Bayesian Framework for Forecasting International Migration: Selected Options

Jakub BIJAK, Central European Forum for Migration Research, Warsaw, Poland

Background and Rationale

- Uncertainty is immanent in all forecasts. In demography, the most difficult variable to predict is migration - a very complex process.
- In population forecasts, the main sources of uncertainty include:
  - Problems with data quality and availability
  - **Selection of the forecasting model** [1] ← **Current study**
  - Judgemental model assumptions
- How is uncertainty usually dealt with in migration forecasting?
  - Ignored (judgemental deterministic models)
  - Acknowledged, but not quantified (variant projections)
  - **Acknowledged and quantified (stochastic forecasts)** ← **Current study**

Aims and Scope

- **Aim 1:** To contribute to the advancement of migration forecasting methodology using standard tools of Bayesian statistics
- **Aim 2:** To prepare forecasts of long-term migration between Germany and three European countries (Italy, Poland and Switzerland) for 2005-2010
- **Data:** yearly migration rates (*MR*)
  MR = Migrants / Population x 1,000
  Economic variables: GDP *per capita* in PPP (Y), unemployment rates (U)

Sources: Eurostat, World Bank, UN ECE, NSIs (1985-2004, Poland: 1991-2004)

Option 1: Bayesian Model Selection and Averaging (Non-Nested Models)

Methodology [2]

Given data vector $\mathbf{x}$, forecast vector $\mathbf{x}^{\mathrm{F}}$, and model space $\mathbf{M}=\{M_i\}$, such that $\bigcup_{i\in\mathbf{M}}\{M_i\}=\mathbf{M}$ and $i\neq j \Leftrightarrow M_i \cap M_j=\varnothing$, one can obtain:

(1) Bayesian model selection criteria, based on the comparison of posterior probabilities of models: $p(M_i \mid \mathbf{x}) = \dfrac{p(M_i)\cdot p(\mathbf{x} \mid M_i)}{\sum_{k:M_k\in\mathbf{M}} p(M_k)\cdot p(\mathbf{x} \mid M_k)}$;

(2) Averaged forecasts: $\bar{p}(\mathbf{x}^{\mathrm{F}} \mid \mathbf{x}) = \sum_{k:M_k\in\mathbf{M}} p(M_k \mid \mathbf{x})\cdot p(\mathbf{x}^{\mathrm{F}} \mid \mathbf{x}, M_k)$.

Assumptions

Let $\mathbf{M}$ be the ARMA[1,1] model class with constraints, e_t~iid $\mathrm{N}(0,\sigma^2)$:

- M_1: $\ln(MR_t)=c+e_t$ (oscillations),
- M_2: $\ln(MR_t)=c+\ln(MR_{t-1})+e_t$ (random walk with drift),
- M_3: $\ln(MR_t)=c+\phi\ln(MR_{t-1})+e_t$, $\phi\neq 0$, $\phi\neq 1$ [AR(1)],
- M_4: $\ln(MR_t)=c-\theta e_{t-1}+e_t$, $\theta\neq 0$ [MA(1)],
- M_5: $\ln(MR_t)=c+\phi\ln(MR_{t-1})-\theta e_{t-1}+e_t$, $\phi\neq 0$, $\theta\neq 0$ [ARMA(1,1)].

Prior probabilities for models: (**a**) $p(M_i)$=0.2 for all i [flat prior], and (**b**) $p(M_i)\propto 2^{\wedge}(-l_i)$ [*Occam's razor* favouring M_i with less parameters l_i].

Prior distributions for parameters: c~$\mathrm{N}(0,100^2)$ [hardly informative], ϕ,θ~$\mathrm{N}(0.5,1^2)$ [likely stationary], $\tau=1/\sigma^2$~$\Gamma(0.5,0.5)$ [low precision].

Selected Results

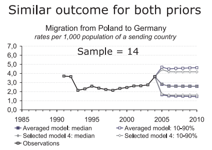
- Model choice is sensitive on prior probabilities $p(M_i)$, but often yields random walks
- Median forecasts for (**a**) and (**b**) are plausible and similar; prediction intervals can differ
- Weak fit (**small samples**): role of prior information

Estimated model posterior probabilities

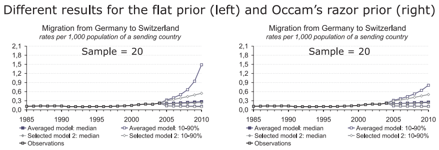
Model (M_i)	M_1	M_2	M_3	M_4	M_5
Prior probabilities					
(a) Flat prior	0.200	0.200	0.200	0.200	0.200
(b) *Occams razor* prior	0.308	0.308	0.154	0.154	0.077
Migration from Italy to Germany, $m_{IT\text{-}DE}$					
$p(M_i\|x)$, prior (a)	0.000	0.347	0.205	0.007	**0.441**
$p(M_i\|x)$, prior (b)	0.000	**0.616**	0.181	0.007	0.196
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
$p(M_i\|x)$, prior (b)	0.000	**0.456**	0.356	0.016	0.171
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
$p(M_i\|x)$, prior (b)	0.272	0.168	0.175	**0.275**	0.111
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
$p(M_i\|x)$, prior (b)	0.135	**0.361**	0.249	0.147	0.108
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
$p(M_i\|x)$, prior (b)	0.187	**0.431**	0.173	0.128	0.081
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
$p(M_i\|x)$, prior (b)	0.000	**0.684**	0.232	0.002	0.081

Sample MR forecasts: medians and 80% prediction intervals (MCMC, WinBUGS)

Similar outcome for both priors



Different results for the flat prior (left) and Occam's razor prior (right)



Option 2: Hybrid Bayesian-LR Nested VAR Modelling 'General to Specific'

Methodology [3]

Given the k-dimensional VAR(1) model M_0: $\mathbf{y_t}=\mathbf{c}+\mathbf{A}\mathbf{y_{t-1}}+\mathbf{e_t}$, $\mathbf{A}=[a_{ij}]_{k\times k}$, the significance of the impact of the i-th component of $\mathbf{y_{t-1}}$ ($y_{i,t-1}$) on the remaining k-1 components of $\mathbf{y_t}$ is tested by imposing restrictions $a_{ji}=0$ ($j\neq i$), defining a new model, M_i. Estimation of the parameters is Bayesian, while the selection is based on the log-likelihood ratio test:

$$\mathrm{LR}(M_0,M_i) = -2\,(\ln p(\mathbf{x}|M_0,\theta_0) - \ln p(\mathbf{x}|M_i,\theta_i)) \quad [\theta\text{: model parameters}]$$

Similarly, restrictions for an extended nested model structure $M_0 \supset M_i \supset M_{ij}$..., and for feedback effects ($a_{ij}=a_{ji}=0$, $j\neq i$), can be tested.

Assumptions

Let M_0 be such that $\mathbf{y_t}=[\ln(MR_t), \ln(Y_t^R/Y_t^S), \ln(U_t^S)]'$, where R and S indicate respectively the receiving and sending country.

The following models are considered: M_0, M_2, M_3, and M_{23} [AR(1)].

Error term is a 3-dimensional Gaussian white noise: $\mathbf{e_t}$~iid $\mathbf{N}(\mathbf{0},\mathbf{T})$, where $\mathbf{T}$ denotes the precision matrix.

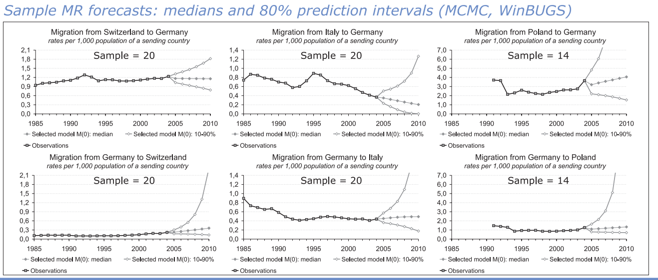
Prior distributions for parameters in M_0: c_i~$\mathrm{N}(0,\sqrt{1000}^2)$, a_{ii}~$\mathrm{N}(1,1^2)$, a_{12}, a_{13}~$\mathrm{N}(0.5,1^2)$ [from the migration *push-and-pull factors* theory], a_{21}~$\mathrm{N}(0,0.1^2)$, a_{31}~$\mathrm{N}(0,\sqrt{0.1}^2)$, a_{23}, a_{32}~$\mathrm{N}(0.1,\sqrt{0.5}^2)$ [weak relations], $\mathbf{T}$~Wishart($\mathbf{P}$,3), $\mathbf{P}=[p_{ij}]_{3\times 3}$, p_{ii}=1 and p_{ij}=0.005 for $i\neq j$ [low precision].

For M_2, M_3, and M_{23}, appropriate elements of $\mathbf{A}$ and $\mathbf{P}$ are set to 0.

Selected Results

- The LR tests did not reduce the initial models M_0 despite weak fits
- Median forecasts of *MR* are plausible, but 80% predictive intervals are mostly too wide and exploding ($\mathbf{y_t}$ close to 3D random walks)

Sample MR forecasts: medians and 80% prediction intervals (MCMC, WinBUGS)



Conclusions

- Both approaches account for the uncertainty of model specification, but in most models the errors of *MR* forecasts are implausibly high, due to: (**1**) small samples, (**2**) a likely random-walk (unpredictable) character of *MR*, and (**3**) the uncertainty of Y and U in the VARs
- Bayesian inference is suitable for forecasting migration based on short series, which are the only available data in most of Europe
- Possible paths of further research:
  - Use of fully Bayesian Lindley-type HPD tests for VAR restrictions
  - Wider classes of time series models; other explanatory variables
  - Analysis of robustness against changes in prior distributions

Notes and Acknowledgements

[1] Ahlburg DA. 1995. *Math. Popul. Stud.* **5** (3), 281-290.
[2] Hoeting JA, Madigan D, Raftery AE, Volinsky CT. 1999. *Stat. Sci.* **14** (4), 382-417.
[3] Hendry DF. 1995. *Dynamic Econometrics*. Oxford: OUP.

Acknowledgements:
Study prepared within the research grant 03-34 of the Foundation for Population, Migration and Environment (BMU-PME) from Zurich, with the support of the stipend for young scientists of the Foundation for Polish Science (FNP). I am grateful to Prof. Jacek Osiewalski and Prof. Marek Kupiszewski for their comments. The usual disclaimer applies.

Contact: ul. Twarda 51/55, 00-818 Warsaw, Poland, tel. +48 22 697 88 34, fax +48 22 697 88 43, e-mail: j.bijak@cefmr.pan.pl, web: www.cefmr.pan.pl

Central European Forum for Migration Research (CEFMR) is a research partnership of the Foundation for Population, Migration and Environment (BMU-PME), Institute of Geography and Spatial Organization of the Polish Academy of Sciences, and International Organization for Migration (IOM)