Journal of the
Royal Statistical Society
SERIES A
Statistics
in Society

J. R. Statist. Soc. A (2010)
173, *Part* 4, *pp.* 775–796

Bayesian forecasting of immigration to selected European countries by using expert knowledge

Jakub Bijak

University of Southampton, UK, and Central European Forum for Migration and Population Research, Warsaw, Poland

and Arkadiusz Wiśniowski

Central European Forum for Migration and Population Research, Warsaw, and Warsaw School of Economics, Poland

[Received January 2009. Final revision November 2009]

Summary. The aim of the paper is to present Bayesian forecasts of immigration for seven European countries to 2025, based on quantitative data and qualitative knowledge elicited from country-specific migration experts in a two-round Delphi survey. In line with earlier results, most of the immigration processes under study were found to be barely predictable in the long run, exhibiting non-stationary features. This outcome was obtained largely irrespectively of the expert knowledge input, which nevertheless was found useful in describing the predictive uncertainty, especially in the short term. It is argued that, under the non-stationarity of migration processes, too long forecasts horizons are inadequate, which is a serious challenge for population forecasts in general.

Keywords: Bayesian forecasts; Delphi method; Expert knowledge; Immigration; Model selection

1. Introduction

The aim of this paper is to present Bayesian forecasts of immigration for seven European countries to 2025, based on quantitative data and qualitative knowledge elicited from country-specific migration experts in a two-round Delphi survey. Such a horizon was selected to assess the overall plausibility of migration forecasts in a longer term. The geographical scope of the study covers two 'old' immigration countries of western Europe (Austria and France), two 'new' immigration countries of southern Europe (Italy and Portugal) and three transition countries of central Europe (the Czech Republic, Hungary and Poland).

According to Willekens (1994) and Kupiszewski (2002), the main difficulties in forecasting international migration include

(a) inherent randomness of the processes and their susceptibility to barely predictable factors,
(b) a lack of coherent definitions of migration across countries and time,
(c) a lack of comprehensive migration theories and
(d) a lack of data or incomplete data, including short time series.

However, migration forecasts with suitable assessments of uncertainty are crucial for ob-

Address for correspondence: Jakub Bijak, Division of Social Statistics and Southampton Statistical Sciences Research Institute, School of Social Sciences, University of Southampton, Southampton, SO17 1BJ, UK.
E-mail: J.Bijak@soton.ac.uk

0964–1998/10/173775

taining credible population predictions, especially in developed countries. As a partial remedy to these problems, Willekens (1994) suggested the formal inclusion of expert knowledge in migration forecasts. A natural methodology for handling the combination of expertise and the data is the Bayesian approach. Expert judgement can be represented by the prior probability distributions and combined with data reflected in the likelihood function by means of the Bayes theorem. This path is followed in this paper.

Apart from this introduction, the paper is structured in five sections. Section 2 briefly addresses the data issues that are specifically related to the current forecasting task. Section 3 focuses on the forecasting models and their estimation, including the numerical procedures that are used. Elicitation of expert knowledge by means of a Delphi survey is described in Section 4, and Section 5 presents some forecast results. Finally, Section 6 summarizes the outcomes and offers some tentative conclusions and recommendations for forecast makers and forecast users.

More detailed technical information about the methods and results can be found in two background working papers. Bijak and Wiśniowski (2009) provide forecast results, assumptions on the parameterization of prior distributions, data support for particular models and WinBUGS code that is used in the computations, whereas Wiśniowski and Bijak (2009) present a sample Delphi survey that is used for the elicitation of expert opinion. A non-technical summary of findings that are presented in this paper is forthcoming in a book summarizing the results of the research project 'IDEA', on which this study was based (Wiśniowski *et al.* (2010); for more details, see the acknowledgements).

2. Sources of data and preparation

The forecasting exercise concentrates on the registered total inflow of migrants into the seven European countries. The primary sources of the data on immigration include Eurostat, United Nations Statistics Division, national statistical institutes and the Council of Europe's *Demographic Yearbooks*. The longest series were available for Italy, starting in 1981; the shortest—12 observations each—for Austria and France. In addition, the most recent figures for Portugal required recalculation, as described in more detail in Section 3.3.

In general, the data on international migration in Europe lack comparability owing to differences in the definitions, especially with respect to the required duration of stay of migrants (Kupiszewska and Nowok, 2008). Moreover, in some countries the definition of who is an immigrant changed, as for example in the Czech Republic in 2001. Since the harmonization of migration data was beyond the scope of this study, the consequence is the lack of direct comparability of flows that are analysed for various countries. The perspective adopted is thus of country-specific population balance equations, with the aim of ensuring within-country consistency of volumes of immigration forecasted and population stocks registered, rather than to enable between-country comparisons. The definitions and temporal coverage of the data for particular countries are summarized in Table 1, together with the statistical sources that were used in each case.

As an alternative to the research design presented, from the point of view of the demographic balance equation alone, the focus of analysis and forecasts could be on net migration. However, data on net migration, calculated from annual changes in population stocks less natural increases, exhibit most of the same problems as data on flows, since both flows and stocks usually share a common underlying definition. Besides, post-census corrections, if any, are often labelled as 'administrative corrections' rather than attributed to particular migration flows. Moreover, since Rogers (1990) criticized net migration as an analytical concept that obscures the patterns of underlying social processes (immigration and emigration), the methodology of

Table 1. Definitions of immigration flow and sources of data for the seven European countries under study†

Country	*Sample*	*Duration of stay*	*Comments*	*Sources*
Austria	1996–2007	3 months	Includes both short- and long-term immigrants	Eurostat and national statistical institutes
Czech Republic	1993–2007	Permanent‡ 3 months§ 1 year§§	Includes permanent immigrants; duration-of-stay criteria were introduced in 2001	Eurostat, United Nations Statistical Division and Council of Europe *Demographic Yearbooks* and national statistical institutes
France	1994–2005	1 year	Includes foreigners with long-term settlement permits	Institut National d'Etudes Démographiques
Hungary	1990–2006	1 year	Includes registered long-term immigrants	Eurostat, United Nations Statistical Division and joint migration questionnaire
Italy	1981–2005	Not specified‡§ 6 months§§	Includes immigrants recorded by population registers	Eurostat, Council of Europe *Demographic Yearbooks* and national statistical institutes*
Poland	1990–2007	Permanent	Includes immigrants for permanent residence only	Eurostat, United Nations Statistical Division and Council of Europe *Demographic Yearbooks* and national statistical institutes
Portugal	1992–2006	1 year	Includes people with permits of stay for at least 1 year	Eurostat and national statistical institutes** (recalculated)

†Source: Kupiszewska and Nowok (2008), pages 54–62, and own elaboration.
‡Refers to nationals.
§Refers to citizens of European Economic Area countries.
§§Refers to non-European Economic Area citizens.
*Data obtained by courtesy of Dr Frank Heins, Institute of Research on Population and Social Policies, Rome.
**Data obtained by courtesy of Professor João Peixoto, Centre for Research on Economic Sociology and the Sociology of Organisations, Lisbon.

population forecasting began to shift towards modelling *rates* of emigration, relative to the population at risk, and *volumes* of immigration, in which case the population at risk cannot be properly defined. This paper deals therefore only with total immigration, whereas Bayesian models for forecasting emigration rates have been proposed and discussed elsewhere (Bijak, 2008).

3. Methodological framework for forecasting immigration

3.1. General forecasting framework applied in the current study

Many researchers argue that probabilistic predictions are the future of population forecasting, owing to their ability to quantify uncertainty in a proper and coherent manner (for an example, see Lutz and Goldstein (2004)). Migration forecasts are here by no means an exception, regardless of the difficulties that are associated with the task, i.e. of high predictive uncertainty. There are several arguments for using the Bayesian approach in the migration context. Firstly, in many cases the data series that are available in Europe, unlike in some other regions, such as Australia and New Zealand (Gorbey *et al.*, 1999), may be too short to allow for a meaningful classical inference. Secondly, for many flows, traditional methods can underestimate the uncertainty of forecasts (Bijak, 2008). Besides, including expert judgement in Bayesian statistical inference allows for accommodating qualitative migration scenarios in the forecasts.

So far, the Bayesian approach in forecasting migration has been successfully applied in a handful of studies (Gorbey *et al.*, 1999; Brücker and Siliverstovs, 2006; Bijak, 2008), where it proved to yield similar *ex post* errors, yet more realistic predictive intervals than the classical approach. Bayesian inference can be also seen as a coherent alternative to the purely expert-based framework that was proposed by Lutz *et al.* (2004). The latter also relies on expert judgement but makes use of neither the inferential mechanism nor the full information from the data sample.

3.2. Specification of the forecasting models

With respect to model specification, several aspects of the forecasting exercise have been considered. Firstly, the analysis has been limited to total immigration and the main forecasted variable is thus a (log-transformed) volume of migration rather than any of the related intensity measures (rates or ratios). The main rationale is that it seemed much more natural to elicit expert knowledge on the absolute size of flows rather than on relative indicators. Besides, as mentioned in Section 2, for such intensity measures the populations at risk cannot be properly defined.

To start with, consider four models: M_1–M_4. Let two of them (M_1 and M_3) be auto-regressive (AR) models of the first order, AR(1), additionally containing a deterministic trend, whereas the remaining models (M_2 and M_4) are simple random-walk (RW) models with drift. Out of each pair, let the first model be characterized by constant variability (CV), and the second by conditional variance changing according to the simplest 'stochastic volatility' (SV) scheme. The model equations are listed below for models M_1, M_2, M_3 and M_4 respectively (see Greene (2003)):

$$\ln(m_t) = c_1 + \gamma_1 \,\text{trend}_t + \phi_1 \ln(m_{t-1}) + \varepsilon_{1t}, \qquad \varepsilon_{1t} \overset{\text{IID}}{\sim} N(0, \sigma_1^2); \tag{1a}$$

$$\ln(m_t) = c_2 + \ln(m_{t-1}) + \varepsilon_{2t}, \qquad \varepsilon_{2t} \overset{\text{IID}}{\sim} N(0, \sigma_2^2); \tag{1b}$$

$$\begin{aligned} \ln(m_t) &= c_3 + \gamma_3 \,\text{trend}_t + \phi_3 \ln(m_{t-1}) + \varepsilon_{3t}, \qquad & \varepsilon_{3t} &\sim N(0, \sigma_{3t}^2), \\ \ln(\sigma_{3t}^2) &= K_3 + \psi_3 \ln(\sigma_{3t-1}^2) + \xi_{3t}, & \xi_{3t} &\overset{\text{IID}}{\sim} N(0, \upsilon_3^2); \end{aligned} \tag{1c}$$

$$\begin{aligned} \ln(m_t) &= c_4 + \ln(m_{t-1}) + \varepsilon_{4t}, \qquad & \varepsilon_{4t} &\sim N(0, \sigma_{4t}^2), \\ \ln(\sigma_{4t}^2) &= K_4 + \psi_4 \ln(\sigma_{4t-1}^2) + \xi_{4t}, & \xi_{4t} &\overset{\text{IID}}{\sim} N(0, \upsilon_4^2). \end{aligned} \tag{1d}$$

In expressions (1a)–(1d), m_t universally denotes registered migration inflow according to a country-specific definition, which is then log-transformed to ensure positive migration flows and asymmetry of predictive distributions (heavier upper tails). Further, c_i are constants, t is the time, trend_t is a country-specific trend function and ϕ_i are auto-regression coefficients, whereas ε_{it} are noise terms, assumed to follow normal distributions with zero mean and variance σ_{it}^2, $N(0, \sigma_{it}^2)$. Furthermore, for the simple SV models M_3 and M_4, the log-variances are assumed to follow a first-order AR process with means K_i, auto-regression coefficients ψ_i and normal errors with zero mean and variance υ_i^2. In this case, the stochastic processes for variance are additionally assumed to be stationary (with $|\psi_i| < 1$), to avoid the exploding volatility of forecasts. In general, the model-specific indices i of parameters $\theta \in \{c, \gamma, \phi, K, \psi, \sigma, \upsilon\}$ denote, not different parameters, but rather a given parameter in the ith model, $\theta_i = (\theta | M_i)$. Further particulars of the models for specific countries are presented in Section 3.3.

It is worth noting that, in simple models, additionally assuming *a priori* gamma-distributed precision (inverse of the variance), a normal likelihood yields a Student t predictive distribution. This holds especially for linear regression models, but it need not be exactly so under non-linearity. For the above models the predictive distributions for $\ln(m_t)$ have heavier tails than the Gaussian distribution, especially for such small samples as those which are available in this study.

The ultimate forecasting models are selected from M_1–M_4 on the basis of the posterior odds criterion. The general idea behind the model selection is also Bayesian and consists in defining prior probabilities $p(M_i)$, for models M_i, $i=1,\ldots,4$, assuming that the models are mutually non-nested. To fulfil this assumption, $\phi_1 \neq 1$ and $\phi_3 \neq 1$ additionally need to hold for models M_1 and M_3. Subsequently, the prior probability of each model M_i is combined with the marginal density of the data x under the model. As outcomes, posterior model probabilities given the data are provided by the Bayes rule (e.g. Osiewalski (2001), page 21, and Hoeting *et al.* (1999)):

$$p(M_i|x) = p(M_i)\ p(x|M_i)/p(x). \tag{2}$$

For forecasting, the model with the highest value of $p(M_i|x)$ is finally selected. In such cases, when more than one model has a relatively high posterior probability, arbitrarily assumed as over 0.05, forecasts can be obtained as weighted averages from the predictive distributions. Averaged distributions are mixtures of predictive distributions yielded by particular models, with mixing obtained by using a categorical distribution defined by the posterior probabilities (2), treated here as 'weights'. More on Bayesian model selection and forecast averaging (also known as 'inference pooling') can be found for example in Hoeting *et al.* (1999) and Osiewalski (2001).

3.3. Country-specific issues

During the modelling process several country-specific issues had to be resolved, focusing on two areas:

(a) deterministic trends, which had to be incorporated in the models in such a way as to remain coherent with the information elicited from the experts in the Delphi survey (see Section 4), and
(b) dummy variables used to handle some specific characteristics of the data.

In general, the deterministic trends were incorporated in AR models, but not in the RW models. This was motivated by the non-stationary characteristic of the RW process. As noted by Wu *et al.* (2007), properly defined trends for non-stationary and non-linear series would have to be intrinsic to the data and thus time varying. Otherwise, global time invariant trends would be barely possible to distinguish from the effects of non-stationarity or cycles (Wu *et al.* (2007), page 14890). Given the limitations of the data discussed before, however, the time varying trends were not considered in this study.

For all countries except France and Poland the trend for immigration flows that was indicated by the experts was logarithmic for immigration totals (or double-logarithmic for log-transformed volumes of flows): $\text{trend}_t = \ln[\ln\{g(t)\}]$, where g is an expert-based function of time t. In the Polish case, the trend followed a logistic pattern, as suggested by almost all the Polish experts. The general formula that is used for the logistic trend, after log-transformation, is

$$\text{trend}_t = \ln(\alpha) - \ln\{1 + \beta \exp(-\gamma t)\}. \tag{3}$$

The parameter α handles the asymptotic value of the future immigration flows, and parameters β and γ are responsible for the curvature and the inflection point. The second part of the right-hand side of equation (3) was transformed by using the formula for the inflection point: $t^* = \ln(\beta)/\gamma$. Thus the trend that was ultimately adopted in models M_1 (expression (1a)) and M_3 (expression (1c)) was

$$\text{trend}_t = -\ln[1+\exp\{\gamma(t^* - t)\}], \qquad \text{for } t = 2, 3, \ldots, 36, \tag{4}$$

with only one parameter γ being subject to inference. The $\ln(\alpha)$ in equation (3) was set to be equal to the model-specific constant c.

For France the trend was assumed to be linear for totals (or logarithmic for log-transformed flows). In fact, 40% of the experts indicated no trend and 40% indicated a linear trend. Although such heterogeneity may seem surprising, it can possibly be attributed, to a certain degree, to the respondents' fields of expertise, which could encompass such different processes as migration within the European Union and inflows from the Maghreb countries or Francophone Africa. Ultimately a linear trend was used, as it seemed to fit the data at hand better. Moreover, the sensitivity analysis that was performed with respect to models with and without trend showed that it virtually did not affect the results on the model selection—in both cases the best model was an RW with probability greater than 0.99 (see Section 5.3).

Specifically in the case of the Czech Republic, the trend additionally accommodated the 2001 change of definition of immigration. In the RW models an additional dummy variable was included for that purpose, equalling 1 in the period until 2000 and 0 thereafter. Besides, the future trend was enhanced by including additional information from qualitative comments of two Czech experts, indicating an increase in total immigration levels until 2012 and a decrease thereafter. In this particular instance the descriptive expert judgement proved to be very helpful.

For Portugal also a dummy variable was introduced, which handled the mostly unwelcome variation in the data in the very last period of the sample (2006), the high value of which resulted from a recent regularization of irregular migrants held in Portugal. The number of 62 332 immigrants in 2006 included people who applied for a residence permit for the first time, as well as people who arrived in Portugal in previous years. The latter number (31 605 people) was distributed over the years 1997–2005, the period between the two regularizations of irregular immigrants (Sabino and Peixoto, 2008). Moreover, by using a single-period dummy for 2006, the last observation was deprived of influence in the forecasts.

3.4. Numerical procedures

To obtain the posterior characteristics of the parameters, as well as the predictive densities, a Gibbs sampling algorithm was used, being a version of the Markov chain Monte Carlo simulation technique (Casella and George (1992), page 168, and Osiewalski (2001), page 39). The algorithm consists in iterative sampling from the conditional distributions for the parameters, starting from some predefined initial values. Usually, an initial sample is discarded (the so-called *burn-in* phase) to assure convergence to the posterior distribution and to eliminate the starting point effects. The rest of the sample is then used to estimate the characteristics of the posterior distribution. Having estimated the densities of the parameters, in the first phase of computations the samples from predictive distributions are generated.

For the problem of model selection, the algorithm of Carlin and Chib (1995) was applied, constituting the second phase of the calculations. This technique allows the computation of the posterior probabilities (2) of particular models (see Section 3.2) within the Gibbs sampling procedure. The method consists of an iterative sampling from full conditional distributions for model-specific parameters (some characteristics of which are obtained in the first phase), as well as for the model indicator. To ensure a smooth passage of the algorithm between various models, additional linking distributions ('pseudopriors') are applied, set close to the first-round posterior estimates (Carlin and Chib, 1995). In the case of Hungary, Poland and Portugal the algorithm for averaging forecasts that were yielded by different models was finally applied, which comprised the third phase of computations.

The computations were performed in version 3.0.3 of the WinBUGS (OpenBUGS) software (Spiegelhalter *et al.*, 2007). From the point of view of applied research, WinBUGS provides a convenient, and quite an intuitive, high level language of programming for Bayesian statistical problems, with possible extensions offering an interface with other statistical packages, such as R (via the library BRugs). Another advantage of using WinBUGS is its flexibility, as the package allows for specifying various models without having to programme the underlying numerical algorithms, such as the Gibbs sampler (see also Congdon (2003)). Sample WinBUGS code that was used in this study can be found in Bijak and Wiśniowski (2009).

For the verification of convergence, a heuristic method was applied, consisting of observing the quantiles and auto-correlations of Markov chain Monte Carlo samples. The number of samples differed across models and countries. In the first phase, the usual burn-in sample size was 50 000 and the number of samples from the assumed posterior distribution was 150 000 (in the case of Poland the burn-in sample size was 850 000 owing to the non-linearity in the model). To reduce the auto-correlation of the samples for some parameters (usually, the constants, trends and parameters in the AR models), so-called *k-thinning* was applied. This means that only every kth iteration from each simulation was selected for the calculation of the posterior density characteristics. The parameter k was selected judgementally, depending on the shape of the auto-correlation functions, and ranged from 10 to 25 (except for Poland, where thinning was not used). The sample sizes were enough to ensure the reasonable stability of the quantiles of the parameters, as well as of the predictive densities.

In the second (Carlin–Chib) phase, the burn-in and posterior samples were of the same size as in the first phase, except for Poland, where they equalled respectively 100 000 and 900 000 iterations. For the forecast averaging (third phase), in the models for Hungary and Portugal the sample sizes were the same as in the first and second phases, whereas in the case of Poland they totalled 750 000 and 250 000. The choice of k was the same in all three phases for all countries. In all instances, the computations involving model selection and hierarchical models, such as the SV models, appeared to be both time and central processor unit demanding, with an average run length equalling about 3 h on a 2-GHz Windows personal computer.

The Monte Carlo sampling errors in most of the models were reasonably small, as far as the model parameters are concerned. However, in some forecasts that were produced by several models, the Monte Carlo error grew together with the forecasting horizon (for example this was so for the AR–SV model for Poland). This was the result of the presence of outliers sampled in earlier iterations. This motivated making all the inferences in terms of location parameters (i.e. the median and quantiles of the distributions), which are unlikely to be affected by outliers, rather than using means and standard errors.

4. Expert knowledge elicitation

4.1. Introduction to the Delphi method

The elicitation of *a priori* expert knowledge on immigration processes concerning the countries under study, which is required by the proposed methodology for the forecasting exercise, was carried out by means of a simple Delphi survey.

In general, the Delphi survey is a technique that obtains data and opinions through surveys carried out via mail, which originally stems from the applications in the US military (see for example Dalkey (1969)). Respondents in the Delphi survey should be experts in a given field and answer anonymously. The key feature of the Delphi method is that the judgements are obtained iteratively and the respondents are provided with the aggregated statistical feedback concerning the results of the preceding round. The respondents providing extreme answers may be asked to justify their views. In the subsequent rounds the respondents can reformulate their opinions to reach a consensus (Armstrong, 1985; Rowe and Wright, 1999), which can be seen as an informed consensus.

The technique requires preparation of a survey according to the rules that take into account insights from cognitive psychology, to ensure unambiguous answers. Rowe and Wright (2001) have provided evidence for a strong influence of question formulation on the answers that are obtained.

The particularity of the research task that is presented in this paper is that several questions concerned subjective probabilities. That means that the experts were asked how they perceived the future in terms of their subjective beliefs or convictions about the behaviour of the inflow of migrants to the expert's country. The literature provides numerous examples of improper assessment of the probabilities in this context: viewing the uncertainty only in terms of frequencies (Gigerenzer, 1994; Kadane and Wolfson, 1998), inconsistencies between perceiving the probabilities expressed either directly or indirectly (Goodwin and Wright, 1998), forgetting about the context of the problem (Gigerenzer, 1994) and finally overconfidence and improper assessment of the spans of uncertainty (Armstrong, 1985; Rowe and Wright, 2001; Kadane and Wolfson, 1998; O'Hagan, 1998).

Another issue concerns the selection of experts for the Delphi survey. They should have appropriate heterogeneous knowledge, encompassing the whole domain of the problem. The group should not be too numerous but should depend on the resources that are available and the quality of feedback that is expected from them. It is argued that more respondents may cause information overload, conflicting opinions or irrelevant arguments (for details, see Rowe and Wright (2001)).

Although the evidence suggests that expert judgement alone is not of great value in forecasting (Armstrong, 1985), in the current novel application of the Delphi approach, the expert knowledge served the purpose of formulating the prior probabilities that were further combined with the quantitative data, to obtain forecasts. Originally, the Delphi survey alone was used as a tool for prediction making (for a migration forecasting example, see Drbohlav (1996)).

4.2. Application of the Delphi survey—general information

In this study, the survey-based elicitation process consisted of two rounds, to allow for corrections and possible convergence of the initial judgements. The *a priori* expert knowledge was elicited from between six (for Austria and Hungary) and 14 (for Poland) respondents per country. The experts were selected by the national teams participating in the project and represented a variety of fields of expertise, involving both academics from various disciplines (including demography, human geography, sociology, economics and political science), as well as public officials (such as representatives of the statistical offices and state agencies).

The survey concerned process characteristics (parameters of the forecasting models), rather than the processes as such (future values of volumes of migration). This solution was found to be more straightforward, as it did not require additional recalculations to transform the expert-based predictive probability distributions into prior distributions.

Unlike in the tacit assumptions that are made in many Bayesian literature examples (e.g. Kadane and Wolfson (1998) and Dey and Liu (2007)), in the present study expert knowledge was elicited from migration specialists of various backgrounds, but predominantly from non-statisticians (for a thorough overview of elicitation issues in this context, see O'Hagan (1998) and O'Hagan *et al.* (2006); another example is provided by Szreder and Osiewalski (1992)).

In this study, the background of experts implied very strong limitations on the use of formal terms in a survey (e.g. ‘distribution’, ‘variance’, ‘probability’, ‘stationarity’ and ‘quantile’). The aim was to elicit the opinions and judgements by using a natural language rather than formal terms, and providing the experts with some intuition by means of the visualizations of certain concepts. However, with respect to migration research, the area seems still uncharted. As has been noted by O’Hagan (1998), page 22,

> ‘...to elicit a genuine prior distribution ... is a complex business demanding a substantial effort on the part of both the statistician and the person whose prior beliefs are to be elicited. A Bayesian who wishes to take this task seriously finds little guidance in published work that is directly relevant to the task that he or she faces.’

Attempts that were undertaken in this study are outlined below.

4.3. Questionnaire construction

The questionnaires for each of the countries under study were prepared with the same layout, which allowed for the handling of country-specific information, such as definitions of migration or data collection practices and the possible effect of some economic and demographic variables on immigration, as well as main directions of migration inflows.

The first-round questionnaire consisted of 15 questions, five of which (1–5) are described in detail below. They concerned the general tendency of immigration to a particular country, the shape of the process and its volatility. In addition, questions 6–11 were related to the possible effect of demographic and economic variables, and questions 12 and 13 referred to main directions of immigration. These issues remain beyond the scope of this paper.

(a) Question 1 concerned the long-term (until 2025) general tendency (direction) of future immigration flows. Stylized figures indicating the shape of the trend (constant, as well as increasing or decreasing linear, logarithmic and logistic), which the flows might follow, were presented. The experts were asked to choose one from the figures or to describe another type of trend themselves.
(b) Question 2 aimed at the elicitation of the (covariance) stationarity characteristics of the immigration process. Three figures presented the examples of immigration processes that indicated the stationary, RW-type non-stationary and explosive characteristics. The experts were asked to provide the subjective chances of occurrence of a given process or to describe their own characteristic and to assign a subjective probability to it.
(c) Question 3 concerned the volatility of the future immigration process. Two stylized figures presented the idea of constant volatility and SV. The experts were asked to provide their subjective estimates of the probabilities of occurrence with the possibility of describing their own characteristic.
(d) In question 4 the experts were asked for estimates of the average future deviations of the immigration processes from the assumed average immigration levels for a given country, to be chosen from the range 10–1000%, or to provide their own. This question was aimed at providing information on the standard deviation of the process.
(e) The aim of question 5 was to assess the volatility of the standard deviation that was provided in the preceding question. This was achieved by asking the experts for their level of certainty concerning the answer given in question 4, measured on an 11-degree scale ranging from 0 (very uncertain) to 10 (almost sure).

Moreover, the survey contained an open-ended question, where the experts could provide additional explanations, justifications for answers or comments concerning either the merit or the questionnaire itself. A sample second-round questionnaire for Austria can be found in the appendix of Wiśniowski and Bijak (2009).

In addition to the above, the second-round questionnaire contained summaries of first-round answers in the form of histograms (questions 1, 4 and 5) or probabilities expressed as percentages (questions 2 and 3). In addition two new questions were included.

(a) Question N1 concerned the characteristics of the logistic trend (only for Poland), namely the upper bound of the future level of immigration and the inflection point of the trend.
(b) Question N2 referred to the effect of the immigration policy on immigration flows, following the first-round indications of its importance provided by the experts.

4.4. Translation of the answers into probability distributions

The experts' answers to the questions were summarized and then used to formulate the prior probability distributions for the models that were described in Section 3.2. Firstly, the constants c were included in every model to handle the mean value of the (log-transformed) levels of immigration. The priors for every country except Poland (a detailed explanation follows below) were assumed to be normal with zero mean. In the case of the AR models the priors were diffuse, although with additional information concerning the expected immigration policy. The tighter the policy the majority of experts indicated in their answers to question N2, the greater the precision that was set. As it was difficult to formulate the policy-related question and to operationalize the answers, an arbitrary, yet simple and hopefully sensible approach (tighter policy—less room for change) was proposed.

We are aware that this assumption can be seen as problematic, e.g. owing to the presence of illegal migration and alternative destinations for migrants, if a more restrictive policy should come into force. A partial solution to this problem can consist of modelling migration flows within a system of countries jointly, including various interdependences, policy-related feedback effects and additional characteristics, such as sizes of population and economic setting. Although such an analysis remains far beyond the scope of this study, it is definitely worth addressing in future research.

Notwithstanding this, except for several cases of the RW models, the priors for c were relatively diffuse and the data changed them anyway. The priors in the RW models were, in almost all cases, concentrated around zero, owing to the undesirable characteristics of the process and the resulting exploding forecasts. On one hand, this can be seen as a drawback of the analysis but, on the other hand, as confirmed by the results (see Section 5), the nature of an RW allows the capture of the specific variability in the immigration data reasonably well.

As mentioned in Section 3.3, the deterministic trend that was indicated by the vast majority of experts in all countries except Poland and France was logarithmic. The trend was included in the AR models only, mainly owing to the reasons that were mentioned in the previous paragraph. As almost all experts in all countries pointed out the increasing tendencies of the immigration flows, the priors that were set for the parameters γ in the AR models were normal with mean 0.5 and precision 1, resulting in a distribution with about 30% of the probability mass below 0. The exception was the Czech Republic, where this prior was set to be diffuse, again to ensure reasonable results.

In the case of Poland, the logistic trend that was suggested by the experts was described in detail in Section 3.3. The $\ln(\alpha)$ in equation (3) served as a constant, with the normal prior elicited from the answers to question N1: the upper bound for immigration was set to 90 000; thus the prior mean was ln(90 000), and the precision was defined according to the precision that was estimated from the experts' answers. The value of t^* in equation (4) was elicited from question

N1 as 2019. The prior for the coefficient γ, because of the computational and convergence issues, was assumed to be informative and to follow a beta distribution with parameters 20 and 2.

For the auto-regression parameters ϕ, normal priors were set according to the information from the answers to question 2, indicating a stationary process with $|\phi| < 1$ and an explosive process with $\phi > 1$. The processes with $\phi \leqslant -1$ appeared to have negligible probability mass attached to them and therefore the priors did not require left truncation. Subsequently, the answers were normalized to represent the experts' beliefs on probability mass below and above 1. Finally, the values of the mean and precision were selected by using a grid searching algorithm. Note that under such a distribution the RWs with $\phi = 1$ have zero probability mass and therefore were accounted for by separate models M_2 and M_4.

The priors for the precision parameter, $\tau = \sigma^{-2}$, were assumed to follow the gamma distributions $\Gamma(r, \mu)$. The shape parameter r was arbitrarily set to 2, which was underlain by the answers given by the experts to question 5: the average degree of experts' certainty was medium and oscillated in every country between 4 and 6. Hence, it was justifiable to use $r = 2$ for each of them and then to control for the expected value of the precision by using the scale parameter μ. If the answers had fallen on average into the range 0–3 or 7–10, then either $r = 1$ or $r = 3$ would have been used, respectively depicting relatively high or low uncertainty.

The expected value of the precision was derived from question 4. The estimate of the standard deviation (the weighted mean of the answers) was multiplied by 1.25 to eliminate the bias resulting from the usual confusion of the average absolute and standard deviations (see for example Goldstein and Taleb (2007)). Subsequently, assuming a log-normal distribution of m_t, the expected value for the precision τ was derived as

$$E(\tau) = -\log\left\{1 + \frac{\mathrm{var}(m_t)}{E(m_t)^2}\right\}.$$

Finally, the value of the scale parameter μ of the gamma distribution was calculated by using the formula $E(\tau) = r/\mu$.

The procedure proposed is merely one of the available options for operationalizing the expert judgement with respect to precision. In the sensitivity analysis, an alternative hyperparameterization was investigated (see Section 5.3). Moreover, after examining the correlation between the answers to questions 4 and 5, it could be also possible to base the results on the outcome of question 4 alone, which would then indicate both the mean and the standard deviation of τ.

The values of the hyperparameters for priors of the SV model-specific parameters, namely K, ψ and precision $\rho = \upsilon^{-2}$, were quite informative to assure the convergence of the numerical algorithms. As these parameters depict rather complex, non-intuitive statistical notions, they were not subject to elicitation from the experts but were instead judgementally assumed by the authors. Moreover, the prior for a dummy variable in the models for the Czech Republic was assumed to be normal, concentrated at -1, corresponding to a more rigorous definition of an immigrant that was in use until 2000. For Portugal a diffuse prior for the dummy was used. The prior distributions for all parameters of the forecasting models are summarized in Table 2.

The prior probabilities that were set for particular models (see Section 3.2) were elicited from the answers to questions 2 and 3. First, the marginal probabilities for AR models, $p(M_1, M_3)$, and RW models, $p(M_2, M_4)$, were calculated from the answers to question 2. The probabilities for the CV property, $p(M_1, M_2)$, as well as for SV, $p(M_3, M_4)$, were derived from the averaged answers to question 3. In both cases the information from open-ended options was included. The final probabilities $p(M_i)$ were calculated assuming the independence of AR–RW and CV–SV models.

Table 2. Prior distributions for the model of total immigration flows—a summary

Parameter	*Distribution*	*Informative*	*Expert knowledge*
C	Normal	Yes or no†	Yes or no†
ϕ	Normal	Yes	Yes
γ	Normal or beta (PL)‡	Yes	Yes or no (PL)‡
$\tau=\sigma^{-2}$	Gamma	Yes	Yes
K	Normal	Yes	No
ψ	Normal	Yes	No
$\rho=\upsilon^{-2}$	Gamma	Yes	No
Dummy	Normal	Yes or no†	No

†Depending on the model type and/or stability of forecasts.
‡Prior distribution for the logistic trend coefficient.

5. Results of forecasts: an overview

5.1. Modification of prior beliefs in the light of data

With respect to the interpretation of the results of Bayesian forecasts, a key issue is how, and to what extent, did the quantitative data modify the prior knowledge that was elicited from the experts via the Delphi survey? In this context, two problems related to the prior beliefs are crucial: firstly, concerning the selection of a probabilistic model M_i driving the processes under study, and, secondly, referring to the model parameters θ.

The most important observation with respect to the selection from models M_1–M_4, defined in expressions (1a)–(1d) in Section 3.2, is that in a majority of cases the data gave clear preference to a simple CV RW model with drift (model M_2). For Austria, the Czech Republic, France and Italy the results were unambiguous, with the *a posteriori* probability of its selection given the data, $p(\mathrm{M}_2|x)$, exceeding 0.95. For Hungary and Portugal, the AR models with trend (model M_1) also appeared to play some role, with an almost 50–50 split between the two models in the former case, and with $p(\mathrm{M}_2|x)=0.63$ and $p(\mathrm{M}_1|x)=0.37$ in the latter. In this case $p(\mathrm{M}_1|x)$ was slightly smaller than the respective prior probability $p(\mathrm{M}_1)=0.38$. For Poland, in turn, two RW models won the competition: the model with the SV (model M_4) and with CV (model M_2). The respective posterior model probabilities were 0.53 and 0.47. Clearly, these results are specific to the very limited space of the models that were considered, but at the same time they appear relatively robust to the prior probabilities. The prior and posterior probabilities of various models for particular countries are presented in Table 3.

With respect to selected parameters of the forecasting models, in many cases the prior distributions have been visibly modified by the data. For example, with the exception of Portugal, the precision parameters τ generally acquired higher values *a posteriori* than *a priori*. This suggests that in many cases the expert knowledge allows for higher assessments of uncertainty than the data alone. A similar conclusion was reached by Bijak (2008), who advocated that a lower predictive precision of migration might be in many cases more realistic than a higher one, yielded by the models exclusively based on the trend extrapolation and devoid of expert knowledge. As to the AR terms in the trend models, the posterior distributions of ϕ were more concentrated than the priors. These distributions were also mostly shifted either towards 1 (especially for Austria, but also to some extent for Hungary, Italy and Portugal) or towards -1 (for Poland), with some non-negligible probability of non-stationarity.

Table 3. Prior and posterior probabilities for models M_1–M_4

Country	*Prior probabilities (averaged beliefs of the experts)*				*Posterior probabilities (modified by the data)*			
	M_1	M_2	M_3	M_4	M_1	M_2	M_3	M_4
Austria	0.390	0.260	0.210	0.140	0.042	0.958	0.000	0.000
Czech Republic	0.433	0.217	0.233	0.117	0.001	0.999	0.000	0.000
France	0.300	0.135	0.390	0.175	0.001	0.998	0.000	0.000
Hungary	0.392	0.083	0.433	0.092	0.498	0.502	0.000	0.000
Italy	0.174	0.106	0.446	0.274	0.001	0.994	0.000	0.005
Poland	0.297	0.153	0.363	0.187	0.000	0.469	0.000	0.531
Portugal	0.384	0.096	0.416	0.104	0.367	0.633	0.000	0.000

For the trend coefficients γ, no radical changes between the prior and posterior distributions could be observed, with the exception of Poland, for which, however, γ played a different role owing to the presence of the logistic trend (see Sections 3.3 and 4.4). In the Polish case, a beta(20,2) prior for γ was applied, strongly concentrated just above zero, to ensure a smooth passage from the lower to the upper asymptote of the logistic function. Finally, the RW constants (drift terms) c have been either identified by the data themselves from vague prior distributions or, in exceptional cases (the Czech Republic and Italy), the data slightly modified the priors upwards. In such instances, the priors were concentrated around 0 and assumed to follow normal distributions $N(0, 0.05^2)$. Under this assumption, it was highly unlikely *a priori* (with probability less than 0.05) that the average annual change of immigration flows would exceed $\pm 10\%$. Both exceptions concerned countries with exploding migration trends. The future levels of immigration were thus kept at reasonable levels by applying very strong prior assumptions, which under other circumstances could seem somewhat artificial.

5.2. *Migration forecasts until 2025 for selected European countries*

The forecasts that were prepared in this study have been ultimately obtained either by using the RW models M_2 having the highest probability *a posteriori* (for Austria, the Czech Republic, France and Italy) or by averaging of forecasts that were provided by various models (models M_2 and M_1 for Hungary and Portugal, and models M_2 and M_4 for Poland). For illustration, this section presents results for four countries, each of them representing one type of outcome: Austria (models M_2), and Portugal and Hungary (mixtures of models M_2 and M_1, in various proportions), as well as Poland (models M_2 and M_4). Although each of these countries represents a different pattern of migration (one western European, one southern European and two central European), the empirical results do not allow any regional generalization or typology of patterns to be made.

Throughout this section, the predictions are presented in terms of central tendencies, which are medians from the respective predictive distributions. As mentioned in Section 3.4, location statistics, such as medians or quantiles, are much more robust than the moment-based statistics, e.g. means or standard deviations. Thus, here the spans of uncertainty are based on symmetric quantiles from the predictive distributions. Table 4 presents the results for the median trajectories, as well as for the upper and lower quartiles. The forecasts are also illustrated in Fig. 1, with 50% intervals depicted by using the darkest shading, and the lighter shading corresponding to 80% and 90% predictive intervals.

Table 4. Predictive medians and 50% intervals: Austria, Hungary, Poland and Portugal

Year	*1st quartile*	*Median*	*3rd quartile*	*Year*	*1st quartile*	*Median*	*3rd quartile*
(a) Austria				*(b) Hungary*			
2007†	—	106905	—	2006†	—	21520	—
2008	93901	111302	131926	2007	17086	21311	26635
2009	90219	115844	148747	2008	15899	21354	28567
2010	87553	120572	164391	2010	14574	21440	31571
2015	80017	145801	265667	2015	12810	22026	37049
2020	74608	176310	416649	2020	11430	22697	42193
2025	69564	215346	660003	2025	10280	23156	46630
(c) Poland				*(d) Portugal*			
2007†	—	14995	—	2006†	—	30727	—
2008	13849	16075	18883	2007	23156	30031	38949
2009	13836	17257	21982	2008	20889	29733	43045
2010	13989	18527	25336	2010	18583	30031	50514
2015	15522	26370	47572	2015	16220	33190	70263
2020	17730	37421	86682	2020	14750	36680	92967
2025	20455	53104	157945	2025	13413	40135	120572

†The last observation in the sample (see the sources in Table 1).

Fig. 1 indicates that for Austria, with respect to the central (median) tendency, the total yearly volume of immigration would still increase, and the related 50% predictive intervals will visibly broaden. The forecasts for Hungary, in turn, depict the expectations of quite steady developments of future inflows, at least in terms of median trajectories and the 50% intervals. The results for Poland also seem moderate, albeit steadily increasing, with relatively narrow, though widening, predictive intervals. The median results that were obtained for Portugal indicate that immigration is expected first to decrease slightly (by around 2009), and then to increase continuously until the end of the forecast period, with visibly increasing assessments of uncertainty. In all cases, the results seem plausible from a demographic point of view.

From the results for the remaining countries the following conclusions can be drawn. For the Czech Republic, the expectations of relatively high uncertainty among the experts, coupled with the increasing trend that was observed in the data, suggest that especially the more distant future of immigration to the Czech Republic is really uncertain, as seen from the wide predictive intervals. Also, for France, the forecasted immigration flows are characterized by an increasing tendency and wide intervals, giving a clear indication of an increasing uncertainty towards the forecasting horizon. Finally, the predictions for Italy are characterized by extreme uncertainty. This is due to two major factors: a long steady increase in immigration that has been observed in the past and dramatic expectations of the experts resulting in their putting much weight on the explosive nature of the process (24% of answers). Such forecasts should not be interpreted in terms of precise values, which quickly become implausible, but at most in terms of orders of magnitude. Given the prior knowledge of the experts and data trends, the results can be seen as mainly indicating the extremely high degree of uncertainty with respect to the future migration inflows to Italy.

In general, the forecast results that were obtained for the seven countries under study can be summarized as follows. In all cases, migration appeared to be much more likely to be represented by non-stationary processes—at best RWs—than by models that are equipped with AR features and trend. This outcome is not surprising, given the nature of the phenomena under study, migration probably being the most uncertain component of population dynamics. In all instances, it

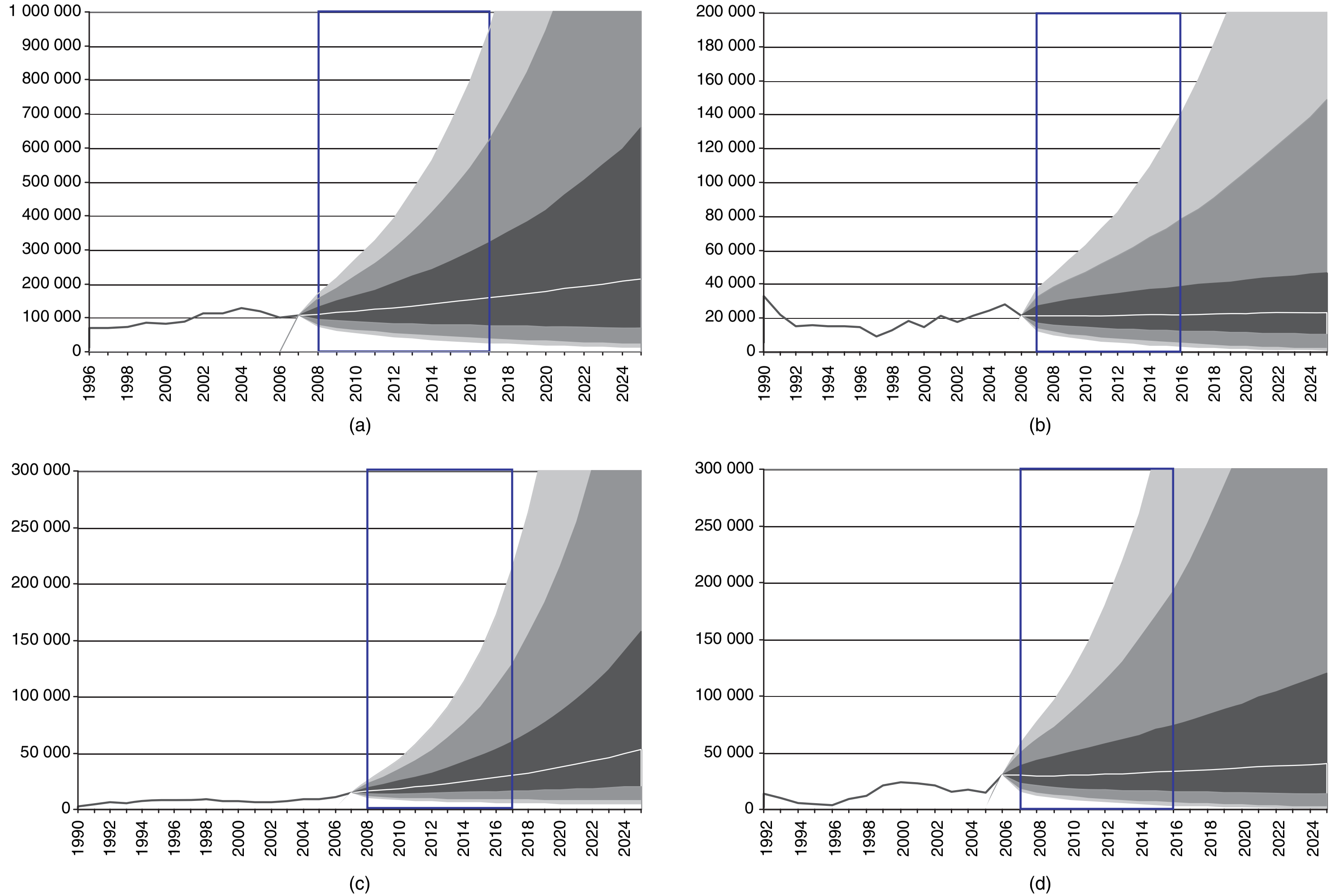

Fig. 1. Forecasts of immigration to (a) Austria, (b) Hungary, (c) Poland and (d) Portugal up to 2025 (the data are as in Table 1; the frame indicates a 10-year forecast horizon): ▢, 90%; ▢, 80%; ■, 50% (with median)

must be borne in mind that the variables under study concern registered total immigration and thus do not take into account the counter-flow of emigrants from particular countries.

One of the features of the 'long memory' RW model is that it cumulates all the random shocks—or uncertainty (embodied in the error term; for technical details see for example Greene (2003), page 593)—from the starting point of the process. This accumulation of uncertainty results in the widening of the predictive intervals over time. This feature strongly limits the length of a sensible forecast horizon, which ideally, in our opinion, should cover one decade at most. After 10 years, the upper bounds of predictive intervals (80%, and in many cases also 50%) become in many cases too high to offer any meaningful information for the forecast users. For this reason, the interpretation of the forecasts that are yielded by the current research should be limited to the 10 years' horizon, and the remaining period should be used merely as an illustration of the increase in the uncertainty that is associated with migration over time. Interestingly, the postulate that demographers should not make migration forecasts for a longer period than 10 years ahead was made by J. Z. Holzer already half a century ago (Holzer, 1959). This postulate also resembles the later suggestions of Keyfitz (1981), who put general limits on plausible demographic forecasts at a maximum of 20 years ahead.

5.3. Sensitivity analysis: selected aspects

The Bayesian approach that was adopted for preparing the forecasts requires at least some basic sensitivity analysis of the results to the assumptions made. The complexity of the problem at hand, however, precludes such an analysis of every aspect of the exercise. Here, the results of three sensitivity analyses are presented, with respect to the data (for Italy), the time trend in the AR model (for France) and the effect of expert judgement on the variability of the future immigration processes (questions 4 and 5 of the Delphi survey).

For Italy, three sets of data were available:

(a) final data for 1981–2005,
(b) provisional data for 1981–2005, where the last two observations were provisional with possible overestimation of the figures, and
(c) data for 1981–2007, where the last four observations were provisional and were possibly affected by the regularization process (for details on the regularization process see Cangiano (2008)).

When comparing the forecasts from the first and second data sets, a slight increase in the median and interquartile ranges can be observed. In the case of the third data set in comparison with the first two, an increase in uncertainty, together with the shift in median, can be seen. In terms of parameters, the difference between the first two and the third data set is visible in the precision τ; the posterior distribution is slightly more concentrated and shifted towards zero. Hence, the expected precision is smaller, resulting in greater uncertainty of the forecasts.

For France, the robustness of the model fit to the data (and thus of the choice of the model by the Carlin–Chib procedure) against the inclusion of the trend in the AR(1) model was additionally assessed. Two sets of models M_1–M_4 were estimated, one of them with AR models M_1 and M_3 equipped with a linear trend, and the other set with no trend at all. The Carlin–Chib procedure was applied to both sets separately to find the best model, assuming the same prior probabilities for particular models. In both cases the best fit to the data, in terms of the posterior probabilities, was achieved by the RW–CV model, the probability of which being around 0.99. Moreover, the final results were insensitive to the starting model choice of the procedure, i.e. to the model in which the algorithm started to generate samples.

Finally, a sensitivity analysis was also carried out with respect to the operationalization of the expert knowledge for the construction of prior densities for precision parameters τ. The assumptions concerning precision were crucial for the assessment of the uncertainty of the forecasts; thus the analysis was performed for each country. Three different hyperparameterizations were employed:

(a) the original parameterization, as described in Section 4.4;
(b) an alternative parameterization with hyperparameters of the gamma distribution assuming that the expected value of the precision equals the value that was suggested by the experts, and that the relative standard error of precision equals the relative standard error of the expert answers to question 4;
(c) a more diffuse prior, here assumed to be $\Gamma(0.01, 0.01)$, implying rather vague assumptions about precision (Congdon, 2003).

As far as the precision parameters are concerned, the posterior densities resulting from these priors were very sensitive in the case of Austria, France and Poland, with differences observed for both the shape and the location parameters. For the Czech Republic and Hungary, only the shape parameters differed, whereas for Italy it was only the case for the location parameter. Nevertheless, in the Italian case the posteriors in (a) and (b) did not differ significantly from each other. Finally, in the case of Portugal, the posteriors were insensitive to the priors that were used.

Sensitivity of forecasts in terms of differences in the medians of the predictive distributions was observed predominantly for the Czech Republic, France and Italy. In turn, sensitivity of the predictive uncertainty was highest in the case of Austria and Poland. In most of the other countries except the Czech Republic, the forecasts were not very sensitive to the parameterization of expert knowledge in the form of the priors (a) and (b). For the Czech Republic, the results proved to be insensitive to any of the prior structures assumed. Generally, for all countries except Portugal, expert knowledge increased the predictive uncertainty. Fig. 2 shows examples of the sensitivity analysis for the Czech Republic, France, Poland and Portugal, using dark grey shading for the diffuse prior (c), light grey for (b) and the in-between shading for the original prior (a).

5.4. *Ex post errors one step ahead*

Since writing the first version of the paper, an additional year of observations became available. This allowed a simple *ex post* analysis of forecasting performance of various models. Table 5 illustrates the absolute errors of one-step-ahead forecasts, except for Italy, where the analysis was performed 2 years into the forecasting period, as the interim data were not available. An additional column indicates the errors of either the formally selected (usually from model M_2) or averaged forecasts (for Hungary, Portugal and Poland, see Section 5.2), with an indication whether the real value fell within, below or above the 50% predictive interval.

From Table 5 it can be seen that on average the smallest absolute errors were obtained for the RW models; except for Hungary and Poland the averaging of forecasts helped further to reduce the errors from model M_2. The highest errors were obtained for the AR(1)–SV models. There is no clear overall pattern of overprediction or underprediction, with actual values falling either within (three cases), below or above the 50% intervals (two cases each). Similar experiments have been conducted for forecasts of emigration rates based on truncated series (Bijak, 2008). Their results indicated that plain RW models can indeed outperform more complex models in terms of *ex post* errors.

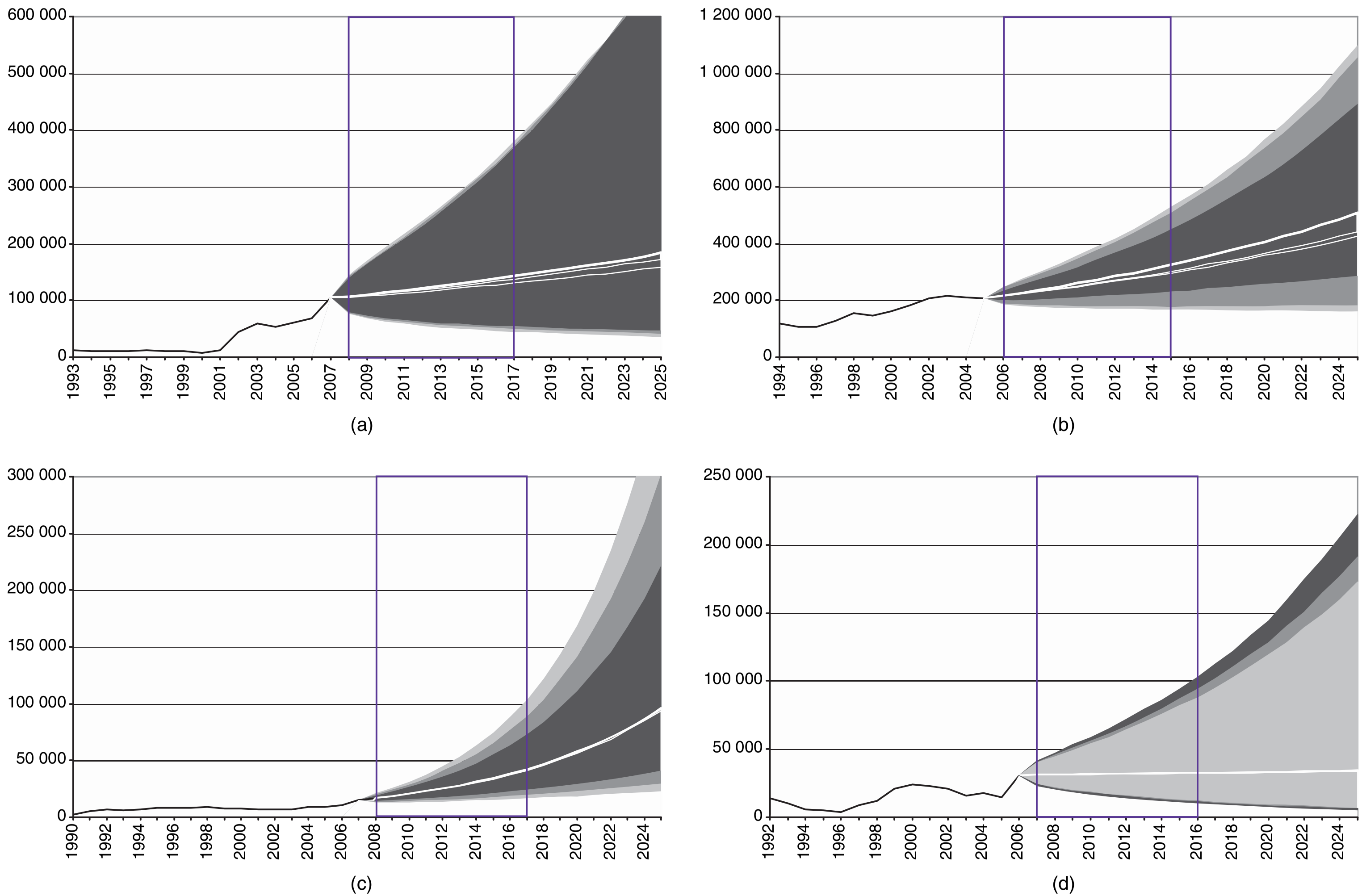

Fig. 2. Sensitivity of immigration forecasts for (a) the Czech Republic, (b) France, (c) Poland and (d) Portugal (the data are as in Table 1; the frame indicates a 10-year forecast horizon): □, medians; ▒, alternative parameterization; ▒, original parameterization; ■, diffuse parameterization

Table 5. Absolute *ex post* errors one step ahead for models M_1–M_4, and for the averaged or selected models†

Country and year	*Absolute errors for the following models (%):*				*Averaged or selected model*	
	M_1	M_2	M_3	M_4	*Absolute error*	*50% interval?*
Austria, 2008	1.1	1.1	0.1	1.1	1.1	Within
Czech Republic, 2008	69.5	37.4	56.5	38.8	37.4	Below
France, 2006	22.9	18.1	21.7	18.1	18.1	Below
Hungary, 2007	10.8	13.9	8.7	12.2	12.5	Within
Italy, 2007‡	48.8	43.4	53.2	48.8	43.4	Above
Poland, 2008	19.1	8.8	312.1	3.6	5.2	Within
Portugal, 2007	38.3	33.2	41.9	37.7	35.1	Above
Average for 7 countries	30.1	22.3	70.6	22.9	21.8	—

†Source: national statistical institutes (Austria, Czech Republic and Poland) and Eurostat, and own elaboration.
‡2 years ahead (data for 2006 were not available).

6. Concluding remarks and lessons learnt

In the methodological aspect, the forecasting exercise that was presented in this study was aimed at moving towards greater synergy and coherence in migration predictions, which could potentially be obtained by combining qualitative and quantitative information within a formal framework of Bayesian probabilistic models. To achieve this, two methods were applied jointly: a Delphi survey, which was intended to yield expert-based information *a priori*, as well as formal econometric and time series modelling in a subjectivist Bayesian setting. The key outcomes of this exercise can be summarized in the following points.

6.1. Migration is barely predictable

In line with the prior intuition of some migration researchers, the processes of population inflows under study appeared to be barely predictable (for more discussion, see for example Pijpers (2008)). Here, the data supported, almost always, non-stationary RW models over the AR models with trend. Treating non-stationary variables as if they were stationary seems a serious methodological flaw which can lead to very high forecast errors and misguide forecast users. A recent example is the forecast of immigration to the UK after the 2004 enlargement of the European Union, prepared for the British Home Office, which explicitly assumed stationarity of the process (Dustmann *et al.* (2003), pages 28 and 68). The inflow of immigrants from the new European Union member states was foreseen as 'between 5,000 and 13,000 immigrants per year up to 2010' (Dustmann *et al.* (2003), page 58), which underestimated true inflows by well over an order of magnitude (Institute for Public Policy Research, 2008).

6.2. Uncertainty matters

Given the above conclusion on the nature of migration, any attempt at its precise prediction in numerical terms is doomed to fail. Nevertheless, following what is currently becoming the state of the art in demographic forecasting, the predictive uncertainty can be embraced by using the stochastic approach (Keilman, 1990), which presents and quantifies the randomness in an explicit manner. Presenting deterministic forecasts instead can be seen as merely hiding the problem of predictive uncertainty, which, however, does not make the problem disappear. Presenting migration predictions in a deterministic fashion gives the decision makers a false sense of certainty which should definitely be avoided if informed decisions are to be made. Interestingly, as indicated by the simple *ex post* analysis, the RW processes or their combinations yielded, on average, the smallest relative errors 1 year ahead.

6.3. Expert knowledge matters—but not everywhere

In this study, the effect of judgmental information elicited from country-specific experts in the Delphi survey, aimed at supplementing weak sample-based information from short data series, was varied. On one hand, the *a priori* expert knowledge appeared to be very important in the estimation of the model parameters, especially with respect to the expected precision (or variability) of forecasts. On the other hand, the effect of subjective expertise was much less profound with respect to the model selection between the possible alternatives, i.e. to the determination of the nature of the processes under study. The dominant selection of RW models, often having quite low probabilities *a priori*, was to a large extent independent of the expectations of experts. It seems to suggest that the uncertain and barely predictable character of migration flows may be their inherent, more general feature, rather than just a characteristic of a particular forecasting model. This conclusion also coincides with the results of earlier studies (Bijak, 2008).

6.4. Forecasts with too long horizons are useless

Confirming some earlier suggestions (Holzer (1959); see Keyfitz (1981)), the sensible horizon of migration forecasts should be limited to 5–10 years at the most. This is due to the nature of the processes under study—if they are indeed non-stationary (which may often be so, as in the RW examples), then their uncertainty is increasing over time. After a decade, the predictive intervals become too large to offer meaningful information to the decision makers. A resulting challenge, remaining far beyond the scope of the current study, is what to do with the necessary migration assumptions in the population predictions that are prepared for a longer horizon, for example, of half a century.

For the forecast users, these conclusions imply that migration is indeed uncertain, but at least attempts can be made to assess more or less adequately the size of this uncertainty by using the statistical data enhanced by expert knowledge. The degree of variability of migration processes is itself an important piece of information. The decisions that are made on the basis of forecasts, such as those presented in this study, will strongly depend on specific problems that the decision makers must face. In particular, they must assess what will have more profound costs or consequences: the underestimation or overestimation of future migration flows?

As noted by Keilman (2008), page 148,

> 'In the short run, forecast uncertainty is not critical, at least for most forecast results at the country level. But in the long run, users should be aware of the costs attached to employing a forecast result that turns out to be too high or too low later on.... In case an immediate decision is required, they should check whether overpredictions are more costly than underpredictions, and base their decisions on such an assessment.'

Needless to add, irrespectively of the decision, the uncertain character of the forecasts must be borne in mind. In the public sphere, where some caution and degree of moderation should be required whenever the spending of public money is involved, these *caveats* are especially vital.

Acknowledgements

The research underlying this paper has been carried out in the project 'IDEA' ('Mediterranean and Eastern European Countries as new immigration destinations in the EU'), funded from the European Communities sixth framework programme (SSP CIS8-CT-2004-44446) and executed by 11 teams from nine European countries, co-ordinated by the Centre of Migration Research, University of Warsaw. More information on the project can be found at the Web site `www.idea6fp.uw.edu.pl`.

For assistance in elicitation and data collection we are very grateful to Heinz Fassmann and Ursula Reeger (Austria), Dušan Drbohlav, Dita Čermáková and Lenka Lachmanová (Czech Republic), Bernard Dinh and Giovanna Tattolo (France), Ágnes Hárs (Hungary), Frank Heins and Corrado Bonifazi (Italy), Dorota Kupiszewska (Poland) and João Peixoto (Portugal). As the study was to a large degree based on expert information, we wholeheartedly thank all the migration experts from the European countries that were relevant to the IDEA project, who decided to devote their time and judgement and participated in the Delphi survey.

In addition we are grateful to Izabela Koryś for her help in questionnaire design and testing. We also thank our colleagues from the Centre of Migration Research: credit for many remarks that helped to improve the earlier drafts go to Marek Okólski, Agata Górny, Izabela Grabowska-Lusińska, Magdalena Lesińska, Marta Mioduszewska, Monika Szulecka and Anna Żylicz, and the implementation of the survey owes much to Małgorzata Timoszuk. For modelling advice, we owe credit to Jacek Osiewalski, Mateusz Pipień and Błażej Mazur of the Chair of Econometrics, Cracow University of Economics. Finally, we are indebted to two reviewers and the Joint Editor for the comments that helped to improve the earlier versions of the paper. Any remaining errors and omissions are ours.

References

Armstrong, J. S. (1985) *Long-range Forecasting*, 2nd edn. New York: Wiley.

Bijak, J. (2008) Forecasting international migration in Europe: the Bayesian approach. *PhD Thesis*. Warsaw School of Economics, Warsaw.

Bijak, J. and Wiśniowski, A. (2009) Forecasting of immigration flows for selected European countries using expert information. *Working Paper 1/2009*. Central European Forum for Migration and Population Research, Warsaw. (Available from `www.cefmr.pan.pl/docs/cefmr_wp_2009-01.pdf`.)

Brücker, H. and Siliverstovs, B. (2006) On the estimation and forecasting of international migration: how relevant is heterogeneity across countries. *Empir. Econ.*, **31**, 735–754.

Cangiano, A. (2008) Foreign migrants in Southern European countries: evaluation of recent data. In *International Migration in Europe: Data, Models and Estimates* (eds J. Raymer and F. Willekens), pp. 89–114. Chichester: Wiley.

Carlin, B. P. and Chib, S. (1995) Bayesian model choice via Markov chain Monte Carlo methods. *J. R. Statist. Soc.* B, **57**, 473–484.

Casella, G. and George, E. (1992) Explaining the Gibbs sampler. *Am. Statistn*, **46**, 167–174.

Congdon, P. (2003) *Applied Bayesian Modelling*. Chichester: Wiley.

Dalkey, N. C. (1969) The Delphi method: an experimental study of group opinion. *Research Memorandum RM-5888-PR*. RAND, Santa Monica.

Dey, D. K. and Liu, J. (2007) A quantitative study of quantile based direct prior elicitation from expert opinion. *Bayesn Anal.*, **2**, 137–166.

Drbohlav, D. (1996) The probable future of European east-west international migration—selected aspects. In *Central Europe after the Fall of the Iron Curtain; Geopolitical Perspectives, Spatial Patterns and Trends* (eds F. W. Carter, P. Jordan and V. Rey), pp. 269–296. Frankfurt: Lang.

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

Gigerenzer, G. (1994) Why the distinction between the single event probabilities and frequencies is important for psychology (and vice-versa). In *Subjective Probability* (eds G. Wright and P. Ayton), pp. 129–161. Chichester: Wiley.

Goldstein, D. G. and Taleb, N. N. (2007) We don't quite know what we are talking about when we talk about volatility. *J. Prtfol. Mangmnt*, **33**, 84–86.

Goodwin, P. and Wright, G. (1998) *Decision Analysis for Management Judgment*, 2nd edn. Chichester: Wiley.
Gorbey, S., James, D. and Poot, J. (1999) Population forecasting with endogenous migration: an application to trans-Tasman migration. *Int. Regl Sci. Rev.*, **22**, 69–101.
Greene, W. H. (2003) *Econometric Analysis*, 5th edn. Upper Saddle River: Prentice Hall.
Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999) Bayesian model averaging: a tutorial. *Statist. Sci.*, **14**, 382–417.
Holzer, J. Z. (1959) *Prognoza Demograficzna Polski na Lata 1960–1975 wg Województw*. Warsaw: Polskie Wydawnictwo Gospodarcze.
Institute for Public Policy Research (2008) Floodgates or turnstiles?: post-EU enlargement migration flows to (and from) the UK. Institute for Public Policy Research, London.
Kadane, J. B. and Wolfson, L. J. (1998) Experiences in elicitation. *Statistician*, **47**, 3–19.
Keilman, N. (1990) *Uncertainty in National Population Forecasting: Issues, Backgrounds, Analyses, Recommendations*. Amsterdam: Swets and Zeitlinger.
Keilman, N. (2008) European demographic forecasts have not become more accurate during the past 25 years. *Popln Devlpmnt Rev.*, **34**, 137–153.
Keyfitz, N. (1981). The limits of population forecasting. *Popln Devlpmnt Rev.*, **7**, 579–593.
Kupiszewski, M. (2002) *Modelowanie Dynamiki Przemian Ludności w Warunkach Wzrostu Znaczenia Migracji Międzynarodowych*. Warsaw: Polish Academy of Sciences.
Kupiszewska, D. and Nowok, B. (2008) Comparability of statistics on international migration flows in the European Union. In *International Migration in Europe: Data, Models and Estimates* (eds J. Raymer and F. Willekens), pp. 47–71. Chichester: Wiley.
Lutz, W. and Goldstein, J. R. (2004) How to deal with uncertainty in population forecasting? *Int. Statist. Rev.*, **72**, 1–4.
Lutz, W., Sanderson, W. C. and Scherbov, S. (eds) (2004) *The End of World Population Growth in the 21st Century: New Challenges for Human Capital Formation & Sustainable Development*. London: Earthscan.
O'Hagan, A. (1998) Eliciting expert beliefs in substantial practical applications. *Statistician*, **47**, 21–35.
O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E. and Rakow, T. (2006) *Uncertain Judgements: Eliciting Experts' Probabilities*. New York: Wiley.
Osiewalski, J. (2001) *Ekonometria Bayesowska w Zastosowaniach*. Cracow: Cracow University of Economics.
Pijpers, R. (2008) Problematising the 'orderly' aesthetic assumptions of forecasts of East–West migration in the European Union. *Environ. Plann*g A, **40**, 174–188.
Rogers, A. (1990) Requiem for the net migrant. *Geogr. Anal.*, **22**, 283–300.
Rowe, G. and Wright, G. (1999) The Delphi technique as a forecasting tool: issues and analysis. *Int. J. Forecast.*, **15**, 353–375.
Rowe, G. and Wright, G. (2001) Expert opinions in forecasting: role of the Delphi technique. In *Principles of Forecasting: a Handbook of Researchers and Practitioners* (ed. J. S. Armstrong), pp. 125–144. Boston: Kluwer Academic.
Sabino, C. and Peixoto, J. (2008) Immigration, labour market and policy in Portugal: trends and prospects. *IDEA Project Report*. Research Centre in Economic Sociology and the Sociology of Organisations, Lisbon.
Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2007) *OpenBUGS Users Manual*, *Version 3.0.2*. Cambridge: Medical Research Council Biostatistics Unit.
Szreder, M. and Osiewalski, J. (1992) Subjective probability distributions in Bayesian estimation of all-excess-demand models. *Discussion Paper in Economics 92-7*. University of Leicester, Leicester.
Willekens, F. (1994) Monitoring international migration flows in Europe: towards a statistical data base combining data from different sources. *Eur. J. Popln*, **10**, 1–42.
Wiśniowski, A. and Bijak, J. (2009) Elicitation of expert knowledge for migration forecasts using a Delphi survey. *Working Paper 2/2009.* Central European Forum for Migration and Population Research, Warsaw. (Available from `www.cefmr.pan.pl/docs/cefmr_wp_2009-02.pdf`.)
Wiśniowski, A., Bijak, J., Kupiszewski, M. and Kupiszewska, D. (2010) Uncertain future of immigration in Europe: insights from expert-based stochastic forecasts for selected countries. In *Europe: the Continent of Immigrants: Trends, Structures and Policy Implications* (ed. M. Okólski). Warsaw: Centre of Migration Research. To be published.
Wu, Z., Huang, N. E., Long, S. R. and Peng, C.-K. (2007) On the trend, detrending, and variability of nonlinear and nonstationary time series. *Proc. Natn. Acad. Sci. USA*, **104**, 14889–14894.