

# Mixed-frequency VAR: a new approach to forecasting migration in Europe using macroeconomic data

Emily R. Barker  and Jakub Bijak 

Department of Social Statistics and Demography, University of Southampton, Southampton, UK

**Corresponding author:** Emily R. Barker; Email: [E.R.Barker@soton.ac.uk](mailto:E.R.Barker@soton.ac.uk)

**Received:** 13 November 2023; **Revised:** 15 August 2024; **Accepted:** 23 September 2024

**Keywords:** Bayesian forecasting; migration; mixed-frequency models; panel VAR

## Abstract



Forecasting international migration is a challenge that, despite its political and policy salience, has seen a limited success so far. In this proof-of-concept paper, we employ a range of macroeconomic data to represent different drivers of migration. We also take into account the relatively consistent set of migration policies within the European Common Market, with its constituent freedom of movement of labour. Using panel vector autoregressive (VAR) models for mixed-frequency data, we forecast migration in the short- and long-term horizons for 26 of the 32 countries within the Common Market. We demonstrate how the methodology can be used to assess the possible responses of other macroeconomic variables to unforeseen migration events—and vice versa. Our results indicate reasonable in-sample performance of migration forecasts, especially in the short term, although with varying levels of accuracy. They also underline the need for taking country-specific factors into account when constructing forecasting models, with different variables being important across the regions of Europe. For the longer term, the proposed methods, despite high prediction errors, can still be useful as tools for setting coherent migration scenarios and analysing responses to exogenous shocks.

## Policy significance statement

Forecasting international migration is a research and policy challenge due to the presence of hardly predictable drivers, problematic data as well as unforeseeable events. In this article, we examine short- and long-term forecasts of migration flows for 26 European countries, making use of similarities between countries and of data at different frequencies. The models are broadly successful in the short-run, with the long-run ones showing high levels of uncertainty, as expected. Unsurprisingly, the countries that have had stable flows exhibit smaller levels of forecast uncertainty. The method has potential to enhance support for migration policies by utilising the available information more fully.

## 1. Introduction

There is much uncertainty about the size, timing, type, duration, and impact of future migration flows and their interplay within the economy and the society. Being able to anticipate some of the migration changes and provide adequate response is a paramount contemporary challenge, which is already recognised in

  This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

many policy areas, notably at the European level<sup>1</sup>. At the same time, the drivers of migration are highly variable and are easily subject to change, sometimes at a very short notice and with hardly any warning signs (Bijak and Czaika, 2020). Moreover, the data for migration and its different drivers may come at various frequencies, from annual to quarterly or monthly. In that context, the aim of this proof-of-concept paper is to explore a new method of forecasting migration flows both for shorter and longer-term horizons, that makes the best possible use of various available data sources with different time granularity.

One key advantage of modern macroeconomic modelling in the context of migration forecasting is its dynamic nature. Here, existing possibilities include gravity (e.g. Beine et al., 2016) or dynamic stochastic general equilibrium models. Of course, there are reasons that people migrate which cannot be modelled with migrant host country data. Such examples include the war in Ukraine or the so-called “asylum crisis” of 2015–2016. The inability for macroeconomic data to capture this additional uncertainty makes forecasting migration challenging. Importantly, they include unpredictable political events or migration policy changes (tightening or relaxation of legislation, see Czaika et al., 2021), and people’s reactions to them at the micro level. That notwithstanding, macroeconomic approaches can offer coherent description of at least *some* of the relevant drivers, help set plausible migration scenarios and analyse their uncertainty and sensitivity.

The approach proposed in this paper relies on panel vector autoregressive (VAR) models using mixed-frequency data. The method can be appealing as a forecasting tool in the short term, and a way for setting macroeconomically coherent “what-if” migration and driver scenarios for the longer horizons. Here, some canonical economic drivers of migration include improved employment opportunities, salary levels, and other job-related factors, and the same holds for push factors, including economic downturns, low wages, or high unemployment (Massey et al., 1993). Such factors can be modelled with macroeconomic data, which may come at different frequency, so that the models do not need to lose information through having to aggregate all data to the least frequent common denominator.

Of course, in addition to economic drivers, there are also other macro-level circumstances in which migration can change unexpectedly, and which cannot be adequately explained or predicted by changes in economic structures and processes alone. Even though anticipating such changes to migration processes is out of scope of this study, we demonstrate how the proposed mixed-frequency VAR modelling can shed light on the impact of rapid migration changes on other areas of the economy. Besides, the proposed methodology is flexible enough to allow for including other drivers of migration in the models as well, as long as the relevant data series are available.

In this article, Section 2 presents the econometric analysis of short-term forecasts based on the panel VAR models for mixed-frequency data. Section 3 includes reflections and some empirical evaluation for longer-range scenario setting, extended to the analysis of the possible economic impacts of uncertain migration and economic events. Section 4 provides further discussion and concludes.

## 2. Short-term forecasts: mixed-frequency panel VAR

### 2.1. General remarks

As mentioned before, forecasting migration is a challenge due to a sheer number of drivers and factors for different types of migration, as well as the irreducible (*aleatory*) uncertainty of the future (Bijak and Czaika, 2020). In this section, we take a look at migration forecasting through the prism of macroeconomic drivers, using Bayesian panel VAR models, extending the previous work of Gorbey et al. (1999) and Bijak (2010). We apply these models to data at different frequencies, quarterly and annual, and evaluate the forecasts by predicting migration indicators for eight quarters, 2018Q1–2019Q4, and comparing them *ex post* with the values observed in the data for that period. By so doing, we are also

<sup>1</sup> See, e.g. the EU migration “Blueprint”: Commission Recommendation (EU) 2020/1366 of 23 September 2020 on an EU mechanism for preparedness and management of crises related to migration, OJ L 317, 1.10.2020, p. 26–38, <http://data.europa.eu/eli/reco/2020/1366/oj>.

trying to assess to what extent including theoretical macroeconomic relationships and regularities encoded in the models can shed some light on the *epistemic* migration uncertainty, related to imperfect knowledge.

In this analysis, we use a Bayesian panel vector autoregressive (VAR) model to present forecasts for 26 European countries from the European Union (EU), the European Free Trade Agreement (EFTA), as well as the United Kingdom<sup>2</sup>. The focus on macroeconomic variables among the many theoretical considerations about migration drivers is of course to some extent arbitrary (see, e.g. Czaika and Reinprecht, 2020; Massey et al., 1993). Our motivation here is twofold. First, these drivers form a very important part of the complex driver environments (Czaika and Reinprecht 2020), especially given the prominent role of labour migration among all population flows. Secondly, even though the examples presented here focus on the economy, the tools used in the analysis can be transferable to other drivers and their environments, as long as appropriate data are available.

In our examples, the macroeconomic data, covering the period 2002Q1–2019Q4, include GDP, wages and salaries, unemployment rates, and employment indicators. Our dataset features both annual (migration) and quarterly data, so that the models need to be analysed using mixed-frequency tools proposed by Canova and Ferroni (2020) and Dieppe et al. (2016). From the national accounts variables, we use GDP, sourced from the OECD data, and wages and salaries, collected from Eurostat together with the labour market data. The wages and salaries are deflated using the GDP deflator. For national accounts, the variables are expressed in real terms, per working-age population, and are log-transformed. The wage premium is calculated as the rate of wages and salaries (in Euros) per employed person relative to the EU-15. The migration estimates are sourced from the IMEM project (Raymer et al., 2013) and the QuantMig Migration Estimates Explorer (<https://bit.ly/quantmig-estimates>). As discussed later in the paper, many of these series, notably including migration, are likely non-stationary, which means that their characteristics change over time.

To assess model performance, we first conduct an in-sample forecasting exercise over the period 2018Q1–2019Q4. The evaluation time frame is limited to these 2 years to avoid potential problems with the effects of the COVID-19 pandemic on the migration measures used. The indicators being forecast are relative measures of migration intensity, expressed per 1000 residents of a given country. Immigration and net migration measures are labelled as “rates,” to reflect that they do not correspond to the correct populations at risk. In addition, we conduct an out-of-sample forecasting exercise, for the period 2018Q1–2025Q4 and long-run horizons up to 2050Q4. Table 1 lists the different variables used in the forecasting models, their brief description, source and any transformation that was made to the data before their inclusion.

## 2.2. Formulation and evaluation of panel VAR models

In the forecasting exercise, we predict immigration, emigration, as well as net migration series separately. In this section we define the model and present illustrative results for the model variant that performed well enough across all groups of countries under study. To formulate and estimate the forecasting models, we use toolboxes by Canova and Ferroni (2020) for mixed-frequency data transformations, and Dieppe et al. (2016) for Bayesian panel VAR forecasting, the latter via the BEAR toolbox: Bayesian Estimation, Analysis and Regression. Ideally, it would be preferable to perform both parts in one step, as in our approach the part of uncertainty that is related to data transformations is not directly propagated into the forecasts. However, no widely used toolbox is currently available to simultaneously estimate MF-PVAR models and use them for forecasting<sup>3</sup>.

<sup>2</sup> These countries were all members of the European common market in 2019. Croatia, Cyprus, Iceland, Liechtenstein, and Malta were excluded from the analysis due to data limitations, and the Netherlands are excluded due to issues with data uncertainty; see Appendix for details.

<sup>3</sup> Recent work by Schorfheide and Song (2015); Kapetanios et al. (2021), and Cipollini and Parla (2023) has made attempts at using MF-PVAR by using different methods. Schorfheide and Song (2015) use a Bayesian approach whilst the others use frequentist methods. Our preliminary experiments with these tools, however, have not yielded satisfactory results.

**Table 1.** Data variables and descriptions

Variable	Description	Source	Transformation
<i>Migration</i>			
Immig	Immigration “Rate”	IMEM, QuantMig	Per 1000 Residents
Emig	Emigration Rate	IMEM, QuantMig	Per 1000 Residents
NM	Net Migration “Rate”	IMEM, QuantMig	Per 1000 Residents
<i>National accounts—expenditure</i>			
GDP	Gross Domestic Product	OECD—CARSA, DNBSA	Real, Per WA, Logged
WageSal	Wages and Salaries	Eurostat	Real, Per WA, Logged
<i>Labour market</i>			
Emp	Employment 15–64	Eurostat	Millions
Unemp	Unemployment Rate	Eurostat	
WagePre	Wages Premium to EU–15	Own calculations	Real, Per WA
<i>Miscellaneous</i>			
TtlCmnLbrMkt	Total size of the EU+ labour market	Barker (2022)	Millions

Note. WA = working-age population; CARSA = national currency, current prices, annual levels, seasonally adjusted; DNBSA = deflator, national base year, seasonally adjusted.

As a result, having initially decoupled the two stages of the process, we then propagate the uncertainty about the mixed-frequency estimation into panel VAR models by using 10,000 random Monte Carlo replicates of the mixed-frequency estimates and summarising the panel VAR results across them<sup>4</sup>. In our case, only one variable (migration) required transforming from lower to higher frequencies, so the resulting additional errors are not consequential. At the same time, the computational expense of this approach proved considerable, even on supercomputer, with the estimations for 10,000 Monte Carlo replicates across all 26 countries and three migration indicators taking several days.

As a proof of concept, we first employ the approach of Canova and Ferroni (2020) relying on mixed-frequency VAR (MF-VAR), based on data from the national accounts, government accounts, unemployment, and migration statistics. In the toolbox, the Gibbs sampler is used in the reduced-form VAR, to estimate the quarterly observations of the variables (Canova and Ferroni, 2020, 54). For each of the 10,000 Monte Carlo replicates, the BEAR toolbox of Dieppe et al. (2016) is subsequently used for panel VAR modelling. The analysis contains  $N$  countries,  $n$  variables,  $p$  lags, and covers  $T$  quarters. The element of the panel VAR model related to country  $i$  ( $i \in 1, 2, \dots, N$ ) is specified as:

$$y_{i,t} = \sum_{j=1}^N \sum_{k=1}^p \Psi_{ij}^{(k)} y_{j,t-k} + \epsilon_{i,t} \quad (1)$$

where,  $y_{i,t}$  is a  $n \times 1$  vector of  $n$  endogenous variables for country  $i$  at time  $t$ . The matrix of coefficients is given by  $\Psi_{ij}$ , of size  $n \times n$ , and  $\epsilon_{i,t}$  is a vector of  $n \times 1$  vector white noise error terms with  $\epsilon_{i,t} \sim N(0, \Sigma_t)$  as specified by Dieppe et al. (2016). The model is estimated with four lags,  $p=4$ , corresponding to 1 year. The contemporaneous mutual impacts of different variables for each country are introduced through the covariance matrix  $\Sigma_t$ , allowing for reverse (reinforcing or dampening) feedback effects to occur simultaneously in the same period. To take advantage of the panel structure, the borrowing of strength between

<sup>4</sup> Note that in statistical terms the result is a compound (mixture) distribution, with panel VAR results integrated over the latent space of all possible mixed-frequency trajectories. This allows applying the Law of Total Variance to calculate the variance of the mixture, as the sum of the mean predictive variance of the panel VAR and variance of the means generated by different mixed-frequency trajectories. We use this property in the analysis presented in this paper to amalgamate the two sources of uncertainty.

different countries occurs not only through  $\Sigma_t$ , but also via the matrices of vector autoregressive parameters  $\Psi_{i,j}$ .

The methodology for the mixed-frequency VAR (MF-VAR) is based on estimation along the lines of Schorfheide and Song (2015) implemented in the toolbox by Canova and Ferroni (2020). In this paper, we use annual and quarterly observations. The toolkit requires the annual observation to be given in the last quarter of the year, with the first to third quarters left as “unobserved.” All the values for GDP are used at a quarterly frequency, with migration being the only annual series in this paper. The figures are already adjusted for seasonality: as such, the quarterly migration rates are obtained from annual ones by simply dividing by four.

In methodological terms, the MF-VAR is represented as a state-space model to manage the unobserved values. The definition of the state-space model used for mixed-frequency forecasting, adapted from Schorfheide and Song (2015, 368–369), is as follows. Defining an auxiliary variable  $z_{i,t} = [y'_{i,t}, \dots, y'_{i,t-p+1}]'$  allows for rewriting Equation (1) in the state-space form, by transforming the original VAR(p) model to its VAR (1) companion form, as:

$$z_{i,t} = \sum_{j=1}^N F(\Psi_{i,j}) z_{j,t-1} + v_{i,t} \tag{2}$$

For each country pair  $(i,j)$ ,  $F(\Psi_{i,j})'$  is an  $np \times np$  block matrix with  $\Psi_{i,j} = [\Psi_{i,j}^{(1)}, \dots, \Psi_{i,j}^{(p)}]$  in the first  $n$  rows, and sub-diagonal block identity matrices  $I_n$  in the remaining  $n(p-1)$  rows. The error term is i.i.d. multivariate normal,  $v_{i,t} \sim N(0, \Omega(\Sigma_i))$ , where the block matrix  $\Omega(\Sigma_i)$  contains  $\Sigma_i$  in the top-left corner and zeros elsewhere.

Let  $y_{i,t}$  be partitioned into variables observed at the quarterly (higher) frequency,  $y_{q,i,t}$ , and those observed at the lower (annual) frequency, such that  $y_{i,t} = [y'_{q,i,t}, y'_{a,i,t}]'$ . The model for the actual observations  $x_{i,t}$  can then combine the processes observed quarterly and annually by converting all data to the common (higher-frequency) “denominator”:

$$x_{i,t} = [y'_{q,i,t}, \tilde{y}'_{a,i,t} \cdot M_{i,t}]', \tag{3}$$

where  $\tilde{y}_{a,i,t} = \frac{1}{4}(y_{a,i,t} + y_{a,i,t-1} + y_{a,i,t-2} + y_{a,i,t-3})$ , and  $M_{i,t}$  is an identity matrix when  $t$  is the last quarter of the year, and zero otherwise. In the representation above, Equation (2) is the state equation, and Equation (3) is the observation equation. For further details on the methodology and Bayesian estimation, see Schorfheide and Song (2015)<sup>5</sup>.

For panel VAR forecasting, as we are using Bayesian methods, we also need to make a number of prior assumptions. The approach relies on a pooled estimator, with data for all countries pooled together to estimate a single, homogenous VAR model, with four lags and no constant in each model, with estimation based on 5000 iteration runs of the Gibbs sampler (following a burn-in of 500 iterations). The parameters and hyperparameters follow standard values from macroeconomic literature encoded in the BEAR package with the conjugate multivariate normal-inverse Wishart model structure: the normal priors for the autoregressive parameters are centred around 0.8, indicating a belief a priori in relatively large autocorrelations, whereas the marginal priors for the residual and factor variances are assumed to follow (tightening) inverse Gamma distributions with the shape parameter 1000 and scale parameter 1, to prevent the forecasts from exploding too fast (Dieppe et al., 2021).

<sup>5</sup> An alternative interpretation, for which we are very grateful to an anonymous reviewer of this articlepaper, is the following: the series that end up used in forecasting are  $[y'_{q,i,t}, (C_{a,z} \hat{z}_{i,t})']'$ . In this notation,  $C_{a,z}$  is a matrix selecting those rows of  $z_{i,t}$  that correspond to annual variables—in our case, the migration series—and  $\hat{z}_{i,t}$  are estimates of  $z_{i,t}$  conditional on the whole sample of length  $T$ . It has to be stressed that the estimates  $\hat{z}_{i,t}$  are generated by the model and are not a part of the original dataset.

As many migration (and indeed other macroeconomic) processes are likely non-stationary, an additional advantage of Bayesian methods is that they do not impose any requirements of stationarity (for a discussion, see, e.g. Koop et al., 2012, Chapters 17 and 18). Of course, alternative model specifications, such as based on differenced or de-trended series, or using the Vector Error Correction (VEC) mechanisms, are possible and could be used instead. The predictive characteristics, such as median trajectories or credible (predictive) intervals are then obtained by aggregating across the 10,000 Monte Carlo simulations, to include the uncertainty of multiple-frequency disaggregation.

In our panel VAR setup, we model four groups of countries: high-income and highly positive-net migration Western European countries (Group 1); high-income and medium positive-net migration European countries (Group 2); medium income and low-positive net migration (Group 3), and countries with low income and negative net migration (Group 4). The inclusion of a group of countries with *negative net migration* in a macroeconomic context is novel, as such migration has not been investigated before in detail across a number of countries.

We group the countries in the analysis to overcome problems arising from the short time period covered by the data. Using a panel VAR with six, seven, five and eight countries in Groups 1–4 respectively, increases the number of observations (the sample size) used to estimate the parameters, while allowing to take the advantage from the similarities in macroeconomic and migration patterns between the countries in each group. The list of countries and a data snapshot are provided in Table 2.

There are additional considerations we take when deciding on the placement of countries within the groups. One notable example, not shown in Table 2, is the number of years that net migration flows were positive. For all countries in Group 1, all years were positive. For Group 2, only Germany and Slovenia experienced one negative year each. Group 3 is more mixed, with a range of 9 (Ireland) to 14 (Slovakia) years with positive net migration. For Group 4, Greece experienced 10 positive years, but with a largely negative gradient, while Hungary experienced two and Estonia 3 years with positive net migration values, otherwise all country-year estimates were negative.

For the economic and labour market variables, Table 2 shows the averages and 2019 (base) values. The reason is to give a sample average, while also indicating the status of the economy at the beginning of the forecasting period, prior to the COVID-19 pandemic. For instance, the economies of Greece, Ireland, Italy, Portugal and Spain, suffered heavily during the global financial crisis of 2008–2011, as evident through large distortions in data, particularly for unemployment rates, and impacting net migration. The “new” EU member states have shown significant economic development since joining the EU, some at a greater pace than others. The economic growth in the former Eastern bloc countries has started to close the wage premium gap. We have also used both GDP per capita, as well as wages and salaries, as they provide different information. On the labour market side, we use employment rates in Table 2 rather than levels of employment, to aid comparison between the countries<sup>6</sup>.

To calculate the forecasts, we employ a vector of endogenous variables  $y_t$  describing the macroeconomy and labour market. Predictions for immigration “rates”, emigration rates, and net migration “rates” are calculated separately, as in Equation (4):

$$y_t = [\text{MIG}_t, \text{GDP}_t, \text{WageSal}_t, \text{Emp}_t, \text{WagePre}_t, \text{Unemp}_t]' \quad (4)$$

$$\text{MIG}_t = \text{Immig}_t \text{ or } \text{Emig}_t \text{ or } \text{NM}_t \quad (5)$$

The non-migration variables include some of the economic drivers of migration identifiable at the aggregate level in data (see, e.g. Massey et al., 1993), with higher levels of wages and employment representing some of the conventional “pull factors,” and lower levels—“push factors.” The GDP represents the overall state of the economy, with an economy above the trend being an attractive

<sup>6</sup>There are structural differences which can make employment rates differ, even between countries perceived as similarly advanced, so this comparison still needs to be seen as relative.



**Table 2.** Case study: selected summary statistics for selected 26 European countries

Country	Net migration		Real GDP p.c.		Wages		EU-15 wage		Unemp.		Emp.	
	Rate		(x 1000 €)		(x 1000 €)		Premium		Rate %		Rate %	
	Avg.	2019	Avg.	2019	Avg.	2019	Avg.	2019	Avg.	2019	Avg.	2019
<i>Group 1: High positive net migration and income</i>												
AUT	5.5	7.2	58.3	63.5	31.0	32.9	1.13	1.13	5.0	4.5	70.3	73.6
CHE	4.5	3.5	92.4	119.7	60.5	65.1	2.20	2.24	4.6	4.4	78.7	80.5
LUX	9.8	13.4	116.3	120.9	55.5	56.4	2.02	1.94	5.1	5.6	65.0	67.9
NOR	6.6	5.9	109.2	99.4	46.1	50.1	1.67	1.72	3.6	3.7	75.5	75.3
SWE	7.1	8.0	68.3	68.5	34.9	38.2	1.27	1.31	7.1	6.9	74.3	77.1
UK	4.3	6.4	55.6	53.7	34.0	34.7	1.24	1.19	5.7	3.8	71.8	75.2
<i>Group 2: Medium positive net migration and high income</i>												
BEL	3.5	4.4	55.2	60.4	32.3	33.3	1.17	1.15	7.7	5.4	61.9	65.3
DEU	2.4	4.8	53.5	60.4	28.6	31.4	1.04	1.08	6.8	3.2	71.0	76.7
DNK	2.0	1.8	69.4	76.8	44.4	47.6	1.61	1.63	5.8	5.0	73.9	75.0
FIN	2.6	3.2	54.0	60.2	32.4	33.4	1.18	1.15	8.1	6.7	69.5	72.9
FRA	2.2	2.8	50.5	55.6	29.8	31.8	1.08	1.09	8.8	8.1	64.1	65.6
ITA	3.3	2.1	44.0	44.7	19.6	19.9	0.71	0.68	9.4	9.9	57.2	59.0
SVN	3.8	6.9	25.1	30.2	17.0	18.9	0.62	0.65	6.9	4.5	66.3	71.8
<i>Group 3: Low positive net migration and medium income</i>												
CZE	-0.2	0.7	20.6	27.6	9.8	12.2	0.35	0.42	5.8	2.0	67.9	75.1
ESP	2.8	7.7	34.3	39.2	20.3	21.2	0.74	0.73	16.3	14.1	60.4	63.3
IRL	0.9	1.6	72.6	106.7	33.8	35.5	1.23	1.22	8.6	5.0	65.5	69.5
PRT	0.1	1.3	25.7	29.4	13.6	14.2	0.50	0.49	9.8	6.5	66.2	70.5
SVK	0.4	1.2	19.5	25.5	9.1	11.8	0.33	0.41	12.6	5.8	61.1	68.4
<i>Group 4: Negative net migration and low income</i>												
BGR	-5.4	-7.3	8.7	11.7	3.8	5.4	0.14	0.19	9.6	4.3	60.6	70.1
EST	-2.6	2.0	22.0	29.0	11.0	13.5	0.40	0.46	8.5	4.5	67.9	74.8
GRC	0.4	-1.4	27.8	26.7	12.6	12.2	0.46	0.42	16.2	17.3	56.1	56.5
HUN	-1.6	-2.2	12.3	13.9	9.0	9.9	0.33	0.34	7.3	3.4	59.9	70.1
LTU	-9.5	-3.5	16.7	24.2	8.4	14.0	0.30	0.48	10.0	6.3	64.6	73.0
LVA	-3.8	-4.4	16.4	22.3	9.0	12.1	0.32	0.42	11.1	6.3	65.0	72.3
POL	-2.6	-1.8	14.1	18.5	8.0	10.6	0.29	0.36	10.5	3.3	59.2	68.2
ROU	-5.5	-5.2	9.9	12.6	5.0	8.4	0.18	0.29	6.5	3.9	60.4	65.8

*Note.* Average (Avg) uses values for 2002–2019, with the values for 2019 either annual or averaged over Q1–Q4. Please note that for non-stationary data, such as migration rates, their averages should be interpreted with caution, as they cannot be interpreted as the estimators of the means of the underlying processes. The values for real GDP per capita and Wages are in 1000s of Euros. Unemployment rate is for 15–64 year olds. Source: Authors' calculations using data from Eurostat, IMEM database, QuantMig Estimates OECD, and national statistical institutes.

destination (pull factor). There are also other variables in the dataset in [Table 1](#), which could individually describe different migration drivers, but for consistency we use the same models for emigration and immigration<sup>7</sup>.

<sup>7</sup> The variables which are excluded via block exogeneity (i.e. exogenous variables) are not listed here. The exogenous variable employed is the size of the total common labour market at that time. The size of the common labour market refers to, in theory, the total source of labour supply that is in the free movement area. For example, the total common labour market increases when there is an EU expansion. These are all calibrated using the specific dates collected as part of the research included in [Barker \(2022\)](#).

2.3. Short-term migration forecasts: summary of the results

In this section we present a selection of forecasts from models shown in Equations (4) and (5). We focus on the forecasts of the immigration and net migration “rates,” as well as emigration rates. Plotting the forecasts for immigration and emigration separately helps identify, in which case the process dynamics was the most uncertain and volatile.

In terms of specific results, Figure 1 presents the forecasts of immigration, Figure 2 emigration, and Figure 3—net migration “rates” produced by models (4) and (5). All the *in-sample* forecasts presented in this section are unconditional. On the whole, the presented forecasting models have performed reasonably well. The predictive or confidence intervals are formed from the 67% percentile intervals from the 10,000 Monte Carlo replicates. The models are generally able to predict values which were close to the 67% predictive intervals for most countries. The *out-of-sample* forecasts contain the “best-possible” current information set. For long-term horizons, the forecasts are again unconditional: after three decades, the impact of the available current data would become almost irrelevant. This suggests an important role of the unpredictable (aleatory) uncertainty in shaping the overall forecast errors, especially for countries undergoing dynamic shifts of their migration patterns.

To show how the forecast errors compare, we present a selection of the error measures for each forecast in Table 3 and the calibration of the predictive intervals in Table 4. As the values are expressed per capita, there is not a need to additionally weigh any of the estimates. For each group, an arithmetic average is taken for each of the scores across the eight quarters analysed and all the countries in that group.

The errors in Table 3 show that, from that point of view alone, smaller percentage errors (MAPE) are obtained for the forecasts of immigration and emigration than for net migration, which supports the notion of forecasting each flow separately rather than net migration (Bijak, 2010; Rogers, 1990), and using flow-specific data for that purpose. There are some countries that distort their group overall with Luxembourg having the largest errors in Group 1, that can be partially explained by the high flows they receive. Group 2’s errors are distorted mainly by Slovenia which is not unexpected due to its change in migratory profile.

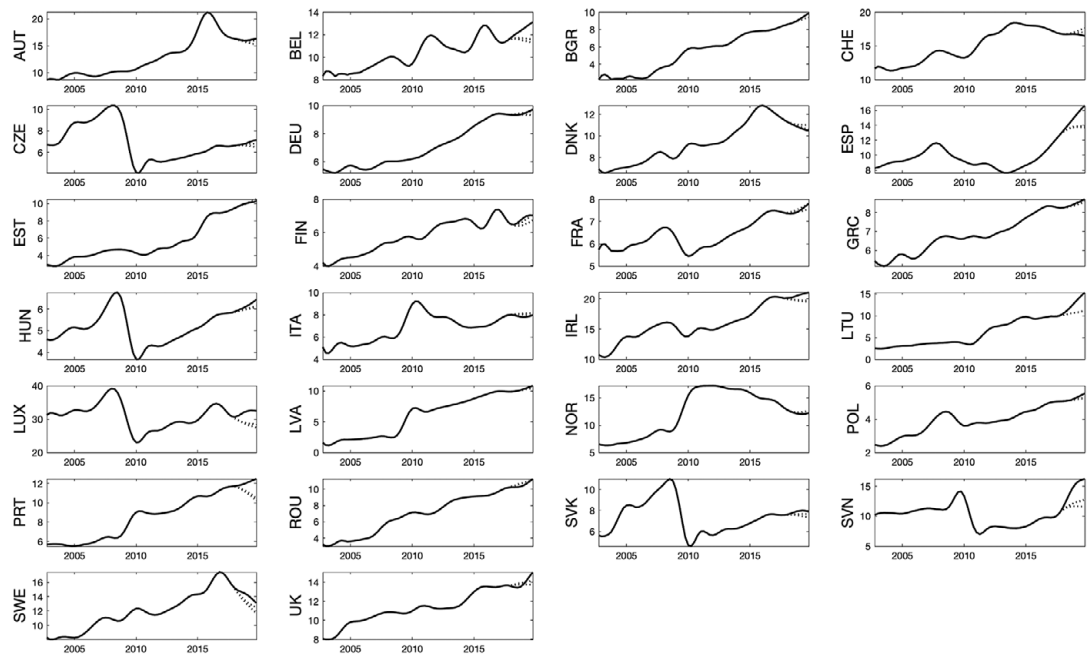
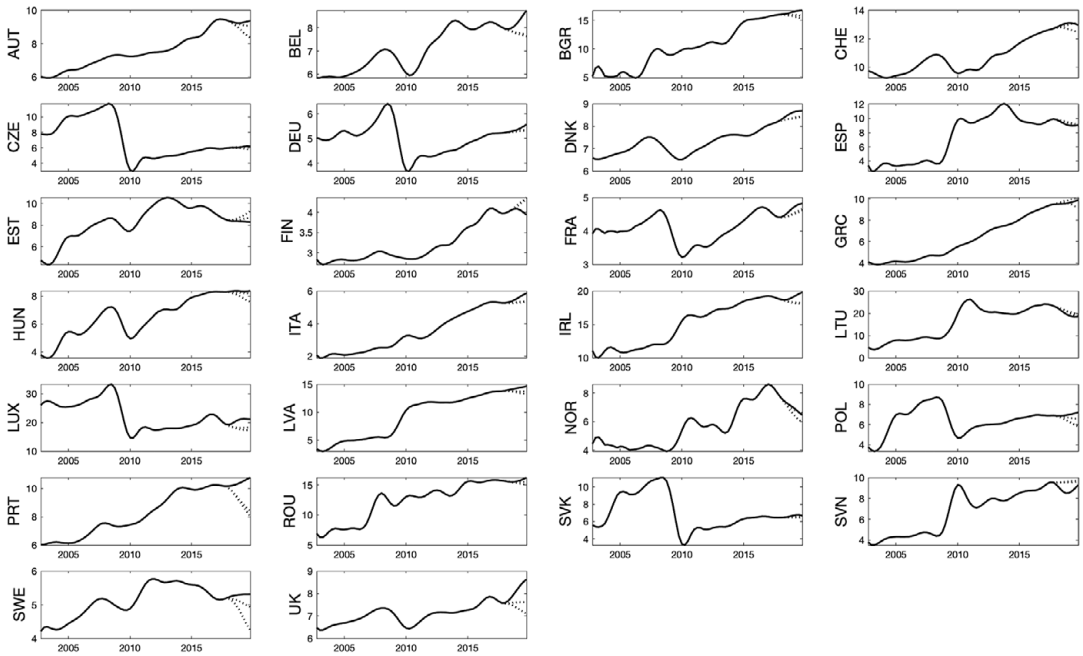
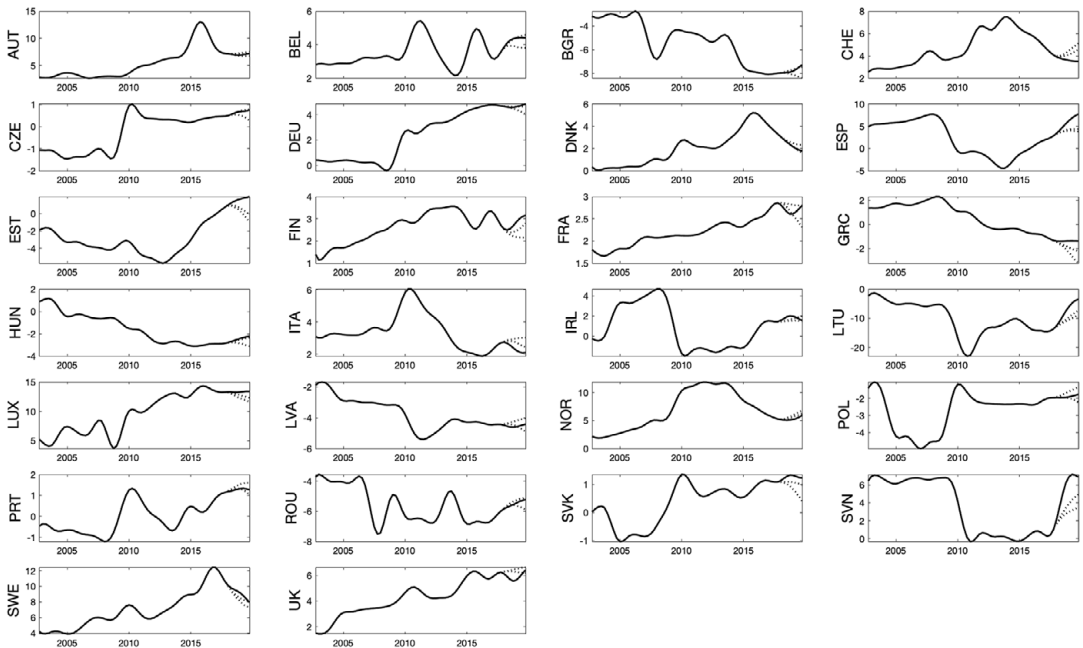


Figure 1. In-sample forecasting exercise for immigration “rates”, 2018–2019. Note: The solid black line represents the data used for estimation. The dotted lines depict the mean forecast and the 67% predictive intervals.





**Figure 2.** In-sample forecasting exercise for emigration rates, 2018–2019. Note: The solid black line represents the data used for estimation. The dotted lines depict the mean forecast and the 67% predictive intervals.



**Figure 3.** In-sample forecasting exercise for net migration “rates”, 2018–2019. Note: These forecasts are produced directly by model (5). The solid black line represents the data used for estimation. The dotted lines depict the mean forecast and the 67% predictive intervals.

**Table 3.** Analysis of forecast errors, 2018–19: selected summary measures

	Group 1			Group 2		
	Immigration	Emigration	Net migration	Immigration	Emigration	Net migration
RMSE	0.5365	0.3852	0.3649	0.3864	0.1930	0.3830
MAE	0.4698	0.3388	0.3139	0.3414	0.1662	0.3414
MAPE	2.1676	2.3614	5.3764	2.7858	2.2175	9.3299
Theil's <i>U</i>	0.0125	0.0139	0.0298	0.0166	0.0130	0.0556
	Group 3			Group 4		
	Immigration	Emigration	Net migration	Immigration	Emigration	Net migration
RMSE	0.3956	0.2893	0.3191	0.2426	0.2700	0.4275
MAE	0.3321	0.2442	0.2776	0.2082	0.2276	0.3715
MAPE	2.4184	2.1786	9.9789	1.8737	1.7204	11.4535
Theil's <i>U</i>	0.0151	0.0131	0.0615	0.0117	0.0101	0.0677

Note. RMSE = root mean square error; MAE = mean absolute error; MAPE = mean absolute percentage error; Theil's *U* = Theil's *U* statistic

**Table 4.** Forecast calibration: observation shares within 67% predictive intervals

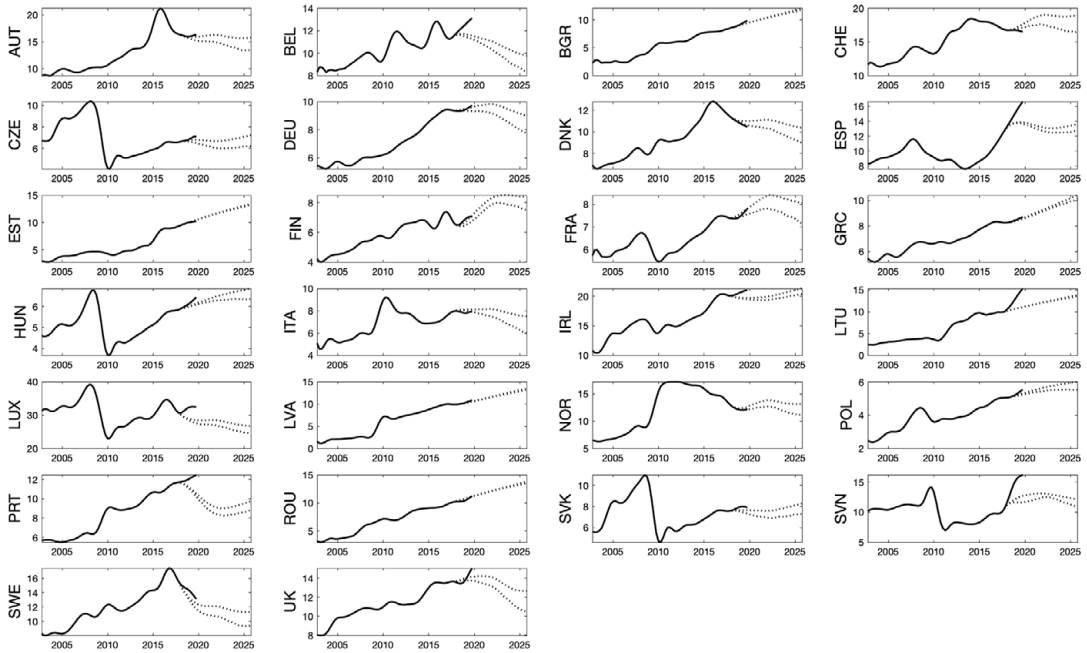
	Group 1			Group 2		
	Immigration	Emigration	Net migration	Immigration	Emigration	Net migration
Coverage	0.33	0.28	0.25	0.30	0.25	0.43
Abs. diff.	0.34	0.40	0.42	0.37	0.42	0.24
	Group 3			Group 4		
	Immigration	Emigration	Net migration	Immigration	Emigration	Net migration
Coverage	0.18	0.22	0.45	0.03	0.50	0.47
Abs. diff.	0.50	0.45	0.22	0.64	0.17	0.20

Note. Coverage = the average proportion of quarterly forecasts where the observation was within the bounds of 67% predictive intervals; Abs. diff. = absolute difference from the nominal coverage probability, 0.67.

In a similar way, Portugal's distortion for Group 3 and for Group 4 Estonia and Lithuania has different reasons for their larger error values. Estonia is on upwards trajectory, while Lithuania has the largest net emigration flows.

There are some countries for which the observed average values can be misleading, though, and be largely outside of the 67% predictive intervals with Luxembourg, Slovenia, Spain, Portugal, Estonia and Lithuania having the fewest observations within the bounds. The calibration results presented in Table 4 show that the errors are relatively small. For immigration and emigration, the empirical coverage of the 67% predictive intervals (the fraction of the observed data points that fell within the intervals) was too small, most notably for Group 3, and for immigration, Group 4.

Overall, in terms of the ex-post analysis of errors, short-term forecasts utilising macroeconomic drivers have yielded results with acceptable errors and mostly too narrow 67% predictive intervals between 2018Q1 and 2019Q4, with immigration and emigration forecasts, especially for the traditionally receiving countries, generally outperforming net migration forecasts. This is expected due to the models including additional variables for the home economies, rather than the respective origin or destination



**Figure 4.** Out-of-sample forecasts of immigration “rates”, 2018–2025. Note: The solid black line represents the data used for estimation. The dotted lines depict the mean forecast and the 67% predictive intervals.

countries of migrants. A more detailed investigation of the model performance for individual countries could additionally involve country-specific detail on microeconomic, political, policy or other reasons that could not be modelled by using macro-level variables, which goes beyond the scope of this study.

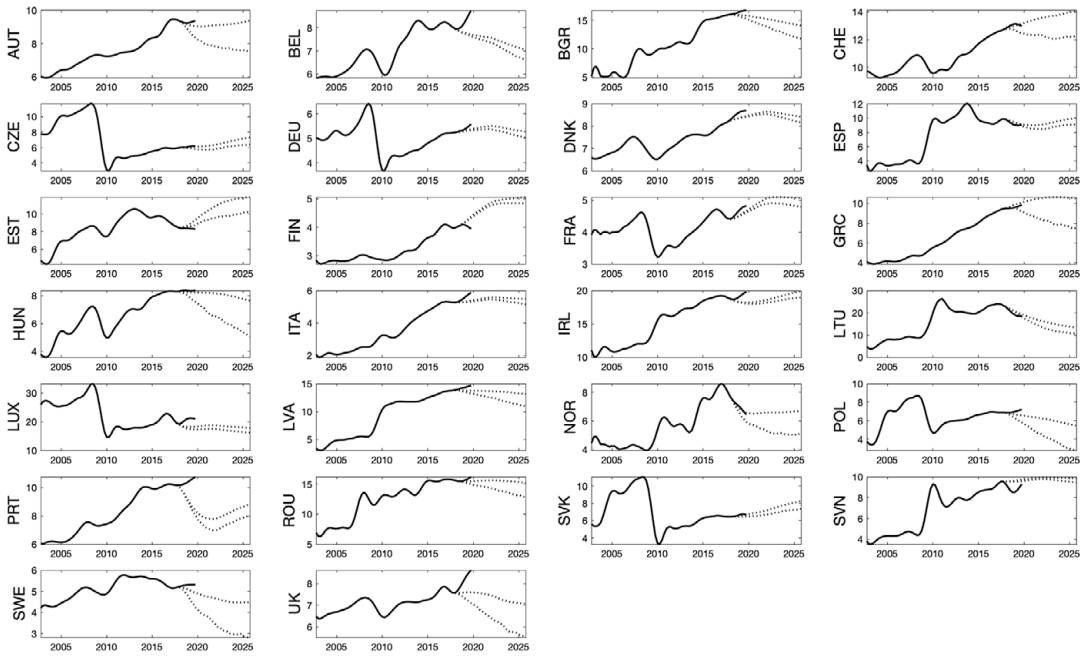
#### 2.4. Out-of-sample analysis of migration forecasts

This study relies on a set of quarterly data available up to 2019Q4. In this section, we present out-of-sample forecasts, with migration predicted for 2018–2024, so with a 5-year horizon. Conditional forecasts are possible at this horizon, however they are not presented here<sup>8</sup>. We use the same vectors of endogenous variables as in Section 2.2. Figure 4 presents the predictions for immigration “rates”, Figure 5 for emigration rates and Figure 6 for net migration “rates”. These forecasts are unconditional and are not corrected for the impact of the COVID pandemic.

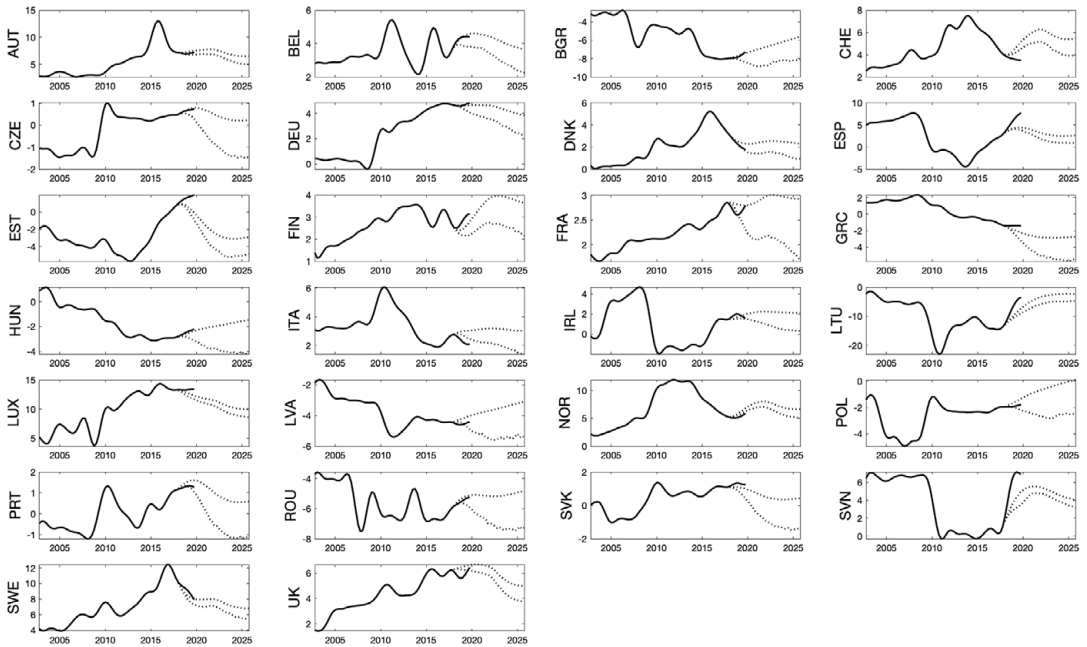
As before, the forecasts presented in Figures 4, 5 and 6 have tighter predictive intervals for some countries than others. Here again, this has to do predominantly with the presence of shocks in the past data series. Luxembourg is an understandable anomaly: the volatility of migration rates is driven by the small size of the country being at the same time at the “core” of the EU migration system. Some other instances can be down to significant economic and migration developments in these countries over a relatively short time period, whether it be joining the European common labour market as in the CEE countries, or prolonged effects from the financial crises in Southern European countries<sup>9</sup>.

<sup>8</sup> This is due to computational limitations and the high volatility of the covariates during the pandemic which restricts efficient learning process.

<sup>9</sup> Greece, Italy, Portugal, and Spain suffered longer effects to the macroeconomy as a result of the Global Financial Crisis 2007–2008 than other countries did, and experienced these effects on the labour markets for a number of years after the majority of countries had recovered to pre-crisis levels.



**Figure 5.** Out-of-sample forecasts of emigration rates, 2018–2025. Note: The solid black line represents the data used for estimation. The dotted lines depict the mean forecast and the 67% predictive intervals.



**Figure 6.** Out-of-sample forecasts of net migration “Rates”, 2018–2025. Note: The solid black line represents the data used for estimation. The dotted lines depict the mean forecast and the 67% predictive intervals.

Another major question is continuing emigration from the CEE countries from Group 4. These countries have experienced large emigration since joining the European common labour market, however, in some cases the emigration rates (for example from Estonia and Poland) have been diminishing, so already now they are becoming net receivers, so moving to Group 3. These economies have also closed the gap towards Southern Europe, which makes them more attractive destinations when compared to some other countries of the region. Other members of Group 4 still have highly negative net migration, which makes them distinct from the rest of Europe—only potential new members of the EU, such as Ukraine, Moldova, or Western Balkan countries, would have similar characteristics.

Even more importantly than providing uncertainty statements (predictive intervals), short-term forecasts presented in this section can help us make a preliminary assessment, what might happen after some exogenous shocks, such as a further expansion of the EU, for example, to include Ukraine. One of the exogenous variables that could be used for that purpose is the size of the common labour market, where significant changes only arise when new countries are added—or removed, as in the case of Brexit.

### 3. Towards long-term migration scenarios

#### 3.1. Extending VAR forecast horizons: results and limitations

By using the Bayesian panel VAR models introduced in Section 2, we also produced long-range forecasts for immigration, emigration and net migration rates up to 2050Q4, so with the horizon of 30 years—around one generation ahead. The extensions of forecasts for the 26 countries included in the analysis in Section 2 are presented in Figures 7, 8, and 9 for immigration, emigration and net migration, respectively, alongside the 67% predictive intervals.

As expected, the predictability of migration over such long horizons appears to be generally poor (Bijak, 2010; Bijak et al., 2019), not least in terms of too narrow predictive intervals. This is especially visible for emigration, and even more so for net migration. This seems to be a problem continuing from the short-term forecasts presented in Section 2, only exacerbated by a longer prediction horizon. The problems are visibly higher for countries that are net senders of migrants, such some Central and Eastern

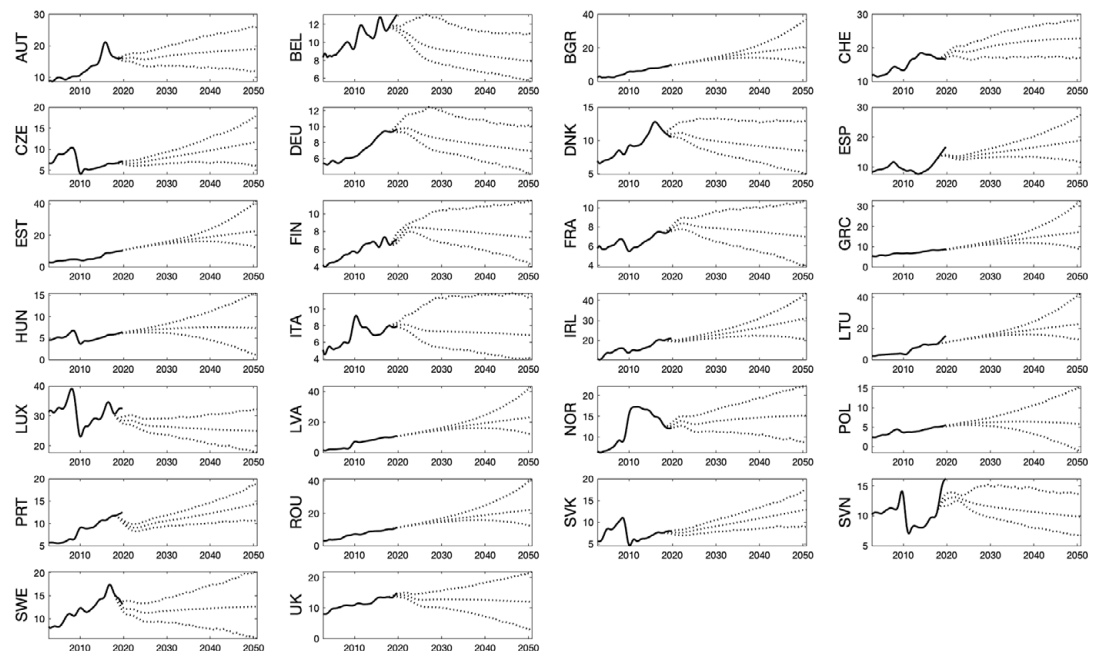
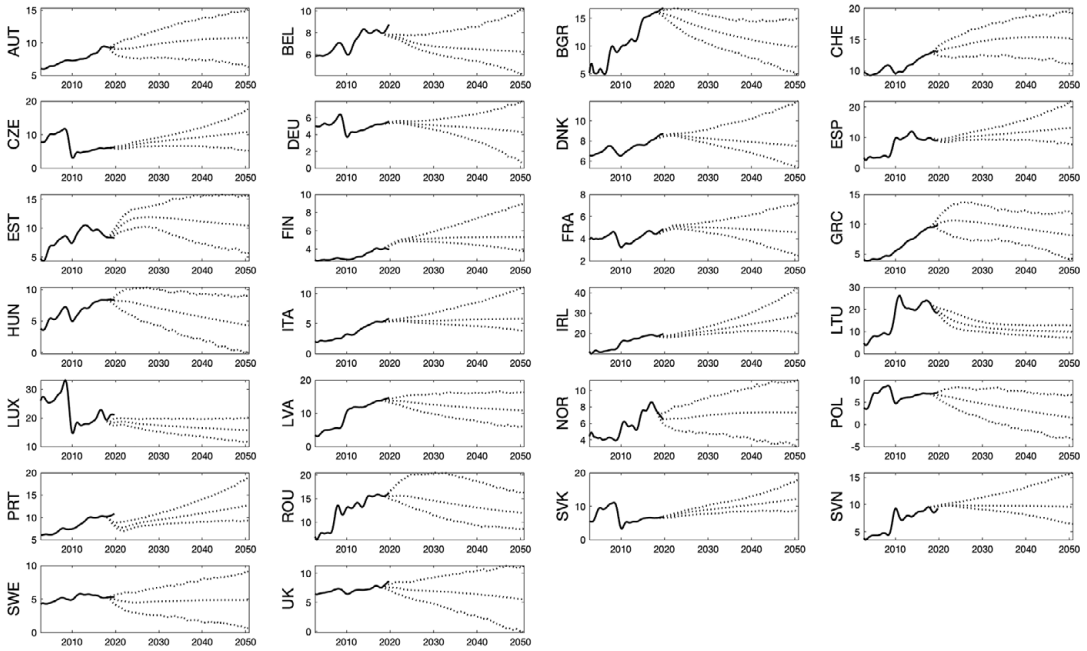
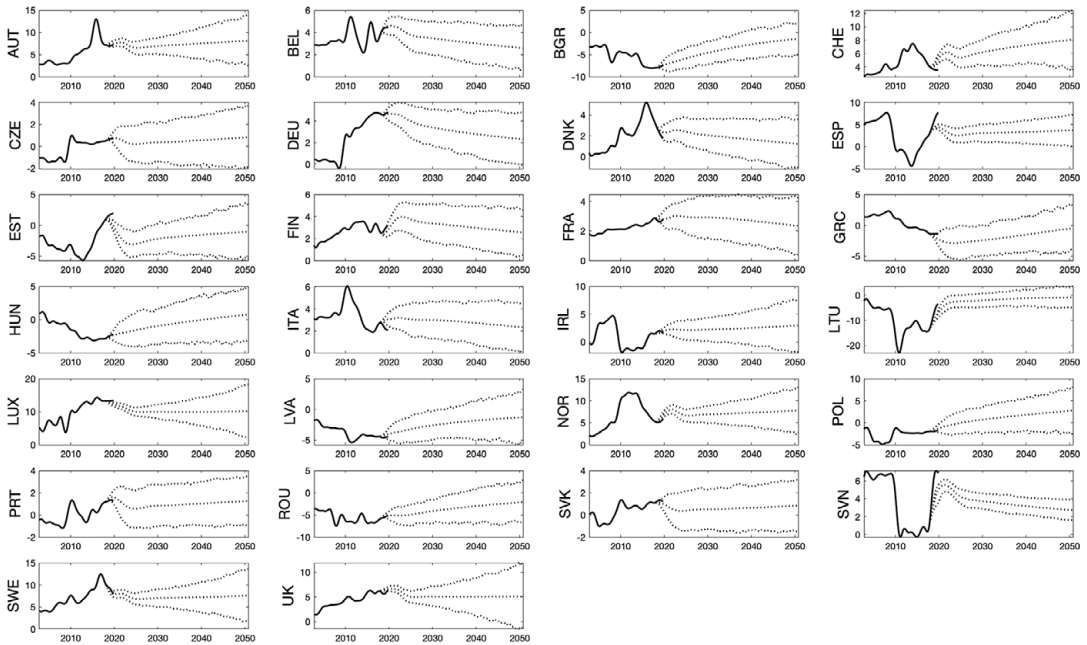


Figure 7. Long-range forecasts of immigration “rates”, 2020–2050.



**Figure 8.** Long-range forecasts of emigration rates, 2020–2050.



**Figure 9.** Long-range forecasts of net migration “rates”, 2020–2050.

European states. More generally, without controlling for population or, equivalently, the total labour market for the entire forecast period, the predicted emigration (and thus net migration) processes exhibit great variability. Especially the transition countries in Group 3 have typically experienced high volatility in migration flows over the sample period.



In general, how the population picture will look in 10, 20, or 30 years' time is hard to predict. It is possible to create conditional forecasts, as in macroeconomics, though these tend to be only short-term, based on external predictions of drivers where available. One example is GDP, with forecasts available from the OECD in a horizon of two or so years, which need to be replaced by model-generated forecasts thereafter, compounding the predictive uncertainty. In the longer term, the role of additional factors is becoming increasingly more likely—one example of such a factor is job automation, as discussed in the study by Barker and Bijak (2021). Attempting to predict migration with variables that may be even less predictable in the long run is futile, but at least the VAR models can be argued to offer a toolkit for creating long-range scenarios that retain macroeconomic coherence.

In the examples presented above, some countries have significantly large error bands towards the end of the forecast horizon. These are mainly countries which have had volatile periods of migration in the sample period, not fully captured by the macroeconomic or labour market data. Interestingly, the forecasts of net migration exhibit relatively most stable long-term tendencies in terms of their uncertainty bounds, similar to those obtained by Azose and Raftery (2015) based on univariate hierarchical models.

In general, the VAR models presented in this paper seem to have been able to identify different patterns of predictive uncertainty for various groups of countries. In particular, there are visible differences between countries, where the presence of high-magnitude unpredictable events in otherwise relatively stable trends is increasing the predictive errors comparatively slowly (Groups 1–3), as opposed to countries with visible changes in trends implying clearly non-stationary processes, as well as wide and rapidly increasing predictive intervals (mainly Group 4). This analysis allowed identifying, even if in qualitative sense, in which countries the predictive uncertainty is the greatest, as they may be (or soon become) at the verge of a migration transition from being net senders to receivers.

Somewhat predictably, the predictive performance of models equipped with macroeconomic drivers is mixed, and in many cases, such as for Central and Eastern European countries, undergoing a transition from net senders to net receivers, outright unsatisfactory. Whenever good data are available, and the underlying processes are relatively stable, as in the pioneering study of Gorbey et al. (1999) of migration between Australia and New Zealand, Bayesian VAR models offer an appealing way of providing coherent scenario trajectories describing the future uncertainty, including theoretical insights into structural mechanisms shaping population flows. In a general case, however, difficulties with predicting migration beyond a five- to 10-year horizon are well known (e.g. Bijak and Wiśniowski, 2010), which leaves the question of how to predict migration in the longer term to a large extent open. Still, the applied Bayesian perspective at least allows for handling time series irrespective of whether they are stationary (Koop et al., 2012)—an advantage for migration, where non-stationarity can often be expected (Bijak, 2010).

The short- and long-run forecasts presented here have extended the previous attempts of using Bayesian VAR methods to forecast migration (Gorbey et al., 1999; Bijak, 2010). The novel contribution has been the use of a much broader set of macroeconomic and labour market data, including data that come at different frequencies. These models have proved more successful for some countries than others—unsurprisingly, especially for those who were relatively unaffected by the shocks of economic and migration nature. Still, the large residual uncertainty indicates that the macro-level picture of future migration is far from complete, and that the only realistic solutions can be approximate.

### 3.2. *The impacts of uncertain migration and economic events*

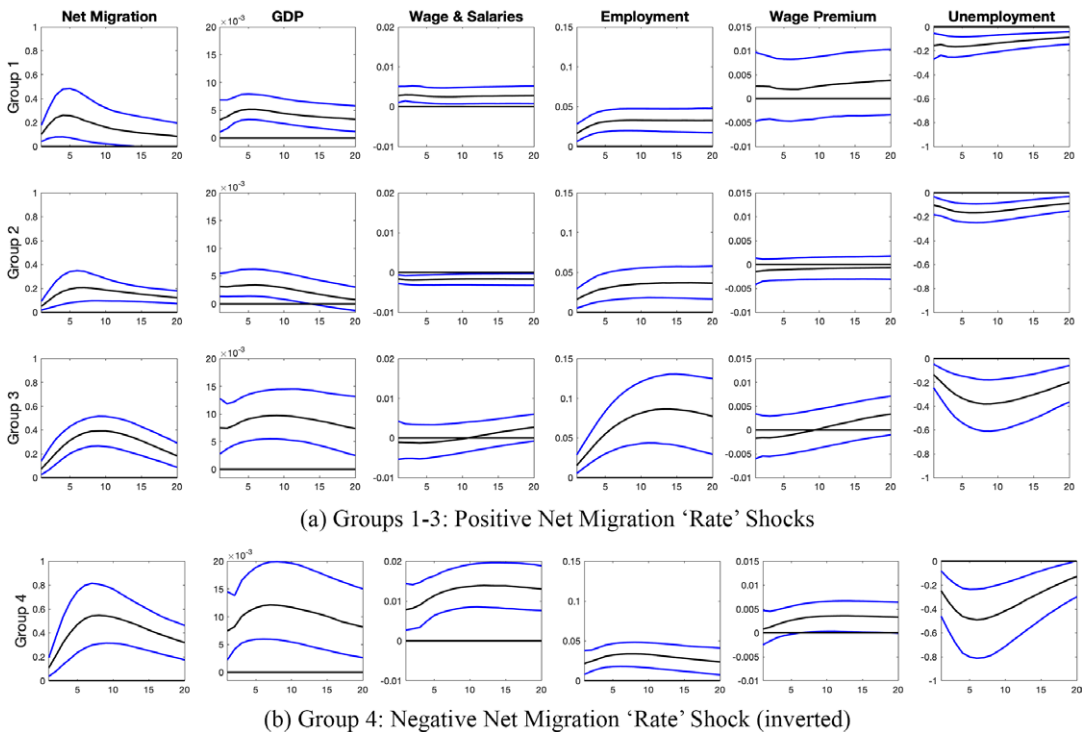
By using the panel VAR methodology described in Section 2, we can also analyse broader macroeconomic impacts of uncertain migration events. The approach uses the same model as above, only examined through the lens of impulse response functions (IRFs), which show the responses of individual model variables *conditional* on a one-off change to migration (impulse). The IRFs presented in this section are shown together with their 67% confidence (credible) intervals, demonstrating the uncertainty of the responses of individual variables to migration events in a particular scenario. These events are of the magnitude of one standard deviation estimated for the observed series, but of course for a policy analysis, this parameter can be arbitrarily changed, depending on the user needs. As is standard in the macroeconomic literature, we look at

one-time events in the first period under study. As before, default priors are used for estimating the Bayesian panel VAR models. For additional results and further models, see Barker (2021).

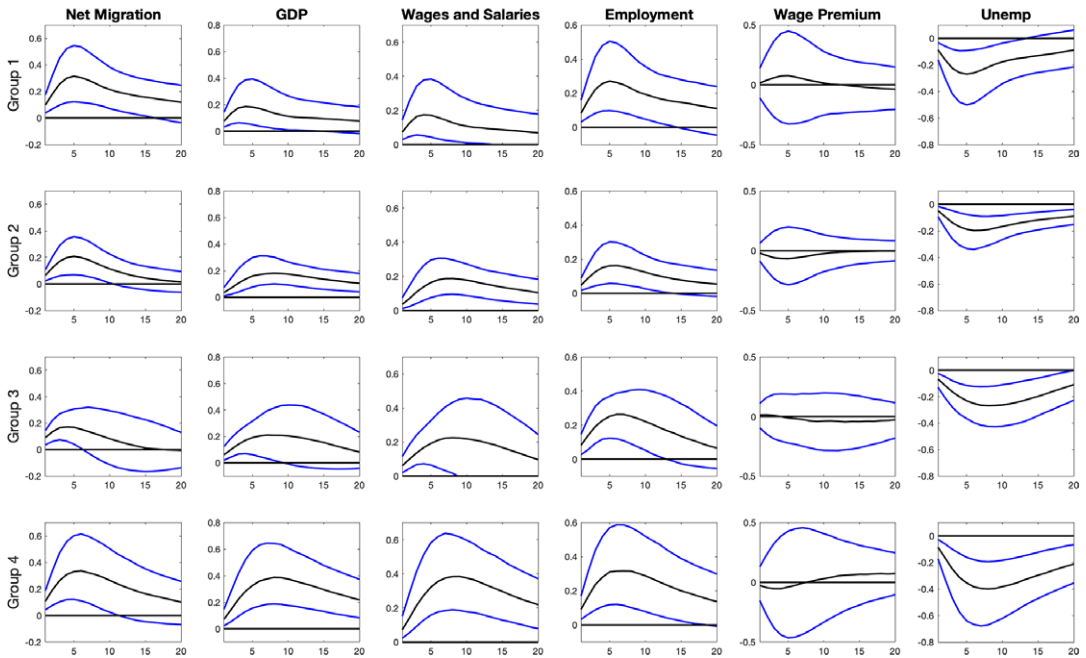
In this part of the paper, we assess the effects of a net migration shock on macroeconomic variables, and vice versa. The explanatory variables represent various migration drivers, which on their own can be subject to shocks of varying degree of predictability and susceptibility to being influenced for example by political and policy changes. The model we use for this analysis is based on the following vector of endogenous variables  $y_t$ , with net migration “rate” for the net receiving country Groups 1–3, and its negative for the net sending Group 4, so that in all cases the modelled values remain positive.

$$y_t = [\pm NM_t, GDP_t, WageSal_t, Emp_t, WagePre_t, Unemp_t]’ \tag{6}$$

The IRFs presented in Figure 10 show the impact a one-standard deviation increase to the net migration rate for the respective country groups. The results show that in Groups 1–3, increases to net migration are expansionary for the economy: GDP, wages, and salaries increase in real per capita terms, with the employment rate increasing and unemployment decreasing. Conversely, in Group 4, decline in (negative) net migration is contractionary, with the opposite effects for macroeconomic variables. This provides an interesting, policy relevant point for Group 4, suggesting that these are not the unemployed workers who



**Figure 10.** Impulse responses for net migration events. Impulse responses to a one standard deviation increase in the positive net migration “rate” (Groups 1–3) and decrease in the negative net migration “rate” (Group 4). The vertical axis identifies the responses in percentage deviations from trend. For the logged variables, the responses are provided in percentages, while for the unemployment rate, the response is in percentage points. The horizontal axis identifies the quarter after the shock, up to 5 years (20 quarters). Column headings identify the responding variables, and the row headings correspond to the country groups. The responses for Groups 1–3 are to a positive net migration shock, while the responses to a negative net migration shock to the variables for Group 4 have been visually inverted to aid comparison.



**Figure 11.** Impulse response of net migration to economic shocks. Note: Impulse responses of net migration to the one-standard-deviation shock increases of macroeconomic variables. The vertical axis identifies the responses in percentage deviations from trend. For the logged variables, the responses are provided in percentages, while for the unemployment rate, the response is in percentage points. The horizontal axis identifies the quarter after the shock, up to 5 years (20 quarters). Column headings identify the variables subject to shocks, and row headings—country groups. Note that for Group 4 of countries, the response to the negative of net migration is presented.

emigrate, but rather the employed workers. The axes have been normalised to reflect the relative size of the response between groups.

In Figure 11, we show the responses of the non-migration variables in the vector  $y_t$  to net migration “shocks.” Again, the responses are given by country group, and variables showing positive responses can be interpreted as “pull factors” (positive drivers) of migration, while negative responses indicating “push factors” (negative drivers). The responses show that increases to GDP, wages and salaries, employment increase immigration, with a rise in unemployment increasing emigration. As argued above, the results of this analysis can additionally aid anticipatory analysis of migration and its macroeconomic impacts, through providing a template for creating coherent “what-if” scenarios, with the methodology easily extendable to other sets of drivers, beyond macroeconomics. Increases to the wage premium have insignificant effects as the main focus is on the wages and salaries.

#### 4. Discussion and conclusion

The research presented in this paper has put forth a new approach to forecasting migration through macroeconomic models using mixed-frequency data. As a proof of concept, the results have indicated reasonable, yet variable performance in the short term, depending on the indicator used—with emigration bearing on average the lowest relative errors, and net migration the highest—as well as on the countries, with uncertainty larger in the countries undergoing migration transition than in the more established economies. In the light of previous work (e.g. Bijak et al., 2019), these findings are not surprising, as they confirm intuitions as to the key role of the stability of the predicted migration processes.

In general, the forecasts presented in this paper, both for the long- and short-term horizons, offer new prospects to forward-looking migration studies. Unlike for other components of population change, births and deaths, there are important macroeconomic drivers which represent the push and pull factors of migration. This allows us to obtain at least some indications of the future direction of change of migration flows, albeit often with very wide prediction intervals, by applying macroeconomic tools and techniques. There are, of course, outstanding issues related to systemic shocks (policy changes, lowering or raising of migration and trade barriers, or political crises), which are foreseeable to a very limited extent, and only at very short horizons.

VAR models and the associated the uncertainty analysis they offer can be also further extended by an explicit modelling of *uncertain shocks* and responses to them. In this way, empirically grounded predictive models, such as panel VARs, can shed light on different aspects of migration uncertainty. They can at least approximate the intrinsic randomness of the future through probabilistic description and contribute to reducing this uncertainty by looking at relationships between migration processes, their drivers and impacts. Some analytical tools, such as IRFs, can also help stress-test different aspects of migration scenarios or policy decisions. In all cases, using mixed-frequency models and data additionally helps make fuller use of all empirical information available on various migration drivers.

Future methodological work can additionally explore hybrid solutions, which could start from driver-based forecasts and scenarios in the short- to mid-term horizons, derived from time series models, and then use dynamic weighting to morph the associated probability distributions into purely expert-based scenarios for the long run. In addition, the relative advantages and limitations of using net migration rather than individual flows in the long run (as in Azose and Raftery 2015 or Section 3 above) warrant further exploration. In any case, opening up the modelling toolkits to approaches allowing the use of time series data at various frequencies is likely to help advance these methods further.

In the otherwise very uncertain world of migration and its forecasts and scenarios, one thing seems relatively certain: the assessment of migration uncertainty over longer time horizons remains a heroic exercise, and at the current state of knowledge, the best solutions are only approximate. Still, the analysis presented in this paper allows for making pragmatic choices: shorter-term forecasts and analysis of migration impacts, especially for countries with more stable migration processes and history, allow for using more complex models describing the economic and other aspects in finer detail, while longer-term scenarios, where the past information is less relevant, call for expert input at a very high level of abstraction. In either case, the crucial epistemological limits of any statements about future migration flows, and the associated caveats for the forecasters and forecast users alike (e.g. Bijak and Czaika, 2020), remain fully in force.

**Acknowledgements.** We are very grateful to Mathias Czaika, Valentina Di Iasio, Peter WF Smith, Jason Hilton, Joanne Ellison, and two anonymous referees, for their comments that helped improve earlier drafts. We acknowledge the use of the IRIDIS High Performance Computing Facility and associated support at the University of Southampton, in the completion of this work, with special thanks to Alister Boags. All the remaining errors and inaccuracies are ours. An early version of this paper is available as Barker and Bijak (2021).

**Author contribution.** E.B. performed the following: methodology, software, validation, formal analysis, investigation, data curation, writing (original draft, review, and editing), and visualisation. J.B. performed the following: conceptualisation, validation, writing (review and editing), supervision, funding acquisition, and project administration.

**Data availability statement.** The data that support the findings of this study are openly available in Zenodo at <https://doi.org/10.5281/zenodo.7709443>.

**Funding statement.** This work has received funding from the European Union's Horizon 2020 research and innovation programme, grant No. 870299 QuantMig: Quantifying Migration Scenarios for Better Policy, and a Horizon Europe project No. 101094741 FutuRes: Towards a Resilient Future of Europe, with this piece of work funded by the UK Research and Innovation through Horizon Europe Guarantee, grant no. 10066259. This document reflects the authors' views, and the Research Executive Agency of the European Commission is not responsible for any use that may be made of the information it contains.

**Competing interest.** No interest declared.

## References

- Azose JJ and Raftery AE (2015) Bayesian probabilistic projection of international migration. *Demography* 52(5): 1627–1650.
- Barker ER (2021) A Panel VAR Analysis of Migration in Europe. Working Paper, University of Southampton.
- Barker ER (2022) A timeline of freedom of movement in the European Economic Area. *Open Research Europe* 2: 133.
- Barker ER and Bijak J (2021) Uncertainty in Migration Scenarios. QuantMig Project Deliverable D9.2, University of Southampton, Southampton. Available at [http://www.quantmig.eu/project\\_outputs](http://www.quantmig.eu/project_outputs) (accessed 2 July 2024).
- Beine M, Bertoli S and Moraga JF-H (2016) A practitioners' guide to gravity models of international migration. *The World Economy* 39(4): 496–512.
- Bijak J (2010) *Forecasting International Migration in Europe: A Bayesian View*. Dordrecht: Springer.
- Bijak J and Czaika M (2020) Assessing Uncertain Migration Futures: A Typology of the Unknown. QuantMig Project. Technical Report D1.1, University of Southampton and Danube University Krems. Available at [http://www.quantmig.eu/project\\_outputs](http://www.quantmig.eu/project_outputs) (accessed 2 July 2024).
- Bijak J, Disney G, Findlay AM, Forster JJ, Smith PW and Wiśniowski A (2019) Assessing time series models for forecasting international migration: lessons from the United Kingdom. *Journal of Forecasting* 38(5): 470–487.
- Bijak J and Wiśniowski A (2010) Bayesian forecasting of immigration to selected European countries by using expert knowledge. *Journal of the Royal Statistical Society Series A* 173(4): 775–796.
- Canova F and Ferroni F (2020) A hitchhiker guide to empirical macro models. CEPR Discussion Papers 15446, C.E.P.R. Discussion Papers.
- Cipollini A and Parla F (2023) Temperature and Growth: A Panel Mixed Frequency VAR Analysis Using NUTS2 Data. Working Paper 155, RECent: Center for Economic Research.
- Czaika M, Erdal MB and Talleraas C (2021) Theorising the Interaction Between Migration-Relevant Policies and Migration Driver Environments. QuantMig Project. Technical Report D1.4, Danube University Krems and Peace Research Institute Oslo. Available at [http://www.quantmig.eu/project\\_outputs](http://www.quantmig.eu/project_outputs) (accessed 2 July 2024).
- Czaika M and Reinprecht C (2020) Drivers of Migration: A Synthesis of Knowledge. IMI Working Paper No. 163, International Migration Institute, University of Amsterdam, Amsterdam.
- Dieppe A, van Roye B and Legrand R (2016) The BEAR Toolbox. Working Paper Series 1934, European Central Bank.
- Dieppe A, van Roye B and Legrand R (2021) The Bayesian Estimation, Analysis and Regression (BEAR) Toolbox: User Guide. Version 5.0. Mimeo, European Central Bank. Available at [https://raw.githubusercontent.com/european-central-bank/BEAR-toolbox/master/documentation/BEAR\\_User\\_guide\\_v5.pdf](https://raw.githubusercontent.com/european-central-bank/BEAR-toolbox/master/documentation/BEAR_User_guide_v5.pdf) (accessed 2 July 2024).
- Gorbey S, James D and Poot J (1999) Population forecasting with endogenous migration: an application to Trans-Tasman migration. *International Regional Science Review* 22(1): 69–101.
- Kapetanios G, Marcellino M, Martino A, Papailias F and Mazzi GL (2021) Panel VAR Models for Nowcasting. Working Paper 2021/5, King's Business School, King's College, London.
- Koop G, Poirier DJ and Tobias JL (2012) *Bayesian Econometric Methods*. Cambridge: Cambridge University Press.
- Massey DS, Arango J, Hugo G, Kouaouci A, Pellegrino A and Taylor JE (1993) Theories of international migration: A review and appraisal. *Population and Development Review* 19(3): 431–466.
- Raymer J, Wiśniowski A, Forster JJ, Smith PWF and Bijak J (2013) Integrated modeling of European migration. *Journal of the American Statistical Association* 108(503): 801–819.
- Rogers A (1990) Requiem for the net migrant. *Geographical Analysis* 22(4): 283–300.
- Schorfheide F and Song D (2015) Real-time forecasting with a mixed-frequency VAR. *Journal of Business & Economic Statistics* 33(3): 366–380.

## Appendix

### A.1. Details on data

**Exclusions:** The study excludes Netherlands due to data quality issues leading to inability to classify the country well, as well as to the presence of non-macroeconomic rules and policies that make migration difficult to explain with macro-level data. The Netherlands is one of the most advanced EU+ economies, yet, prior to 2009, according to the harmonised IMEM estimates; the Netherlands might have experienced a period of negative net migration, which is inconsistent with the official published data. Croatia, Cyprus, Iceland, and Malta are excluded due to lack of availability of some macroeconomic data.

**Independent variables:** For the variable that describes the increase in the size of the common labour market, the country is only included in its first full period of inclusion. For example, the 2004 EU enlargement took place on 1 May 2004, which is part way through the second quarter of the year. In this case, in our model, the increase in the size of the labour market is assumed from the third quarter of 2004 onwards, as due to job search and matching frictions in the labour market, full effects of the enlargement were unlikely to be seen in the second quarter.

---

**Cite this article:** Barker ER and Bijak J (2025). Mixed-frequency VAR: a new approach to forecasting migration in Europe using macroeconomic data. *Data & Policy*, 7: e3. doi:10.1017/dap.2024.82