Central European Forum for Migration Research

CEFMR ŚFBM

Środkowoeuropejskie Forum Badań Migracyjnych

Bayesian Model Averaging in Forecasting International Migration

Jakub Bijak

Central European Forum for Migration Research

European Population Conference 2006, Liverpool, United Kingdom, 21–24 June 2006

Session 63 (904): Projections

Project financed from the research grant of the Foundation for Population, Migration and Environment (BMU-PME) from Zurich

Plan of the presentation

1. Uncertainty in migration forecasting
2. Bayesian statistics: introductory notes
3. Bayesian model selection and averaging
4. Empirical examples of migration forecasts between Germany, and Italy, Poland and Switzerland by 2010
5. Concluding remarks

1. Uncertainty in migration forecasting

- Uncertainty is immanent in every forecast about the future
- Besides, the sources of uncertainty usually include:
  - Data quality and availability;
  - Selection of the forecasting model;
  - Subjectivity of the assumptions.
- How uncertainty is dealt with in migration forecasting?
  - Ignored (deterministic models)
  - Acknowledged, but not quantified (variant projections)
  - Acknowledged and quantified (stochastic forecasts, with uncertainty measured in terms of probability)

2. Bayesian statistics: introductory notes

The Bayes Theorem

Let θ denote unknown model parameters, and x – data (observations).

Then [Bayes, 1763; Laplace, 1812]:

$$p(\boldsymbol{\theta} \mid x) = \frac{p(x \mid \boldsymbol{\theta}) \cdot p(\theta)}{p(x)}$$

Posterior distribution → $p(\theta \mid x)$

Prior distribution → $p(\theta)$

Likelihood of the data, given θ ('traditional') → $p(x \mid \theta)$

Marginal likelihood of x (independent from θ) → $p(x)$

- **In Bayesian statistics, probability is interpreted subjectively, as a measure of belief**

2. Bayesian statistics: introductory notes

Bayesian forecasting

Let x denote observed (past) values, and x^F – forecasted (future) values

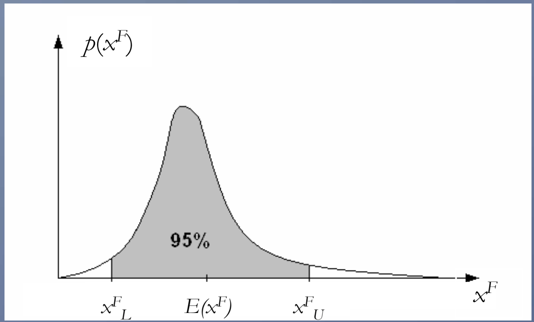
$$p(x^F \mid x) = \int_{\theta \in \Theta} p(x^F \mid x, \theta) \cdot p(\theta \mid x) d\theta$$

Result: predictive distribution

Sample-based predictions

Posterior distribution of θ

The outcome of a Bayesian forecast is the whole predictive distribution, and not a single value



3. Bayesian model selection and averaging

Bayesian Model Selection

- Let $M_1, \ldots, M_m$ be mutually exclusive (not nested) models adding up to the whole finite space of possible models, **M**
- Let $p(M_1), \ldots, p(M_m)$ be the models' prior probabilities, e.g.:
  - Flat prior (equal probabilities): $p(M_1) = \ldots = p(M_m)$
  - "Occam's razor" prior, favouring simpler models with smaller numbers of parameters, l_i: $p(M_i) \propto 2 \wedge (-l_i)$
- For forecasting, a model with highest posterior probability is selected on the basis of the Bayes Theorem:
  [Hoeting et al., 1999; Osiewalski, 2001]

$$p(M_i \mid x) = \frac{p(M_i) \cdot p(x \mid M_i)}{\sum_{k \in \mathbf{M}} p(M_k) \cdot p(x \mid M_k)}$$

3. Bayesian model selection and averaging

Bayesian Model Averaging

- Under the same assumptions, the forecasted vector $\mathbf{x}^{\mathrm{F}}$ given the data $\mathbf{x}$, averaged over the model space **M**, is:

$$\bar{p}(\mathbf{x}^{\mathbf{F}} \mid \mathbf{x}) = \sum_{i \in \mathbf{M}} p(M_i \mid \mathbf{x}) \cdot p(\mathbf{x}^{\mathbf{F}} \mid \mathbf{x}, M_i)$$

Averaged predictive distribution — Model posterior probability — Predictive distribution in the *i*-th model

- Rationale for use in migration (population) forecasts:
  - There is no evidence, whether simpler or more complex models perform better, but the forecast accuracy can be potentially improved by combining various forecasts [cf. Ahlburg, 1995; Smith, 1997]
  - Existing Bayesian migration forecasts are scarce [Gorbey et al., 1999]
  - There is a non-Bayesian example of an averaged migration forecast in the recent projections of the Eurostat [Lanzieri / EUROPOP, 2004]

4. Empirical examples of migration forecasts

Aim

To forecast long-term migration between Germany and three countries: Italy, Poland and Switzerland by 2010

Data

- Forecasted variable: logarithms of emigration rates per 1,000 population of the sending country (denoted by m_t)
- Data series for 1985–2004 (to / from Poland: 1991–2004)
- Sources of data: population – Eurostat; migration – data of a country with higher numbers (usually Germany)
- Population stocks include post-census adjustments

4. Empirical examples of migration forecasts

Models

- M_1: random oscillations around a constant
  $m_t = c + \varepsilon_t$
- M_2: random walk with drift
  $m_t = c + m_{t-1} + \varepsilon_t$
- M_3: autoregressive process AR(1)
  $m_t = c + \phi\, m_{t-1} + \varepsilon_t$ $\phi \neq 0,\ \phi \neq 1$
- M_4: moving average process MA(1)
  $m_t = c - \theta\, \varepsilon_{t-1} + \varepsilon_t$ $\theta \neq 0$
- M_5: autoregressive moving average process ARMA(1)
  $m_t = c + \phi\, m_{t-1} - \theta\, \varepsilon_{t-1} + \varepsilon_t$ $\phi \neq 0,\ \theta \neq 0$

4. Empirical examples of migration forecasts

Other remarks

- Sample distribution: Normal, $\varepsilon_t \sim N(0, \sigma^2)$
- Prior distributions for parameters:
  - Constants: diffuse, $N(0, 100^2)$ – hardly informative
  - Parameters of the AR / MA components: $N(0.5, 1^2)$
  - Variance (σ^2): Gamma(0.5, 0.5) – low precision assumed
- Estimation: numerical simulation using Markov chain Monte Carlo (MCMC), with 10,000 iterations in the burn-in phase and further 100,000 used in the estimation
- Software: WinBUGS 1.4 [Spiegelhalter *et al.*, 2003]
- Convergence assessment: visual inspection of quantiles

4. Empirical examples of migration forecasts

Model probabilities: prior and posterior (estimated)

Model (M_i)	M_1	M_2	M_3	M_4	M_5	Σ
Prior probabilities						
(A) Non-informative prior, $p(M_i) \propto const.$	0.200	0.200	0.200	0.200	0.200	1
(B) 'Occam's razor' prior, $p(M_i) \propto 2^{(-l_i)}$	0.308	0.308	0.154	0.154	0.077	1
Migration from Italy to Germany, $m_{IT\text{-}DE}$						
$p(M_i\|x)$, prior (A)	0.000	0.347	0.205	0.007	**0.441**	1
$p(M_i\|x)$, prior (B)	0.000	**0.616**	0.181	0.007	0.196	1
Migration from Germany to Italy, $m_{DE\text{-}IT}$						
$p(M_i\|x)$, prior (A)	0.000	0.249	**0.367**	0.018	0.366	1
$p(M_i\|x)$, prior (B)	0.000	**0.456**	0.356	0.016	0.171	1
Migration from Poland to Germany, $m_{PL\text{-}DE}$						
$p(M_i\|x)$, prior (A)	0.155	0.092	0.198	**0.313**	0.241	1
$p(M_i\|x)$, prior (B)	0.272	0.168	0.175	**0.275**	0.111	1
Migration from Germany to Poland, $m_{DE\text{-}PL}$						
$p(M_i\|x)$, prior (A)	0.079	0.207	**0.291**	0.171	0.252	1
$p(M_i\|x)$, prior (B)	0.135	**0.361**	0.249	0.147	0.108	1
Migration from Switzerland to Germany, $m_{CH\text{-}DE}$						
$p(M_i\|x)$, prior (A)	0.119	**0.283**	0.224	0.166	0.208	1
$p(M_i\|x)$, prior (B)	0.187	**0.431**	0.173	0.128	0.081	1
Migration from Germany to Switzerland, $m_{DE\text{-}CH}$						
$p(M_i\|x)$, prior (A)	0.000	**0.469**	0.311	0.003	0.217	1
$p(M_i\|x)$, prior (B)	0.000	**0.684**	0.232	0.002	0.081	1

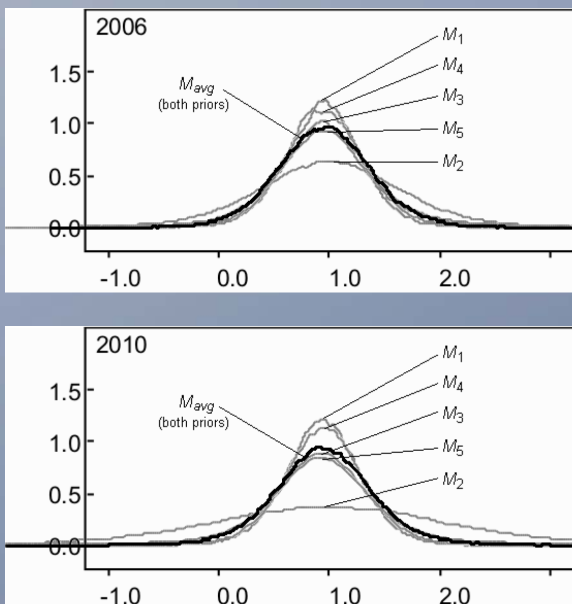
Central European Forum for Migration Research

Środkowoeuropejskie Forum Badań Migracyjnych

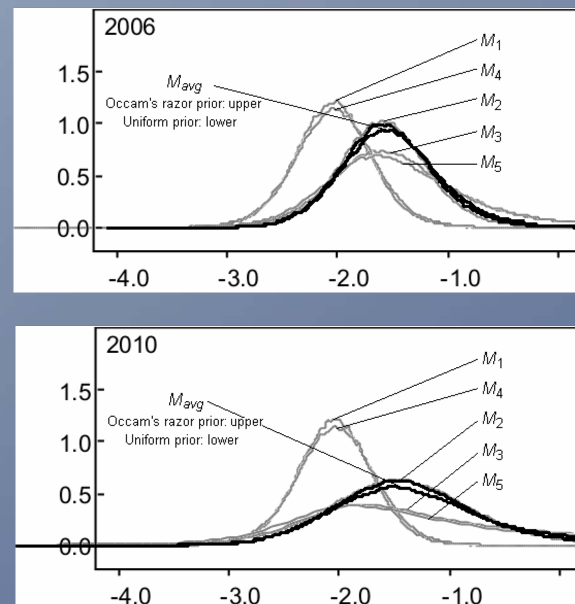
4. Empirical examples of migration forecasts

Results: predictive distributions (selected)

a) Logs of emigration rates Poland-Germany



b) Logs of emigration rates Germany-Switzerland



Central European Forum for Migration Research

Środkowoeuropejskie Forum Badań Migracyjnych

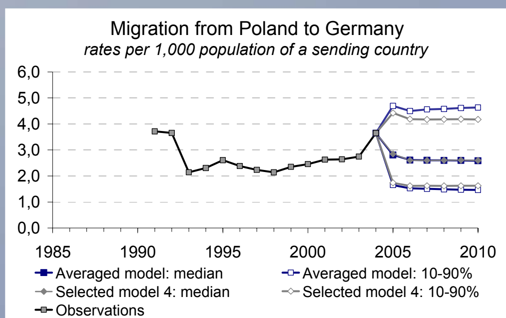
4. Empirical examples of migration forecasts

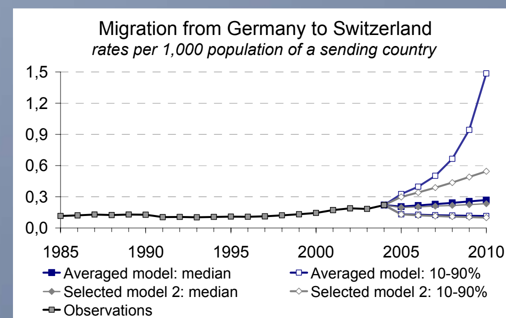
Results: quantiles from predictive distributions

a) Emigration rates Poland-Germany | *b) Emigration rates Germany-Switzerland*

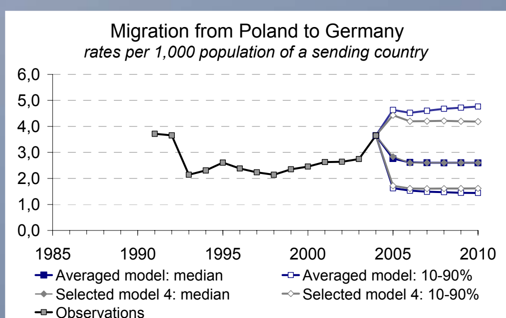
Model prior:

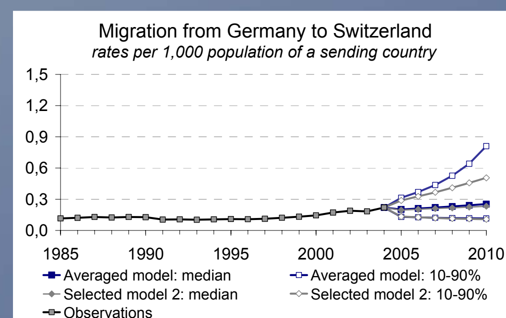
flat





Occam's razor





5. Concluding remarks

- **Bayesian model selection and averaging techniques allow to reduce uncertainty of model specification**
- General advantages of Bayesian approach in forecasting:
  - Inherent analysis of forecasts' uncertainty: predictive distributions
  - Formal and explicit incorporation of expert judgement in stochastic forecasts in the form of prior distributions
  - Methodology suitable for small samples (short series)
  - Subjective interpretation of probability allows for avoiding some interpretation problems, including the repeatable sample assumption
- The major disadvantage: computational complexity

  Solution: numerical methods (MCMC), software freely available
- Possible paths of further research:
  - A wider class of models, including other explanatory variables
  - Robustness against changes in prior distributions

Thank you for your attention!