transition data). Participants are usually resident in the UK.	People living in communal establishments are not sampled; therefore, under-coverage of some students. UK-wide survey. Possible undercount	Potential non-response of certain migrant groups, especially recent migrants. Possible undercount (see Wiśniowski 2013)	Sampling variability, especially under disaggregation

Table 1: Data Audit and Assessment (continued)

Source	Data Characteristics							Data Assessment			
	Data Collection	**Years Available**	**Citizen-ship**	**Country of Birth**	**Previous residence**	**Flow/ Stock**	**Availability**	**Definition**	**Coverage**	**Bias**	**Accuracy**
NINo (DWP) – Immigration	Administra-tive Registration Data	2002–2013 2014 data incomplete	Y	N	N	Proxy for flow	Publicly available	Does not match the UN definition. No duration of stay criteria so includes many short term migrants. Probable overcount.	Does not include British migrants, full time students and children. Includes people who work formally and/or claim social security. Doesn't include children. Probable undercount.	Only discernible bias comes from only counting circular and repeat migrants once. Possible undercount.	Accurately measures number of new NINo registrations. Error will be administrative. Non-random error in registration lag.
HESA – Immigration	Administra-tive Data Collected by HESA	2002–2010	Y	N	N	Proxy for flow	Permission obtained	Does not match exactly, but closer match than NINo. Includes students who drop out but not short courses. Majority will study for at least 1 year.	Just public HE students. Doesn't cover private universities and FE Colleges. No British, children or older people. Probable undercount	There is no discernible bias	Accurately measures number of non-UK domiciled students by citizenship; error will be administrative
Worker Registration Scheme – Immigration	Administra-tive Registration Data There was a £90 registration fee for migrant workers.	2004–2011	Y	N	N	Proxy for flow	Publicly available	Does not match the UN definition. There is no requirement to de-register. Information on duration of stay is limited.	Poor coverage as it only includes migrants from A8 countries and only includes people who migrate to work. Self-employed aren't required to register	The cost could be a disincentive for migrants to register, especially low paid Possible undercount	Accurately measures the number of migrants registering to work

Table 1: Data Audit and Assessment (continued)

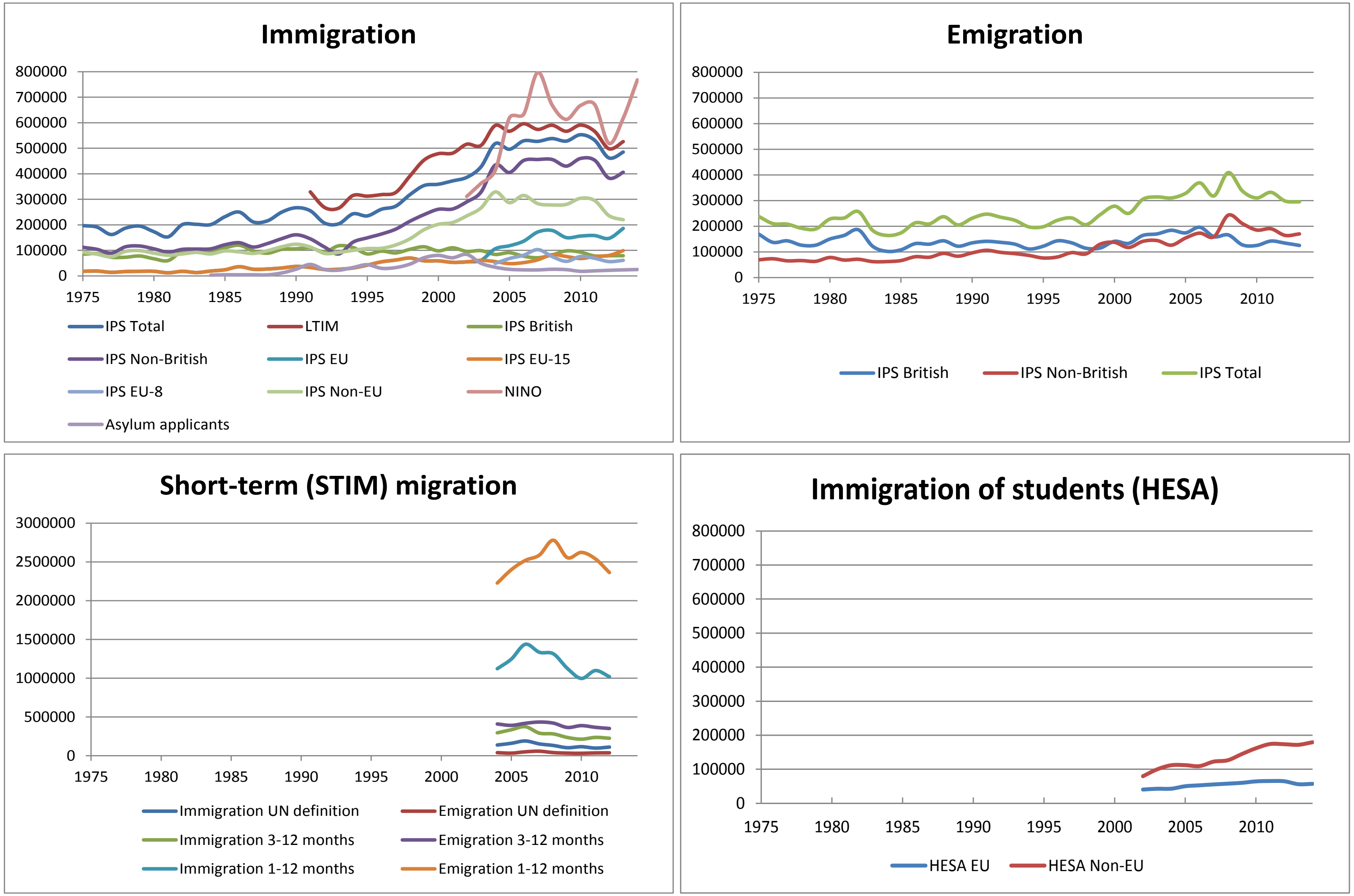
Source	Data Characteristics							Data Assessment			
	Data Collection	Years Available	Citizen-ship	Country of Birth	Previous residence	Flow/ Stock	Availability	Definition	Coverage	Bias	Accuracy
Home Office Entry Clearance Visa Data – Immigration	Administra-tive data	2004–2013	Y	N	N	Flow data	Publicly available	There is not clear information on duration of stay. Some limited infor-mation on length of visa is published, but this may not reflect length of stay as people may leave early or be granted an extension.	Does not include EU migrants or British return migrants. Does include all non-EU who require a visa for entry (and residence)	There is no discernible bias	Accurately measures the number of people who require a visa to enter UK; error will administrative
2001 Census Special Migration Statistics. Table CO711B - Immigration	Population Census Data Transition data indicating residence 12 months prior to Census night.	2001	N	N	Y	Flow data (tran-sitions)	Specially commi-ssioned census table.	Does not match UN definition exactly, transition data means duration of stay is unknown Respondents are usually resident in UK	Only includes migrants who are usually resident in England and Wales. Theoretically has complete coverage of England and Wales Probable undercount	Non-response bias of Census form. Hard to count groups – students, young people. Probable undercount	Accurately records people who answer question in survey. Error is administrative
2011 Census Data – Immigration	Population Census Data, Transition Data	2011	N	N	Y	Flow data (tran-sitions)	Specially commi-ssioned census table.	As above	As above	As above	As above

Table 1: Data Audit and Assessment (continued)

Source	Data Characteristics							Data Assessment			
	Data Collection	**Years Available**	**Citizen-ship**	**Country of Birth**	**Previous residence**	**Flow/ Stock**	**Availability**	**Definition**	**Coverage**	**Bias**	**Accuracy**
Home Office Asylum Seeker Data – Immigration	Administra-tive Data Applications, decisions and leave to remain grants Application numbers from Home Office match IPS adjustment.	1980s onwards	Y	N	N	Data can be used to derive stocks and flows	Publicly available	People will live in the UK for a period of time whilst they are awaiting their asylum decision. Most decisions are made within 6 months (Home Office) Cross-tabulated data on timings are not currently available, therefore assumptions about duration of stay are made	For the flow of asylum seekers coverage is good.	There is no discernible bias	Accurately records number of asylum seekers given leave to remain. Error is administrative

Notes: Green cells indicate a close match to true flow with relatively low uncertainty, orange cells denote a reasonable match to true flow with medium levels of uncertainty, and red cells indicate a poor match to true flow with large levels of uncertainty. For a full discussion for this approach to assessing UK immigration data please refer to Disney (2014).

Figure 1: Data on migration in the United Kingdom



Source: Office for National Statistics; Higher Education Statistics Authority; Home Office (various years)

When using multiple sources of data, especially data whose primary purpose is not to measure migration, a key task is to establish the extent to which the true flow is distorted and the associated uncertainty resulting from this distortion (Disney 2014). However, carrying out this assessment in a coherent way, and formally including it in the empirical exercise, remains beyond the scope of this study.

4. Review of Forecasting Methods

In this section, the main approaches to forecasting migration are reviewed. They are summarised in the table below and are reviewed in detail later in this section. Broadly speaking, there are two main types of migration forecasts. The first are deterministic forecasts (often called 'projections'; cf. Keilman 1990) and the second are probabilistic (stochastic) forecasts. In this section both types of models are defined and illustrative examples of each forecasting approach are given.

4.1. Review of Past Reviews

The existing recent reviews of migration forecasting methods include the studies by Howe and Jackson (2005), Bijak (2010, 2012), as well as – in the wider context of population projections or forecasts – by Wilson and Rees (2005), Keilman (2007, 2008) and Shaw (2007). In Bijak (2010), the theoretical discussion of the methods largely follows the grouping into the deterministic and probabilistic (stochastic) classes, which are discussed in more detail in Sections 4.2 and 4.3 below.

Amongst the various methods reviewed in those sources, some (e.g. ethno-surveys or event history analysis) require extensive micro-level information about individual migrants and migrations, which is currently available only for a handful of very specific contexts. As such, the micro methods remain outside of the scope of this review. Some other approaches, such as 'sociodynamics', with interlinked systems of dynamic differential equations, remain too complex and have too high data requirements to be useable in practice. Similarly, macro-level demo-economic models are quite complex and require bespoke specifications driven by the data availability, and hence have not been widely used in practice. Some other methods, such as assessments of 'migration potential' are criticised for not measuring migration, but rather a general state of dissatisfaction with the situation in the home country (Bijak 2010). Still, such variables can be potentially used as covariates in econometric models.

From the various methods, the ones most commonly used, both in academic literature and in official statistical applications, include judgemental scenarios, econometric models, and statistical time series extrapolations – with or without expert input, the former either as purely expert-based or Bayesian approaches. The variables being predicted are typically either stocks or shares of foreign-born populations, volumes or rates of flows, or net migration. Examples of the particular studies making use of the various methods are available in Bijak (2010).

In a vast majority of countries and international institutions (Eurostat, or the 2012 round of the UN World Population Prospects), argument-based scenarios are utilised, sometimes based on past trends, or additionally including additional parameters (target levels). This situation has not changed much since the early 1990s (as reported by Keilman and Cruijsen 1992, see also

Howe and Jackson 2005 for an update, specifically focusing on migration). The number of projection scenarios differs between the countries, usually between one and three; and most of the forecasts assume a constancy of migration (either volumes or rates) after some relatively short time. As such, variant projections do not include a formal assessment of uncertainty and it is not possible to evaluate their quality following the principles applied in this paper. This is their fundamental weakness: we know that single variant projections will be wrong, but there is nothing to tell us by how much.

Since migration forecasts prepared by the official statistical agencies are intended mainly as input for population forecasting, they are specified in terms of flows – usually net volumes (number of immigrants minus the number of emigrants), except for a handful of countries. Individual flows are being modelled in Canada, The Netherlands, Poland and Spain, with rates being used in Canada and – partially – The Netherlands. In the Netherlands, the assumptions additionally distinguish different groups of migrants (van Duin and Stoeldraijer 2014). In general, assumptions specified in terms of net migration volumes can be seen as problematic, as they conflate different processes (immigration and emigration) that are susceptible to entirely different drivers, and exhibit different patterns, for example by age (Rogers 1990).

The existing attempts to assess the quality of projections are scarce, with the notable exception of Keilman (2007, 2008) for Europe, and Shaw (2007) for the UK, whose findings are in line with the arguments presented in this report. In particular, Keilman (2007) noted that migration was historically underestimated; that the errors of various migration projections made in different European countries increase with the projection horizon, are very volatile, and vary between different countries. For Northern Europe, including Nordic countries, the UK, and the Benelux, the errors were markedly smaller than for countries from Central and Southern Europe, such as Germany, Austria, Luxembourg, Portugal, and Switzerland (Keilman 2007: 25). The differences are largely attributed to the underlying assumptions rather to the argument-based methodology, which was similar in most cases – so that lower forecast errors could be attributed to luck rather than to using superior methods.

Only in a few countries – notably, The Netherlands and New Zealand – probabilistic methods are used to predict migration (Bijak 2012). In both cases the models combine statistical extrapolations with expert arguments, based on the analysis of such factors as migration drivers or policies. Statistics New Zealand assumes that the median level of net migration stabilises after 2017. The errors around these assumptions are estimated and forecast by using an autoregressive integrated moving average (ARIMA), in this case ARIMA (0,1,2) model (Statistics New Zealand 2014, see also Section 4.3.1). Several variants are subsequently derived from these distributions, in particular, following the 5th, 25th, 75th and 95th percentiles.

A similar method is applied in the forecasts prepared by Statistics Netherlands, only that the error distributions are expert-based, following de Beer and Alders (1999), and different percentiles are chosen for the derived variants: (2½th, 33rd, 67th and 97½th; van Duin and Stoeldraijer 2014). To our knowledge, however, the models used in those countries have not been subject to the ex-post analysis of how well their uncertainty ranges cover the observed data.

The UN is expected to follow suit with their prototype migration forecasts (Azose and Raftery 2013). Their model already includes the calibration of the uncertainty assessment as a part of its construction. Hence, it is guaranteed to produce as well calibrated forecasts as possible, especially given the paucity of migration data available for most of the countries of the world. Still, a more thorough analysis of ex-post errors of the forecasts would require longer data series, and ideally more examples of forecasts prepared according to the same methodology.

Out of the entire range of migration forecasting methods and models, the most popular ones in practice are either deterministic judgemental scenarios, or time-series extrapolations, the latter possibly also equipped with covariate information or expert knowledge (Bijak 2010). One reason for this is that there is a time lag between statistical forecasting approaches being developed, mainly through research, and their application in national statistical offices. Other reasons that have been cited in the literature include: a perceived "misleading sense of precision" of the probabilistic methods; a "mechanistic" nature of the time series extrapolations; and a lack of appropriate training available for the practitioners of official statistics (Lutz and Goldstein 2004: 3–4). More contemporarily, the key challenges involve the forecasters' and users' attitudes towards uncertainty and the associated cognitive challenges; specificity of the various user requirements; the need to deal with incomplete or conflicting information; and some technical challenges (Bijak et al. 2015). Both deterministic and probabilistic approaches are discussed in more detail below, alongside some of the more recent examples of migration forecasts that are not included in the two reviews mentioned above.

4.2. Deterministic Approaches

The main sources of information in deterministic demographic models are judgemental scenarios. They describe possible future trajectories of components of population change (Bijak 2010). These scenarios are based on both quantitative and qualitative evidence about plausible changes in the quantities of interest and often produce low and high bounds for them.

Deterministic scenarios of international migration are used by the vast majority of the statistical institutes in the developed countries, especially Europe, in the production of migration forecasts (Bijak 2010). An example of deterministic projections comes from the ONS (2014), who fix their short term net migration assumptions at 165,000 per year. This assumption is based on past demographic trends and does not attempt to predict the impact of new or future government policies, changing economic circumstances or other migration factors (ibid: page 1).

The main criticism of deterministic methods is that they do not allow for a coherent and explicit quantification of uncertainty in their estimation. Traditionally, the uncertainty of a migration forecast has been taken into account through the production of uncertainty variants, namely 'high', 'medium' and 'low' (Wilson and Rees 2005). This, however, leads to problems with interpretation of the results and associated uncertainty as some users will, understandably take the medium trajectory as the most probable forecast. The high, medium and low scenarios simply refer to the quantity of the determined trajectories of the forecast and not their probability. Hence, it is not possible to assess whether the deterministic forecasts are well calibrated – there is nothing to calibrate the errors against. For this reason, the deterministic approaches are not included in the empirical assessment of methods presented in this study.

4.3. Probabilistic Models

Probabilistic forecasts specify the likelihood that a particular future population (or migration flow) value will occur given a set of assumptions about the underlying probability distributions (Abel et al. 2010, page 1; Wiśniowski et al. 2015). In other words, one can assign a probability to a range in which the quantity of interest (such as population size or a given migration flow) is expected to lie at some point in the future. In this section, we review probabilistic extrapolations of time series, probabilistic forecasts based on expert judgement, Bayesian forecasts, econometric models with covariate information, and extrapolation of time series through propagation of historical forecast errors.

4.3.1 Probabilistic Extrapolation of Time Series

The standard approach to time series extrapolation is to apply various ARIMA models, typically within the frequentist (likelihood-based) statistical paradigm (see Bijak 2010 for an overview in the context of migration). A particular strength of ARIMA modelling, in terms of application, is that forecasters can draw on a large body of existing statistical theory, applications and software (Wilson and Rees 2005). Since ARIMA models consider the time series as realisations of a stochastic process with uncertainty, construction of the predictive intervals is possible (Keilman et al. 2001).

A non-seasonal ARIMA model is classified as an ARIMA (*p,d,q*) where *p* is the number of autoregressive terms, *d* is the number of non-seasonal differences needed for a series to be stationary and *q* is the number of lagged forecast errors in the prediction equation. For a full account of the family of ARIMA models refer to Greene (2000). In demographic applications, the order of the ARIMA models used for forecasting usually does not go beyond (1,1,1) (Keilman et al. 2001).

ARIMA models can have a longer or a shorter 'memory', depending on the parameters of the process. For example, a simple benchmark long-memory model for extrapolating migration five to ten years ahead is a random walk with drift (Bijak 2012). This is a special case of ARIMA of order (0,1,0) where the logarithm of predicted migration, m_{t+1}, depends on the value from the preceding period, m_t, a drift constant c, and an error term, ε_t, usually assumed to follow independent normal distributions (Bijak 2012):

$$\ln(m_{t+1}) = \mathrm{c} + \ln(m_t) + \varepsilon_t. \tag{1}$$

In applications to migration forecasting, the random walk model, given its simplicity, was found to be favoured by formal model selection criteria (Bijak 2010, Bijak and Wiśniowski 2010). Additionally, one of its statistical features, namely non-stationarity, reflects the volatile character of the contemporary migration flows (ibid; see also Section 2). Furthermore, when applied to long-run forecasting, the uncertainty in the forecast increases, which is evident in widening of the prediction intervals (Keilman et al. 2001). This gradual increase of predictive uncertainty over the forecast horizon as well as the permanent effects of policy shocks for example, are also desirable features of this particular simple benchmark model (Bijak 2012).

The main criticism of this approach is that the forecasts are based on data alone and it is possible that this may lead to unreasonable predictions. For example, if one considers immigration to the UK from EU-8 countries following the expansion of the EU, using just an extrapolation of a time series of data, that includes the significant increase of immigration post-enlargement, then it is possible to obtain a forecast that shows ever-increasing immigration.

4.3.2 Probabilistic Expert-Based Forecasts

In comparison to the time series extrapolation outlined in 4.3.1, Lutz et al. (see 2004) developed 'expert-based probabilistic population projections'. Subjective expert judgement alone is used to prepare expert forecasts. The model has the form:

$$\upsilon_t = \overline{\upsilon_t} + \varepsilon_t \tag{2}$$

where υ_t is the migration flow under study, $\overline{\upsilon_t}$ is the average trajectory of the process which is an a priori assumption taken from the subjective expert judgement, and ε_t follows the chosen stochastic process.

If a structural break can be anticipated, such as the enlargement of the EU, the average trajectory may be modified (e.g. increased) at the time point of the break. The expert-based modification can relate to a structural change of the entire flow or only of its component, which can be reflected in the model by a common average trajectory and an additional term for the period when the structural change occurs. Expert opinion can be potentially of great importance if such breaks are anticipated to lead to a permanent change of migration levels and the historic data carry no information of similar events in the past.

In the case when the structural break cannot be predicted, such as an economic crisis or a war, experts may provide their opinion on the possible consequences to migration after the first signs of the break. However, the aftermath of any unpredictable breaks is even harder to forecasts due to their very nature.

An algorithm for expert-based forecasting of population components, including net migration, within the Bayesian paradigm is provided by Billari et al. (2012). This approach is further extended to include emigration and immigration separately as well as account for possible correlations between experts and population components in Billari et al. (2014).

However, a purely expert based approach does not make use of any data, which is a major limitation. It is solely reliant on subjective expert judgement. Experts also tend to be over confident in their judgements, and if one is basing the forecast on subjective information from a panel of experts, there could be issues with bimodality if there are a range of views on the panel. A Bayesian approach, described in the next section, allows inclusion of expert opinion within the statistical model based on data. An example of using subjective opinion together with data within a statistical model is further described in section 5.1.

4.3.3 Bayesian Forecasts

Different sources of information can be brought together using a Bayesian approach. Following the suggestions of Bijak (2010), historical trends, expert judgements and various models can be combined in a probabilistically coherent way. Hence, all these three sources of uncertainty can be included in the forecast. Furthermore, in circumstances where the data series are too short to allow for a meaningful classical inference, then one can adopt a Bayesian approach where formal elicitation of expert judgment about future trends is included in the model (Bijak and Wiśniowski 2010). The expert judgement could be included as prior distributions of different parameters of the forecasting model. The parameters are then subject to updating from data.

Azose and Raftery (2013) propose a method for probabilistic projection of global net international migration rates. Being fully Bayesian, their model provides a natural quantification of uncertainty from the posterior distributions of the forecasts. Only demographic variables are used as inputs in their models. They argue it leads to long-term projections without explosion in the degree of uncertainty (ibid).

Azose and Raftery (2013) fit a Bayesian hierarchical first-order autoregressive, or AR(1), model to net migration rate data for all countries. The (net) migration rate $r_{c,t}$, in country c and time period t is modelled as:

$$(r_{c,t} - \mu_c) = \phi_c(r_{c,t-1} - \mu_c) + \varepsilon_{c,t}\ , \quad (3)$$

where $\varepsilon_{c,t}$ is a normally distributed random deviation with mean zero and variance σ_c^2. Although expert-based priors are not elicited in this model, some parameters are constrained by means of the prior distribution, e.g. ϕ_c is assumed to take values from the range between 0 and 1, which ensures stationarity – the 'well-behaved' nature of the forecasted processes. The data used has only 12 time points per country. As such, a non-Bayesian inference could be difficult. However, taking a Bayesian approach alleviates this problem somewhat, as information from some countries can be used to augment the inference for other countries ('borrowing of strength').

The data used by Azose and Raftery (2013) is taken from the United Nations Population Division's biennial World Population Prospects report (United Nations Population Division 2011). A potential limitation of this approach is that the estimates of net migration for different countries are likely to be of varying quality and this is not taken into account in this model. This could lead to estimates of uncertainty that are not realistic. However, acknowledging the differences in the sources of data for all countries in the world is a substantial task and it is perhaps more realistic for broad groups of countries.

In response to rising user expectations about accuracy, timeliness and detail, statistical agencies have begun to widen the range of data sources they use. With few exceptions, methods to combine sources of data have largely been labour intensive, complex and provide little information about uncertainty. Amongst the exceptions, Bryant and Graham (2011) introduce a Bayesian framework for subnational population estimates and projections in New Zealand. The framework allows evaluation of data quality, estimation of historical demographic rates and counts, the projection of future rates and counts, and the assessment of uncertainty. For example, if data are taken from a sample survey, such as a labour force survey, then prior distributions could include information about the design of the survey (ibid).

Bijak and Wiśniowski (2010) present Bayesian forecasts of immigration for seven European countries to 2025 based on both quantitative data and qualitative knowledge elicited from country-specific migration experts in a two round Delphi survey. Their analysis is limited to total immigration, which means that the main forecasted variable is a (log transformed) volume of migration rather any of the related intensity measures, such as rates and ratios. It is also argued that eliciting expert knowledge on the actual size of the flows rather than relative indicators is a more natural approach. It also overcomes the problem with the difficulty of identifying the population at risk in a migration rate estimate (ibid).

Four models are specified. The first two are auto-regressive models of order 1, AR(1), one with a constant variance, and the other one with a random (stochastic) variance. The second set consists of two random walk models with drift (1), again one with a constant and one with a random variance. The expert opinion was found to be useful in describing the predictive uncertainty especially in the short term. Most of the immigration processes under study were found to be barely predictable in the long run and exhibited non-stationary features.

Abel et al. (2013) forecast possible migration inflows to the UK which could be driven by environmental factors elsewhere. Expert judgement is explicitly included in the forecast model, and this is elicited via a Delphi survey. This expert judgement is combined with time series data sets to produce a Bayesian forecast of so-called environmental migration up to 2060.

It is argued by Abel et al. (2013) that the key contribution of their paper, rather than the empirical dimensions of their forecast, is the approach taken by the research team. Firstly, they suggest, that there is value in seeking expert opinion in areas where other evidence is lacking. Secondly, they advocate that uncertainty should be coherently estimated in migration forecasts. They do not propose that predicting the future precisely is possible, especially in the case of unforeseeable events – in this case climate change and migration.

A similar approach to Abel et al. (2013) was adopted by Wiśniowski et al. (2014) to forecast how the outcome of the Scottish constitutional referendum that took place on 18 September 2014 may impact migration to and from Scotland. Historical data on Scottish migration have been combined with the expert opinion in an autoregressive model. The general conclusion is that the uncertainty about migration itself, especially international immigration, is relatively high, and this would not depend on the referendum outcome. Any incentives for immigrants offered by the independent Scotland would likely be balanced by an increase in internal and international emigration.

With this in mind, it is important to consider the utility of migration theories for estimation. Migration theories are at present too fragmented and too vague to be able to support forecasting, besides being used as possible justifications for the construction of argument-based scenarios taking selected push and pull factors into account (see Arango 2000, Bijak 2010, Kupiszewski 2013). Besides, there remains a problem of how to predict the time-varying predictors – their uncertainty will propagate into the migration forecasts, additionally compounded by the error in their mutual relationships. If treated within a joint statistical framework, where all sources of uncertainty are accounted for, the resulting forecast has extremely wide predictive intervals, which renders them hardly useful in practice (Bijak 2010).

4.3.4 Econometric Models with Covariate Information

Econometric models are a natural tool to both predict migration and to also verify particular economic theories on the basis of empirical data, as they typically include covariate information. The recent propensity for researchers to apply econometric models for forecasting migration dates back to the 1990s and almost exclusively focuses on population flows to Western Europe following the enlargement of the European Union.

An example of a simple econometric model comes from Fertig and Schmidt (2000) who estimated immigration rates to Germany m, from the four other candidate countries at the time of preparing the study (Czech republic, Estonia, Hungary and Poland). This model includes country specific, time specific and cross-sectional effects, in addition to the overall mean migration rate. Country of origin is denoted by i and time (year) by t. The model assumes that $\varepsilon_i \sim \mathrm{N}(0, \sigma_i^2)$, $\varepsilon_{it} \sim \mathrm{N}(0, \sigma_{it}^2)$ and ε_t is a Gaussian autoregressive process AR(1).

(4) $$m_{it} = \mu + \varepsilon_i + \varepsilon_t + \varepsilon_{it}$$

This approach was further developed by Dustmann et al. (2003), who forecasted European migration after the EU enlargement in 2004. They included covariate information in the form of relative income per capita. Their approach assumed stationarity of the time series and, as a result, did not take into account the effect of lifting the freedom of movement restrictions for the new EU citizens. Such specification of the model led to large ex-post errors in the first years of Dustman et al.'s (2003) forecast horizon for migration into the United Kingdom, but also yielded relatively accurate predictions for Germany – one of the countries that imposed transitory restrictions on the access to its labour market. This illustrates the importance of treating migration as a non-stationary process where systemic 'shocks' are expected.

Another example of an econometric post-enlargement forecast is taken from Alvarez-Plata et al. (2003). Migration to the EU-15 countries from ten countries of Central and Eastern Europe is forecasted. Specifically, they model the share of migrants from country *i* residing in country *j* expressed as a percentage of the total population of country *j* ($ms_{i,j}$):

(5) $$ms_{i,j,t} = \alpha + (1-\delta)ms_{i,j,t-1} + \beta_1 \ln\left({}^{w_{j,t}}/_{w_{i,t}}\right) + \beta_2 \ln(w_{i,t}) + \beta_3 \ln(e_{i,t}) + \beta_4 \ln(e_{j,t}) + \beta_5 \ln(P_{i,t}) + Z_{i,j} + u_{i,j,t}$$

where $u_{i,j,t} = u_{i,j} + v_{i,j,t}$ and $v_{i,j,t}$ denotes Gaussian white noise. The covariates are w – real income levels, e – employment rates, P – population sizes, and Z – dummy variables denoting geographical and cultural proximity of particular countries. The forecast assumed long-term convergence of the economic explanatory variables detailed above, to the average EU-15 levels. This results in an exponential decline of the net migration forecasts from 367,000 people per year shortly after expansion of freedom of movement to below zero by 2030 (ibid: page 60).

The main criticism of the econometric forecasts focuses on the shortcomings of the model specification, especially with respect to demographic variables, which are often missing.

Ideally, the forecast should control for basic demographic characteristics such as the size and the age structure of the population (Kupiszewki 2002). Furthermore, when migration rates are forecasted in a model which uses population size as an explanatory variable, such as the Alvarez-Plata et al. (2003) example given above, migration flows should increase the population at destination and decrease the population at origin by the same number (Bijak 2010, Cohen 2012). Treating the population size as exogenous is a source of bias in the model.

An example of a modelling approach which does include demographic considerations is a gravity model, where population sizes act as 'masses' drawing people over spatial distance – the more populous being the origin and destination countries, the higher volumes of migration being generated, other factors being equal (see Cohen et al. 2008 and Cohen 2012). Cohen et al. (2008) propose a generalised linear model (GLM) based on Zipf's (1946) gravity model of intercity migration. Their approach utilises demographic independent variables only, such as population and the area of origin and destination, as well as distance between origin and destination. This approach can be criticised as the time-invariant predictors, such as distance or area, can inform forecasts about the structure of the future flows, but not about their magnitude.

A model by Cohen (2012) provides point projections of net migration counts from 'less developed' to 'more developed' countries, as well as from the rest of the world to the USA, by using a gravity model based exclusively on the population size. The main criticism here is the above-mentioned dependence of the population size on the net migration. According to Cohen (idem), this effect is rather small. Nevertheless, to alleviate the problem he advocates carrying out projections in a step-wise manner. Second, the approach may be a reasonable approximation for the flows between two large regions of the world, but country-specific results would require adjustments of the methodology to fit particular data requirements. Last but not least, both approaches in Cohen et al. (2008) and Cohen (2012) rely on data as they are, that is, regardless of the definitions used in particular countries and differences in quality of data in terms of coverage and undercount.

4.3.5 Extrapolation – Historical Forecast Errors

According to Alho and Spencer (1985), errors in population forecasts arise from errors in the jump-off population and errors in the predictions of future vital rates. They argue that historical vital rates (including, in this case, migration) can be viewed as realisations of some, possibly non-stationary, stochastic process. The specification of future vital rates is treated as being derived from the predictions of some assumed parametric model of the past stochastic process (ibid). This allows for forecasting future migration with error, assuming that the actual forecast errors from the past can be extrapolated into the future.

The features of various approaches to probabilistic population forecasts have been synthesised within the framework of the EU funded project "Uncertain Population of Europe" (UPE) (Alho et al. 2005). The UPE predictions combine cohort-component models of population dynamics with probabilistic forecasts of fertility, mortality and migration based on the analysis of time series, expert opinion, and importantly historical forecast errors (ibid). There has been a comprehensive empirical analysis of the correlations between forecast errors for components of population change (including international migration), and between the countries under study.

4.4 Matching Forecasting Methods to Available Data

To guide the selection of appropriate methods for the empirical evaluation of different forecasting methods, it is necessary to determine what methods can realistically be applied to the available data. Furthermore, only certain models are appropriate for forecasting based on relatively short time series, or for predicting immigration flows disaggregated by the region of origin. On the other hand, some other models may only be suitable for forecasting the total UK immigration, or even net migration, as for methods based on the extrapolation of the past errors, depending on data availability. Consequently, a matrix of possible methods and data combinations, based on theoretical considerations as well as practicality given the data availability, is presented in Table 2. The disaggregations detailed in the table include 'total', 'UK citizens', 'Other EU', and 'non-EU'.

In summary, it is clear from the table that certain approaches can be applied to more of the sources of data and at greater levels of disaggregation than others. For example, if there are enough observations, one can apply a frequentist model to extrapolate a time series using an ARIMA model. On the other hand, it is only possible to apply the extrapolation of time series based on the propagation of historical forecast errors approach (Alho and Spencer 2005) to data for which forecasts have been produced in the past, in this case, to net overall migration.

With regard to the econometric models with covariate information, further investigation of the available covariate information is required. The main difficulty lies in identifying useable forecasts of economic and demographic covariates. For instance, it is likely that this is more readily available for the EU immigration flows and specific lower level flows, e.g. USA.

Finally, expert-based forecasting is conditional on the subjective information that can be elicited from an appropriate group of heterogeneous experts. It can be expected that subjective opinion is relatively easier to be obtained for the aggregate flows, such as total immigration, rather than disaggregated ones, or even specific flows.

Table 2: Data and Method Matrix

Data Source	IPS: Long-term				IPS: Short-term				LFS			NINo (DWP)			HESA			Home Office data: Visas issued	Home Office data: Asylum seekers
Method/Disaggregation	Total	UK citizens	Other EU	Non-EU	Total	UK citizens	Other EU	Non-EU	Total	Other EU	Non-EU	Total	Other EU	Non-EU	Total	Other EU	Non-EU	Total (non-EU only)	Total (non-EU only)
Extrapolation of Time Series, ARIMA model (eg. Bijak 2012)									*	*	*	**	**	**	**	**	**	**	
Probabilistic Expert Based Forecasts (Lutz et al. 2004, 2014)									*										
Bayesian Expert Based (eg. Bijak and Wiśniowski 2010, Abel et al. 2013)					F	F	F	F	*	*	*								
Econometric Models with Covariate Information (eg. Alvarez-Plata et al. 2003)	In-flows (net)	In-flows (net) ***	In-flows (net)	In-flows (net) ***					*	*	*								
Extrapolation of Time Series, Propagation of Historical Forecast Errors (eg. Alho and Spencer 2005)	Net mig. only																		

Notes: green cells denote that a given method for respective data set is readily applicable; for pale yellow cells elicitation of additional expert opinion is required; orange denotes that the application is still possible, but not recommended for the reasons set forth below, and red denotes that the application is unrealistic

* Given a large conceptual overlap between the LFS and IPS, the modelling of the IPS is preferred, given that it better reflects the underlying process (Table 1)

** Depending on the length of the series available; if the series are short, Bayesian methods should be used instead. NB: NINo is available for 12 observations, HESA for 9–13 observations depending on series

*** Depending on the level of disaggregation by country of origin and availability of the covariates; a simple model with only UK covariates size may be used instead

F – Data suitable only for a forward-look forecasting exercise (very short series)

5. Backward- and Forward-Look Empirical Analysis

This section summarises the empirical analysis undertaken, both with respect to the backward-look (based on truncated series), as well as forward-look (into the future) forecast exercises. We develop a framework for assessing the quality of various models based on the measures of error and calibration of forecasts. The key results of the analysis are also reported, while the more detailed numerical outcomes are included in the Appendices A and B to this report.

5.1 Empirical Research – Framework

There is no clear agreement, either amongst practitioners in national statistics offices or in the academic literature, about which type of probabilistic (stochastic) forecasts produce the 'best' results (Bijak 2010). Consequently, and given the main types of stochastic models reviewed and subsequently matched to the data in Section 4, a range of models has been applied in this study. This has allowed for an assessment and comparison of the various approaches of producing UK immigration forecasts, given the data available. The empirical analysis has been carried out whenever the combinations of data sources and forecasting methods have been highlighted in Table 2 as 'readily applicable' (green shading), not requiring additional information such as the elicitation of expert opinion.

(i) Extrapolation of time series using ARIMA models

For long enough series (as a rule of thumb, at least 20 observations), we have examined a suite of five different ARIMA models, listed below:

- Random walk with drift (1) for log-transformed volumes of migration flows;
- A general, unconstrained autoregressive model of the first order, AR(1):

(6) $$\ln(m_{t+1}) = \mathrm{c} + \varphi \ln(m_t) + \varepsilon_t;$$

- An ARMA(1,1) model, with a moving average element added to the AR(1) model above. Here, the model equation is:

(7) $$m_t = c + \varphi\, m_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1};$$

- An AR(1) model estimated on differences in log-transformed volumes of migration flows;
- An AR(1) model explicitly assuming the underlying linear trend (hence, 'de-trended'):

(8) $$m_t = c + bt + \varphi\,(m_{t-1} - c - b\,\{t-1\}) + \varepsilon_t.$$

Both in (7) and (8) the error terms ε_t are assumed to be independent and identically – in these cases normally, distributed.

(ii) Bayesian models (ARIMA) with expert prior distributions

The second group of models is essentially the same as the one listed above under category (i), with the addition of expert-based information through prior distributions, as detailed in Section 4.3.2. The models in this group have been estimated by using Bayesian, rather than frequentist (likelihood-based) statistical methods, in order to allow for a coherent and fully probabilistic integration of expert information with the observed data, which is one of the key features of Bayesian statistical approaches. For longer series, all five models listed above have been used; while for shorter series, for the sake of brevity, the exercise was limited to the first three models (random walk, general AR(1) and ARMA(1,1)).

For the time series models that are purely data-driven and that do not have additional covariates, the prior assumptions relate to the statistical properties of the time series of interest. For example, for the autoregressive AR(1) model, the prior distribution which is set in relation to the autoregressive term expresses a prior judgement about how the forecasted values depend linearly on the previous observed values in the data series.

So, any expert judgement included in a Bayesian time series model via the priors relates to a prior interpretation of the model terms. For example, if an expert believes a given flow is non-stationary, then for an autoregressive term in the mode, a prior distribution that allows a given parameter to take a value of 1 or above could be appropriate. In Section 5.3.2, we demonstrate the use of expert opinion on the category of flows from the EU. The definition of this category changes over time due to the enlargement of the EU in 2004 and 2007, when flows from the new EU member states are added to the flows from the EU-15.

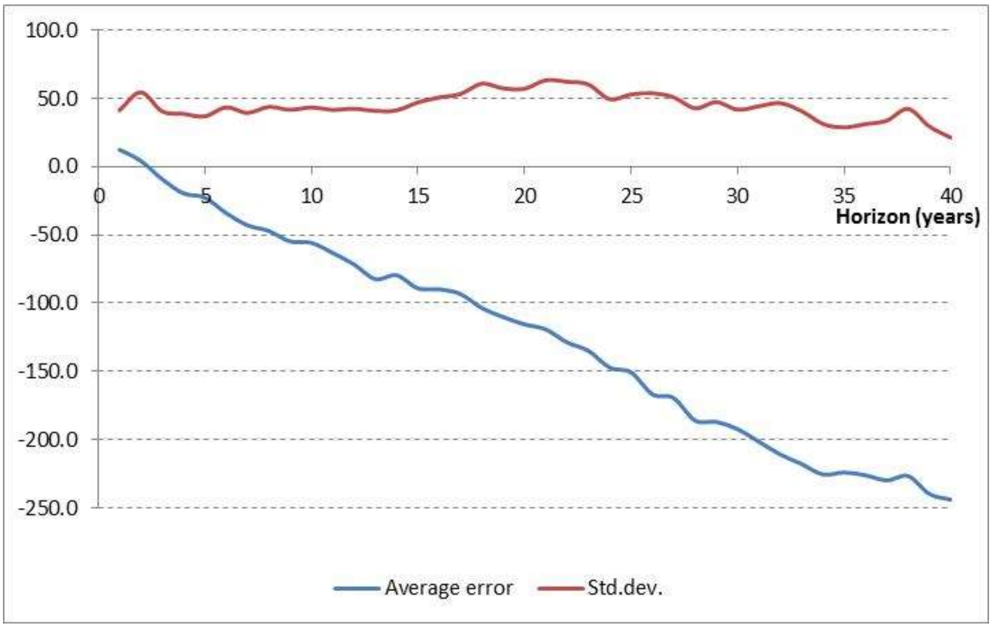
(iii) Econometric models with covariates (ADL)

Autoregressive distributed lag (ADL) models are extensions of the autoregressive (AR) models described in (i) and (ii). They utilise past values of migration, as well as current and past values of explanatory variables, such as Gross National Income, Gross Domestic Product, or the unemployment rate, to predict current values of migration. To forecast future values of migration, ADL models require point forecasts of the explanatory variables. Since more than one of such point forecasts can be fed into the model, this approach can be considered as scenario-based forecasting. In this exercise, we assumed ‘perfect foresight’, and have used the actual values of the explanatory variables (Gross National Income and unemployment rates) in the models.

(iv) Extrapolation of time series through propagation of historical forecast errors

Here, the past errors have been estimated by looking at the past ONS net migration assumptions for various editions of National Population Projections since 1970, and comparing them with the IPS-based net migration estimates across a range of projection horizons, from one to ten years. There is an apparent regularity in the forecast errors, which increase (in absolute terms) almost linearly with the forecast horizon (Figure 2).

Figure 2: Average error and its standard deviation by projection horizon, past official net migration assumptions for National Population Projections, 1970-based to 2012-based



Source: Data from Government Actuary's Department and Office for National Statistics

In order to reflect this particular trend, three models have been considered:

- Random walk with drift (1) for empirical errors from the past, and a constant forecast of net migration, assuming an average value from the last five years of observations, hence, akin to (2)[1]. In this case, the error drift has not been propagated into the forecasts.
- Random walk with drift (1) for the past errors, and another random walk with drift for net migration, with error in the forecast period assumed to follow the estimated model for past errors (1), allowing for drift;
- Linear trend for errors by horizon, and a random walk with drift for net migration, with error in the forecast period assumed to follow the error model, with a trend.

All net migration models have been estimated using a natural (not log-transformed) scale, as net migration can assume negative, as well as positive, values.

5.2 Empirical Research – Results

The models above have been compared across a range of indicators. For errors, the following summary measures have been calculated:

- **Mean Percentage Error (MPE):** an average difference between the observed values and median forecasts, expressed as percentages of the observed values. MPE measures the relative magnitude of errors, allowing for errors of different signs to cancel each other out.

[1] Interestingly, the migration assumptions of this model are very similar to the ones underpinning the recent edition of pure expert-based projections (Lutz et al. 2014), where it is assumed that an average net migration in the UK would equal around +184 thousand per year in 2010–15 (with a range between +152 and +214 thousand). The IPS average for the five year period 2008–12 was +185 thousand.

- **Mean Absolute Error (MAE):** an average value of absolute differences between the observed values and median forecasts. MAE measures the absolute magnitude of error without taking the direction into account.
- **Root Mean Square Error (RMSE)**: a square root of the average value of squared differences between the observed values and median forecasts. In comparison with MAE, the RMSE gives higher weight to larger errors.

In addition, the empirical coverage of the nominal **50-percent** and **80-percent** intervals has been computed, that is, the *ex post* frequency with which the actual observations fall into the respective *ex ante* error intervals. In well-calibrated models, we expect the empirical frequencies to be close to the respective nominal coverage probabilities – 50% and 80%, respectively. If the models are too conservative, the predictive intervals are too wide: more than 50% (80%) actual observations would be falling into the respective intervals, which should be narrower. Conversely, if the models are too optimistic, the predictive intervals are too narrow, and less than 50% (80%) actual observations fall into the respective intervals, which indicates that they should be wider. This second situation is potentially more problematic, as it may lead to too risky decisions.

As previously mentioned, a key theme of the report is how to safeguard against making bad migration forecasts under uncertainty of future events, data and models. In particular the truncation exercise is an opportunity to evaluate forecasting approaches against the inherent uncertainty of future events – the main source of uncertainty outlined in Section 2. For that reason, two different truncation points were chosen where events could lead to significant structural breaks in migration flows, and therefore where the effect on the magnitude of migration is uncertain.

In particular, the analysis has been performed on the series truncated in 2008, with five years of comparable data (2009–2013), and, for longer data series only – with observations truncated in 2003, thus, allowing to calculate the *ex post* errors and frequencies for ten years of empirical observations (2004–2013). Detailed results across the above set of five indicators are reported in Appendices A and B. This enables a comparison of the model estimate from that given point with the observed data after truncation

The first truncation point, 2003, is the last year in the time series before the expansion of freedom of labour movement within the EU in 2004 and the second, 2008, is the year of the global financial crisis and subsequent recession. Importantly, these two truncation points are different in nature. In 2003 it was safe to assume that there was going to be an increase in immigration from the EU countries, with the main source of uncertainty being the magnitude of increase in this flow. On the other hand, the effects of the financial crisis on the magnitude of immigration and emigration were far more uncertain, and the various disaggregations of the migration flows in question – for example by groups of countries of origin – were affected in different ways. In most cases, we would expect the models to perform better in the periods of higher stability, with fewer and smaller structural breaks.

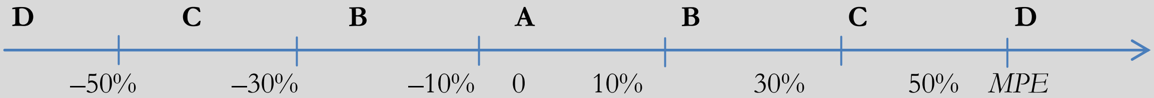
The algorithm applied to summarise the results has been as follows. First, the errors and coverage measures have been summarised by a range of qualitative codes related to the MPE and to calibration of both 50-percent and 80-percent intervals, as shown in Box 1.

Box 1. Conversion of error and calibration measures into quality classes: Examples

Model	MPE (Mean Percentage Error)	MAE (Mean Absolute Error)	RMSE (Root Mean Square Error)	Coverage: 50-percent interval	Coverage: 80-percent interval	Quality Class
General AR(1)	–6%	32,190	42,310	80%	100%	**A<**
	Error classes: **A, B, C, D**			Calibration classes: =, ~, <, >, <<, >>		

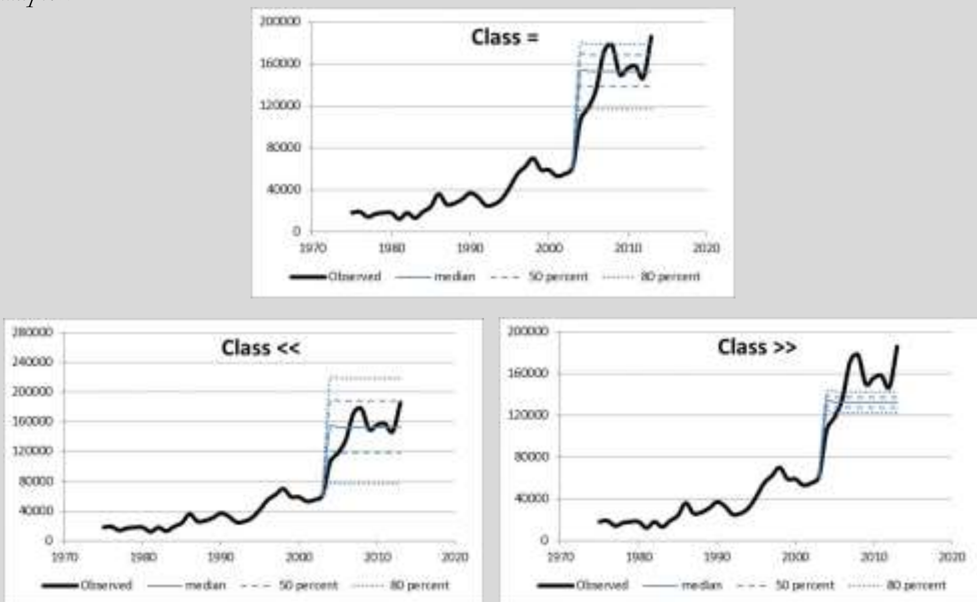
The quality class is based on two elements: average error, including the direction of the bias, as measured by the MPE, and calibration of the 50 and 80-percent intervals. The assumed error and calibration classes are as follows:

Error classes
A: MPE between –10% and 10%
B: MPE either between –30% and –10%, or between 10% and 30%
C: MPE either between –50% and –30%, or between 30% and 50%
D: MPE below –50% or over 50%



Calibration classes
= Well calibrated models, both empirical coverages within ±10 percent from the nominal values
~ Slightly miscalibrated models: one value above and one below the nominal ±10 percent
< Too conservative model: too wide intervals, both values above the nominal +10 percent
> Too optimistic model: too tight intervals, both values below the nominal –10 percent
<< Extremely wide intervals: both 50- and 80-percent intervals cover 90–100% of observations
>> Extremely tight intervals: both 50- and 80-percent intervals cover 0–10% of observations

Stylised examples:



The following models were included in the analysis: random walks, general AR(1) and ARMA(1,1) models from equations (6) and (7), AR(1) estimated on differenced series, as well as de-trended AR(1) models introduced in equation (8)– all in the frequentist and Bayesian versions. Furthermore, for long non-stationary series, ADL(1) with predicted covariates and with 'perfect foresight', as well as three models for past errors propagation were estimated: with random walk errors based on constant and random walk migration forecasts, and with a linear trend in errors.

The exercise was carried out both for five years and ten years of data available *ex post* (on series truncated in 2008 and 2003), except for the short series, where the evaluation could have been performed only for five years. Similarly, the ADL(1) models with forecasted predictors were only included for series truncated in 2008, as no similar data on the predictions of the economic variables were available earlier years. Overall, the assessment of performance is based on the results of 198 models. A summary of the models and series used is shown in Table 3.

Table 3: Summary of data series and models analysed in the empirical exercise

Data Source	**Long series**									**Short series**	
	IPS: Long-term immigration				**IPS: Long-term emigration**				**Asylum seekers**	**NINo (DWP)**	**HESA**
Method/Disaggregation	Total	UK citi-zens	Other EU	Non-EU	Total	UK citi-zens	Other EU	Non-EU	Total (non-EU only)	Total	Total
Extrapolation of Time Series											
Random Walk	+	+	+	+	+	+	+	+	+		
AR(1)	+	+	+	+	+	+	+	+	+		
ARMA(1,1)	+	+	+	+	+	+	+	+	+		
AR(1) on differences	+	+	+	+	+	+	+	+	+		
AR(1) de-trended	+	+	+	+	+	+	+	+	+		
Bayesian Expert Based											
Random Walk	+	+	+	+	+	+	+	+	+	*	*
AR(1)	+	+	+	+	+	+	+	+	+	*	*
ARMA(1,1)	+	+	+	+	+	+	+	+	+	*	*
AR(1) on differences	+	+	+	+	+	+	+	+	+		
AR(1) de-trended	+	+	+	+	+	+	+	+	+		
Econometric Models											
ADL(1) projected covariates	*		*								
ADL(1) perfect foresight	+		+								
Error propagation (net migration)											
RW errors, const.	+										
RW errors, RW	+										
Linear errors, RW	+										

Notes: '+' indicates that a given combination of a data series and forecasting model was included in both parts of the empirical exercise; and '*' that it was only included for series truncated in 2008

Once the quality classes have been obtained for all models, they have been converted into numerical scores, penalising the models with high errors and/or those being miscalibrated. Two conversion tables have been used: one 'symmetric', where the errors and calibration have been deemed similarly important and penalised to a similar degree; and one 'asymmetric', where the magnitude of errors was assumed to be more important than calibration. The asymmetric table reflects an assumption that the magnitude of errors may have more direct policy consequences than the assessment of uncertainty. Both score tables are detailed in Appendix C.

Finally, for both scoring tables, average scores have been calculated separately for each of the three categories of data: (i) long stationary series, (ii) long non-stationary series, and (iii) short series, for appropriate models under study, as listed in Tables A1–A5 and B1–B5 in Appendices A–B. Those average scores have been ultimately given a colour rating: green for the relatively most appropriate methods for data series exhibiting particular features; orange where the application of such models needs to proceed with caution; and red for those that are definitely not recommended, based on high likelihood of very high errors and/or problems with calibration. The summary of the results of the empirical analysis for both rating tables – symmetric and asymmetric – is offered in Tables 4 and 5, respectively. Note that for short series, only Bayesian models can be realistically used, as the data ideally need to be augmented by expert judgement.

Table 4: Backward look summary table – *symmetric* scores, errors and calibration similarly important

(i) Extrapolation methods

Type of data series	Random walk	General AR(1)	ARMA(1,1)	AR(1) on differences	AR(1), de-trended
Longer series, stationary features					
Longer series, non-stationary features					

(ii) Bayesian methods

Type of data series	Random walk	General AR(1)	ARMA(1,1)	AR(1) on differences	AR(1), de-trended
Short series of underlying data				N/A	N/A
Longer series, stationary features					
Longer series, non-stationary features					

(iii) Econometric models **(iv) Past errors propagation**

Type of data series	ADL(1) with projected	ADL(1) perfect foresight	RW errors, constant	RW errors, RW migration	Linear errors, RW migration
Longer series, non-stationary features					

Table 5: Backward look summary table – *asymmetric* scores, errors more important than calibration

(i) Extrapolation methods

Type of data series	Random walk	General AR(1)	ARMA(1,1)	AR(1) on differences	AR(1), de-trended
Longer series, stationary features					
Longer series, non-stationary features					

(ii) Bayesian methods

Type of data series	Random walk	General AR(1)	ARMA(1,1)	AR(1) on differences	AR(1), de-trended
Short series of underlying data				N/A	N/A
Longer series, stationary features					
Longer series, non-stationary features					

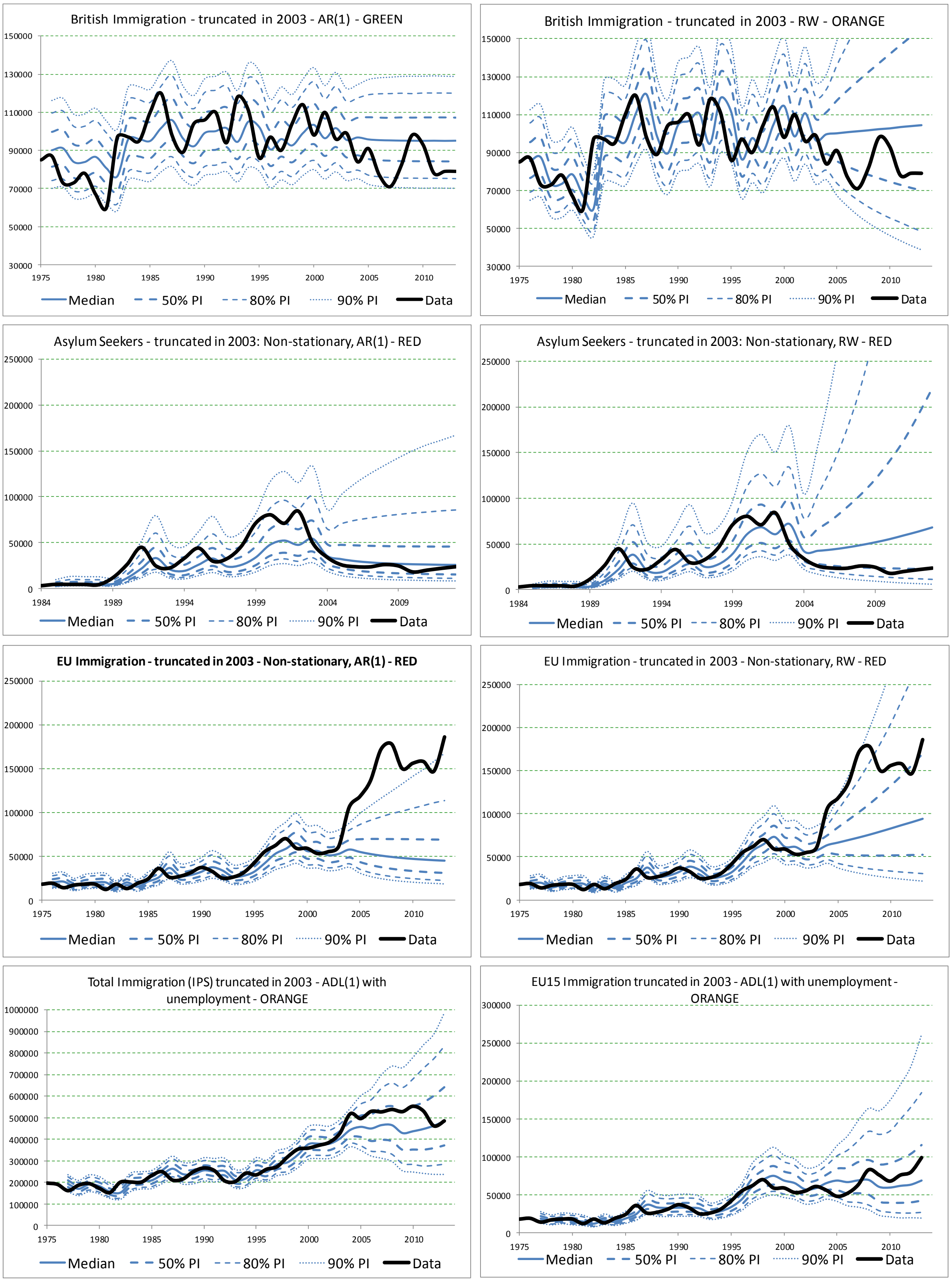
(iii) Econometric models **(iv) Past errors propagation**

Type of data series	ADL(1) with projected	ADL(1) perfect foresight	RW errors, constant	RW errors, RW migration	Linear errors, RW migration
Longer series, non-stationary features					

In Figure 3 we present selected examples of the backward-look exercise from various groups of models (green, orange and red). We observe that the forecasts can vary greatly in terms of median forecasts, which are reflected in the error measures (for example, the Mean Percentage Error, MPE), and uncertainty around the median, which results in different calibration scores.

For longer series that exhibit stationary characteristics, the AR(1) or ARMA(1,1) models lead to relatively small forecast errors and well-calibrated predictive intervals. This is a direct consequence of the fact that these models yield forecasts that converge to some constant value over time with a constant uncertainty, which is a requirement for stationarity. A random walk with drift model introduces ever increasing uncertainty and the ever increasing (or decreasing) level of future migration, which may result in relatively large errors and poor calibration.

Figure 3: Selected examples of backward-look exercise. First row: stationary series; second row: non stationary series; third row: non-stationary series with a structural break in 2003; fourth row: econometric models with 'perfect foresight'



Source: Author's calculations based on Office for National Statistics; Home Office (various years)

As an example of forecast errors in the results obtained for non-stationary series, the more reliable models for migration from the new EU member states, such as the AR(1) models with additional expert input (Section 5.1), predicted that there was a 50 per cent chance that the average annual immigration between 2004 and 2013 would range from 100 to 200 thousand people; in reality, this proved to be just above 150,000.[2]

The high errors are best illustrated by forecasts of British immigration based on 1975–2003 data (top plot of Figure 3). The AR(1) model produces a constant forecast at around 95,000 with the uncertainty intervals remaining at a similar level for each year of the forecast. For the RW model, the width of the predictive intervals and the level of migration increase with every year of the forecast. For non-stationary series, such as Asylum Seekers (second row of Figure 3) and immigration for the EU countries (third row of Figure 3), the RW model may provide more realistic assessment of uncertainty but still does not guarantee small errors in forecasts. It is due to the fact that non-stationary series may change the direction "randomly".

To sum up, only when the underlying series were relatively stable, such as for migration of the UK nationals, were some models able to produce relatively small errors – in other situations, the applicability of various methods was either limited, or outright inappropriate, depending on the exact circumstances. In particular, no model was able to predict migration well if the underlying data series were short, or in the presence of shocks (structural breaks), such as the enlargement of the European Union.

5.3 Sensitivity Analysis

One of the three main sources of uncertainty outlined in Section 2 is the uncertainty associated with the forecasting models themselves. It is clear from the results in section 5.2 that applications of different models produce forecasts of varying quality for given time-series and data. Furthermore, each forecasting model is based on certain assumptions; both statistical and related to expert opinion. There is no single, objective and 'correct' approach to forecasting migration as each approach is based on different assumptions and these assumptions can affect the calibration of the uncertainty in the model, the error, and also the magnitude of the forecast.

As such, it is important to test the sensitivity of the forecasts to model assumptions. In this section, brief illustrative examples of forecast sensitivity to prior assumptions for the Bayesian time series models, expert knowledge for EU immigration flows and applications of the econometric models to specific migration flows are assessed. Given the large number of models for different disaggregations of migration flows and sources of data, the aim of this section is to provide examples of how alternative model assumptions can affect migration forecasts.

5.3.1 Sensitivity to Prior Assumptions

As mentioned in Section 5.1, the prior distributions in Bayesian time-series models relate to a judgement of the statistical properties of the particular time series to be modelled. Forecast sensitivity to changing prior assumptions of the AR(1) model for total inflows was tested for time series truncated both in 2003 and 2008.

[2] See, for example, the IMEM estimates (Raymer et al. 2013), available at http://www.imem.cpc.ac.uk.

We introduced various assumptions about the prior distributions for both the autoregressive term and the overall error term in the model. We assumed such prior distributions that give preference to data, as well as distributions that impose stationarity on the forecasts. For the total migration flow measured by the IPS, the forecasts are largely insensitive to the prior distributions. It seems possible that for the forecast to be significantly affected by expert judgement, the judgement about particular statistical properties (e.g. stationarity) incorporated in the prior distribution would have to be very strong and the corresponding prior distribution would need to be very certain, allowing minimum error. However, this finding should not be generalised, as this may not be the case for each different disaggregation and source of data. In general, the more observations in the data set are available, the less important expert opinion is. If only scarce data points are available, subjective opinions will play a more prominent role. Consequently, it is recommended that the sensitivity of forecasts to prior assumptions is carried out in any future Bayesian forecasting of migration by the Home Office.

5.3.2 Sensitivity to Expert Knowledge

The potential of expert opinion is demonstrated on the category of flows from the EU. The definition of this category changes over time due to the enlargement of the EU in 2004 and 2007, when flows from the new EU member states are added to the flows from the EU-15. We focus on the first of the breaks as it brought about a significant change in the series.

In particular, we supplement the five time series models described in Section 6.1 (i) with basic "expert knowledge" based on the population size of the EU before and after the 2004 enlargement. Forecasts for the period 2004-2013, based on the 1975-2003 series, are rescaled by multiplying the drift term c in the model by the ratio of the EU population size after the 2004 enlargement to the population size before 2004, measured in 2003. This reflects the change in the expected level of the EU flow. Also, as of 2003, the date of the next enlargement in 2007 is treated as unknown.

The results vary depending on the model. An improvement in accuracy of the forecasts is observed in the AR(1) and ARMA(1,1) models, where the MPE is significantly reduced compared with the MPE for models without expert knowledge. In the random walk, de-trended AR(1), and AR(1) models on differences, reduction of error is, however, only minimal. Nonetheless, in a general case, expert opinion should be used with caution, as it may not always lead to a reduction in forecast error. Expert opinion may also be anchored in the past data and introduce a false sense of conviction in such a forecast. A comprehensive review of issues related to quantification of expert opinion within Bayesian paradigm and caveats involved is in O'Hagan et al. (2006). Following the recommendations of Bijak and Wiśniowski (2010), we advocate using diversified opinions obtained from experts with various backgrounds, and averaging forecasts over various models, by applying, for instance, Bayesian Model Averaging (idem).

5.3.3 Sensitivity of the econometric models

The results of the econometric models suggest that there is a link between the unemployment rate and total flows as measured by the IPS, as well as flows from the EU-15, and that unemployment rates can help predict migration. To assess the quality of these forecasts, we test the model with various configurations of the simultaneous and lagged unemployment rate together with the Gross National Income (GNI) as a macroeconomic proxy measure of wages dynamics and economic performance (see Abel 2010). The results suggest that the GNI is not significant, thus in the final forecast it has been removed from the model.

Incorporation of only (1) simultaneous, (2) lagged, and (3) both simultaneous and lagged unemployment rates lead to different paths of the forecasted migration. The forecasting errors, however, remain similar in all three configurations.

The fact that the unemployment rate can be used as a predictor for migration triggered a hypothesis that it actually influences only migration related to labour. To analyse this hypothesis, we utilise the IPS data on total labour migration (1977–2013). On the one hand, the results confirm the explanatory power of the unemployment rate. On the other hand, forecasts based on the series truncated in 2003 and 2008 become biased as they fail to predict the downturn in labour migration after the economic crisis in 2008.

This result confirms that migration is a complex process that may be influenced by various social and economic circumstances in different periods of time and taking place in various places of the world. Even if a covariate, such as the unemployment rate, can explain its past behaviour, it does not guarantee unbiased and precise forecasts. Moreover, different types of migration are likely to have different and interrelated drivers. For example, migration for family reasons usually follows labour migration, which may be driven not only by the relative economic situation of the sending and receiving countries, but also by the existing networks in the receiving country. Therefore, we advocate using econometric models with caution, and applying them to specific flows, rather than aggregates. Also, due to the nature of predictions based on projected values of the covariates, the forecasting horizon should remain short, for example, five years at most.

5.3 Forward-Look Exercise

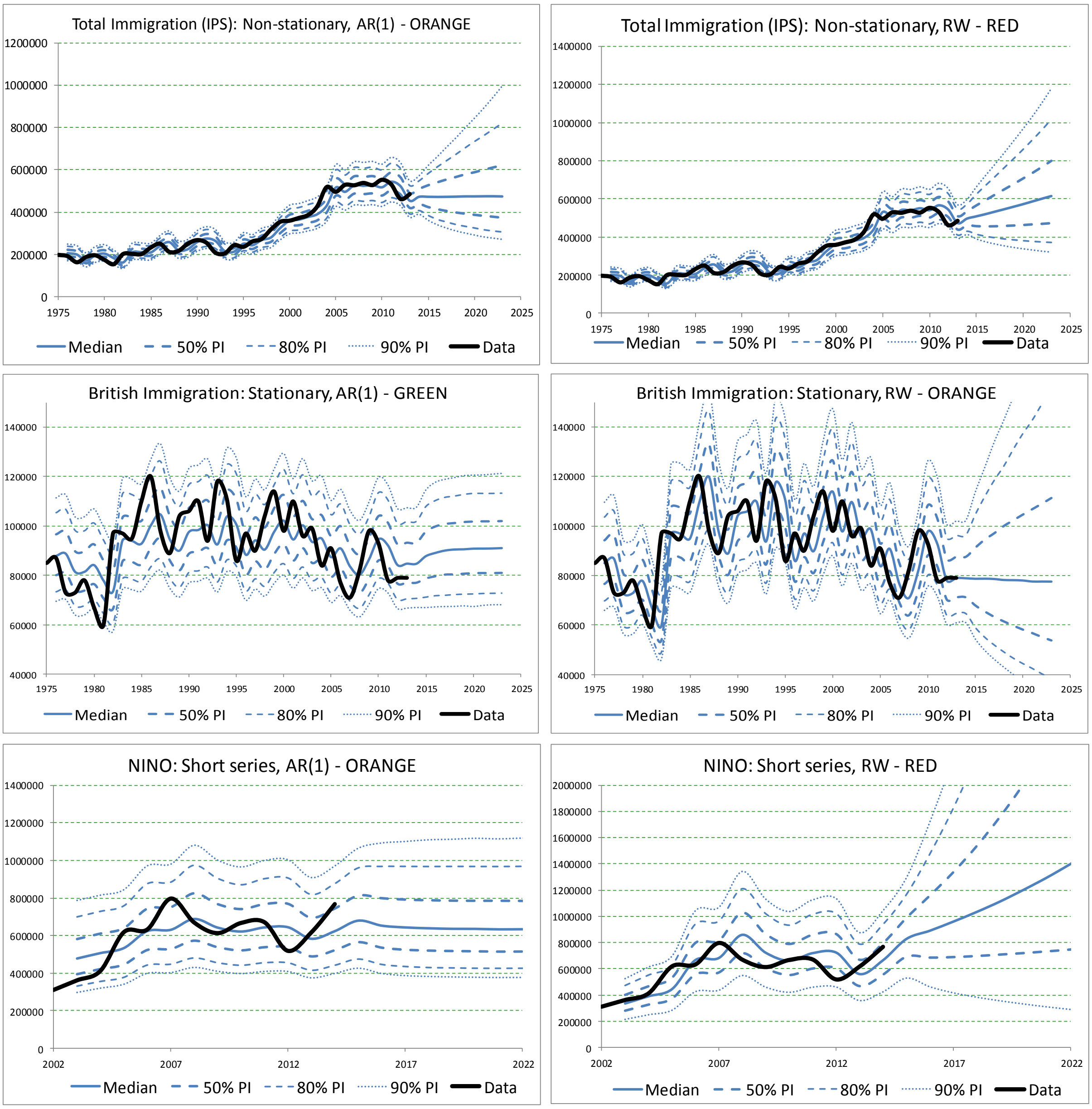
In this section, we present sample results produced with the series ending in 2013, that is, as if we were to forecast migration into the future. Here, we utilise Bayesian models without any expert judgement included. These results are presented as an illustration of the possible outcomes of various time series models and how they may impact the predicted trajectories of future migration. Hence, these results should not be interpreted as actual predictions of future migration. As an example, in Figure 4, we show two results of the forecasting models based on the stationary series (immigration of British nationals), non-stationary series (total immigration as measured by the IPS), and short time series (NINO data), which lead to relatively wide predictive intervals. These forecasts represent also the different scores of the model: green, orange and red.

For example, for total immigration (first row of Figure 4), we observe that the AR(1) and RW models yield similar forecasts in terms of both central tendency (median) and uncertainty. However, based on the backward-look performance of both models, we assign them orange and red categories. In another example, forecasts of the immigration of British nationals (second row) produced by the same types of models differ greatly in terms of uncertainty. The RW model from the orange category yields very wide predictive intervals comparing to the AR(1), which may be deemed impractical from the point of view of policymakers. Similar conclusions can be made about the short time series such as NINO data (third row).

The explanation of such behaviour lies in the nature of the processes represented by the RW and AR(1) models. If a given migration series is non-stationary, such as total immigration (first row), the AR(1) model without any strong expert-based prior input will still be able to pick up non-stationarity. However, if a series seems to be stationary (British immigration in the second row), the application of the RW model may be misleading in terms of the uncertainty.

The model for forecasting can be selected by using the backward-look exercise and comparison of *ex-post* errors (e.g. MPE) and calibration, such as those performed in Section 5. Alternatively, an ensemble of models can be used and their forecasts can be averaged by using, e.g., Bayesian Model Averaging (Bijak and Wiśniowski 2010; see also Section 5.3.2).

Figure 4: Examples of forward-look exercise. First row: non-stationary series; second row: stationary series; third row: short series



Source: Author's calculations based on the Office for National Statistics; Home Office (various years)

6. Conclusions and Recommendations

This section summarises the main findings and limitations of the study, and makes suggestions for key further research directions and policy recommendations. We conclude by discussing the use of statistical decision analysis and risk management tools to support migration related policies and decisions made under uncertainty.

6.1 Key findings and limitations of the study

Migration is a very complex and multi-dimensional process, responding to many different drivers, so its forecasting is extremely difficult. Migration forecasts have the highest levels of error amongst the main components of population change, and the longer the forecast horizon, the higher the errors. In many cases, forecast errors become too high to be useful beyond the horizon of five-ten years into the future (Bijak and Wiśniowski 2010).

Besides, no good or comprehensive migration theories exist (Arango 2000), which could otherwise be helpful in forecasting. Explanations put forward for migration processes are very fragmented and entrenched in particular disciplines of social sciences, such as economics, sociology or human geography. Even if there exist credible theoretical explanations of past migration flows, their tenets are very difficult to extrapolate into the future. Moreover, such extrapolations are likely to contribute additional uncertainty to the migration forecasts due to the inherent uncertainty of the social, geographic and economic processes used to predict migration.

Errors in data compound the already very high uncertainty about the future migration. As discussed in Section 3, in different sources of data, migration is defined in different ways. There is also a wide variation in what groups of migrants are covered by the particular sources, and how accurately migration is measured. Often the data series are short, which precludes the use of many forecasting models, especially those without additional expert input.

Many methods have been used to predict migration, which are reviewed in Section 4. The existing forecasting methods usually involve deterministic scenarios or probabilistic models – the former do not attempt to measure uncertainty in any way, while the latter seek to assess the probability of various future migration trajectories. Even though most of the official migration projections prepared by the statistical offices of other countries or international organisations (such as the UN or Eurostat) remain deterministic, some countries (The Netherlands and New Zealand) have moved towards probabilistic methods, and the UN is expected to follow suit with their prototype (Azose and Raftery 2013). Their model already includes the calibration of the uncertainty assessment already as a part of its construction.

The probabilistic methods are either based on the past data (including errors from the past forecasts), expert opinion, additional variables, or a combination of those sources of information. Amongst the probabilistic methods, no single approach is universally superior to others, and their applicability depends on the particular migration flow being forecast, and on the features of data, such as the length of the series and stability of trends.

When tested on the empirical data from the past, all models produced considerable uncertainty, but for some the prediction errors were much larger than for others. Still, even in such cases some models performed better than the other ones: models that did not assume stability of trends, when none was to be expected, at least described the forecast uncertainty more accurately – and thus more honestly.

6.2 Migration forecasting: a three-point modelling process

In this section we recommend a general approach to migration forecasting. This is based on the three types of uncertainty outlined earlier and the following assessment of the publicly available data; main methodological approaches; and then the empirical forecast results. Importantly, the recommendations focus on the process of making migration forecasts rather than recommending a single model. As shown in the analysis of the forecast results, different models appear to perform relatively well (or relatively badly) depending on the nature of the data series. Consequently, it would not be prudent to recommend the 'best model'.

The forecasting process we recommend should safeguard against very unrealistic forecasts of migration and should ensure appropriate expressions of forecast uncertainty. The importance of uncertainty in migration forecasts is stressed throughout the report and this section on recommendations is no different.

- **A thorough understanding of what one is trying to forecast, and of the features of the particular migration flow**

With regard to the three categories of uncertainty outlined in Section 2 this relates to the inherent uncertainty of migration itself. For example, the migration flow of interest could be susceptible to external political or economic shocks, or particularly influenced by changes in government policy or other interventions, or, conversely, could be a flow which is relatively stable over time. Understanding the potential future nature of the flow, will help guide appropriate selection of a forecasting model(s).

By the same token, there is inherent uncertainty in the assessment of a migration flow's potential volatility to changes in policy. As such the forecast uncertainty should allow for the possibility for future changes in the migration flow of interest. This will aid policy makers in any decision making based on a migration forecast. For example, asylum flows, generated by war and conflict in other parts of the world, can be expected to be less stable than return flows of the UK nationals, and the respective policy impacts of these two flows will also differ.

- **The available data need to be assessed, with their relative strengths and weaknesses taken into consideration**

This consideration relates to the second source of uncertainty outlined in Section 2 – the uncertain nature of migration data itself. For example, one of the conclusions from the empirical analysis is that migration forecasts based on short time series are problematic. Where the forecasts are estimated using a Bayesian times series approach, with a low number of observations, the forecasts are strongly influenced by the specification of the priors.

Also, there tends to be a high level of uncertainty in forecasts which are based on short time series, and so are predictions of such migration flows, which were subject to shocks in the past. This does not mean that the forecasts are of no use as they illustrate, given the available evidence, the high level of forecast uncertainty that is present when one only has a limited number of observations, and the sensitivity of the results to model specification.

Harmonising the different data sources is beyond the scope of this report; however, in the interpretation of results, it is recommended that one considers the assessment of each of the sources of data in relation to true flow. This should help make sense of some of the discrepancies between the forecasts based on different sources of data.

- **An appropriate modelling approach needs to be selected given the characteristics of the migration flow in question and the available data**

The final recommendation relates to selecting appropriate models to forecast the flow of interest, taking into account in particular the length of the available data series, as well as its character – especially, whether the series exhibit non-stationary features. In particular, a data series with non-stable characteristics should not be forecast by using models which assume stationarity of the process, and vice versa. Short data series may ideally require additional expert input concerning the future migration flows.

Following the process outlined above cannot guarantee that the resulting forecasts will exhibit no or only small errors, but would help safeguard against making poor forecasts and thus against radically incorrect decisions. It is unrealistic to expect that there will be no uncertainty: even in good forecasts, errors are inevitable. It is thus especially vital that the forecasters do not offer forecasting methods that strive to be producing too certain predictions, as they will most likely fail, but neither should policy-makers expect or require them.

6.3 Concluding remarks: From forecasts to decisions

The main findings of this study suggest that, given the high levels of uncertainty of migration forecasts, this uncertainty should be stated explicitly, ideally in terms of probabilities. Presenting the predictive uncertainty needs to be an important part of any migration forecast outcomes, as it can help decision makers safeguard against the less expected developments. Further work in this area, instead of trying to do the impossible and design the 'best possible' migration forecasting method, should rather focus on translating uncertain forecasts into decisions, creating early warning systems, and providing risk management strategies. The prerequisite is an honest reporting not only of forecasting uncertainty, but also of the related features of the forecasting models, including their past performance and susceptibility to shocks.

More research needs to be done on early warning models which would seek to detect the signs of changes in migration trends in response to the dynamics of some other variables, for example macroeconomic indicators (unemployment, job vacancies) or policies (migration caps, visa regulations, etc.), which could signal changes in migration trends and herald upcoming structural breaks in long-term trends. Such models could be also used to test the possible responses of migration flows to different policies by allowing the decision makers to compare the results of different policy interventions. The outcomes could be subsequently analysed by using the risk management tools – combining the potential policy impacts of such interventions with their uncertainty – to help the policy makers make prudent and robust decisions. The examples of policy areas influenced by various migration flows include labour markets; services provided to migrants and asylum seekers; access to benefits, health and social care; access to naturalisation, etc. An example of such a risk management matrix for different migration flows is provided in Table 6. The key policy focus should be especially on the red and yellow areas – those with either a medium or high impact of different types of migration on a range of policy areas, or having substantial uncertainty.

Table 6: A stylised example of a risk management matrix related to migration forecasts, from the point of view of possible policy impacts

Uncertainty (risk) / Impact	Low	Medium	High
Low		Long-term migration of UK nationals*	Short-term non-EU migration
Medium		Long-term migration of other EU nationals: old EU (Western Europe)* Long-term migration of non-EU nationals	Long-term migration of other EU nationals: Central & Eastern Europe* Short-term EU migration* Student migration
High		Visas issued, by type	**Refugees and asylum seekers**

Notes: Asterisks (*) denote flows, for which not too many policy controls exist. No migration flows are characterised by low uncertainty.

Furthermore, the whole forecasting process could also become more interactive, with forecasters providing bespoke decision advice related to specific user needs (Bijak 2010). For example, it is possible to utilise a formal statistical decision analysis to support migration-related policies and decisions under uncertainty (Alho and Spencer 2005, Bijak et al. 2015). Here, the advice given to policy makers based on forecasts would additionally include information on the relative costs of under-predicting and over-predicting migration when making specific policy or other decisions (Bijak 2010). This is yet another area for further exploration.

Finally, it is clear from the data audit and assessment that there are inconsistencies and uncertainty inherent in each of the data sources. There is therefore a need to extend the research agenda to include the harmonisation of the publicly available data to a common "true flow" denominator before forecasting. Ideally, how each source of data distorts the value to be estimated - the future true migration flow - needs to be taken into account in the forecasts. However, as previously mentioned in Section 2, this is a substantial task and there are only a few examples of this (see Raymer et al. 2013, Wiśniowski 2013 and Disney 2014). Consequently, this is beyond the scope of this study. The need to harmonise data to a common "true flow", before forecasting, is thus an important part of future research recommendations.

References

Abel, GJ (2010) Estimation of international migration flow tables in Europe. *Journal of the Royal Statistical Society: Series A*, 173(4): 797–825.

Abel G, Bijak J and Raymer J (2010) A comparison of official population projections with Bayesian time series forecasts for England and Wales. *Population Trends* 141: 95-114

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

Alho J and Spencer BD (1985) Uncertain population forecasting. *Journal of the American Statistical Association* 80 (390): 306-314

Alho J and Spencer BD (2005) *Statistical Demography and Forecasting* Berlin-Heidelberg: Springer

Alvarez-Plata P, Brucker H and Siliverstovs B (2003) *Potential migration from central and eastern Europe into the EU-15 – an update* Report for the EC DG of employment and social affairs. Deutsches Institut fur Wirtschaftsforschung, Berlin

Arango J (2000) Explaining Migration: A Critical View. *International Social Science Journal*, 52(165): 283–296.

Azose J and Raftery A (2013) *Bayesian Probabilistic Projection of International Migration Rates.* Paper for the 2014 Meeting of the Population Association of America, Boston, MA, 1-3 May

Bijak J (2010) *Forecasting International Migration in Europe: A Bayesian View*. Springer Series on Demographic Methods and Population Analysis, 24. Dordrecht: Springer.

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

Bijak J (2012) Migration Assumptions in the UK National Population Projections: Methodology Review. Report for the Office for National Statistics. Southampton: S3RI.

Bijak J, Alberts I, Alho J, Bryant J, Buettner T, Falkingham J, Forster JJ, Gerland P, King T, Onorante L, Keilman N, O'Hagan A, Owens D, Raftery A, Ševčíková H and Smith PWF (2015) Uncertain Population Forecasting: A Case for Practical Uses. Letter to the Editor. Forthcoming in *Journal of Official Statistics*.

Billari FC, Graziani R, Melilli E (2012) Stochastic population forecasts based on conditional expert opinions. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 175 (2): 491–511.

Billari FC, Graziani R, Melilli E (2014) Stochastic Population Forecasting Based on Combinations of Expert Evaluations Within the Bayesian Paradigm. *Demography* 51 (5): 1933-1954.

Bryant JR and Graham P (2013) Bayesian demographic accounts: subnational population estimate using multiple data sources. *Bayesian Analysis* 8(3): 591-622

Cohen J, Roig M, Reuman D and GoGwilt C (2008) International migration beyond gravity: a statistical model for use in population projections. *PNAS* 105 (40): 15269-15274

Cohen J (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

de Beer J and Alders M (1999) Probabilistic population and household forecasts for the Netherlands. Paper for the European Population Conference, 30 August–3 September 1999, The Hague. Voorburg: Centraal Bureau voor de Statistiek

Disney G (2014) *Model-Based Estimates of UK Immigration* PhD Thesis, University of Southampton

Dustmann C, Casanova M, Fertig M, Preston I and Schmidt CM (2003) *The impact of EU enlargement on migration flows.* Home Office Report 25/03. London: Home Office

Fertig M and Schmidt CM (2000) *Aggregate-level migration studies as a tool for forecasting future migration streams* IZA Discussion Paper 183, Institute zur Zukunft der Arbeit, Bonn

Greene WH (2000) *Econometric Analysis* Upper Saddle River, NJ: Prentice Hall

Howe N and Jackson R (2005) Projecting immigration: A Survey of the Current State of Practice and Theory. Washington DC: Centre for Strategic & International Studies. Via: http://csis.org/files/media/csis/pubs/0504_howe_jacksonprojimmigration.pdf

Keilman N (1990) *Uncertainty in National Population Forecasting: Issues, Backgrounds, Analyses, Recommendations.* Amsterdam: Swets and Zeitlinger.

Keilman N (2001) Demography: uncertain population forecasts. *Nature* 412 (6846): 490-491

Keilman N (2007) UK national population projections in perspective: How successful compared to those in other European countries? *Population Trends* 129: 20–30.

Keilman N (2008) European Demographic Forecasts Have Not Become More Accurate Over the Past 25 Years. *Population and Development Review* 34(1): 137–153.

Keilman N and Cruijsen H (eds.) (1992) *National population forecasting in industrialised countries*. NIDI & CBGS publication 24. Amsterdam: Swets & Zeitlinger.

Keilman N, Pham DQ, and Hetland A (2001) *Norway's Uncertain Demographic Future*. Oslo: Statistics Norway.

Kupiszewski M (2002) How trustworthy are forecasts of international migration between Poland and the European Union? *Journal of Ethnic and Migration Studies,* 28(4): 627-645

Kupiszewski M (ed.) (2013) *International Migration and the Future of Populations and Labour in Europe*. Springer Series on Demographic Methods and Population Analysis, 32. Dordrecht: Springer.

Kupiszewska D and Nowok B (2005) Comparability of statistics on international migration flows in the European Union *CEFMR Working Paper 7/2005 Central European Forum for Migration Research, Warsaw*

Lutz W and Goldstein JR (2004) Introduction: how to deal with uncertainty in population forecasting? *International Statistical Review,* 72 (1): 1-4

Lutz W, Sanderson WC and Scherbov S (eds) (2004) *The end of world population growth in the 21st century: new challenges for human capital formation and sustainable development.* London: Earthscan.

Lutz W, Butz WP, and KC S (eds) (2014) *World Population & Human Capital in the Twenty-First Century*. Oxford: OUP.

Morris DE, Oakley JE, Crowe JA. 2014. A web-based tool for eliciting probability distributions from experts. Environmental Modelling & Software 52: 1–4.

O'Hagan, A., C. E. Buck, A. Daneshkhah, J. R. Eiser, P. H. Garthwaite, D. J. Jenkinson, J. E. Oakley, and T. Rakow (2006). *Uncertain Judgements: Eliciting Experts' Probabilities*. John Wiley, New York.

Office for National Statistics (2014) Chapter 5: Migration, 2012-based NPP Reference Volume. *National Population Projections, 2012-Based Reference Volume* ONS: Titchfield

NRC [National Research Council] (2000) *Beyond Six Billion. Forecasting the World's Population.* National Academies Press, Washington DC

Raymer J, Wiśniowski A, Forster JJ, Smith PWF, and Bijak J (2013) Integrated Modelling of European Migration *Journal of the American Statistical Association* 108: 801 - 819

Rogers A (1990) Requiem for the Net Migrant. *Geographical Analysis*, 22(4): 283–300.

Shaw C (2007) Fifty years of United Kingdom national population projections: How accurate have they been? *Population Trends*, 128: 8–23.

Statistics New Zealand (2014) National Population Projections: 2014(base)–2068. Christchurch: Statistics New Zealand. Via: http://www.stats.govt.nz/~/media/Statistics/Browse%20for%20stats/NationalPopulationProjections/HOTP2014/NationalPopulationProjections2014HOTP.pdf

van Duin C and Stoeldraijer L (2014) Bevolkingsprognose 2014–2060: groei door migratie. *Bevolkingstrends 2014*. Voorburg: Centraal Bureau voor de Statistiek

Willekens F (1994) Monitoring international migration flows in Europe. Towards a statistical data base combining data from different sources. *European Journal of Population* 10 (1): 1-42

Wilson T and Rees P (2005) Recent Developments in Population Projection Methodology: A Review. *Population, Space and Place* 11(5): 337–360.

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: 583 - 607

Wiśniowski A (2013) Bayesian Modelling of International Migration with Labour Force Survey Data. PhD Thesis. Collegium of Economic Analyses, Warsaw School of Economics, Warsaw.

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

Wiśniowski A, Smith PWF, Bijak J, Raymer J, Forster JJ (2015) Bayesian population forecasting: extending the Lee-Carter method. *Demography*. DOI: 10.1007/s13524-015-0389-y

Zipf GK (1946) The P1P2/D Hypothesis: on the intercity movement of persons *American Social Review* 11: 677-686

■

Glossary of the Key Terms

Autoregression: A time series model, where the present and future values of the forecasted variable, such as migration, depend on the past (lagged) values.

Bayesian methods: Methods of statistical inference based on the Bayes Theorem (1763), whereby the initial (prior) knowledge about the model parameters $\boldsymbol{\theta}$, described by a prior distribution $Pr(\boldsymbol{\theta})$, is updated in the light of data $\mathbf{x}$, to yield a posterior distribution $Pr(\boldsymbol{\theta}|\mathbf{x})$. In Bayesian methods all quantities are treated as random, including model parameters.

Calibration: The level to which the assessment of forecast uncertainty is reliable; or the ex-ante and ex-post errors are aligned.

Covariates: Additional explanatory variables included in the model, for example related to macroeconomic indicators, geographic features, etc.

Coverage: The level to which a particular predictive interval is designed to cover (**nominal coverage**), or actually covers (**empirical coverage**) the observed values of the variable being predicted (e.g. migration).

Ex-ante error: The errors that are expected to be made during the forecasting process; usually assumed in the specification of the properties of the time series model used for forecasting.

Ex-post errors: The errors that are actually made during the forecasting process, which can be calculated, once the actual observations become available, as a difference between the point forecasts and the observations.

Frequentist methods: Methods of statistical inference based exclusively on information from the data sample $\mathbf{x}$, where model parameters are treated as constant, yet unknown quantities

Likelihood: Probability that a given sample of data $\mathbf{x}$ has been generated by a model with a particular set of parameters $\boldsymbol{\theta}$, $Pr(\mathbf{x}|\boldsymbol{\theta})$.

Point forecast: In the Bayesian framework, typically a mean or a median ('middle value') of the predictive distribution. In the frequentist framework, a single value produced by the forecasting model, such as a time series, for a specific year beyond the data series.

Posterior distribution: A probability distribution of the model parameters $\boldsymbol{\theta}$ given data $\mathbf{x}$, $Pr(\boldsymbol{\theta}|\mathbf{x})$, obtained by combining the information from the prior distributions and the likelihood.

Predictive distribution: In the Bayesian framework, a probability distribution of the future values of the forecasted variable, such as migration, based on the predictive model, such as the time series.

Prior distribution: A probability distribution of the model parameters $\boldsymbol{\theta}$ without taking data into account, $Pr(\boldsymbol{\theta})$. Prior distributions can be *informative*, for example by including some expert judgement about the parameters in question, or *vague* / *hardly informative*.

Predictive interval: In the Bayesian framework, an interval within which the predicted variable is expected to lie with a given probability (e.g. 0.75 or 0.99).

Probability distribution: A description of uncertainty of a given random quantity expressed in the language of probabilities (measures of uncertainty bounded between 0 and 1). Probability distributions can be discrete, assuming values from a countable set, or continuous, taking any real values from a given set. Examples of continuous distributions include the Normal for all real numbers, Gamma for positive numbers, or uniform for a numbers from a given interval.

Rate (in demographic sense): The number of demographic events under study (e.g. migrations, births, deaths), divided by the size of the population at risk of experiencing such events.

Stationarity: A property of the time series models, where all the observed values are generated from the same probability distribution (or, in a weaker sense, have the same mean and the same variance). Stationary series are more predictable than non-stationary ones, as in the latter case, the predictive distributions for different time periods will vary.

Time series model: Statistical model to describe and predict a process that changes in time.

Trend: A common tendency in the time series of data, which can be described by using a mathematical function (e.g. linear or non-linear).

'True flow': A hypothetical construct related to the true, yet unknown, level of migration for a given definition, depending on a purpose; for example, the 12-month length of stay stipulated by the UN in its definition of long-term migration; or the 3-month short-term migration, etc.

Uncertainty: Everything that is not known about a process, phenomenon, or a variable. The way of describing and measure uncertainty is by using probability distributions.

Vital events: key demographic events, such as births, marriages, deaths, civil partnerships and divorces (Office for National Statistics); which may or may not include migration, depending on the convention. Vital rates refer to the demographic rates of vital events.

Appendix A. Backward-look exercise, series truncated in 2008

Model	MPE (Mean Percentage Error)	MAE (Mean Absolute Error)	RMSE (Root Mean Square Error)	Coverage: 50-percent interval	Coverage: 80-percent interval	Quality Class*
IPS: Long-term migration estimates (LTIM) – total inflows						
(i) Extrapolations						
Random walk	–6%	32,200	42,330	80%	100%	**A<<**
General AR(1)	–6%	32,190	42,310	80%	100%	**A<<**
ARMA(1,1)	1%	19,200	24,870	100%	100%	**A<<**
AR(1), *differences*	–6%	32,390	42,530	100%	100%	**A<<**
AR(1), *de-trended*	–9%	47,362	62,975	60%	100%	**A<**
(ii) Bayesian models						
Random walk	–14%	68,860	87,671	60%	100%	**A<**
General AR(1)	–7%	38,500	51,221	100%	100%	**A<<**
ARMA(1,1)	11%	59,020	71,471	60%	100%	**A<**
AR(1), *differences*	–15%	72,820	91,732	60%	100%	**B<**
AR(1), *de-trended*	–4%	42,280	51,501	80%	100%	**A<**
(iii) Econometric models with covariates						
ADL(1), *projected covariates*	–3%	55,420	61,724	40%	100%	**A~**
ADL(1), *perfect foresight*	–2%	49,480	53,594	100%	100%	**A<<**
(iv) Past errors propagation (net migration only)						
RW errors, const.	10%	26,880	34,180	80%	100%	**A<**
RW errors, RW	20%	43,260	54,340	60%	80%	**B=**
Linear errors, RW	29%	61,000	70,020	60%	80%	**B=**
IPS: Long-term migration estimates (LTIM) – total outflows						
(i) Extrapolations						
Random walk	–30%	94,600	96,160	0%	60%	**C>**
General AR(1)	–10%	31,870	35,040	60%	100%	**A<**
ARMA(1,1)	–12%	37,350	38,910	40%	100%	**A~**
AR(1), *differences*	–25%	78,210	80,270	0%	80%	**B>**
AR(1), *de-trended*	–12%	37,071	39,027	20%	100%	**B~**
(ii) Bayesian models						
Random walk	–37%	115,980	118,877	0%	40%	**C>**
General AR(1)	–5%	17,800	22,034	100%	100%	**A<<**
ARMA(1,1)	9%	27,220	29,617	60%	100%	**A<**
AR(1), *differences*	–34%	106,320	109,883	0%	40%	**C>**
AR(1), *de-trended*	–28%	87640	89986	0%	40%	**B>**

Table A1: Results for the backward-look exercise based on series truncated in 2008 (five years): IPS totals

* The quality classes are as follows:

A small errors (MPE < 15%) **B** medium errors (MPE 15–30%) **C** large errors (MPE 30–50%) **D** very large errors (MPE ≥ 50%)
= well-calibrated model ~ slightly miscalibrated model **>>** extremely tight intervals **<<** extremely wide intervals
> too optimistic model (too tight intervals) **<** too conservative model (too wide intervals)

Model	MPE (Mean Percentage Error)	MAE (Mean Absolute Error)	RMSE (Root Mean Square Error)	Coverage: 50-percent interval	Coverage: 80-percent interval	Quality Class*
IPS: Long-term migration estimates (LTIM) – inflows of UK citizens						
(i) Extrapolations						
Random walk	3%	7,400	9,070	80%	100%	**A<**
General AR(1)	–5%	10,290	10,630	20%	100%	**A~**
ARMA(1,1)	–2%	9,900	10,180	40%	100%	**A<**
AR(1), *differences*	6%	6,980	10,490	60%	80%	**A=**
AR(1), *de-trended*	–12%	12,780	14,614	20%	100%	**B~**
(ii) Bayesian models						
Random walk	3%	7,296	9,079	80%	100%	**A<**
General AR(1)	–7%	10,416	10,976	20%	100%	**A~**
ARMA(1,1)	–7%	10,366	10,988	20%	100%	**A~**
AR(1), *differences*	6%	6,714	10,319	60%	80%	**A=**
AR(1), *de-trended*	–6%	10,698	11,113	80%	100%	**A<**
IPS: Long-term migration estimates (LTIM) – outflows of UK citizens						
(i) Extrapolations						
Random walk	–27%	35,400	36,010	20%	60%	**B>**
General AR(1)	–14%	18,380	20,400	40%	80%	**A=**
ARMA(1,1)	–14%	17,460	19,800	40%	80%	**A=**
AR(1), *differences*	–27%	34,930	35,550	20%	60%	**B>**
AR(1), *de-trended*	–21%	27,220	28,193	20%	40%	**B>**
(ii) Bayesian models						
Random walk	–27%	35,160	35,770	40%	60%	**B>**
General AR(1)	–11%	14,040	16,584	40%	100%	**A~**
ARMA(1,1)	–12%	15,060	17,557	40%	80%	**A=**
AR(1), *differences*	–29%	38,000	38,570	20%	60%	**B>**
AR(1), *de-trended*	–27%	35,320	35,927	0%	40%	**B>**

Table A2: Results for the backward-look exercise based on series truncated in 2008 (five years): UK citizens

* The quality classes are as follows:

A small errors (MPE < 15%) **B** medium errors (MPE 15–30%) **C** large errors (MPE 30–50%) **D** very large errors (MPE ≥ 50%)
= well-calibrated model ~ slightly miscalibrated model **>>** extremely tight intervals **<<** extremely wide intervals
> too optimistic model (too tight intervals) **<** too conservative model (too wide intervals)

Model	MPE (Mean Percentage Error)	MAE (Mean Absolute Error)	RMSE (Root Mean Square Error)	Coverage: 50-percent interval	Coverage: 80-percent interval	Quality Class*
IPS: Long-term migration estimates (LTIM) – inflows of other EU citizens						
(i) Extrapolations						
Random walk	–12%	21,800	23,210	80%	100%	**A<**
General AR(1)	–1%	15,270	19,800	100%	100%	**A<<**
ARMA(1,1)	–2%	15,420	19,600	100%	100%	**A<<**
AR(1), *differences*	–12%	21,660	23,020	80%	100%	**A<**
AR(1), *de-trended*	–5%	11,147	12,685	100%	100%	**A<<**
(ii) Bayesian models						
Random walk	–39%	61,040	63,156	20%	100%	**C~**
General AR(1)	–23%	36,080	37,579	80%	100%	**B<**
ARMA(1,1)	52%	83,102	83,841	0%	40%	**D>**
AR(1), *differences*	–49%	78,120	80,119	20%	100%	**C~**
AR(1), *de-trended*	–18%	27,560	29,353	80%	100%	**B<**
(iii) Econometric models with covariates						
ADL(1), *projected covariates*	3%	4,214	5,443	100%	100%	**A<<**
ADL(1), *perfect foresight*	4%	5,378	7,179	100%	100%	**A<<**
IPS: Long-term migration estimates (LTIM) – outflows of other EU citizens						
(i) Extrapolations						
Random walk	–143%	18,600	18,680	0%	0%	**D<<**
General AR(1)	–59%	7,950	8,240	20%	60%	**D>**
ARMA(1,1)	–57%	7,420	7,500	0%	60%	**D>**
AR(1), *differences*	–97%	12,460	12,680	0%	0%	**D<<**
AR(1), *de-trended*	–45%	5,658	6,081	20%	60%	**C>**
(ii) Bayesian models						
Random walk	–198%	25,412	25,911	0%	0%	**D>>**
General AR(1)	–31%	4,272	4,827	80%	100%	**C<**
ARMA(1,1)	26%	3,664	4,043	40%	100%	**B~**
AR(1), *differences*	–159%	20,480	20,866	0%	0%	**D>>**
AR(1), *de-trended*	–54%	6,860	7,162	80%	100%	**D<**

Table A3: Results for the backward-look exercise based on series truncated in 2008 (five years): Other EU citizens

* The quality classes are as follows:

A small errors (MPE < 15%) **B** medium errors (MPE 15–30%) **C** large errors (MPE 30–50%) **D** very large errors (MPE ≥ 50%)
= well-calibrated model **~** slightly miscalibrated model **>>** extremely tight intervals **<<** extremely wide intervals
> too optimistic model (too tight intervals) **<** too conservative model (too wide intervals)

Model	MPE (Mean Percentage Error)	MAE (Mean Absolute Error)	RMSE (Root Mean Square Error)	Coverage: 50-percent interval	Coverage: 80-percent interval	Quality Class*
IPS: Long-term migration estimates (LTIM) – inflows of non-EU citizens						
(i) Extrapolations						
Random walk	–6%	29,000	34,920	80%	100%	**A<**
General AR(1)	–6%	29,000	34,920	60%	100%	**A<**
ARMA(1,1)	1%	25,340	27,750	80%	100%	**A<**
AR(1), *differences*	–5%	29,180	34,550	80%	100%	**A<**
AR(1), *de-trended*	-18%	43,160	61,064	60%	60%	**B~**
(ii) Bayesian models						
Random walk	–18%	43,300	61,070	60%	80%	**B=**
General AR(1)	–12%	33,660	48,119	60%	100%	**A<**
ARMA(1,1)	–7%	34,160	40,658	80%	100%	**A<**
AR(1), *differences*	–23%	53,900	72,035	60%	100%	**B<**
AR(1), *de-trended*	-28%	65,440	87,175	60%	60%	**B~**
IPS: Long-term migration estimates (LTIM) – outflows of non-EU citizens						
(i) Extrapolations						
Random walk	–37%	3,200	3,290	20%	60%	**C>**
General AR(1)	–25%	2,130	2,300	40%	60%	**B>**
ARMA(1,1)	–30%	2,580	2,690	0%	40%	**C>**
AR(1), *differences*	–30%	2,560	2,680	20%	40%	**C>**
AR(1), *de-trended*	–57%	4,944	5,059	0%	0%	**D<<**
(ii) Bayesian models						
Random walk	–46%	3,982	4,070	0%	40%	**C>**
General AR(1)	–27%	2,262	2,404	40%	60%	**B>**
ARMA(1,1)	–45%	3,882	4,022	0%	40%	**C>**
AR(1), *differences*	–40%	3,428	3,541	0%	40%	**C>**
AR(1), *de-trended*	–124%	10,804	10,909	0%	0%	**D>>**

Table A4: Results for the backward-look exercise based on series truncated in 2008 (five years): non-EU citizens

* The quality classes are as follows:

A small errors (MPE < 15%) **B** medium errors (MPE 15–30%) **C** large errors (MPE 30–50%) **D** very large errors (MPE ≥ 50%)
= well-calibrated model ~ slightly miscalibrated model **>>** extremely tight intervals **<<** extremely wide intervals
> too optimistic model (too tight intervals) **<** too conservative model (too wide intervals)

Model	MPE (Mean Percentage Error)	MAE (Mean Absolute Error)	RMSE (Root Mean Square Error)	Coverage: 50-percent interval	Coverage: 80-percent interval	Quality Class*
NINo (DWP data)						
(ii) Bayesian models						
Random walk	–101%	640,394	677,086	0%	100%	**D~**
General AR(1)	2%	69,087	84,417	100%	100%	**A<<**
ARMA(1,1)	–1%	64587	80906	100%	100%	**A<<**
HESA student data (non-UK citizens only)						
(ii) Bayesian models						
Random walk	–1%	16,851	18,638	60%	100%	**A<**
General AR(1)	24%	55,429	56,009	0%	0%	**B>>**
ARMA(1,1)	26%	60,369	60,853	0%	0%	**B>>**
Asylum seekers (non-EU citizens only)						
(i) Extrapolations						
Random walk	–19%	3,831	4,590	100%	100%	**B<<**
General AR(1)	–5%	2,670	3,326	100%	100%	**A<<**
ARMA(1,1)	–6%	3,016	3,630	100%	100%	**A<<**
AR(1), *differences*	–25%	5,094	5,700	100%	100%	**B<<**
AR(1), *de-trended*	–149%	33,046	36,601	17%	50%	**D>**
(ii) Bayesian models						
Random walk	–65%	14,054	14,920	66%	100%	**D<**
General AR(1)	–39%	8,314	8,703	66%	100%	**C<**
ARMA(1,1)	–39%	8,212	8,597	83%	100%	**C<**
AR(1), *differences*	–55%	11,842	12,448	83%	100%	**D<**
AR(1), *de-trended*	4%	3,597	4,456	100%	100%	**A<<**

Table A5: Results for the backward-look exercise based on series truncated in 2008 (five years): administrative data

* The quality classes are as follows:

A small errors (MPE < 15%) **B** medium errors (MPE 15–30%) **C** large errors (MPE 30–50%) **D** very large errors (MPE ≥ 50%)
= well-calibrated model ~ slightly miscalibrated model **>>** extremely tight intervals **<<** extremely wide intervals
> too optimistic model (too tight intervals) **<** too conservative model (too wide intervals)

Appendix B. Backward-look exercise, series truncated in 2003

Model	**MPE (Mean Percentage Error)**	**MAE (Mean Absolute Error)**	**RMSE (Root Mean Square Error)**	**Coverage: 50-percent interval**	**Coverage: 80-percent interval**	**Quality Class***
IPS: Long-term migration estimates (LTIM) – total inflows						
(i) Extrapolations						
Random walk	17%	89,700	93,410	30%	90%	**B~**
General AR(1)	28%	145,290	148,600	0%	10%	**B>**
ARMA(1,1)	30%	156,400	160,500	0%	10%	**C>**
AR(1), *differences*	17%	89,230	92,960	30%	90%	**B~**
AR(1), *de-trended*	18%	93,580	102,465	20%	30%	**B>**
(ii) Bayesian models						
Random walk	3%	51,090	56,914	90%	100%	**A<<**
General AR(1)	25%	129,900	132,839	20%	90%	**B~**
ARMA(1,1)	27%	141,660	144,839	20%	90%	**B~**
AR(1), *differences*	2%	51,820	58,775	90%	90%	**A<<**
AR(1), *de-trended*	4%	59020	64836	70%	90%	**A<**
(iii) Econometric models with covariates						
ADL(1), *perfect foresight*	12%	62,600	72,548	70%	100%	**A<**
(iv) Past errors propagation (net migration only)						
RW errors, const.	48%	95,780	99,190	20%	80%	**C>**
RW errors, RW	12%	38,580	45,530	70%	90%	**A<**
Linear errors, RW	29%	61,000	70,020	60%	100%	**B<**
IPS: Long-term migration estimates (LTIM) – total outflows						
(i) Extrapolations						
Random walk	4%	25,200	37,100	80%	100%	**A<**
General AR(1)	26%	87,610	95,020	0%	10%	**B<<**
ARMA(1,1)	27%	92,900	100,180	0%	10%	**B<<**
AR(1), *differences*	4%	25,350	37,450	80%	100%	**A<**
AR(1), *de-trended*	17%	58,442	68,967	20%	50%	**B>**
(ii) Bayesian models						
Random walk	–1%	28,420	37,473	100%	100%	**A<<**
General AR(1)	28%	95,480	102,140	0%	10%	**B<<**
ARMA(1,1)	28%	94,990	101,686	0%	10%	**B>>**
AR(1), *differences*	–4%	33,170	41,398	100%	100%	**A<<**
AR(1), *de-trended*	7%	34,800	46,496	80%	100%	**A<**

Table B1: Results for the backward-look exercise based on series truncated in 2003 (ten years): IPS totals

* The quality classes are as follows:

A small errors (MPE < 15%) **B** medium errors (MPE 15–30%) **C** large errors (MPE 30–50%) **D** very large errors (MPE ≥ 50%)
= well-calibrated model ~ slightly miscalibrated model **>>** extremely tight intervals **<<** extremely wide intervals
> too optimistic model (too tight intervals) **<** too conservative model (too wide intervals)

Model	MPE (Mean Percentage Error)	MAE (Mean Absolute Error)	RMSE (Root Mean Square Error)	Coverage: 50-percent interval	Coverage: 80-percent interval	Quality Class*
IPS: Long-term migration estimates (LTIM) – inflows of UK citizens						
(i) Extrapolations						
Random walk	–20%	15,800	17,670	70%	100%	**B<**
General AR(1)	–15%	12,170	13,860	30%	80%	**B~**
ARMA(1,1)	–15%	12,430	14,140	30%	80%	**B>**
AR(1), *differences*	–19%	15,160	17,100	60%	90%	**B<**
AR(1), *de-trended*	–41%	33,058	34,443	0%	20%	**C>**
(ii) Bayesian models						
Random walk	–24%	18,796	20,517	70%	100%	**B<**
General AR(1)	–16%	12,744	14,507	30%	90%	**B~**
ARMA(1,1)	–16%	12,864	14,611	30%	90%	**B~**
AR(1), *differences*	–23%	17,966	19,744	70%	100%	**B<**
AR(1), *de-trended*	–38%	30,590	32,198	0%	30%	**C>**
IPS: Long-term migration estimates (LTIM) – outflows of UK citizens						
(i) Extrapolations						
Random walk	–14%	25,600	30,010	80%	100%	**A<**
General AR(1)	7%	21,180	26,350	50%	60%	**A>**
ARMA(1,1)	8%	22,320	28,090	50%	60%	**A>**
AR(1), *differences*	–14%	25,590	29,990	80%	100%	**A<**
AR(1), *de-trended*	10%	22,283	28,824	50%	60%	**A>**
(ii) Bayesian models						
Random walk	–14%	25,850	30,320	80%	100%	**B<**
General AR(1)	9%	22,520	28,946	50%	60%	**A>**
ARMA(1,1)	9%	22,410	28,715	50%	60%	**A>**
AR(1), *differences*	–19%	31,560	37,443	70%	100%	**B<**
AR(1), *de-trended*	2%	21,530	24,677	50%	90%	**A=**

Table B2: Results for the backward-look exercise based on series truncated in 2003 (ten years): UK citizens

* The quality classes are as follows:

A small errors (MPE < 15%) **B** medium errors (MPE 15–30%) **C** large errors (MPE 30–50%) **D** very large errors (MPE ≥ 50%)
= well-calibrated model **~** slightly miscalibrated model **>>** extremely tight intervals **<<** extremely wide intervals
> too optimistic model (too tight intervals) **<** too conservative model (too wide intervals)

Model	MPE (Mean Percentage Error)	MAE (Mean Absolute Error)	RMSE (Root Mean Square Error)	Coverage: 50-percent interval	Coverage: 80-percent interval	Quality Class*
IPS: Long-term migration estimates (LTIM) – inflows of other EU citizens						
(i) Extrapolations						
Random walk	58%	89,700	92,890	0%	0%	**D<<**
General AR(1)	67%	103,010	106,840	0%	0%	**D<<**
ARMA(1,1)	66%	101,440	105,140	0%	0%	**D<<**
AR(1), *differences*	59%	90,600	93,750	0%	0%	**D<<**
AR(1), *de-trended*	40%	60,954	63,494	0%	10%	**C<<**
(ii) Bayesian models						
Random walk	47%	72,538	74,993	10%	50%	**C>**
General AR(1)	66%	101,135	104,772	0%	0%	**D<<**
ARMA(1,1)	66%	101,203	104,820	0%	0%	**D>>**
AR(1), *differences*	48%	73,518	75,967	0%	50%	**C>**
AR(1), *de-trended*	37%	56,251	58,970	0%	40%	**C>**
(iii) Econometric models with covariates (EU 15)						
ADL(1), *perfect foresight*	2%	14,928	16,291	80%	100%	**A<**
IPS: Long-term migration estimates (LTIM) – outflows of other EU citizens						
(i) Extrapolations						
Random walk	–22%	4,700	6,120	70%	90%	**B<**
General AR(1)	27%	6,100	8,320	20%	70%	**B>**
ARMA(1,1)	11%	3,910	6,610	70%	80%	**A<**
AR(1), *differences*	–14%	3,860	5,890	80%	80%	**A<**
AR(1), *de-trended*	–13%	4,273	6,370	50%	60%	**A>**
(ii) Bayesian models						
Random walk	–65%	9,571	10,404	60%	90%	**D=**
General AR(1)	31%	6,408	8,613	20%	90%	**C~**
ARMA(1,1)	31%	6,413	8,620	20%	90%	**C~**
AR(1), *differences*	–50%	7,969	8,780	50%	90%	**D=**
AR(1), *de-trended*	–17%	4,495	6,460	50%	60%	**B>**

Table B3: Results for the backward-look exercise based on series truncated in 2003 (ten years): Other EU citizens

* The quality classes are as follows:

A small errors (MPE < 15%) **B** medium errors (MPE 15–30%) **C** large errors (MPE 30–50%) **D** very large errors (MPE ≥ 50%)
= well-calibrated model **~** slightly miscalibrated model **>>** extremely tight intervals **<<** extremely wide intervals
> too optimistic model (too tight intervals) **<** too conservative model (too wide intervals)

Model	MPE (Mean Percentage Error)	MAE (Mean Absolute Error)	RMSE (Root Mean Square Error)	Coverage: 50-percent interval	Coverage: 80-percent interval	Quality Class*
IPS: Long-term migration estimates (LTIM) – inflows of non-EU citizens						
(i) Extrapolations						
Random walk	5%	31,900	35,690	80%	90%	**A<**
General AR(1)	14%	42,730	48,500	50%	90%	**A=**
ARMA(1,1)	19%	56,110	60,140	50%	90%	**B=**
AR(1), *differences*	–1%	24,390	32,410	90%	90%	**A<<**
AR(1), *de-trended*	7%	38,851	41,711	70%	90%	**A<**
(ii) Bayesian models						
Random walk	–19%	60,560	78,498	70%	90%	**B<**
General AR(1)	–36%	99,050	127,933	70%	100%	**C<**
ARMA(1,1)	-21%	64,840	83,370	80%	100%	**B<**
AR(1), *differences*	–26%	74,740	95,392	70%	100%	**B<**
AR(1), *de-trended*	–76%	200,410	256,297	20%	80%	**D>**
IPS: Long-term migration estimates (LTIM) – outflows of non-EU citizens						
(i) Extrapolations						
Random walk	–44%	4,300	4,870	20%	60%	**C>**
General AR(1)	–18%	2,380	2,590	30%	90%	**B=**
ARMA(1,1)	–29%	3,240	3,550	20%	50%	**B>**
AR(1), *differences*	–41%	4,110	4,630	10%	40%	**C>**
AR(1), *de–trended*	–55%	5,468	6,353	10%	20%	**D>**
(ii) Bayesian models						
Random walk	–76%	7,129	8,192	20%	40%	**D>**
General AR(1)	–16%	2,321	2,560	40%	100%	**B~**
ARMA(1,1)	–21%	2,680	2,912	40%	90%	**B=**
AR(1), *differences*	–75%	7,018	8,090	20%	30%	**D>**
AR(1), *de–trended*	–58%	5,659	6,606	10%	30%	**D>**

Table B4: Results for the backward-look exercise based on series truncated in 2003 (ten years): non-EU citizens

* The quality classes are as follows:

A small errors (MPE < 15%) **B** medium errors (MPE 15–30%) **C** large errors (MPE 30–50%) **D** very large errors (MPE ≥ 50%)
= well-calibrated model **~** slightly miscalibrated model **>>** extremely tight intervals **<<** extremely wide intervals
> too optimistic model (too tight intervals) **<** too conservative model (too wide intervals)

Model	MPE (Mean Percentage Error)	MAE (Mean Absolute Error)	RMSE (Root Mean Square Error)	Coverage: 50-percent interval	Coverage: 80-percent interval	Quality Class*
Asylum seekers (non-EU citizens only)						
(i) Extrapolations						
Random walk	–110%	25,293	25,590	45%	100%	**D**
General AR(1)	–49%	11,352	12,185	82%	100%	**C<**
ARMA(1,1)	–13%	4,116	4,867	100%	100%	**B<<**
AR(1), *differences*	–70%	15,950	16,286	100%	100%	**D<<**
AR(1), *de-trended*	–1446%	328,091	391,522	0%	0%	**D>>**
(ii) Bayesian models						
Random walk	445%	99,096	113,061	0%	40%	**D>**
General AR(1)	–98%	22,339	22,598	0%	100%	**D~**
ARMA(1,1)	–85%	19,281	19,585	40%	100%	**D~**
AR(1), *differences*	–330%	73,577	82,831	0%	100%	**D~**
AR(1), *de-trended*	–368%	82,425	90,115	0%	100%	**D~**

Table B5: Results for the backward-look exercise based on series truncated in 2003 (ten years): administrative data

* The quality classes are as follows:

A small errors (MPE < 15%) **B** medium errors (MPE 15–30%) **C** large errors (MPE 30–50%) **D** very large errors (MPE ≥ 50%)
= well-calibrated model **~** slightly miscalibrated model **>>** extremely tight intervals **<<** extremely wide intervals
> too optimistic model (too tight intervals) **<** too conservative model (too wide intervals)

Appendix C. Scoring rules for the forecast quality categories

Symmetric

Error \ Calibration	=	~	< or >	<< or >>
A	0	1	3	5
B	1	3	5	7
C	3	5	7	9
D	5	7	9	10

Table C1: Backward Look Summary Table – Symmetric Score Table, with errors and calibration similarly important

Asymmetric

Error \ Calibration	=	~	< or >	<< or >>
A	0	1	2	3
B	2	3	4	5
C	4	6	7	8
D	7	8	9	10

Table C2: Backward Look Summary Table – Asymmetric Score Table, with errors more important than calibration

In both cases, the average scores under 3 fall into the **green category** (safer to use), scores between 3 and 5 – to the **orange category** (use with caution), while the scores above 5 – the **red category** (do not use)

ESRC Centre for Population Change

The ESRC Centre for Population Change (CPC) is a joint initiative between the Universities of Southampton, St Andrews, Edinburgh, Stirling, Strathclyde, in partnership with the Office for National Statistics (ONS) and the General Register Office Scotland (GROS). The Centre is funded by the Economic and Social Research Council (ESRC) grant numbers RES-625-28-0001 and ES/K007394/1.

This report publishes independent research, not funded directly from the core grant of the Centre. The views and opinions expressed by authors do not necessarily reflect those of the CPC, ESRC, ONS or NRS.

Website | Email | Twitter | Facebook | Mendeley