

Received: 1 November 2017 | Revised: 10 September 2018 | Accepted: 31 January 2019

DOI: 10.1002/for.2576

RESEARCH ARTICLE

WILEY

Assessing time series models for forecasting international migration: Lessons from the United Kingdom

Jakub Bijak^{1,2} | George Disney^{1,3} | Allan M. Findlay^{1,4} | Jonathan J. Forster^{1,2} | Peter W.F. Smith^{1,2} | Arkadiusz Wiśniowski^{1,5}¹ ESRC Centre for Population Change, UK² University of Southampton, UK³ University of Melbourne, Australia⁴ University of St Andrews, UK⁵ University of Manchester, UK**Correspondence**

Jakub Bijak, Department of Social Statistics and Demography, University of Southampton, Southampton SO17 1BJ.
Email: j.bijak@soton.ac.uk

Funding information

Economic and Social Research Council, Grant/Award Number: ES/K007394/1; Home Office Science, Grant/Award Number: HOS/14/040

Abstract

Migration is one of the most unpredictable demographic processes. The aim of this article is to provide a blueprint for assessing various possible forecasting approaches in order to help safeguard producers and users of official migration statistics against misguided forecasts. To achieve that, we first evaluate the various existing approaches to modelling and forecasting of international migration flows. Subsequently, we present an empirical comparison of *ex post* performance of various forecasting methods, applied to international migration to and from the United Kingdom. The overarching goal is to assess the uncertainty of forecasts produced by using different forecasting methods, both in terms of their errors (biases) and calibration of uncertainty. The empirical assessment, comparing the results of various forecasting models against past migration estimates, confirms the intuition about weak predictability of migration, but also highlights varying levels of forecast errors for different migration streams. There is no single forecasting approach that would be well suited for different flows. We therefore recommend adopting a tailored approach to forecasts, and applying a risk management framework to their results, taking into account the levels of uncertainty of the individual flows, as well as the differences in their potential societal impact.

KEYWORDS

ARIMA models, Bayesian methods, decision making, forecasting, international migration, uncertainty

1 | INTRODUCTION

Forecasting migration flows is characterised by high levels of error, higher than for the other components of demographic change: fertility and mortality (Bongaarts & Bulatao, 2000), yet these errors are of crucial importance for overall population forecasts (*idem*; Long, 1991). There are many social, economic, political and

environmental drivers which impact migration flows (Massey et al., 1993), yet there is no single, robust migration theory that can be used for forecasting purposes (Arango, 2000). Migration is also susceptible to events that are difficult to predict in terms of timing and impact, such as changes in the economic cycle, policies or political circumstances. Besides, even if credible explanations of past migration flows existed, their tenets

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors Journal of Forecasting Published by John Wiley & Sons Ltd.

would be difficult to extrapolate into the future – for that reason, most of the formal forecasting models rely on time series analysis, be it frequentist or Bayesian (Azose & Raftery, 2015; Bijak, 2010; Bijak & Wiśniowski, 2010; Cappellen, Skjerpen, & Tønnessen, 2015; de Beer, 1993, 2008; Wiśniowski, Smith, Bijak, Raymer, & Forster, 2015).

Given the inherently uncertain nature of future events, and the history of shock changes to migration flows, the main aim of this paper is to provide a blueprint for assessing possible forecasting approaches to help safeguard producers and users of official migration statistics against misguided forecasts. To do so, we evaluate the various existing approaches to forecasting of migration flows, and present an empirical comparison of *ex post* performance of various forecasting methods, applied to international migration to and from the United Kingdom (UK). Even though we focus on international mobility, the findings and general recommendations also apply to internal migration, which typically exhibits more stable and regular features (an observation made since Ravenstein, 1885), yet still can have considerable volatility (see e.g. Raymer, Abel, & Rogers, 2012). Throughout this paper, ‘migration’ is thus used as shorthand for ‘international migration’. Besides, the terms ‘forecast’ and ‘prediction’ are used interchangeably; ‘projections’ being reserved mainly for the results of deterministic calculations of future population size and structure under a set of specific assumptions (Keilman, 1990).

This paper considers three sources of uncertainty in migration forecasts: the inherent uncertainty of future events, errors in the data (Section 2), and uncertainty related to relying on a particular forecasting model (Section 3). In the empirical analysis, various models are compared based on their forecast errors and the accuracy of calibration of the forecast uncertainty (Section 4). The results obtained for forecasts using data leading up to the two major ‘shocks’ observed for UK international migration patterns – the enlargement of the European Union (EU) in 2004; and the economic crisis in 2009 – are subsequently presented (Section 5) and assessed using a forecast assessment algorithm we outline. Finally, we make recommendations related to the usefulness of various forecasting approaches for policy-makers, with focus on the role of uncertainty (Section 6).

2 | UNCERTAIN MIGRATION, UNCERTAIN DATA

A vital consideration in forecasting migration is how to incorporate uncertainty into the estimates. There are three broad sources of uncertainty we consider in this

paper. 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 (Alho & Spencer, 1985).

The second source of uncertainty is associated with *migration data*. Sources of migration data from different countries are often based on differing definitions (Raymer, Wiśniowski, Forster, Smith, & Bijak, 2013). The available data are often inaccurate, inconsistent and incomplete. Migration into and out of the UK is no exception. The precise volume of international migration flows are difficult to measure; data collection systems used to record migrants often produce biased and inaccurate estimates (Disney, 2015; Kupiszewska & Nowok, 2008; Willekens, 1994; Wiśniowski, Forster, Smith, Bijak, & Raymer, 2016). Related to that is the uncertainty in how migration is operationalised – as net or gross figures, for each area separately or jointly for a multiregional system, as crude numbers or rates, the latter additionally involving uncertainty in the population at risk (see e.g. Raymer et al., 2012).

The third source of uncertainty comes from the *forecasting models*. Applications of different models to the same data can produce different forecasts, including different assessments of the uncertainty of the predictions. If the forecasts from various competing models are combined using formal criteria, additional uncertainty about the model is introduced (Bijak & Wiśniowski, 2010).

In the UK, the main source of data on long- and short-term migration is the International Passenger Survey (IPS). The IPS is supplemented by a range of administrative data: Home Office statistics on refugees and asylum seekers, new National Insurance Numbers (NINo) issued to foreign nationals by the Department of Work and Pensions, and data on foreign students from the Higher Education Statistics Authority (HESA). At the moment, there is not much quantitative data on emigration from the UK except for the IPS, especially available in the public domain, although the situation is changing, with the increasing availability of data on exit checks. These data are already collected by the Home Office and shared with the Office for National Statistics for analytical purposes, such as those related to the numbers of international students in the UK (Home Office, 2017).

Each of the sources of data can be assessed in relation to the concept of ‘true flow’ (Raymer et al., 2013; Wiśniowski et al., 2016), defined as the unknown number of migrants that is being estimated *under a given definition of a migrant*. It represents the number that one would obtain if one was able to monitor the given definition of immigration perfectly, without bias and undercount, and with complete coverage of the population. For the purpose of this paper, the concept of a true flow

follows the UN (1998) definition of long-term international migration, whereby:

“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.” (UN, 1998, p. 18)

The data quality and uncertainty in each source can be assessed in relation to this ‘true flow’ definition, by using the following analytical categories (Disney, 2015; Raymer et al., 2013):

- Definition – how closely the data align with the standard UN duration of stay criteria for long-term (12 months or more) and short-term (3–12 months) migration;
- Coverage – conceptually, what portion of the total immigration flow the data set can cover, and are there any population excluded from the measurement by design (e.g. students or irregular migrants);
- Bias – whether there is any systematic bias as a result of the way the data are collected, such as an undercount of the number migrants due to the lack of incentives for them to register;
- Accuracy – with regard to its intended purpose, how accurate the data are, for example with respect to the sampling error, or other inaccuracies not covered above.

The main source of the UK migration statistics – the IPS – is a sample survey, thus disaggregation of the data by countries of origin or destination of migrants can have high margins of error resulting from sampling of respondents. This is especially important given the small number of survey contacts declaring migration intentions – just over 3,000 in 2016 – leading to standard errors above 3% for the total immigration and emigration estimates, and much higher for individual flows by sex, age or country of citizenship (ONS, 2018). In addition, there can also be 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, which is another cause of a bias in the estimates. There is a clear trade-off between the typically more accurate administrative sources, which however only offer crude proxies for immigration flows and migrant stocks, and the conceptually clearer survey-based

measures, such as the IPS. Overall, the assessment of the different sources of data on UK immigration according to these four criteria is summarized in Table A1 in the Appendix, and details are provided in Disney, Wiśniowski, Forster, Smith, and Bijak (2015).

3 | REVIEW OF EXISTING METHODS

Existing migration forecasting methods usually involve either deterministic scenarios, or various types of probabilistic models, including time series and econometric models. Most official projections prepared by statistical offices in developed countries or international organisations remain deterministic, and are based on judgemental scenarios (Bijak, 2010; de Beer, 2008). The main criticism of these methods is that they do not allow for a coherent and explicit quantification of uncertainty in their estimation (Wilson & Rees, 2005). For scenarios, even though it is possible to calculate *ex post* errors, it is not possible to assess whether they are well calibrated in the probabilistic sense, i.e. whether the *ex ante* statements about the chances of errors of different magnitudes match the error frequencies observed *ex post*. As in such studies there is typically no *ex ante* statement of the likelihood of specific outcomes, there is nothing to calibrate the distributions of these errors against. For this reason, existing examples of these approaches are not included in the empirical assessment presented in this study.

On the other hand, probabilistic forecasts specify the chances for future migrations to occur, given a set of assumptions about the underlying probability distributions (Alho & Spencer, 1985; Bijak et al., 2015), and quantify at least some of the sources of uncertainty mentioned before. These assumptions also depend on expert judgement, but typically on the likelihood function and model parameters (error variance, autoregression parameters, and so on) rather than future values, as is the case with scenarios. The standard approach to time series extrapolation is to apply ARIMA (autoregressive integrated moving average) models, typically within the frequentist (likelihood-based) statistical paradigm (see de Beer, 1993 and 2008, for examples, and Bijak, 2010 for an overview). Another group of time series models, dating back to Stoto (1983), relies on extrapolating past errors in forecasts, rather than data per se. Examples of analysis of past forecast errors include Shaw (2007) for the United Kingdom and Keilman (2008) for a group of European countries.

The ARIMA models can have a longer or a shorter ‘memory’, depending on the parameters, and can exhibit different properties with respect to their stationarity. In demographic forecasting, the order of the ARIMA models

usually does not go beyond (1,1,1) (Keilman, 2001). ARIMA models include an estimate of forecast uncertainty via confidence intervals. To some extent, this takes into account the second source of uncertainty outlined previously – uncertainty in the data. The main theoretical criticism of frequentist ARIMA models is that the forecasts are based on data alone and that it may lead to unreasonable predictions, especially in the presence of shock events, when the underlying processes are clearly non-stationary.

In comparison to data-driven time series extrapolations, Lutz and Goldstein (2004) developed ‘expert-based probabilistic population projections’, where subjective expert judgement alone is used to prepare expert forecasts. 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. A purely expert-based approach does not make full use of data, which is a major limitation given that data on previous flows is the best source of evidence available to forecasters.

A Bayesian approach allows inclusion of expert opinion within a statistical model. Following Bijak (2010), in this approach, historical trends, expert judgements and various model specifications can all be combined in a probabilistically coherent way (Bijak & Wiśniowski, 2010). This is especially important when the data series are too short to allow for a meaningful classical inference (Bijak & Wiśniowski, 2010). Theoretically, Bayesian approaches allow the forecaster to incorporate all three elements of uncertainty in their estimates. As such, they are the main focus of our empirical testing methods later in the paper. Recently, Azose and Raftery (2015) proposed a method for probabilistic projection of global net international migration rates with Bayesian hierarchical first-order AR(1) models.

Econometric models can both predict migration and verify economic theories on the basis of empirical data and covariate information. One example is a model by Dustmann, Casanova, Fertig, Preston, and Schmidt (2003) forecasting net migration after the EU enlargement in 2004 based on scenarios in relative income per capita. Their approach assumed stationarity of the errors, and did not take into account the effect of allowing freedom of movement for the new EU citizens. Such model specification led to large *ex post* errors for migration into the UK, but yielded relatively accurate predictions for Germany – one of the countries that imposed transitory restrictions on access to its labour market. This illustrates the importance of treating migration as a non-stationary process where systemic ‘shocks’ are expected. An alternative approach for modelling such shocks relies on including dummy variables for such past events (Cappellen

et al., 2015; de Beer, 1993), but even then the key issue with unpredictability of such shocks remains – and the approach can be argued to suppress the predictive uncertainty as a result.

Finally, an example of a modelling approach which includes demographic explanatory variables is a gravity model, where population sizes act as ‘masses’ drawing people over spatial distance (Cohen, Roig, Reuman, & GoGwilt, 2008). Some aspects of this approach can be criticised, as the time-invariant predictors, for example the distance or area, can inform forecasts about the structure of the future flows, but not about their magnitude. On the other hand, forecasting magnitude alone is not sufficient for obtaining robust predictions of inter-country migration. For the purpose of this study, given that we concentrate on a single destination country, these models are structurally identical to other econometric approaches, and are therefore omitted from the empirical evaluation exercise.

4 | METHODOLOGY

4.1 | Framework for the empirical research

There is no clear agreement, both amongst practitioners in national statistics offices and in the academic literature, about which type of probabilistic (stochastic) forecasts produce the ‘best’ results (Bijak, 2010). From the main approaches proposed in the literature, a range of candidate stochastic models has been applied to the data discussed in Section 2.

The empirical analysis has been carried out whenever the combinations of data sources and forecasting methods has been deemed to be ‘readily applicable’, not requiring additional information such as the elicitation of expert opinion. For all the models, the variable being forecast is the annual number of (gross) migrants in year t as measured in a given data source, denoted as m_t , log-transformed to ensure positivity. For presentational simplicity, and also because overall migration flows are more directly relevant to policy decisions than those disaggregated by sex, age, or individual countries of origin (rather than legally-defined groups of countries with varying levels of immigration restrictions), the exercise is based on total migration flows. The following three groups of models are considered.

- 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, for the different migration processes under study, m_t . First, we have estimated a general, unconstrained autoregressive model of the first order, AR(1), given by (1) below, with a constant c , describing a stationary process whenever the autoregression parameter $\varphi \in (-1, 1)$:

$$\ln(m_t) = c + \varphi \ln(m_{t-1}) + \varepsilon_t. \quad (1)$$

Second, we have examined a special case of (1), a non-stationary random walk with drift c , with $\varphi = 1$. Third, we have estimated an ARMA(1,1) model given by (2), with a moving average element added to the AR(1) model above, with an additional parameter θ . Here, the model equation is:

$$\ln(m_t) = c + \varphi \ln(m_{t-1}) + \varepsilon_t + \theta\varepsilon_{t-1}. \quad (2)$$

Fourth, we have estimated another AR(1) model (1), but based on differences in log-transformed volumes of migration flows m_t . Finally, we have examined an AR(1) model explicitly assuming the underlying linear trend (hence, ‘de-trended’), given by (3) below, which can be re-expressed as a modified version of (1), with a new constant and an additional time-dependent term, $b([1 - \varphi]t + \varphi)$:

$$\begin{aligned} \ln(m_t) &= c + bt + \varphi (\ln(m_{t-1}) - c - b[t-1]) + \varepsilon_t \\ &= c(1 - \varphi) + b([1 - \varphi]t + \varphi) + \varphi \ln(m_{t-1}) + \varepsilon_t. \end{aligned} \quad (3)$$

In all variants of (1)–(3) the individual error terms ε_t are assumed to be independent and identically (normally) distributed.

The second group of models is essentially the same as the one listed above, with the addition of expert-based information through prior distributions (Section 5.2). The models in this group have been estimated 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. For longer series, all five models listed above have been used; while for shorter series the exercise was limited to the first three models (random walk, general AR(1) and ARMA(1,1)). Any expert judgement included in a Bayesian time series model via the priors relates to the interpretation of the model parameters. For example, if an expert believes a given flow is non-stationary, then for an autoregressive term in the model, a prior distribution that allows a given parameter to take a value of 1 or above is appropriate.

Clearly, without the expert judgement, any estimates of the structural parameters of ARIMA models based on short time series, not to mention assessments of

stationarity, are problematic and can even be difficult to identify. Besides, such short series are still the norm in quantitative migration studies. For these reasons, instead of discussing the parameter estimates of the time series models, we focus on presenting a detailed comparison of the empirical performance of the resulting forecasts.

ii. 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, in this case, changes in the unemployment rate (u_t) and gross national income (GNI, g_t), to predict current values of migration. The model, with parameters $\beta_0 - \beta_3$ related respectively to the contemporaneous and lagged unemployment rates and GNI, is specified as:

$$\begin{aligned} \ln(m_t) &= c + \varphi \ln(m_{t-1}) + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 g_t \\ &\quad + \beta_3 g_{t-1} + \varepsilon_t. \end{aligned} \quad (4)$$

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 first assumed a ‘perfect foresight’, and have used the actual values of the explanatory variables (unemployment rates) in the models. In parallel, we have also used the forecasts of these variables available from the Office for Budget Responsibility (OBR, 2015).

iii. Extrapolation of time series through propagation of historical forecast errors

The past errors for the UK net migration have been estimated by looking at the previous ONS 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 1). The above-mentioned regularities are observed along the *horizon* dimension, which then requires translating into the *period* dimension. An additional difficulty here is that the raw data have a ‘triangular’ form: for earlier forecasts, more observations across different horizons are available than for the most recent ones. To mitigate that, the analysis has been restricted to forecast errors for up to ten years ahead.

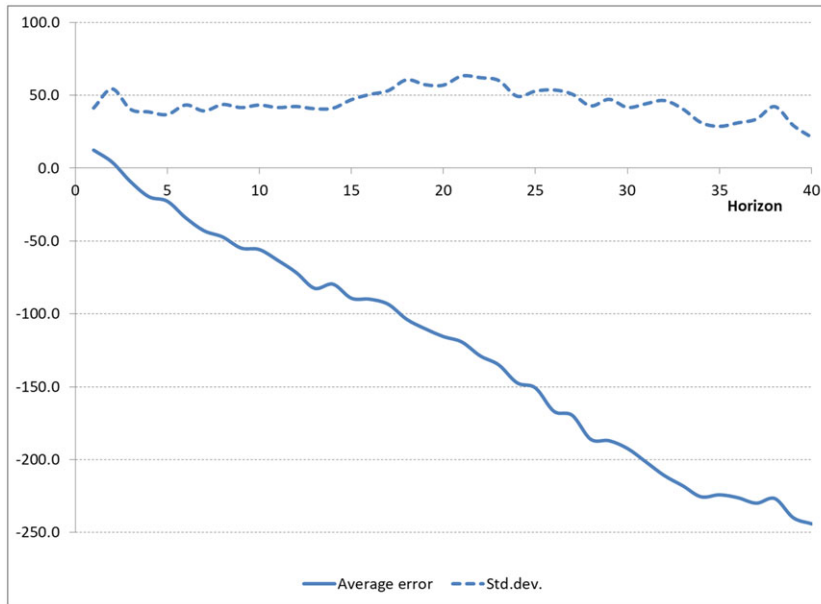


FIGURE 1 Average error and its standard deviation by projection horizon, past official net migration assumptions for National Population Projections, 1970-based to 2012-based Source: Government Actuary's department and Office for National Statistics [Colour figure can be viewed at wileyonlinelibrary.com]

Hence, in order to reflect these trends, three models for net migration (n_t) have been considered, with $e_{t,h}$ denoting empirical errors observed for a forecast made in year t with a horizon of h years, so for the year $t + h$. The models for errors have been estimated alongside the horizon dimension, but assuming common features for different time periods. The formal specification of these models is as follows:

$$e_{t,h} = \gamma + e_{t,h-1} + \xi_{t,h} \text{ and } n_{t+h} = \bar{n} - h\gamma + \sum_{x=1}^h e_{t,x} \quad (5a)$$

$$\begin{aligned} e_{t,h} &= \gamma + e_{t,h-1} + \xi_{t,h} \text{ and } n_{t+h} \\ &= n_t + h(c - \gamma) + \sum_{x=1}^h e_{t,x} \end{aligned} \quad (5b)$$

$$\begin{aligned} e_{t,h} &= b_0 + b_1 h + \xi_{t,h} \text{ and } n_{t+h} \\ &= n_t + [hc - (b_0 + b_1 h)] + \sum_{x=1}^h e_{t,x} \end{aligned} \quad (5c)$$

In the above model equations, \bar{n} is an average value of net migrations from the most recent five years of observations, γ is the drift term for the random walk model for the errors, b_0 and b_1 are the respective linear trend parameters, and $\xi_{t,h}$ is a normal noise term for the model for the errors. Finally, c is the drift term for the random walk for net migration, modelled as in the special case of (1), with $\varphi = 1$. The interpretation of these models is as follows:

- Model (5a) assumes a random walk with drift for empirical errors from the past, and a constant forecast of net migration (\bar{n}), assuming an average value from the most recent five years of observations. In this case, the error drift is not propagated into the forecasts.
- Model (5b) is also a random walk with drift for the past errors, coupled with another random walk with

drift for net migration, with error in the forecast period assumed to follow the estimated model for past errors, allowing for drift;

- Model (5c) contains a 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 on a natural (not log-transformed) scale, as net migration can assume negative, as well as positive values. It has to be noted that models (5a) – (5c) are based on net migration, which is an artificial measure conflating flows in different directions, as this is the variable which is used in the official population projections for the United Kingdom, despite being considered theoretically and conceptually inferior than gross migration (Bijak, 2012).

4.2 | Assessment of models

The models detailed above have been compared across a range of quality indicators. For errors, the following summary measures have been calculated, in the spirit of similar comparisons done in the past (e.g. Keilman & Kučera, 1991):

- Mean Percentage Error (MPE), which measures relative bias;
- Mean Absolute Error (MAE), capturing absolute magnitude of error;
- Root Mean Square Error (RMSE), which additionally penalizes large 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 paper is how to safeguard against making bad migration forecasts under uncertainty of future events, data and models. In particular using the forecasts based on truncated data to reproduce the observed reality 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 (1) on the series truncated in 2008, allowing the calculation of the *ex post* errors and empirical frequencies for five years of observations (2009–2013), and (2) for longer data series only, also with observations truncated in 2003, with ten years of data for comparison (2004–2013). The earlier 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 later one, 2008, is the starting year of the global financial crisis and subsequent recession. In most cases, we would expect the models to perform better in the periods of higher stability, with fewer and smaller structural breaks. Detailed results across the above set of five quality indicators are reported in Appendices A and B.

To assess each of the models and to summarise our empirical results, we have developed the following algorithm. First, the errors and coverage measures have been described by a range of qualitative codes related to the MPE and to calibration of both 50-percent and 80-percent intervals. MPE has been selected here over other measures of error (MAE, RMSE) as an indicator of forecast bias, which is readily comparable across different series. The following models were included in the analysis: random walks, general AR(1) and ARMA(1,1) models, AR(1) estimated on differenced series, as well as de-trended AR(1) models – all in the frequentist and Bayesian

versions. Furthermore, for long non-stationary series, ADL(1) with predicted covariate (unemployment rates) 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 for earlier years. Overall, the assessment of performance is based on the results of 198 models. They included five ARIMA models, each estimated for nine groups of flows (four inflows, four outflows, and asylum seekers), both in the frequentist and Bayesian approaches, for the series truncated in 2003 and 2008. Additionally, three past error models involving net migration and two econometric ADL models were estimated, each for both truncation points, as well as eight other specific series for 2008 only. A detailed summary of the models and series used is shown in Table A2 in the Appendix.

Once the quality classes have been obtained for all models, they have been converted into numerical scores, penalising the models with high overall MPE errors and/or those being miscalibrated. A ‘symmetric’ conversion table has been used, where the error (bias) and calibration have been deemed similarly important and penalised to a similar degree. Finally, 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.

The average scores have been ultimately given a categorical quality rating: high (green) for the relatively most appropriate methods for data series exhibiting particular features; medium (orange) where the application of such models needs to proceed with caution; and low (red) for those that are definitely not recommended, based on high likelihood of high errors and/or problems with calibration. A detailed summary of the scores given to different combinations of errors and calibration outcomes is offered in Table A3 in the Appendix.

5 | RESULTS

5.1 | A summary of empirical results

The summary of the empirical results is offered in Table 1 and selected examples are illustrated in Figure 2. Note

TABLE 1 Results summary table for forecast biases and calibration of errors

(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	Orange	Green	Orange	Orange	Red
Longer series, non-stationary features	Red	Red	Red	Red	Red
(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	Orange	Red	Red	N/A	N/A
Longer series, stationary features	Orange	Green	Green	Orange	Orange
Longer series, non-stationary features	Red	Red	Red	Red	Red
(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	Orange	Orange	Red	Orange	Orange

Key: ■ (Green) low errors and good calibration; ■ (Orange) middle-sized errors and reasonable calibration; ■ (Red) high errors and/or poor calibration.

that for short series, only Bayesian models could be realistically used, as the data need to be augmented by expert judgement. We observe that the forecasts can vary greatly in terms of median forecasts, which is reflected in the error measures (for example, the MPE), and uncertainty around the median, which results in different calibration scores. In the summary table and graphs, the assessments have been assigned ‘colour’ codes, ranging from ‘green’ for models producing relatively small biases and good calibration, to ‘red’ for forecasts that are either highly miscalibrated or are highly biased, with ‘orange’ in between. In this exercise, bias and calibration have been considered similarly important.

The supporting documentation available at Disney et al. (2015) additionally provides detailed results assuming that the errors higher weight over calibration, which are generally similar, if slightly more favourable to the ARMA and ADL models for longer series of data.

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. 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.

As an example, for non-stationary series, e.g. those for migration from the new EU member states, the more reliable models, such as the AR(1) models with additional expert input, 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 was likely just above 150,000.¹

These errors can be illustrated by forecasts of British immigration based on 1975–2003 data (Figure 2, graphs a–h). The AR(1) model produces constant forecast at around 95,000 with the uncertainty intervals remaining at a similar level for each year of the forecast (graph a). For the RW model, the width of the predictive intervals and the level of migration increase with every year of the forecast (graph b). For non-stationary series, such as Asylum Seekers (graphs c and d) and immigration from the EU countries (graphs e and f), the RW model may provide more realistic assessment of uncertainty, but still does not guarantee small forecast errors, as non-stationary series may change the direction “randomly”. Finally, the econometric models with ‘perfect foresight’ of unemployment produced reasonable errors, but with far too pessimistic uncertainty assessment (graphs g and h).

Only when the underlying series were relatively stable, such as for migration of UK nationals, were some models able to produce relatively small errors – in other situations, the applicability of various methods was either limited, or 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.

Full results of the empirical forecast exercise for individual models and data series are reported

¹See, for example, the IMEM estimates (Raymer et al., 2013), available at <http://www.imem.cpc.ac.uk>.

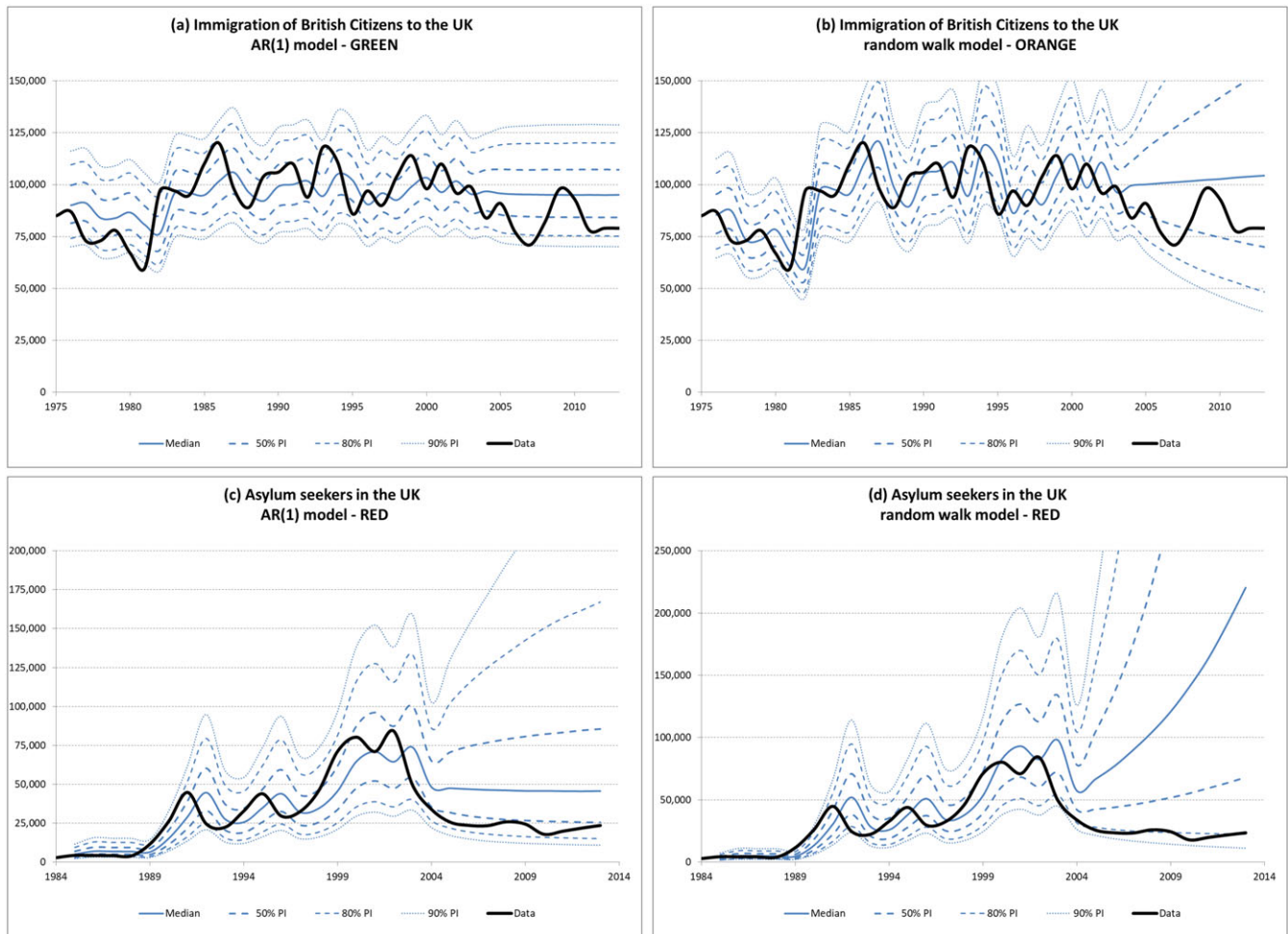


FIGURE 2 Examples of the empirical assessment for selected data series truncated in 2003 Notes: ‘AR(1)’ denotes the autoregressive model given by (1) and ‘ADL(1)’ the autoregressive distributed lag model (4). ‘PI’ refers to the bounds of the respective 50-, 80- and 90-per cent predictive intervals. Data source: Office for National Statistics for graphs (a) (b) (e) (f) (g) and (h); Home Office for (c) and (d), for various years. Predictions computed by the authors [Colour figure can be viewed at wileyonlinelibrary.com]

alongside detailed error and coverage indicators in Disney et al. (2015).

5.2 | Sensitivity analysis

One of the three main sources of uncertainty outlined in Section 2 is the uncertainty associated with the forecasting model themselves, and different models produce forecasts of varying quality for given time-series. 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.

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 dis-aggregations 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.

i. Sensitivity to Prior Assumptions

As mentioned in Section 4, 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.

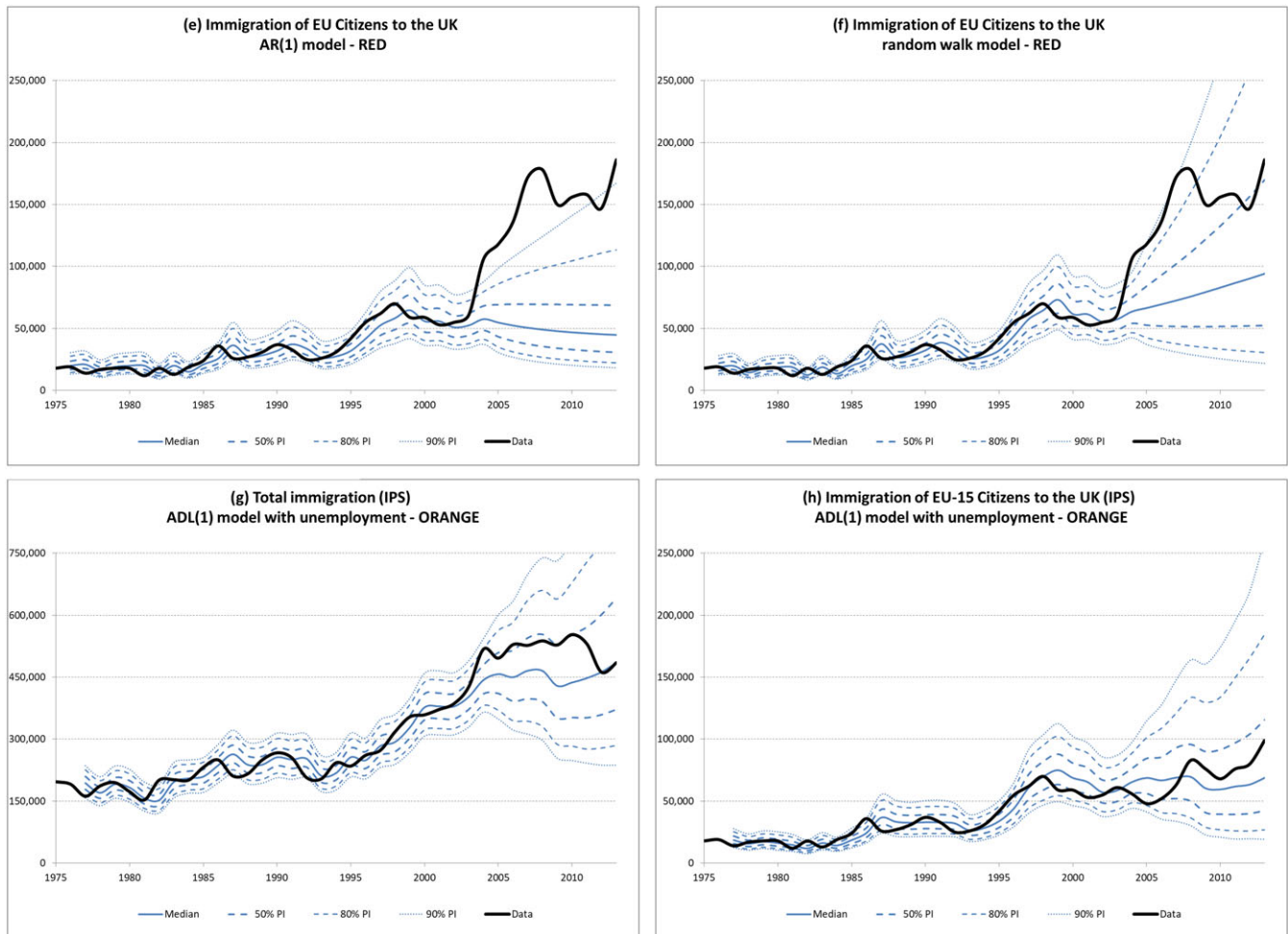


FIGURE 2 Continued.

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 of the time series. For the total migration flow measured by the IPS, the forecasts are largely insensitive to the prior distributions. For the forecast to be significantly affected by the expert knowledge, the judgement about particular statistical properties (e.g. stationarity) incorporated in the prior distribution would have to be strong and well-justified, and the corresponding prior distribution would need to be very tight, allowing minimum error. In particular, making such assumptions should not be driven by convention, analytical convenience or tractability, as the presented comparison of stationary and non-stationary models has shown. Besides, the prior assumptions on stationarity should not be generalised without proper reflection, as this property may not hold 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 relatively few data points are available, subjective opinions will play a more prominent role.

ii. Sensitivity to Expert Knowledge

The potential of expert opinion is demonstrated in relation to 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 old 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 before 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 a crude assumption on the change in the expected level of the EU flow, based on a notion of population gravity. Of course other specifications are possible, including taking advantage of the prior knowledge of which countries would open their labour markets to the citizens of the new EU member states, although this knowledge was not included in the contemporaneous forecasts of Dustmann et al. (2003). 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.

In summary, these results show that expert opinion can be of value as a supplement in migration forecasting. This appears to be especially true in relation to situation where there is a breakpoint in time series data, such as the one which occurred with the 2004 enlargement of the EU, or which shows the first signs of occurring following UK's 'Brexit' from the European Union, with a recent decline in immigration from other EU countries and a sharp increase in emigration (ONS, 2018).

iii. 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 wage dynamics and economic performance (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 (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 one hand, the results

confirm 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, 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, following de Beer (2008), we advocate applying econometric models 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 and depend on the length of the available series. This is in line with earlier suggestions (Bijak & Wiśniowski, 2010), although clearly longer-term projections, such as those produced by national statistical offices, would need to rely on different methods, such as scenario-based ones.

Scenarios might have limited predictive capacity and be strongly dependent upon expert judgement, although statistical models also rely on their underpinning assumptions. Besides, scenario-based methods can be probabilistic as well, involving expert judgement on the magnitude of errors, and in any case should include at least qualitative statements on the likelihood – for example using categories of uncertainty (low, medium, high, extreme) to describe the underlying assumptions.

For short-term forecast horizons, another promising research avenue is related to early warning models, which would seek to detect the signs of structural 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.). 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 interventions. The outcomes could be subsequently analysed by using risk management tools – combining the potential policy impacts of such interventions with their uncertainty – to help policy makers make prudent and robust decisions.

An example of such a risk management matrix for different migration flows is provided in Table 2, with

TABLE 2 A stylised example of a risk management matrix related to migration forecasts, from the point of view of possible policy impacts

Impact \ Uncertainty (risk)	Uncertainty (risk)		
	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

Key: ■ (Green) low policy concern; ■ (Orange) moderate policy concern; ■ (Red) high policy concern.

traffic-lights colour coding this time corresponding to the importance of managing the flows of different characteristics in terms of their uncertainty and impact. The key policy focus should be on the red and orange areas – those with either substantial uncertainty, or having higher impact of the volume of different types of migration on a range of social and policy areas, such as on the resources needed for the processing of applications or for integration efforts. One such ‘red’ area, with the highest impact and uncertainty, is asylum-related migration – in this context, a discussion of the potential for using early-warning models is included in the recent report for the European Asylum Support Office (EASO, 2017).

6 | CONCLUSIONS

Migration is a very complex and multi-dimensional process, responding to many different drivers. Thus, its forecasting is extremely difficult. When tested on empirical data from the past, all models under study produced considerable uncertainty, but for some the prediction errors were much larger than for others. 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.

Following the results of the empirical analysis, we recommend a general *process* for 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 either relatively well or relatively poorly, dependent on the nature of the data series. Consequently, it would not be prudent to recommend any one model as the ‘best model’ for all situations.

The recommended process has three steps:

1. A thorough understanding 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. The distinction can go beyond stability and should ideally encompass other features of migration as well. As suggested by de Beer (2008), different types of flows, including e.g. asylum, labour, family, or student migration have different characteristics, and need to be ideally studied separately. 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 specific migration flow’s potential volatility to changes in policy. In any case, the forecast uncertainty should allow for the possibility for future changes in the migration flow of interest (Pijpers, 2008), the magnitude of which would be dependent on the specific type of flow. This approach would aid policy makers in any decisions based on a migration forecast. One may expect different policy responsiveness of the different types of flows. 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.

2. Formal assessment of the available data, their strengths, weaknesses and uncertainty

This consideration relates to the second source of uncertainty outlined in Section 2 – the uncertain nature of migration data itself. One of the conclusions from

the empirical analysis is that migration forecasts based on short time series, or series that were subject to shocks in the past, 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, and a thorough sensitivity analysis becomes imperative.

In the context of the UK, it is clear from the data assessment that there are inconsistencies and uncertainties inherent in each of the 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 (Disney, 2015; Raymer et al., 2013). 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.

Specifically in the United Kingdom, further advances in improving the quality of international migration statistics may come from: making a fuller use of administrative data sources (Bijak, 2012; Boden & Rees, 2010), including data on border crossings and exit checks, where important developments have taken place in the recent years (Home Office, 2017); use of data from other countries in a harmonised and systematic fashion (Raymer et al., 2012); application of statistical modelling to take advantage of different data sources (Disney, 2015; Raymer et al., 2012). Recent recommendations for UK migration and population data are available in Raymer, Rees, and Blake (2015).

3. Selection of a modelling approach appropriate for the type of migration and the available data

The final recommendation relates to selecting appropriate models to forecast the flow of interest, taking into account its characteristics – especially, what type of flow this is, how stable or susceptible to shocks, whether the series exhibit non-stationary features. The length of the available data series is an important consideration as well. In particular, series with non-stable characteristics, such as asylum flows, should not be forecast by using models which assume stationarity of the process, and vice versa, stable labour migration between two highly developed countries have more orderly features than a non-stationary model would predict. Short data series may require additional expert input concerning the future migration or the features of the processes.

On the whole, the main findings of this paper suggest that, given the high levels of uncertainty of migration forecasts, this uncertainty should be stated explicitly, ideally in terms of probabilities. 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 type of migration under study and of the forecasting models, including their past performance and susceptibility to shocks. For longer horizons, scenario-based methods, especially involving probabilistic scenarios, could provide an alternative to statistical extrapolations, but they too would need to be carefully justified and calibrated.

Furthermore, the whole forecasting process could also become more interactive, with forecasters providing bespoke decision advice related to specific user needs (Bijak, 2010; de Beer, 1993). For example, it is possible to utilise a formal statistical decision analysis to support migration-related policies and decisions under uncertainty (Alho & Spencer, 2006; Bijak et al., 2015). Here, the advice given to policy makers based on forecasts would also include information on the relative costs of under-predicting and over-predicting migration when making specific policy or other decisions (Bijak, 2010). Other elements of the research agenda aimed at enhancing the predictive capacity of various migration models include building computer simulations with solid micro-foundations, whereby migrant decisions in response to policies could be modelled explicitly (see Klabunde & Willekens, 2016 for an overview of agent-based migration models). Such an approach would be particularly helpful in setting realistic and potentially interactive migration scenarios, as discussed above.

Following the process outlined above cannot guarantee that the resulting forecasts will exhibit smaller errors, but would help safeguard against making poor forecasts and thus also radically incorrect decisions. It is especially vital that the forecasters remain honest about the limitations of their product, and do not offer methods producing too certain predictions, as they will most likely fail, but neither should the decision makers expect or require them. The main conclusion from the presented analysis is that the various statistical models, such as time series, should be used with caution, their properties being tailored for the characteristics of particular migration flows and for the available data, with due attention paid to the implications of various explicit and implicit model assumptions. With migration often being a politically-charged topic, these caveats are becoming more important than ever.

ACKNOWLEDGEMENTS

The work underpinning this paper has been prepared for and was funded by the UK Migration Advisory Committee (MAC), under the Home Office Science

contract HOS/14/040, and continued within the ESRC Centre for Population Change (ES/K007394/1). We would like to thank the members of the MAC and the MAC Secretariat, as well as two anonymous reviewers of *Journal of Forecasting* for their comments on the earlier drafts. All the interpretations in this paper are exclusively those of the authors, and do not necessarily represent the views of the Migration Advisory Committee, ESRC, or any institutions with which the authors are or were affiliated.

FUNDING

This work was funded by the Migration Advisory Committee (MAC), UK Home Office, under the Home Office Science contract HOS/14/040, and also supported by the ESRC Centre for Population Change grant ES/K007394/1.

REFERENCES

- Abel, G. J. (2010). Estimation of international migration flow tables in Europe. *Journal of the Royal Statistical Society A*, *173*, 797–825.
- Alho, J. M., & Spencer, B. D. (1985). Uncertain population forecasting. *Journal of the American Statistical Association*, *80*, 306–314.
- Alho, J. M., & Spencer, B. D. (2006). *Statistical demography and forecasting*. New York: Springer.
- Arango, J. (2000). Explaining migration: A critical view. *International Social Science Journal*, *52*, 283–296.
- Azose, J. J., & Raftery, A. E. (2015). Bayesian probabilistic projection of international migration. *Demography*, *52*, 1627–1650.
- Bijak, J. (2010). *Forecasting international migration in Europe: A Bayesian view*. Dordrecht: Springer.
- Bijak, J. (2012). *Migration Assumptions in the UK National Population Projections: Methodology Review*. Report for the Office for National Statistics. Southampton: Centre for Population Change.
- Bijak, J., Alberts, I., Alho, J., Bryant, J., Buettner, T., Falkingham, J. C., ... Smith, P. W. (2015). Probabilistic population forecasts for informed decision making. *Journal of Official Statistics*, *31*, 537–544.
- Bijak, J., & Wiśniowski, A. (2010). Bayesian forecasting of immigration to selected European countries by using expert knowledge. *Journal of the Royal Statistical Society A*, *173*, 775–796.
- Boden, P., & Rees, P. (2010). Using administrative data to improve the estimation of immigration to local areas in England. *Journal of the Royal Statistical Society A*, *173*, 707–731.
- Bongaarts, J., & Bulatao, R. A. (2000). *Beyond Six Billion: Forecasting the World's Population*. Washington DC: National Academies Press.
- Cappellen, Å., Skjerpen, T., & Tønnessen, M. (2015). Forecasting immigration in official population projections using an economic model. *International Migration Review*, *49*, 945–980.
- Cohen, J. E., Roig, M., Reuman, D. C., & GoGwilt, C. (2008). International migration beyond gravity: A statistical model for use in population projections. *Proceedings of the National Academy of Sciences*, *105*, 15269–15274.
- de Beer, J. (1993). Forecast intervals of net migration: The case of the Netherlands. *Journal of Forecasting*, *12*, 585–599.
- de Beer, J. (2008). Forecasting international migration: time series projections vs argument-based forecasts. In J. Raymer, & F. Willekens (Eds.), *International Migration in Europe: Data, Models and Estimates* (pp. 283–306). Chichester: John Wiley.
- Disney, G. (2015). *Model-based estimates of UK immigration*. PhD Thesis. University of Southampton, Southampton.
- Disney, G., Wiśniowski, A., Forster, J.J., Smith, P.W.F., & Bijak, J. (2015). Evaluation of existing migration forecasting methods and models. Report for the Migration Advisory Committee. Southampton: Centre for Population Change. Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/467405/Migration_Forecasting_report.pdf (access on 10 July 2018)
- Dustmann, C., Casanova, M., Fertig, M., Preston, I., & Schmidt, C. M. (2003). *The impact of EU enlargement on migration flows. Home Office Online Report 25/03*. London: Home Office.
- EASO. (2017). Quantitative assessment of asylum-related Migration. In *A survey of methodology*. European Asylum Support Office. Luxembourg: EU Publications Office.
- Home Office. (2017). *Second report on statistics being collected under the exit checks programme*. London: Stationery Office. Available at <https://www.gov.uk/government/statistics/second-report-on-statistics-being-collected-under-the-exit-checks-programme>. access on 16 July 2018
- 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., & Kučera, T. (1991). The impact of forecasting methodology on the accuracy of National Population Forecasts. Evidence from the Netherlands and Czechoslovakia. *Journal of Forecasting*, *10*, 371–398.
- Keilman, N. W. (1990). *Uncertainty in national population forecasting: issues backgrounds analyses recommendations*. Lisse: Swets and Zeitlinger.
- Keilman, N. W. (2001). Demography: Uncertain population forecasts. *Nature*, *412*, 490–491.
- Klabunde, A., & Willekens, F. (2016). Decision-making in agent-based models of migration: State of the art and challenges. *European Journal of Population*, *32*, 73–97.
- Kupiszewska, D., & Nowok, B. (2008). Comparability of statistics on international migration flows in the European Union. In J. Raymer & F. Willekens (Eds.), *International Migration in Europe: Data, Models and Estimates* (pp. 41–71). Chichester: John Wiley.
- Long, J. F. (1991). The relative effects of fertility, mortality, and immigration on projected age structure. In W. Lutz (Ed.), *Future demographic trends in Europe and North America: what can we assume today?* (pp. 503–522). San Diego, CA: Academic Press.

- Lutz, W., & Goldstein, J. R. (2004). Introduction: How to deal with uncertainty in population forecasting? *International Statistical Review*, 72, 1–4.
- Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., & Taylor, J. E. (1993). Theories of inter-national migration: A review and appraisal. *Population and Development Review*, 19, 431–466.
- OBR. (2015). *Economic and fiscal outlook – March 2015*. London: Office for Budget Responsibility. Available online via: <http://budgetresponsibility.org.uk/efo/economic-fiscal-outlook-march-2015/>. access on 1 June 2017
- ONS. (2018). *Long-Term International Migration estimates methodology*. Titchfield: Office for National Statistics. Available online via: <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/methodologies/longterminternationalmigrationestimatesmethodology>. access on 10 July 2018
- Pijpers, R. (2008). Problematising the ‘orderly’ aesthetic assumptions of forecasts of East-West migration in the European Union. *Environment and Planning A*, 40, 174–188.
- Ravenstein, E. G. (1885). The Laws of Migration. *Journal of the Statistical Society of London*, 48, 167–235.
- Raymer, J., Abel, G. J., & Rogers, A. (2012). Does specification matter? Experiments with simple multiregional probabilistic population projections. *Environment and Planning A*, 44, 2664–2686.
- Raymer, J., Rees, P., & Blake, A. (2015). Frameworks for guiding the development and improvement of population statistics in the United Kingdom. *Journal of Official Statistics*, 31, 699–722.
- Raymer, J., Wiśniowski, A., Forster, J. J., Smith, P. W., & Bijak, J. (2013). Integrated modeling of European migration. *Journal of the American Statistical Association*, 108, 801–819.
- Shaw, C. (2007). Fifty years of United Kingdom national population projections: How accurate have they been? *Population Trends*, 128, 8–23.
- Stoto, M. A. (1983). The accuracy of population projections. *Journal of the American Statistical Association*, 78, 13–20.
- UN. (1998). *Report of the United Nations Technical Symposium on International Migration and Development*. New York: United Nations.
- Willekens, F. (1994). Monitoring international migration flows in Europe. *European Journal of Population*, 10, 1–42.
- Wilson, T., & Rees, P. (2005). Recent developments in population projection methodology: A review. *Population, Space and Place*, 11, 337–360.
- Wiśniowski, A., Forster, J. J., Smith, P. W., Bijak, J., & Raymer, J. (2016). Integrated modelling of age and sex patterns of European migration. *Journal of the Royal Statistical Society A*, 179, 1007–1024.
- Wiśniowski, A., Smith, P. W., Bijak, J., Raymer, J., & Forster, J. J. (2015). Bayesian population forecasting: Extending the Lee-Carter method. *Demography*, 52, 1035–1059.

AUTHOR BIOGRAPHIES

Jakub Bijak is Professor of Statistical Demography at the Department of Social Statistics and Demography, University of Southampton, UK, and a joint leader of the Modelling Strand of the ESRC Centre for Population Change.

George Disney is Research Fellow In Social Epidemiology, Melbourne School of Population and Global Health, Melbourne, Australia, and an associate member of the ESRC Centre for Population Change.

Allan M Findlay is Honorary Professor of Human Geography at the School of Geography and Sustainable Development, University of St Andrews, UK, and was a joint leader of the Migration Strand of the ESRC Centre for Population Change.

Jonathan J Forster is Head of Mathematical Sciences and Professor of Statistics at the University of Southampton, UK, and a member of the Modelling Strand of the ESRC Centre for Population Change.

Peter WF Smith is Professor of Social Statistics at the Department of Social Statistics and Demography, University of Southampton, UK, and a joint leader of the Modelling Strand of the ESRC Centre for Population Change.

Arkadiusz Wiśniowski is Lecturer in Social Statistics and Quantitative Social Science at the Social Statistics Department, University of Manchester, UK, and an associate member of the ESRC Centre for Population Change.

How to cite this article: Bijak J, Disney G, Findlay AM, Forster JJ, Smith PWF, Wiśniowski A. Assessing time series models for forecasting international migration: Lessons from the United Kingdom. *Journal of Forecasting*. 2019;1–18. <https://doi.org/10.1002/for.2576>

APPENDIX A

Detailed information on the forecast assessment exercise

TABLE A1 Quality assessment of the main data series on the UK migration flows

Data source	Definition	Coverage	Bias	Accuracy
IPS long-term ^a : Immigration and emigration (ONS)	+	+	-	-
IPS short-term ^b : Immigration and emigration (ONS)	-	+	O	-
LFS: Immigration	-	O	O	-
NINo (DWP): Immigration	-	-	O	+
HESA (students) ^c : Immigration	O	-	+	+
Worker registration scheme: Immigration	-	-	O	+
Home Office entry clearance visa data: Immigration	O	-	+	+
2001 census special migration Statistics: Immigration	O	O	O	+
2011 census data: Immigration	O	O	O	+
Home Office asylum seeker Data: Immigration	-	*	+	+

Key: + close match to the true flow; O moderate distortion of the true flow; - major distortion;

Notes: ^a Long-term migrants encompass those who arrive (or leave) with the intention to stay in the country (or abroad) for twelve months or longer;

^b Short-term migrants include those who arrive (or leave) with the intention to stay in the country (or abroad) for three months or longer; ^c Students who had their domicile outside of the UK before commencing their studies (source: HESA, www.hesa.ac.uk).

*Exact match for the legally-defined category in question (asylum seekers). Source: For details, see Disney et al. (2015).

TABLE A3 Summary of scoring rules for classifying different forecasting models into quality categories

Calibration \ Error MPE	Well-calibrated intervals* score = 0	Slightly miscalibrated* score = 1	Too optimistic or too pessimistic* score = 3	Extremely narrow/extremely wide* score = 5
Small errors within ±15% Score = 0	0	1	3	5
Medium errors ± 15 to ±30% Score = 1	1	3	5	7
Large errors ±30 to ±50% Score = 3	3	5	7	9
Very large errors Beyond ±50% Score = 5	5	7	9	10

Note: The average scores under 3 fall into the ■ (green) category, with low errors and good calibration; scores between 3 and 5 to the ■ (orange) category, with middle-sized errors or reasonable calibration, and scores of 5 or above to the ■ (red) category, with high errors or poor calibration.

*Scores for calibration classes – empirical coverage of the nominal 50-per cent and 80-per cent intervals:

