Forecasting Migration: A Model Averaging Approach

Jakob Gregor Zellmann[a1,4], **Jesus Crespo Cuaresma**[2], **Juan Caballero**[4], **Katharina Fenz**[4], **Teodor Yankov**[3], and **Amr Taha**

[1]Università Alma Mater Studiorum di Bologna, Italy
[2]Vienna University of Economics and Business, Austria
[3]University of Oxford, United Kingdom
[4]Max Planck Institute for Demographic Research, Germany
[5]World Data Lab, Austria

Abstract

Anticipating future migration dynamics is central to forming reasonable expectations of economic, demographic and social developments. However, the discussion around which forecasting methods can provide the most accurate projections of migration flows is still contested. In this paper we propose a model averaging approach that takes into account three non-causal models and a state-of-the-art gravity model of migration to exploit the strengths of all modelling techniques. Using OECD data on bilateral migration flows, we conduct a pseudo out-of-sample validation exercise to gain insights in the predictive power of the individual models and their combinations.

Keywords: Migration, forecasting, autoregressive modelling, gravity model, prediction validation, model averaging

[a]Corresponding author. Università Alma Mater Studiorum di Bologna, Italy *jakob.zellmann@unibo.it.*

1 Introduction

Migration processes significantly affect economic, demographic and social dynamics. Anticipating future changes in migration flows across countries is thus particularly important for the design of migration policy and understanding the effect of human mobility on socioeconomic outcomes. The discussion concerning adequate forecasting methods for (bilateral) migration flows is far from settled in the academic literature. Globally, only few governments share detailed information about inflows and outflows of migrants, and rigorous quantitative research in migration is thus restricted by data availability about migration patterns and the composition of migration flows. Such a problem limits the space of applicable methods and increases the uncertainty of migration forecasts. The existing quantitative literature has evolved around explaining and predicting migration patterns ranging from sub-national to global geographic scales using several methodological frameworks. The statistical specifications employed range from causal models, which aim to understand how changes in drivers of migration affect the outcome, to models that extrapolate migration patterns based on their past dynamics and persistence patterns, both of which come with strengths and weaknesses.

By identifying the link between push and pull factors of migration and observed migration movements, causal models help to understand and quantify the structural behaviour of migration patterns. The complexity of migration processes has prevented the evolution of a unified theory of global migration so far. Gravity models have historically provided an established approach to understanding migration, although their performance in terms of (in-sample) predictive ability has been heavily criticized (see Beyer and Lotze-Campen (2022)). Gravity models identify the connection between economic, socio-demographic and geographic variables, and migration flows (Ramos 2016) and thus allow for the quantitative assessment of scenario-based migration projections, which could be integrated into alternative future trajectories such as those put forward by the Shared Socioeconomic Pathways (Benveniste et al. 2021).

Instead of analysing empirically the push and pull factors driving migration patterns, (non-causal) autoregressive models for migration focus on extrapolating human mobility based on past trends and the stylized facts of their dynamics. Despite the comparatively good forecasting performance of those models (see e.g. Azose and Raftery (2015), Welch and Raftery (2022)), they are not able to generate scenario-based migration projections conditional on particular dynamics in the determinants of mobility. Furthermore, since these models do not help us understand drivers of migration and past shocks in migration patterns, which may be unlikely to appear in the future, can significantly affect the quality of predictions.

In this paper we present a systematic comparison of the out-of-sample predictive ability of existing models for bilateral migration flows and assess the potential improvements that can be obtained from aggregating predictions from specifications making use of model averaging methods (Kapetanios et al. 2006; Raftery et al. 1997). We utilise three non-causal models and a state-of-the-art gravity-type model and benchmark these specifications against each other to predict bilateral migration flows in a pseudo out-of-sample validation exercise. Our results suggest that among the individuals models a pooled autoregression approach is the best performing and a variant of the Bayesian flow model proposed by Welch and Raftery 2022 the worst performing model in terms of MAE and $RMSE$. That result holds irrespective of the forecasting horizon. Furthermore we find, that a gravity model approach performs best when it comes to forecast relative changes of migration flows. Combining the forecasts of the individual model by weights obtain by regressing the realized values on the forecast of the individual models, leads to a improvement of the prediction in terms of MAE. We conclude, that model averaging does improve predictability of bilateral migration flows. The following section describes the data and methods, section 3 presents the main results, and section 4 concludes.

2 Data & Models

Our models for explaining bilateral migration flows across countries and over time are based on the OECD's International Migration Database (OECD International migration database 2022) as source for mobility data. The data in OECD International migration database (2022) provides information about migration inflows from almost 200 origin countries to OECD member states. We only consider migration flows without missing values in the period for 2000 to 2022. This leads to a panel with 2196 cross sectional units and a temporal coverage of 23 years.

We consider three benchmark non-causal models to create bilateral migration flow forecasts: (i) simple country-pair specific autoregressive models, (ii) a pooled autoregression model with persistence parameters which are origin-specific, (iii) a specification in the spirit of Welch and Raftery (2022) that models country-specific emigration and spatial allocation probabilities by which the migration outflows are distributed among the possible destination countries. In the first specification under consideration, we model each individual time series of migration flows from a given origin country to a given destination country by and autoregressive approach. Denoting the migration flow from origin country i to destination country j in period t by $m_{i,j,t}$ the model reads as follows,

$$m_{i,j,t} = \mu_{i,j} + \gamma_{i,j} m_{i,j,t-1} + \varepsilon_{i,j,t}, \qquad \text{with} \qquad \varepsilon_{i,j,t} \overset{iid}{\sim} WN(0, \sigma_{i,j}). \tag{1}$$

We denote this specification as individual autoregression (IAR).

As in both data sets the number of cross sectional units is large in comparison to the number of observations within a time series, exploiting the information about dynamics across country pairs rather then utilizing exclusively the information provided by a single time series of bilateral migration flows appears as an potentially more efficient method to estimate the parameters of the model. We thus adjust the model described in equation (1), so that information across migration flows with the same origin country is shared in order to estimate the persistence parameter and the variance of the model. Formally, this specification is given by

$$m_{i,j,t} = \mu_{i,j} + \gamma_i m_{i,j,t-1} + \varepsilon_{i,t}, \qquad \text{with} \qquad \varepsilon_{i,t} \overset{iid}{\sim} WN(0, \sigma_i^2). \tag{2}$$

By exploiting the pooled structure of the data to estimate the parameters of the model, we lower the information requirements while keeping the model flexible enough to capture complex migration dynamics. We denote this specification as pooled autoregression (PAR).

In addition to the benchmark autoregressive approaches outlined above, we consider a model similar to that proposed by Welch and Raftery (2022). They argue that decomposing a model of bilateral migration flows into a model for outflows and a distribution of those outflows to destination countries increases forecasting performance. In this setting, bilateral migration flows depend on emigration probabilities for a given origin country ,$\delta_{i,t}$, the (time invariant) probability of spatial allocation of a migrant form origin country i to destination country j, $\pi_{i,j}$, and the population of the origin country, $P_{i,t}$. In such a methodological framework, the conditional expected value of the bilateral flow $m_{i,j,t}$ is given by

$$\mathbb{E}[m_{i,j,t}|\pi_{i,j}, \delta_{i,t}, P_{i,t}] = \pi_{i,j}\delta_{i,t}P_{i,t}. \tag{3}$$

To estimate the model and obtain draws form the unconditional probability distribution of $m_{i,j,t}$, we utilize a hierarchical Bayesian prior in a specification that assumes that the logarithmic outflow rates follow an autoregressive process of first order and that the probability vector of destination probabilities $\boldsymbol{\pi}_{i,\_} := (\pi_{i,1}, ..., \pi_{i,C})$ is time invariant. Denoting $\boldsymbol{m}_{i,\_,t} := (m_{i,1,t}, ..., m_{i,C,t})$ the vector of migration flows with origin country i, this specification implies

$$\textbf{Observation}\left\{\ \boldsymbol{m}_{i,\_,t} \quad \sim \mathcal{MN}(\delta_{i,t}P_{i,t}, \boldsymbol{\pi}_{i,\_})\right. \tag{4}$$

$$\textbf{Outflows}\begin{cases} \ln\delta_{i,t} & \sim \mathcal{N}(\alpha_i + \phi\ln\delta_{i,t-1}, \xi^2) \\ (\alpha_i, \phi_{i,j})^\top & \sim \mathcal{N}(0, \tau\boldsymbol{I}_2) \\ \xi_i^2 & \sim \mathcal{IG}(a, b) \end{cases} \tag{5}$$

$$\textbf{Destination Allocation}\begin{cases} \pi_{i,j} & = \exp\eta_{i,j}/\sum_{k\neq i}\exp\eta_{i,k} \\ \eta_{i,j} & = \ln\frac{\pi_{i,j}}{g(\pi_{i,1},...,\pi_{i,C-1})} \\ \eta_{i,j} & \sim \mathcal{N}(\nu_{i,j}, \psi_{i,j}^2) \\ \nu_{i,j} & \sim \mathcal{N}(0, \tau) \\ \psi_{i,j}^2 & \sim \mathcal{IG}(a, b) \end{cases} \tag{6}$$

where, $\mathcal{MN}(\cdot,\cdot)$ denotes the multinomial distribution and $g(\cdot, ..., \cdot)$ the geometric mean. Following Welch and Raftery 2022 we will refer to that model as Bayesian flow model (BFM).

The fourth model under consideration is a causal gravity-type model. Gravity models are an established tool to quantify migration patterns, as shown in the work of Karemera and Davis (2000), Cohen and GoGwilt (2008), Kim and Cohen (2010), Mayda (2010) and Cohen (2012), among others. The underlying idea is to link bilateral migration flows to socio-economic, demographic and geographical variables in origin and destination countries. Typically gravity models rely on the assumption that migration flows depend log-linearly on the regressors under consideration. Following that assumption gravity models are described by,

$$\ln m_{i,j,t} = \mathbf{Z}_{i,t}\boldsymbol{\theta} + \mathbf{Z}_{j,t}\boldsymbol{\phi} + \mathbf{X}_{i,j,t}\boldsymbol{\eta} + \varepsilon_{i,j,t}, \tag{7}$$

where origin-specific variables are summarized in the vector $\mathbf{Z}_{i,t}$, the destination-specific variables in $\mathbf{Z}_{j,t}$, the bilateral factors in $\mathbf{X}_{i,j,t}$ and $\varepsilon_{i,j,t}$ is an error term that is assumed to fulfill the usual assumption of linear models. To be consistent with the assumption of log-linearity $\mathbf{Z}_{i,t}$, $\mathbf{Z}_{j,t}$ and $\mathbf{X}_{i,j,t}$ include the logarithm of migration drivers.

In the specification of the gravity model applied in this paper we consider the total population, GDP per capita, infant mortality and the dependency ratio to be driving origin and destination specific factors for migration. As bilateral migration drivers we consider variables measuring whether origin and destination countries share a common border and whether they share the same official language as well as the spacial distance. The demographic covariates are sourced form the UNDESA (2019) World Population Prospects dataset. Geographic data is sourced form the CEPII (2021) database and economic variables from the World Economic Outlook by the International Monetary Fund (2022).

3 Results

To gain insights about the predictability of the models outlined above as well as linear combinations of their predictions we proceed as follows: the data is split in the time dimension into training, hold-out, and pseudo-out-of-sample data. We estimate the models on the test data to generate predictions for the hold out data. Based on those predictions we calculate weights by different methods that we subsequently use to combined forecast for the pseudo-out-of-sample data.[1] Finally the predictions for the pseudo-out-of-sample data obtained by the individual model and their combination is compared with the realized flows and mean absolute error and root mean squared error are calculated as prediction validation metrics. This procedure is repeated for 1-,2- and 3-steps ahead forecasts. The results are reported in Table (1).

The results suggest that among the individuals models the performance of the pooled and individual autoregression is similar and both outperform the Bayesian flow model which in turn outperforms the gravity model. This holds irrespectively of the forecasting horizon and validation metric. Among the averaged forecasts we find that calculated weights by IC and BMA leads to corner solution were all weight is allocated to the best performing model in hold-out period. Furthermore we find that utilizing MSE or MAE performs similarly and leads to the best (non degenerated) averaged forecasts. This holds irrespectively of the forecasting horizon and validation metric. However, non of the combined forecast outperforms the best performing individual model. Thus in the current state of the project we can not provide evidence that model averaging increases predictive ability. This however can be subject to the specific data set under consideration and further investigation should check the robustness of this finding for different data sets.

[1] The details of the different model averaging approach are outlined in the appendix.

Model	1-step		2-step		3-step	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
Individual Forecasts						
IAR	101.33	**100.00**	103.37	100.11	106.17	100.78
PAR	**100.00**	100.18	**100.00**	**100.00**	**100.00**	**100.00**
GM	271.77	125.04	240.31	124.08	234.65	123.84
BFM	135.09	102.88	123.86	102.01	138.41	106.95
Averaged Forecasts						
M	128.65	102.04	117.85	101.34	114.37	102.43
TM	111.46	101.02	106.57	100.56	110.80	103.57
MSE	100.54	100.03	100.81	100.10	100.39	101.62
MAE	104.26	100.21	104.19	100.07	103.29	101.66
OLS	118.52	101.12	110.72	100.58	108.13	101.78
IC	101.33	100.18	**100.00**	**100.00**	**100.00**	**100.00**
BMA	101.33	100.18	**100.00**	**100.00**	**100.00**	**100.00**

Table 1: Mean absolute error (MAE) and root mean squared error (RMSE) of observed and predicted migration flows for 1-, 2- and 3-steps ahead predictions (minimum normalized to 100.00).

4 Conclusion

In this paper we outline the main ideas behind causal and non-causal modeling techniques for migration flows. As the question of how to model migration is yet far from being settled, forecasts based on individual modeling techniques embed a high amount of model uncertainty into predictions. Following several approach to combine forecasts, we average the forecasts of the models under consideration and evaluate the predictability of individual and combined predictions in a pseudo out-of-sample forecasting exercise.

Applying this approach to bilateral migration flow data provided by the OECD suggests that among the individual models a simple pooled autoregressive model where information is shared across observations with the same origin country performs best. Furthermore, we find, a model averaging approach based on mean squared error and mean absolute error performs similarly and leads to the best averaged forecasts. However the current results do not proved evidence that model averaging leads to better predictive performance than utilizing individual models. This motives robustness checks of the findings for different data set.

References

Azose, J.J. and A.E. Raftery (2015): “Bayesian probabilistic projection of international migration”. In: *Demography* 52.5, pp. 1627–1650. DOI: https://doi.org/10.1007/s13524-015-0415-0.

Benveniste, Hélène et al. (2021): “Tracing international migration in projections of income and inequality across the Shared Socioeconomic Pathways”. In: *Climatic Change* 166.3, pp. 1–22.

Beyer R.M., Schewe J. and H. Lotze-Campen (2022): “Gravity models do not explain, and cannot predict, international migration dynamics”. In: *Humanities and Social Sciences Communications* 9.56. DOI: https://doi.org/10.1057/s41599-022-01067-x.

CEPII (2021): “Gravity Equations: Toolkit, Cookbook, Workhorse”. In: *Handbook of International Economics* 4. ISSN: 01621459. URL: http://www.cepii.fr/CEPII/en/bdd_modele/bdd_modele_item.asp?id=8 (visited on 09/06/2022).

Cohen, J.E. (2012): “Projection of net migration using a gravity model”. In: *Proceedings of the XXVII IUSSP International Population.*

Cohen J.E., Roig M. Reuman D. and C. GoGwilt (2008): “International migration beyond gravity: A statistical model for use in population projections”. In: *Proceedings of the National Academy of Sciences* 105.40, pp. 15269–15274. DOI: https://doi.org/10.1073/pnas.0808185105.

International Monetary Fund (2022): *World Economic Outlook, April 2022: War Sets Back The Global Recovery.* World Economic and Financial Surveys. URL: https://www.imf.org/en/Publications/WEO/Issues/2022/04/19/world-economic-outlook-april-2022.

Kapetanios, George et al. (2006): “Forecasting using predictive likelihood model averaging”. In: *Economics Letters* 91.3, pp. 373–379.

Karemera D., Oguledo V. and B. Davis (2000): “A gravity model analysis of international migration to North America”. In: *Applied economics* 32.13, pp. 1745–1755. DOI: https://doi.org/10.1080/000368400421093.

Kim, K. and J.E. Cohen (2010): “Determinants of international migration flows to and from industrialized countries: A panel data approach beyond gravity”. In: *International migration review* 44.4, pp. 899–932. DOI: https://doi.org/10.1111/j.1747-7379.2010.00830.x.

Mayda, A.M. (2010): “International migration: A panel data analysis of the determinants of bilateral flows”. In: *Journal of population economics* 23.4, pp. 1249–1274. DOI: https://doi.org/10.1007/s00148-009-0251-x.

OECD International migration database (2022): *International Migration Database.* Tech. rep. URL: https://doi.org/10.1787/data-00342-en.

Raftery, Adrian E. et al. (1997): “Bayesian Model Averaging for Linear Regression Models”. In: *Journal of the American Statistical Association* 92.437, pp. 179–191. ISSN: 01621459. URL: http://www.jstor.org/stable/2291462 (visited on 09/06/2022).

Ramos, R. (2016): “Gravity models: a tool for migration analysis”. In: *IZA World of Labor.* DOI: 10.15185/izawol.239.

UNDESA (2019): *World Population Prospects (2019 Revision).* World Demographic Surveys. URL: https://population.un.org/wpp/.

Welch, Nathan G. and Adrian E. Raftery (2022): “Probabilistic forecasts of international bilateral migration flows”. In: *Proceedings of the National Academy of Sciences* 119.35, e2203822119.

DOI: 10.1073/pnas.2203822119. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.2203822119. URL: https://www.pnas.org/doi/abs/10.1073/pnas.2203822119.

5 Appendix

To introduce some notation we assume that the combinations of forecasts are given by,

$$\overline{\mathbf{m}}^s_{\tau+h} = \sum_{k=1}^{n} \omega^s_{\tau+h,k} \widehat{\mathbf{m}}_{\tau+h,k}, \tag{8}$$

where $\widehat{\mathbf{m}}_{t_3,k}$ denotes the vector of h-steps-ahead bilateral migration flow predictions obtained by model k, $\omega^s_{\tau+h,k}$ is the weight of the h-step ahead prediction of model k and the respective method that is used to calculate the weights which is denoted by superscript s. Finally $\overline{\mathbf{m}}^s_{\tau+h}$ denotes the averaged h-steps-ahead predictions based on weight-method s. The approaches that we apply to obtain the weights in equation (8) are,

1. the unweighted mean of the forecasts (M). Here,

$$\omega^{mean}_{\tau+h,k} = \frac{1}{n}, \tag{9}$$

   where n denotes the number of models.

2. the trimmed mean (TM). In this case we take the unweighted average of the models excluding the models that performed best and worst (in terms of RMSE) in predicting the hold-out data. Here,

$$\omega^{tmean}_{\tau+h,k} = \frac{1}{n-2}. \tag{10}$$

3. calculating the weights according to the mean squared error of the individual predictions (MSE). Here,

$$\omega^{MSE}_{\tau+h,k} = \frac{MSE^{-1}_{\tau+h,k}}{\sum_{k=1}^{n} MSE^{-1}_{\tau+h,k}}, \tag{11}$$

   where $MSE_{\tau+h,k} = ||\widehat{\mathbf{m}}_{\tau+h,k} - \mathbf{m}_{\tau+h,k}||_2^2/d$ and d denotes the dimensionality of the vector $\mathbf{m}_{\tau+h,k}$.

4. calculating the weights according to the mean absolute error of the individual predictions (MAE). Here,

$$\omega^{MAE}_{\tau+h,k} = \frac{MAE^{-1}_{\tau+h,k}}{\sum_{k=1}^{n} MAE^{-1}_{\tau+h,k}}, \tag{12}$$

   where $MAE_{\tau+h,k} = ||\widehat{\mathbf{m}}_{\tau+h,k} - \mathbf{m}_{\tau+h,k}||_1/d$ and d is defined as above.

5. estimating the weights by OLS (OLS). Here,

$$\omega^{OLS}_{\tau+h,k} = \frac{exp(\hat{\omega}_{\tau+h,k})}{\sum_{k=1}^{n} exp(\hat{\omega}_{\tau+h,k})} \tag{13}$$

   where $\hat{\omega}_{\tau+h,k}$ corresponds to its OLS estimate obtained by estimating (8) with an additional constant regressor.

6. calculating the weights according to an out-of-sample information criteria as outlined in Kapetanios et al. 2006 (IC). Here,

$$\omega^{IC}_{\tau+h,k} = \frac{exp(-\frac{1}{2}\Psi_{\tau+h,k})}{\sum_{k=1}^{n} exp(-\frac{1}{2}\Psi_{\tau+h,k})}, \tag{14}$$

   where $\Psi_{\tau+h,k} = \phi_{\tau+h,k} - \min_{j\in\{1,\ldots,n\}}\{\phi_{\tau+h,j}\}$ and $\phi_{\tau+h,k}$ is the concentrated log-likelihood of the model based on the root mean squared forecasting error (RMSE), i.e. $\phi_{\tau+h,k} = -\frac{d}{\ln \tilde{\sigma}_{\tau+h,k}}$ with $\tilde{\sigma}_{\tau+h,k} = ||\widehat{\mathbf{m}}_{\tau+h,k} - \mathbf{m}_{\tau+h,k}||_2/\sqrt{d} = RMSE$ and d is defined as above.

7. calculating the weights based on predictive Bayesian model averaging (BMA). Here,

$$\omega^{BMA}_{\tau+h,k} = \frac{d^{\frac{p_1-p_k}{2}} \left(\frac{\tilde{\sigma}_{\tau+h,1}}{\tilde{\sigma}_{\tau+h,k}}\right)^{\frac{d}{2}}}{\sum_{k=1}^{n} d^{\frac{p_1-p_l}{2}} \left(\frac{\tilde{\sigma}_{\tau+h,1}}{\tilde{\sigma}_{\tau+h,k}}\right)^{\frac{d}{2}}}, \tag{15}$$

   where $\tilde{\sigma}_{\tau+h,i}$ is defined as above.