

Bayesian Hierarchical Model for Estimating Age and Sex Patterns of International Migration Using Socioeconomic Covariates

Aysha Basheer^{a,b}, Arkadiusz Wiśniowski^a, Maciej Dańko^b, Emilio Zagheni^b

^a*Department of Social Statistics, University of Manchester, UK*
^b*Max Planck Institute for Demographic Research (MPIDR), Germany*

Introduction

International migration is a major force shaping demographic change across Europe, yet obtaining reliable, age- and sex-disaggregated data remains challenging due to inconsistent definitions, incomplete coverage, and reporting gaps across countries. This lack of data hampers efforts to understand population dynamics, plan labour markets and public services, and inform policy. Detailed information on migrants' age and sex distributions is essential for understanding population heterogeneity, improving forecasts, and ensuring equitable access to services such as healthcare and education. Estimating bilateral migration flows from limited or indirect data has long been central to demographic research. Early stock-based approaches derived age–sex profiles from census data (Rogers et al., 2003; Abel, 2013; Shen et al., 2024), while recent Bayesian hierarchical models have harmonized data sources to generate coherent migration estimates (Abel, 2010; Azose and Raftery, 2015; Raymer et al., 2013; Wiśniowski et al., 2016; Raymer and Wiśniowski, 2018; Welch and Raftery, 2022). These models borrow strength across countries and over time, correcting for inconsistencies and quantifying uncertainty. However, few studies have jointly modelled age- and sex-specific migration alongside socioeconomic drivers within a unified framework. Building on this foundation, the present study develops a Bayesian hierarchical Poisson model with log-normal random effects that jointly estimates bilateral migration flows while integrating socioeconomic covariates such as life expectancy, unemployment, and educational attainment levels. This framework reconciles data reported by both sending and receiving countries, quantifies uncertainty, and enables partial pooling across origin–destination–age–sex combinations. By linking demographic patterns with socioeconomic drivers, the model advances migration estimation methodology and enhances our understanding of how contextual factors shape age–sex migration structures across Europe.

Methodology

This study analyses annual bilateral migration flows among 31 European countries (EU member states and EFTA countries) over the period 2009–2022. The data are structured as contingency tables of migration counts disaggregated by origin country (o), destination country (d), age group (a), sex (s), year (t), and reporting country (k), where k can indicate either the sending country (S) or the receiving country (R). Each observed count $z_{odast}^{(k)}$ represents the number of individuals of sex s in age group a who migrated from country o to country d in year t , as recorded by either the sending country S or the receiving country R . In line with Eurostat definitions, we treat migration as a change of usual residence for at least 12 months. In addition to migration flows, the model includes age- and sex-specific socioeconomic covariates to understand how they influence migration patterns. These include: life expectancy, unemployment rate and the proportions per 1000 individuals who attained ISCED levels of primary ($e1$), secondary ($e2$), and tertiary ($e3$) education. All these covariates are available by country, age, sex and time variables and collected from the Eurostat database.

The Bayesian hierarchical modelling framework provides a coherent approach to harmonize data reported by both sending and receiving countries, while quantifying uncertainty due to measurement differences and missing information. Migration counts are assumed to follow a Poisson distribution, with expected counts modelled through a log-linear structure that incorporates demographic and socioeconomic determinants. The general structure of the model assumes that the expected migration flow between each origin–destination pair is a function of additive and interaction effects for origin, destination, age, and sex. Random effects are introduced hierarchically to borrow strength across countries and demographic groups, ensuring stable estimation even for sparse data combinations.

Specifically, the expected migration flow from origin o to destination d , for age group a , sex s , and year t , is modelled as,

$$z_{odast}^k \sim \text{Poisson}(\lambda_{odast}^k),$$

with mean λ_{odast}^k linked to the underlying true latent migration flow m_{odast} through the log-linear model,

$$\log(\lambda_{odast}^k) \sim N\{\log(m_{odast}) + \alpha_{odt}^k, \tau_k\}$$

where α_{odt}^k captures origin–destination–year-specific interactions. The true latent flow m is modelled as,

$$\begin{aligned} \log(m_{odast}) \sim N(\alpha_{da} + \alpha_{ds} + \alpha_{oa} + \alpha_{os} + \\ \beta_{das}^l l_{dast} + \beta_{oas}^l l_{oast} + \beta_{das}^{ur} ur_{dast} + \beta_{oas}^{ur} ur_{oast} + \\ \beta_{oas}^{e1} e1_{oast} + \beta_{das}^{e1} e1_{dast} + \beta_{oas}^{e2} e2_{oast} + \beta_{das}^{e2} e2_{dast} + \beta_{oas}^{e3} e3_{oast} + \beta_{das}^{e3} e3_{dast}, \tau_m) \end{aligned}$$

Here, l_{oast} , ur_{oast} , $e1_{oast}$, $e2_{oast}$, and $e3_{oast}$ represent age-, sex-, origin-, and year-specific measures of life expectancy, unemployment rate, and the proportions per 1,000 individuals who attained ISCED levels of primary (e1), secondary (e2), and tertiary (e3) education, respectively. Interaction terms such as α_{da} , α_{ds} , α_{oa} , and α_{os} allow flexible partial pooling across demographic and geographic dimensions. Age-based masking is applied for covariates whose effects are expected to be negligible in specific age groups (e.g., education for very young children or the elderly). Weakly informative priors were assigned to the covariate effects and interaction terms to ensure stable estimation while maintaining flexibility. While we initially explored independent normal priors, a multivariate approach may be more appropriate to account for potential correlations among covariates, and will be considered in future refinements.

This covariate-informed model was compared against alternative hierarchical structures that omit covariates (combined-effect model),

$$\log(m_{odast}) \sim N(\alpha_{da} + \alpha_{ds} + \alpha_{oa} + \alpha_{os}, \tau_m)$$

or include simpler interaction terms (main-effect model),

$$\log(m_{odast}) \sim N(\alpha_d + \alpha_o + \alpha_a + \alpha_s, \tau_m)$$

as well as with traditional approaches such as the Rogers–Castro model. These comparisons help assess the robustness and added explanatory power of including socioeconomic covariates in capturing migration patterns. The model estimation was implemented in JAGS (Just Another Gibbs Sampler) through the R interface using MCMC (Markov Chain Monte Carlo) sampling. Convergence was assessed using the Gelman–Rubin R statistic, trace plots, and effective sample sizes. All covariates were standardised (mean = 0, sd = 1) to allow direct comparison of effect magnitudes and to ensure convergence stability in MCMC sampling.

Preliminary Results and Discussions

Metric	Covariate Model	Combined-effect Model	Main-effect model
RMSE	0.004	0.005	0.009
DIC	57963	58437	58883

Table 1: Model Comparison (lower = better)

Preliminary results highlight strong and interpretable associations between migration patterns and socioeconomic conditions across Europe. Table 1 shows that the model with covariates has lower DIC and RMSE than the baseline models, indicating improved fit and predictive accuracy when including covariates. Figure 1 illustrates the average relative effect sizes of life expectancy at destination by age group and sex. Positive values suggest that higher life expectancy at the destination is associated with higher migration intensity. The effect is consistently positive for younger and middle-aged adults with peaks observed in the 25–34 age range. This pattern indicates that longer life expectancy at destination may attract working-age migrants seeking better living conditions and health-care access. Figure 2 shows the posterior means and 95% credible intervals for the relative effect of origin-specific unemployment in Bulgaria on migration by age and sex. For males, unemployment shows a strong positive effect on migration among individuals aged 25–54, aligning with traditional economic theories (Pissarides and McMaster (1990)). For females, the effect is more pronounced and spans a wider age range. Significant positive effects are

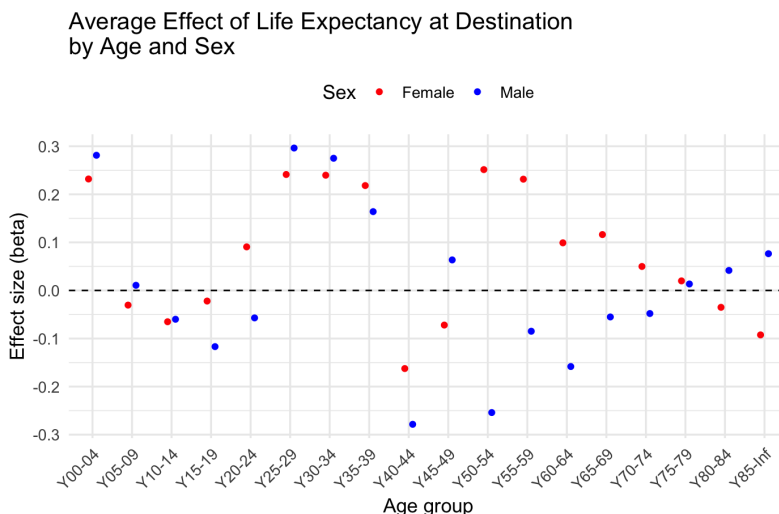


Figure 1: Median relative effect of life expectancy averaged over all destinations by age and sex

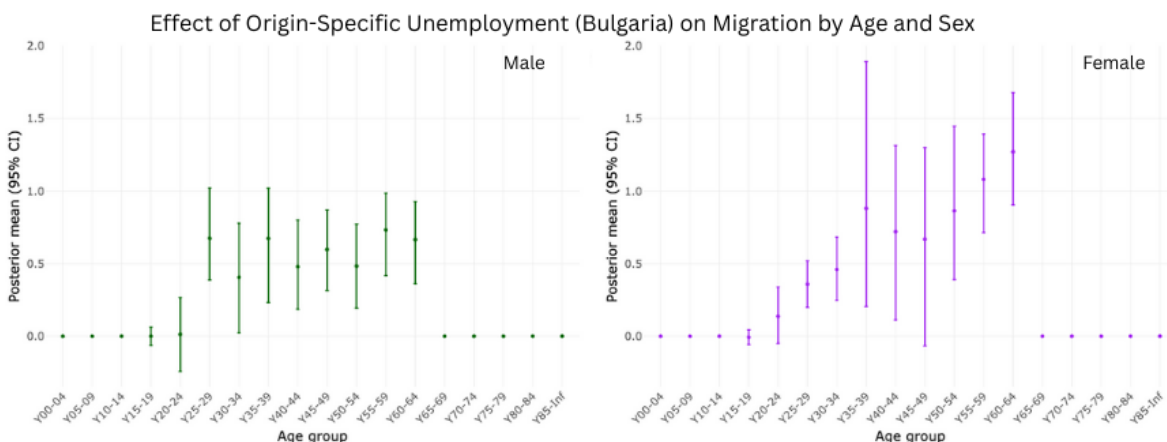


Figure 2: Posterior median estimates for unemployment rate effects (relative) at origin on migration by age and sex for Bulgaria as origin

observed from age 20 to 64, with particularly strong coefficients in the 35–59 range. This may indicate heightened economic vulnerability or increased female labour migration from Bulgaria during periods of economic distress. These findings suggest that while unemployment is a key driver of migration for both sexes, its intensity and age distribution vary substantially.

Figure 3 compares model-based estimates with observed migration counts, showing three curves: the covariate model, two alternative model variations, and the reported data. All curves align closely, indicating that the covariate model reproduces observed migration patterns well. While the overall age patterns are similar for both sexes, there are slight differences in migration proportions between males and females (Figure 4). In general, males show slightly higher migration proportions across most age groups and countries. However, the magnitude of this difference varies by country and year. Figure 5 show that the migration proportions for different age groups remain relatively stable over the observed years for most countries. However, there are some noticeable fluctuations, particularly in the peak age groups (20-34). Both females and males exhibit higher migration proportions during these years, which could be attributed to career opportunities and other socio-economic factors.

These results confirm the study’s main premise, that integrating socioeconomic covariates within a hierarchical Bayesian framework captures key heterogeneities in migration dynamics that would remain hidden in purely demographic models. The observed age–sex patterns align with theoretical expectations while demonstrating that socioeconomic gradients, such as unemployment or life expectancy, shape the intensity and direction of flows.

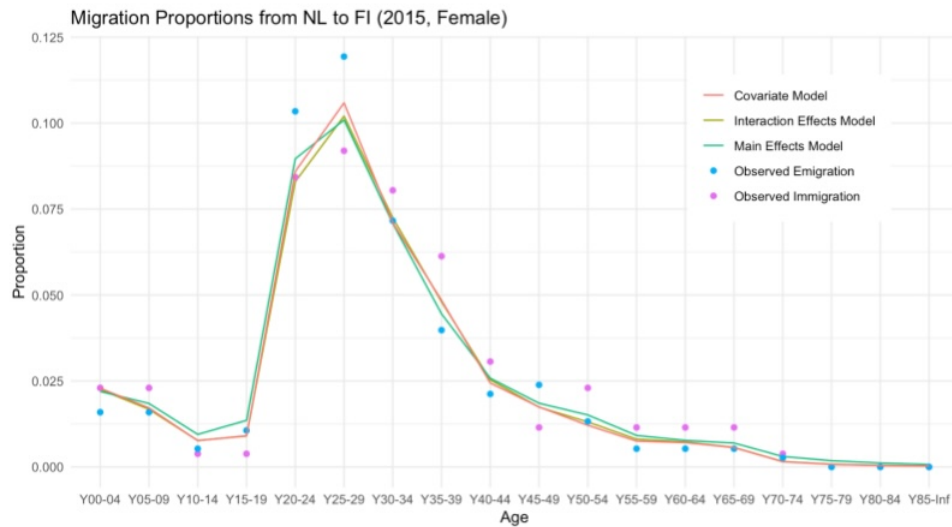


Figure 3: Estimated and observed emigration (reported by the Netherlands) and immigration (reported by Finland)

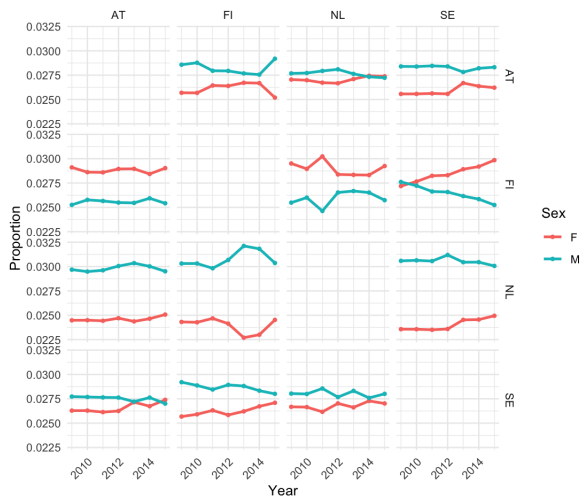


Figure 4: Estimated migration sex patterns, with rows and columns representing origin and destination countries, respectively

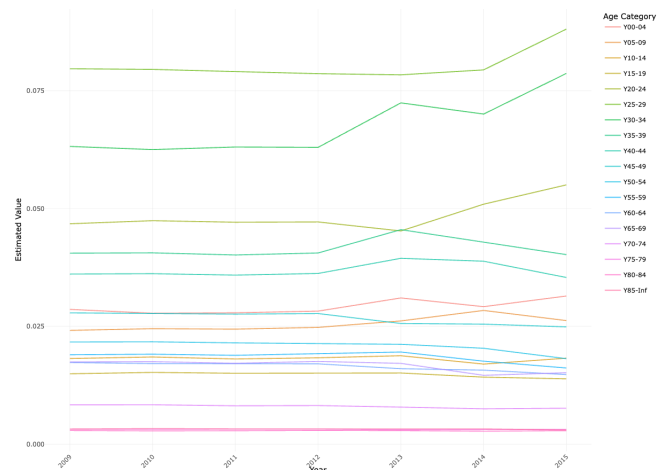


Figure 5: Age-specific proportions over time for female migration from Sweden to Austria

Future Work

Future work will focus on enhancing model robustness and validation. Planned steps include extending the model to capture temporal changes in covariate effects and a systematic comparison with benchmark approaches such as the Rogers–Castro model and hierarchical models without covariates to evaluate goodness of fit and predictive accuracy. Robustness checks will include alternative prior structures and temporal cross-validation. Sensitivity analyses will test the stability of posterior estimates to model assumptions and data sparsity. Model validation will employ posterior predictive checks and out-of-sample evaluation (if possible) to ensure reliability for forecasting and policy applications. These steps will strengthen confidence in the model’s capacity to provide consistent, interpretable migration estimates across Europe.

References

- Abel, G., 2013. Estimating global migration flow tables using place of birth data. *Demographic Research* 28, 505–546. URL: <https://www.demographic-research.org/volumes/vol28/18/>, doi:10.4054/DemRes.2013.28.18, arXiv:<https://www.demographic-research.org/volumes/vol28/18/28-18.pdf>.
- Abel, G.J., 2010. Estimation of International Migration Flow Tables in Europe. *Journal of the Royal Statistical Society Series A: Statistics in Society* 173, 797–825. URL: <https://doi.org/10.1111/j.1467-985X.2009.00636.x>, doi:10.1111/j.1467-985X.2009.00636.x, arXiv:https://academic.oup.com/jrsssa/article-pdf/173/4/797/49525384/jrsssa_173_4_797.pdf.
- Azose, J.J., Raftery, A.E., 2015. Bayesian probabilistic projection of international migration. *Demography* 52, 1627–1650. URL: <https://doi.org/10.1007/s13524-015-0415-0>, doi:10.1007/s13524-015-0415-0, arXiv:<https://read.dukeupress.edu/demography/article-pdf/52/5/1627/866781/1627azose.pdf>.
- Pissarides, C.A., McMaster, I., 1990. Regional migration, wages and unemployment: Empirical evidence and implications for policy. *Oxford Economic Papers* 42, 812–831. URL: <http://www.jstor.org/stable/2663123>.
- Raymer, J., Wiśniowski, A., 2018. Applying and testing a forecasting model for age and sex patterns of immigration and emigration. *Population Studies* 72, 339–355. URL: <https://doi.org/10.1080/00324728.2018.1469784>, doi:10.1080/00324728.2018.1469784, arXiv:<https://doi.org/10.1080/00324728.2018.1469784>, PMID: 29873285.
- Raymer, J., Wiśniowski, A., Forster, J., Smith, P., Bijak, J., 2013. Integrated modeling of european migration. *Journal of the American Statistical Association* 108, 801–819. doi:10.1080/01621459.2013.789435.
- Rogers, A., Willekens, F., Raymer, J., 2003. Imposing age and spatial structures on inadequate migration-flow datasets. *The Professional Geographer* 55, 56–68. URL: <https://doi.org/10.1111/0033-0124.01052>, doi:10.1111/0033-0124.01052.
- Shen, T., Raymer, J., Guan, Q., Wiśniowski, A., 2024. The estimation of age and sex profiles for international migration amongst countries in the asia-pacific region. *Population, Space and Place* 30, e2716. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/psp.2716>, doi:<https://doi.org/10.1002/psp.2716>, arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/psp.2716>.
- Welch, N.G., Raftery, A.E., 2022. Probabilistic forecasts of international bilateral migration flows. *Proceedings of the National Academy of Sciences* 119, e2203822119. URL: <https://www.pnas.org/doi/abs/10.1073/pnas.2203822119>, doi:10.1073/pnas.2203822119, arXiv:<https://www.pnas.org/doi/pdf/10.1073/pnas.2203822119>.
- Wiśniowski, A., Forster, J.J., Smith, P.W.F., Bijak, J., Raymer, J., 2016. Integrated modelling of age and sex patterns of european migration. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 179, 1007–1024. URL: <http://www.jstor.org/stable/44682194>.