

WP. 22
2 October 2007

ENGLISH ONLY

UNITED NATIONS STATISTICAL COMMISSION and STATISTICAL OFFICE OF THE EUROPEAN
ECONOMIC COMMISSION FOR EUROPE COMMUNITIES (EUROSTAT)

CONFERENCE OF EUROPEAN STATISTICIANS

Joint Eurostat/UNECE Work Session on Demographic Projections
(Bucharest, 10-12 October 2007)

Agenda item 8: Specific projection issues

**BAYESIAN MODEL SELECTION IN FORECASTING INTERNATIONAL MIGRATION:
SIMPLE TIME SERIES MODELS AND THEIR EXTENSIONS**

Invited Paper

Submitted by Poland¹

Abstract

The paper aims to present selected ideas concerning the application of Bayesian time series analysis to international migration forecasting. In the methodological part, the discussion focuses on the Bayesian framework for formal selection of forecasting models, which is one way of accounting for the uncertainty of model specification. In this approach, the subjective *a priori* knowledge of the researcher concerns not only the parameters, but also particular models as such. Both types of prior beliefs are subsequently transformed into knowledge *a posteriori* on the basis of information obtained from the sample of observations.

The theoretical discussion is illustrated in the paper by an empirical example of forecasts of both-way migration flows between Poland and Germany for 2005–2015, based on the aggregate data series from German population registers. The analysis covers three sets of forecasting models: simple stochastic processes – sub-models of ARMA(1,1), extensions of an AR(1) model to simplest cases with non-constant conditional variance, as well as propositions assuming a linear analogy to post-accession migratory developments in countries that joined the European Community earlier (Portugal and Spain). In each case, the outcome of the formal model selection in the Bayesian framework allows for the identification of models strongly supported by the data at hand. The Bayesian framework also enables to interpret the results with respect to uncertainty of the forecasted phenomena in coherent, probabilistic terms.

¹ Prepared by Jakub Bijak, Central European Forum for Migration and Population Research (j.bijak@cefmr.pan.pl).

1. Introduction

The paper aims to present selected ideas concerning possible applications of Bayesian time series analysis in forecasting international migration among several European countries. The Bayesian paradigm ensures the formality of inference, while allowing to include the *a priori* expert judgement in the analysis, alongside with the observations. Hence, the former can supplement the data-based information for small samples characterising many time series of within-European migration. Such a research problem fits into the methodological framework of stochastic population forecasts, where uncertainty assessment forms a key issue (Keilman, 1990; Lutz, Sanderson and Scherbov, 2004), only that within the Bayesian approach probability is defined in subjective terms, as a measure of belief.

Additionally, the paper aims to contribute to the debate on simplicity and complexity of models used in population forecasting (e.g., Smith, 1997) through applying formal model selection techniques. Such methods constitute one possible way of assessing and controlling the uncertainty of model specification.

In the first, methodological part (Section 2), the discussion focuses on the Bayesian framework for a formal selection of forecasting models. The theoretical background is illustrated in Section 3 by an empirical example of forecasts of both-way migration flows between Poland and Germany for 2005–2015, for three alternative classes of models. Section 4 briefly touches upon the issue of robustness of forecasts against selected changes in the assumed prior distributions. Finally, in Section 5, the main conclusions from the analysis are offered.

2. Bayesian methodology of formal model selection

In the Bayesian approach to formal model selection, the subjective *a priori* knowledge of the researcher concerns not only the parameters, but also particular models as such. Both types of prior beliefs are subsequently transformed into knowledge *a posteriori* on the basis of information obtained from the sample of observations. The result of the selection procedure consists thus of posterior probabilities of the choice of particular models, given the data.

In forecasting, the Bayesian approach consists in calculating a predictive probability distribution of the vector of future values of the variable of interest, \mathbf{x}^F , yielded by a particular forecasting model M on the basis of an observations vector \mathbf{x} . The estimation involves the *a posteriori* information about the vector $\boldsymbol{\theta}$ of the parameters of M given data \mathbf{x} , $p(\boldsymbol{\theta} | \mathbf{x})$, which is obtained from the Bayes' theorem as:

$$p(\boldsymbol{\theta} | \mathbf{x}) = p(\boldsymbol{\theta}) \cdot p(\mathbf{x} | \boldsymbol{\theta}) / p(\mathbf{x}). \quad (1)$$

In (1), $p(\boldsymbol{\theta})$ denotes subjective knowledge *a priori* about the parameters $\boldsymbol{\theta}$, embodied in a form of a probability distribution; $p(\mathbf{x} | \boldsymbol{\theta})$ is the likelihood of data given $\boldsymbol{\theta}$; and the normalising constant $p(\mathbf{x})$ is found by integrating $p(\mathbf{x} | \boldsymbol{\theta}) \cdot p(\boldsymbol{\theta})$ over the whole parameter space Θ . Given the above, the predictive probability distribution of \mathbf{x}^F is calculated as:

$$p(\mathbf{x}^F | \mathbf{x}) = \int p(\mathbf{x}^F, \boldsymbol{\theta} | \mathbf{x}) d\boldsymbol{\theta} = \int p(\mathbf{x}^F | \boldsymbol{\theta}, \mathbf{x}) \cdot p(\boldsymbol{\theta} | \mathbf{x}) d\boldsymbol{\theta}, \quad (2)$$

integration being carried out over the whole Θ (Zellner, 1971: 29).

Following Osiewalski and Steel (1993), let M_1, \dots, M_m denote m non-nested, mutually exclusive models, adding up to the whole model space \mathbf{M} . Assuming their *a priori* probabilities $p(M_i)$, which reflect the researcher's subjective intuition about plausibility of particular models for a specific forecasting task, the respective *a posteriori* probabilities, given the data vector \mathbf{x} , are obtained from the Bayes' Theorem as:

$$p(M_i | \mathbf{x}) = \frac{p(M_i) \cdot p(\mathbf{x} | M_i)}{\sum_{k \in \mathbf{M}} p(M_k) \cdot p(\mathbf{x} | M_k)}. \quad (3)$$

The models can be assumed *a priori* as equiprobable, with $p(M_i) = 1/m$ for all i , or follow the Occam's razor principle (meaning that "entities should not be multiplied unnecessarily")². In the latter case, simpler models are preferred over more complex ones, which can formally be reflected for example in prior probabilities such that $p(M_i) \propto 2^{-l_i}$, where the symbol ' \propto ' denotes proportionality and l_i is the number of parameters in the i -th model (*idem*).

In the model selection problems, the Bayes' Theorem is applied twice: to update the prior distributions of parameters θ_i for all models M_i given the data \mathbf{x} , following (1), and at the same time to obtain the posterior probabilities of particular models using (3). In order to accommodate the problem within the framework of the Gibbs sampling procedure, which is often used for numerical computations in Bayesian analyses, the *Model Choice via Markov Chain Monte Carlo* (MC³) algorithm of Carlin and Chib (1995) can be applied. In the current study, the method has been implemented within the WinBUGS 1.4 software environment (Spiegelhalter *et al.*, 2003).

The proposed procedure consists in an iterative sampling from the full conditional distributions for model-specific parameters θ_i and the model index μ , repeated sequentially until convergence to an ultimate solution is reached. The full conditional distributions are given by the following equations (Carlin and Chib, 1995: 475–477):

$$\begin{cases} p(\theta_i | \theta_{j \neq i}, \mu, \mathbf{x}) \propto \begin{cases} p(\mathbf{x} | \theta_i, \mu = i) \cdot p(\theta_i | \mu = i) & \text{for } \mu = i \\ p(\theta_i | \mu \neq i) & \text{for } \mu \neq i \end{cases} \\ p(\mu = i | \theta, \mathbf{x}) = \frac{p(\mathbf{x} | \theta_i, \mu = i) \cdot p(M_i) \cdot \prod_{j \in \mathbf{M}} p(\theta_j | \mu = i)}{\sum_{k \in \mathbf{M}} [p(\mathbf{x} | \theta_k, \mu = k) \cdot p(M_k) \cdot \prod_{j \in \mathbf{M}} p(\theta_j | \mu = k)]} \end{cases} \quad (4)$$

The model parameters θ_i are thus either sampled using a standard Gibbs procedure if $\mu = i$ or drawn from pre-defined linking densities ("pseudo-priors") $p(\theta_i | \mu \neq i)$ otherwise. The latter can be for example the preliminary estimates of model-specific posteriors $p(\theta_i | \mu = i, \mathbf{x})$ (*idem*). As in all Markov Chain Monte Carlo methods, the first S iterations of (4), until convergence, are discarded (the "burn-in" phase), while further N are used to estimate the posterior distributions of the parameters and the model index (Casella and George, 1992: 168).

² Merriam-Webster Online Dictionary: [www.m-w.com/dictionary/occam's razor](http://www.m-w.com/dictionary/occam's%20razor), as of 23.07.2007.

3. Empirical application to Polish–German migration forecasts

The empirical example offered in the current section concerns forecasts of both-way migration flows between Poland and Germany for 2005–2015, based on the aggregate data series from the German population registers and from the Eurostat. The variable under study is a log-transformed annual migration rate, that is, the number of migrants per 1,000 inhabitants of the sending country, denoted as $m_t = \ln(\text{Mig}_t / \text{Pop}_t \cdot 1,000)$. For all presented model selection procedures, the Occam's razor priors $p(M_i) \propto 2^{-l_i}$ have been assumed.

The presented analysis covers three distinct sets \mathbf{M} of forecasting models, here treated separately due to different sample vectors \mathbf{x} . Firstly, simple stochastic processes of different complexity and features – subclasses of ARMA(1,1), are studied. The sample is composed of 14 observations from the period 1991–2004, after socio-economic transformations in Poland and East Germany, to ensure an unchanging institutional setting. Five models are considered:

$$\begin{aligned}
 M_1: m_t &= c + \varepsilon_t \text{ [oscillations around a constant];} \\
 M_2: m_t &= c + m_{t-1} + \varepsilon_t \text{ [random walk with drift];} \\
 M_3: m_t &= c + \phi m_{t-1} + \varepsilon_t \text{ [AR(1) process];} \\
 M_4: m_t &= c - \theta \varepsilon_{t-1} + \varepsilon_t \text{ [MA(1) process];} \\
 M_5: m_t &= c + \phi m_{t-1} - \theta \varepsilon_{t-1} + \varepsilon_t \text{ [ARMA(1,1)].}
 \end{aligned} \tag{5}$$

The random term ε_t follows a Normal distribution $N(0, \sigma^2)$. Priors for the constants are diffuse, $c \sim N(0, 100^2)$, while the ones for the AR and MA components are more informative, with $\phi, \theta \sim N(0.5, 1^2)$, reflecting prior beliefs in a likely stationarity or time-reversibility of the relevant processes. Precision $\tau = \sigma^{-2}$ of the random term was assumed to follow a Gamma distribution $\Gamma(0.25, 0.25)$ for migration from Poland to Germany ($E\tau = 1, \text{Var}\tau = 4$, very low precision), and $\Gamma(4, 0.4)$ for the opposite-direction flows ($E\tau = 10, \text{Var}\tau = 25$, higher precision). These assumptions reflect prior beliefs in an uncertain character of migratory processes.

Secondly, extensions of an AR(1) model to cases with non-constant conditional variance are analysed, including ARCH(1), GARCH(1,1) and the simplest stochastic volatility (SV) models. In this case, the sample ranges over 20 years from 1985 to 2004, in an attempt to capture three periods of different institutional background and potentially also migration volatility: before, during and after the system transformation. Four models based on an autoregressive process AR(1), $m_t = c + \phi m_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim N(0, \sigma_t^2)$ are analysed, where:

$$\begin{aligned}
 M_6: \sigma_t^2 &= \sigma^2 \text{ [reference model with constant variance];} \\
 M_7: \sigma_t^2 &= k + \alpha \varepsilon_{t-1}^2 \text{ [AR(1)-ARCH(1)];} \\
 M_8: \sigma_t^2 &= k + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \text{ [AR(1)-GARCH(1,1)];} \\
 M_9: \ln(\sigma_t^2) &= k + \gamma \ln(\sigma_{t-1}^2) + \zeta_t; \zeta_t \sim \text{iin}(0, \rho^2) \text{ [simplest AR(1)-SV].}
 \end{aligned} \tag{6}$$

The priors for c and ϕ are the same as before, while for computational reasons for the remaining parameters it was assumed that: $\alpha, \beta, \gamma \sim \Gamma(10, 20)$, $1/\rho^2 \sim \Gamma(10, 1)$, as well as $k \sim \Gamma(1, 0.1)$. Thus, the proposed class encompasses the simplest processes either with deterministic (ARCH / GARCH) or stochastic (SV) change in conditional variance, compared with the constant-volatility reference model.

Finally, a proposition is examined (Kupiszewski, 1998) that there may exist an analogy between the Polish–German migration following the Poland's 2004 accession to the European

Union, and migratory developments in countries that joined the EEC earlier, in particular, Portugal and Spain. The sample covers 13 years (1992–2004), as such data on respective Iberian migration rates observed 18 years earlier were available. This analogy preserves the timing between the EEC/EU accession and the opening of German labour market for respective foreigners, in the case of Poles envisaged for 2011. Consider four models:

$$\begin{aligned}
 M_{10}: m_t &= c + \varepsilon_t \text{ [reference model with no analogy];} \\
 M_{11}: m_t &= c + a m_{t-18}^{PT} + b z_t + \varepsilon_t \text{ [analogy to Portugal, PT];} \\
 M_{12}: m_t &= c + a m_{t-18}^{ES} + \varepsilon_t \text{ [analogy to Spain, ES];} \\
 M_{13}: m_t &= c + a m_{t-18}^{IB} + b z_t + \varepsilon_t \text{ [analogy to both Iberian countries, IB].}
 \end{aligned}
 \tag{7}$$

In $M_{10} - M_{13}$, autoregressive random components are assumed, $\varepsilon_t \sim \text{AR}(1)$. A binary variable z_t in M_{11} and M_{13} removes an 1984 outlier for the Portuguese migration. The prior distribution for $a \sim \text{N}(0.5, 1^2)$ assumes a likely existence of such analogies.

The results of Bayesian model selection procedures for the three classes of models are summarised in Table 1 using the posterior probabilities $p(M_i|\mathbf{x})$. In a vast majority of cases, the selected models are the only ones, for which the data enhanced the prior beliefs. One important addition is the AR(1)-SV process for German–Polish migration: its posterior probability is over two times higher than the respective $p(M_i)$, also indicating significant data support to this model.

Table 1. Bayesian model selection results: posterior probabilities of various models

| Migration flow | Subclasses of ARMA(1,1) | | | | | Extensions of variance | | | | Models with analogy | | | |
|-----------------------|-------------------------|-------------|-------|-------|-------|------------------------|-------|-------|-------------|---------------------|----------|----------|----------|
| | M_1 | M_2 | M_3 | M_4 | M_5 | M_6 | M_7 | M_8 | M_9 | M_{10} | M_{11} | M_{12} | M_{13} |
| Poland → Germany | 0.42 | 0.29 | 0.14 | 0.11 | 0.03 | 0.12 | 0.10 | 0.05 | 0.74 | 0.69 | 0.07 | 0.14 | 0.10 |
| Germany → Poland | 0.23 | 0.49 | 0.16 | 0.08 | 0.03 | 0.71 | 0.02 | 0.00 | 0.27 | 0.88 | 0.01 | 0.09 | 0.02 |
| Model priors $p(M_i)$ | 0.31 | 0.31 | 0.15 | 0.15 | 0.08 | 0.50 | 0.25 | 0.13 | 0.13 | 0.40 | 0.20 | 0.20 | 0.20 |

Boldface denotes highest $p(M_i|\mathbf{x})$ in a given model class. Probabilities may not add up to unity due to rounding. Source: own elaboration in WinBUGS.

Among the ARMA(1,1) sub-models, high posterior probabilities of the random walk processes and oscillations, with those for AR(1) being just around the respective priors, confirm the research intuition about a hardly predictable character of migratory processes. With respect to similarity to Iberian migration, none of the presented analogies has found enough data support. As the outcome of an analysis for models with different conditional variance indicate high posterior probabilities of the stochastic volatility models, it seems that this is not only migration, but also its own variability that can be perceived as random and uncertain. The SV models indicate that although the Occam’s razor priors favour simpler models, the procedure can also point out to more complex ones if they have enough data support. In such way, the obtained solution to the “simplicity *versus* complexity” dilemma offered by the proposed method is not arbitrary, but based on the information provided by the statistical data at hand.

Selected results of forecasts yielded by particular models are presented in Table 2, which shows the predicted migration rates, $\exp(m_t)$ for 2005, 2010 and 2015, together with the jump-off observations for 2004. Along with the forecasted central tendencies, depicted by medians from the respective predictive distributions (2), uncertainty assessments are indicated using the

80-percent predictive intervals, based on the 10-percent and 90-percent quantiles from these distributions. Such an approach is widely applied in stochastic demographic forecasting³.

From a demographic viewpoint, the obtained median trajectories are plausible and indicate a stabilisation of trends rather than rapid changes. Limits of the 80-percent predictive intervals are also generally reasonable, except for the random walk and AR(1) processes, which indicates a likely non-stationary character of the latter.

Table 2. Forecasts of $\exp(m_t)$ for 2005, 2010 and 2015: various models

| Model | Forecasted $\exp(m_{2005})$ | | | Forecasted $\exp(m_{2010})$ | | | Forecasted $\exp(m_{2015})$ | | |
|---|-----------------------------|-------------|-------------|-----------------------------|-------------|-------------|-----------------------------|-------------|-------------|
| | 10% | 50% | 90% | 10% | 50% | 90% | 10% | 50% | 90% |
| Migration from Poland to Germany; $\exp(m_{2004}) = 3.65$ | | | | | | | | | |
| M_1: oscillation | 1.77 | 2.57 | 3.72 | 1.78 | 2.57 | 3.71 | 1.78 | 2.57 | 3.72 |
| M_2 : RWD | 2.49 | 3.64 | 5.33 | 1.23 | 3.62 | 10.71 | 0.69 | 3.61 | 18.80 |
| M_3 : AR(1) | 1.99 | 3.00 | 4.53 | 1.64 | 2.67 | 5.02 | 1.56 | 2.63 | 5.17 |
| M_4 : MA(1) | 1.80 | 2.80 | 4.21 | 1.72 | 2.60 | 3.94 | 1.72 | 2.60 | 3.96 |
| M_5 : ARMA(1,1) | 1.97 | 3.09 | 4.81 | 1.60 | 2.68 | 5.29 | 1.53 | 2.65 | 5.46 |
| M_6 : AR(1) | 2.33 | 3.63 | 5.66 | 1.53 | 3.60 | 9.01 | 1.37 | 3.60 | 10.73 |
| M_7 : - ARCH | 2.03 | 3.12 | 4.73 | 1.63 | 2.55 | 4.13 | 1.60 | 2.52 | 4.08 |
| M_8 : - GARCH | 2.43 | 3.45 | 4.73 | 1.78 | 2.99 | 5.58 | 1.66 | 2.89 | 6.15 |
| M_9: - SV | 2.64 | 3.42 | 4.41 | 1.74 | 3.31 | 7.18 | 1.60 | 3.28 | 8.40 |
| M_{10}: no analogy | 1.96 | 2.99 | 4.41 | 1.71 | 2.62 | 4.04 | 1.71 | 2.63 | 4.05 |
| M_{11} : Portugal | 1.99 | 2.99 | 4.42 | 1.79 | 2.93 | 4.88 | 1.77 | 3.24 | 6.22 |
| M_{12} : Spain | 1.97 | 2.94 | 4.31 | 1.77 | 2.76 | 4.34 | 1.85 | 2.96 | 4.80 |
| M_{13} : both jointly | 1.98 | 2.97 | 4.37 | 1.81 | 2.90 | 4.72 | 1.85 | 3.28 | 6.03 |
| Migration from Germany to Poland; $\exp(m_{2004}) = 1.27$ | | | | | | | | | |
| M_1 : oscillation | 0.72 | 1.00 | 1.39 | 0.72 | 1.00 | 1.39 | 0.72 | 1.00 | 1.38 |
| M_2: RWD | 0.91 | 1.25 | 1.71 | 0.48 | 1.18 | 2.90 | 0.28 | 1.12 | 4.36 |
| M_3 : AR(1) | 0.82 | 1.14 | 1.58 | 0.64 | 1.02 | 1.79 | 0.59 | 1.00 | 1.89 |
| M_4 : MA(1) | 0.77 | 1.09 | 1.52 | 0.70 | 1.01 | 1.45 | 0.70 | 1.01 | 1.45 |
| M_5 : ARMA(1,1) | 0.80 | 1.14 | 1.61 | 0.63 | 1.01 | 1.72 | 0.59 | 0.99 | 1.76 |
| M_6: AR(1) | 0.90 | 1.26 | 1.76 | 0.69 | 1.24 | 2.36 | 0.63 | 1.24 | 2.71 |
| M_7 : - ARCH | 0.84 | 1.17 | 1.60 | 0.68 | 1.02 | 1.61 | 0.65 | 1.00 | 1.61 |
| M_8 : - GARCH | 0.87 | 1.19 | 1.59 | 0.64 | 1.04 | 1.79 | 0.58 | 1.02 | 1.89 |
| M_9 : - SV | 0.94 | 1.25 | 1.67 | 0.72 | 1.21 | 2.24 | 0.68 | 1.21 | 2.62 |
| M_{10}: no analogy | 0.80 | 1.12 | 1.55 | 0.69 | 1.01 | 1.48 | 0.69 | 1.01 | 1.48 |
| M_{11} : Portugal | 0.73 | 1.12 | 1.69 | 0.66 | 1.01 | 1.56 | 0.63 | 1.00 | 1.58 |
| M_{12} : Spain | 0.74 | 1.06 | 1.50 | 0.60 | 0.91 | 1.40 | 0.65 | 0.96 | 1.42 |
| M_{13} : both jointly | 0.71 | 1.05 | 1.54 | 0.61 | 0.94 | 1.45 | 0.69 | 1.03 | 1.53 |

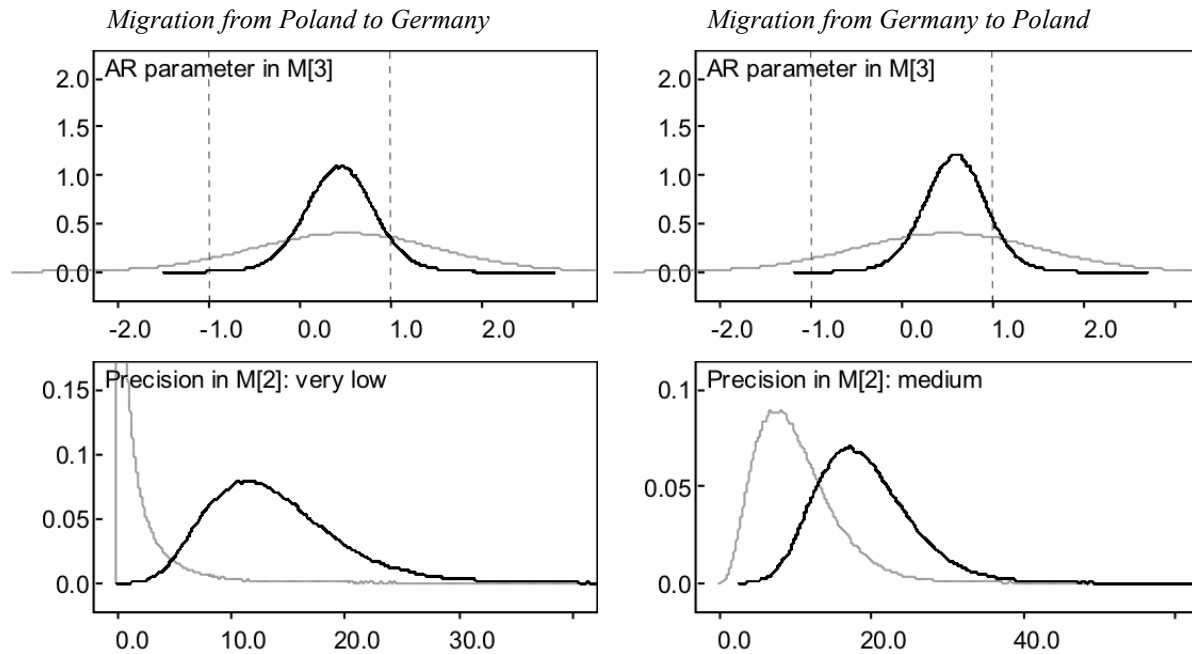
Boldface denotes models with highest posterior probabilities under the Occam's razor priors.

Source: own elaboration; data for 2004: Eurostat / Destatis.

³ Justifications are given for example by Lutz, Sanderson and Scherbov (2004: 37), including that "forecast distributions are themselves uncertain at the extremities. The 80 percent intervals are far more robust to the technicalities in the forecasting methodology than the 95 percent intervals."

Figure 1 illustrates, how the sample data modified the prior distributions for two selected parameters: the autoregression coefficient ϕ in model M_3 , and precision of the random term (τ) in the random-walk process M_2 . Especially in the latter case, a shift towards higher precision (lower variance) of the statistical noise can be observed.

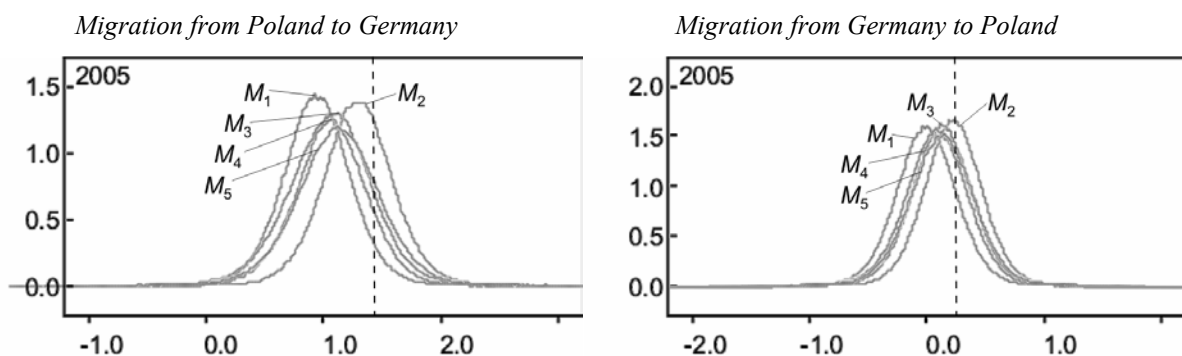
Figure 1. Comparison of prior and posterior distributions for selected parameters



Grey lines depict the distributions *a priori* and black ones – *a posteriori*. Dashed lines are the limits of unit circles.
Source: own elaboration in WinBUGS.

The predictive distributions of log-transformed migration rates yielded for 2005 by various models from the ARMA(1,1) class are illustrated in Figure 2, together with the actual observations for that year. In both cases the highest absolute *ex-post* errors have been obtained for the constant models (M_1), while the lowest ones – for the random walks (M_2). For migration from Germany to Poland a yet smaller error was yielded by the AR(1) model M_6 .

Figure 2. Forecasts of m_t for 2005 yielded by various models from the ARMA(1,1) class, compared with the actual observations



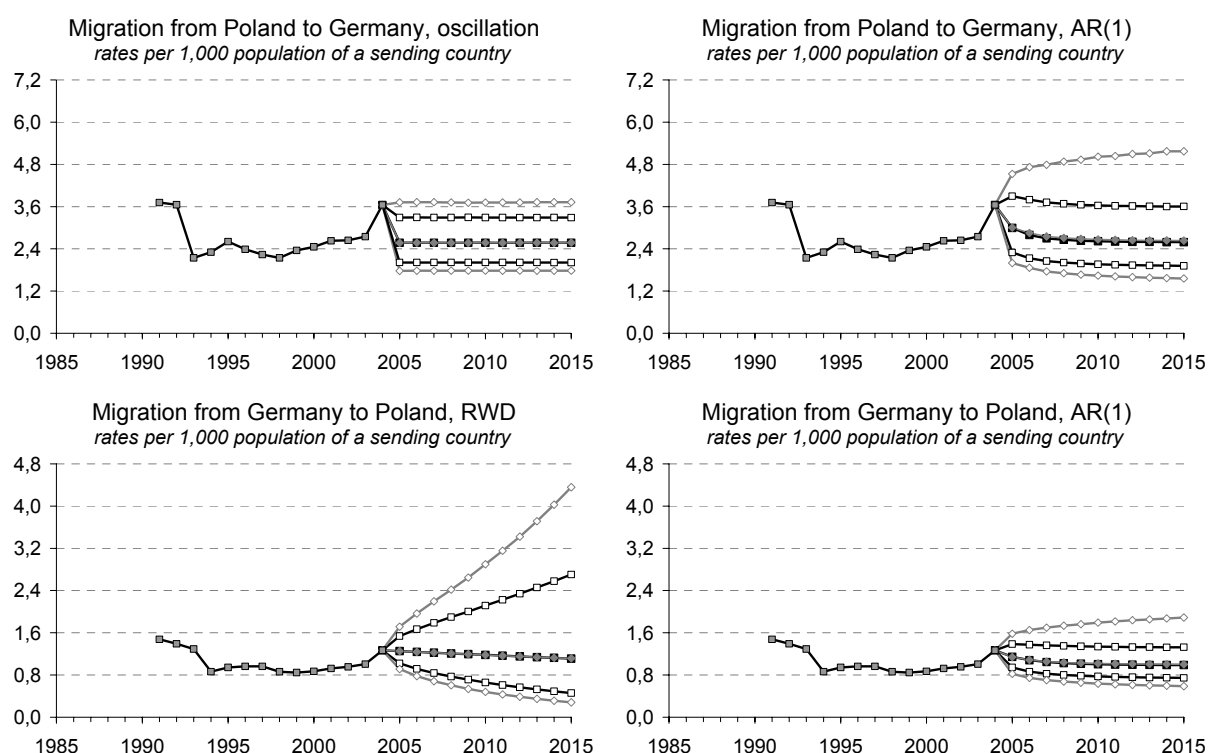
Dashed lines indicate the 2005 values of log-transformed migration rates, with $\exp(m_{PL-DE}) = 4.17$, $\exp(m_{DE-PL}) = 1.28$.
Source: own elaboration in WinBUGS, data for 2005: Eurostat/NewCronos.

4. Sensitivity of forecasts to selected changes in prior distributions

As Bayesian inference involves the use of judgemental prior distributions, it is important to assess the robustness of the outcome against changes in the latter. One possibility is to compare the impact of applying various competing priors on the obtained posterior or predictive distributions. In the Bayesian theory, attention is focused on “non-informative” priors, carrying as little statistical information as possible (Jeffreys, 1961). As an alternative, in practical applications “hardly informative” (vague) distributions can be used, as for example Normal $N(0, D^2)$ for structural parameters, where D denotes an arbitrary large number, or Gamma $\Gamma(a, a)$ for precision parameters, with a carrying small values (Congdon, 2003: 2–3, 21).

In this paper, the analysis is limited to two aspects of sensitivity to changes in the prior distributions: for the parameters and models, both treated separately. Firstly, the robustness of forecasts yielded by a handful of models is assessed, that is, by a random walk with drift (RWD) for migration from Germany to Poland, oscillations around the constant for flows from Poland to Germany, and additionally for AR(1) in both cases. In order to obtain “hardly informative” priors for the parameters mentioned above, it has been assumed that $D = 100$ and $a = 0.001$. The results for the selected models are presented in Figure 3. Black lines depict trajectories for the medians and the 80-percent predictive intervals for the respective migration rates, $\exp(m_t)$, yielded for 2005–2015 under vague prior distributions, while grey lines – trajectories obtained under more informative priors, the same as defined in Section 3.

Figure 3. Forecasts of $\exp(m_t)$ for 2005–2015 under vague and informative priors



Grey lines denote medians, as well as 10-percent and 90-percent quantiles from the predictive distributions obtained under the informative priors for parameters, while black lines – under the vague (“hardly informative”) ones.

Source: own elaboration; data until 2004: Eurostat / Destatis.

The forecasts yielded by particular models are visibly not robust against the suggested changes in priors, especially in the context of uncertainty assessments. Due to the shortness of

time series, this outcome is consistent with the expectations. Nonetheless, the low-precision assumption is a natural premise in migration research, given the uncertain nature of the processes under study. In several instances the 80-percent predictive intervals obtained under vague priors are narrower than the intra-sample variability, and thus hardly realistic. In turn, in the case of the RWD model for migration from Germany to Poland, the intervals under informative priors are too wide, what can indicate that the prior beliefs in relatively low precision were too pessimistic. In all cases, under very modest sample-based information, the role of *a priori* judgement in migration forecasts seems crucial, especially for the *ex-ante* assessments of forecast errors.

Finally, robustness to model priors has been evaluated on the example of the ARMA(1,1) class, by assuming $p(M_i) = 0.2$ as an alternative to the Occam's razor distribution used before. For migration from Poland to Germany, the procedure yielded $p(M_1|\mathbf{x}) = \mathbf{0.32}$, $p(M_2|\mathbf{x}) = 0.22$, $p(M_3|\mathbf{x}) = 0.21$, $p(M_4|\mathbf{x}) = 0.16$ and $p(M_5|\mathbf{x}) = 0.10$, while for flows in the opposite direction, $p(M_1|\mathbf{x}) = 0.17$, $p(M_2|\mathbf{x}) = \mathbf{0.37}$, $p(M_3|\mathbf{x}) = 0.24$, $p(M_4|\mathbf{x}) = 0.12$ and $p(M_5|\mathbf{x}) = 0.10$. Although the selected models (respectively, oscillations and RWD) are the same, the posterior probabilities are slightly different than before. Such an outcome can be also attributed to small sample sizes.

Clearly, the presented basic sensitivity assessments do not exhaust all possible analytical options, which are worth addressing in a separate study focusing for example on the robustness of the outcomes of the model selection procedures against changes in the prior distributions for the parameters and models jointly.

5. Conclusion

The outcome of formal model selection in the Bayesian framework allows for identifying models with relatively highest data support. Results yielded by the selected models, in terms of predictive distributions and intervals, enable in turn to assess the uncertainty of forecasts of the variables under study. This is especially important in predicting international migration flows, given a hardly determinate nature of the processes in question.

Moreover, the Bayesian paradigm allows to incorporate prior expert knowledge into migration forecasts in a formal way, as suggested by Willekens (1994). This seems to be another advantage of the proposed approach in the context of predicting international migration within Europe, where in many cases only short series of data are available, carrying weak sample-based statistical information. As indicated by the analysis of the robustness of forecasts against selected changes in the priors, without judgemental assumptions about the low precision of the random terms, the *ex-ante* predictive intervals would be in many cases implausibly narrow.

The empirical examples also indicate that for Polish–German migration, among the three classes of models under study preference was given either to simple, unstructured processes like random walks or oscillations, or to models without any historical analogies, or finally to models with stochastic conditional variance (SV). These results support a research intuition about hardly predictable nature not only of international migration as such, but also of its uncertainty characteristics, related to heavy tails of the random term distributions. Notwithstanding, the key methodological conclusion from the presented analysis is a full support for including in migration forecasts formal assessments of uncertainty on various levels, including model specification, to which the Bayesian approach provides probabilistically-coherent tools.

Acknowledgements

Paper prepared within the research grant no. 03–34 of the Foundation for Population, Migration and Environment (BMU-PME) from Zurich, first presented during the conference “Dynamic Econometric Models” in Toruń, Poland, 4–6 September 2007 under the title *Bayesian Forecasting of International Migration in Europe: Selected Methodological Approaches*. The author is truly indebted to Prof. Jacek Osiewalski of the Cracow University of Economics for several ideas underlying the paper, many helpful remarks and discussions.

References

- Carlin, B. P., and S. Chib (1995), Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society B*, 57 (3): 473–484.
- Casella, G., and E. George (1992), Explaining the Gibbs sampler. *The American Statistician*, 46 (3): 167–174.
- Congdon, P. (2003), *Applied Bayesian Modelling*. John Wiley, Chichester.
- Jeffreys, H. (1961), *Theory of Probability*, 3rd edition. Oxford University Press, Oxford.
- Keilman, N. (1990), *Uncertainty in national population forecasting: Issues, backgrounds, analyses, recommendations*. Swets & Zeitlinger, Amsterdam.
- Kupiszewski, M. (1998), Can the experience of other countries help forecast international migration after the EU accession? In: P. Korcelli (ed.), *Migration changes as a result of Poland’s accession to the European Union*. Committee of Spatial Organization, Polish Academy of Sciences, Warsaw: 67–78 [in Polish].
- Lutz, W., W. C. Sanderson, and S. Scherbov (2004), The End of World Population Growth. In: W. Lutz, W. C. Sanderson, and S. Scherbov (eds.), *The End of World Population Growth in the 21st Century: New Challenges for Human Capital Formation and Sustainable Development*. Earthscan, London: 17–83.
- Osiewalski, J., and M. F. J. Steel (1993), A Bayesian perspective on model selection. Manuscript, published in Spanish as: Una perspectiva bayesiana en selección de modelos. *Cuadernos Economicos*, 55 (3): 327–351.
- Smith, S. K. (1997), Further thoughts on simplicity and complexity in population projection models. *International Journal of Forecasting*, 13 (4): 557–565.
- Spiegelhalter, D. J., A. Thomas, N. G. Best, and D. Lunn (2003), *WinBUGS Version 1.4 Users’ Manual*. Medical Research Centre, Cambridge; www.mrc-bsu.cam.ac.uk.
- Willekens, F. (1994), Monitoring international migration flows in Europe. Towards a statistical data base combining data from different sources. *European Journal of Population*, 10 (1): 1–42.
- Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*. John Wiley, New York.
