International Journal of Data Science and Analytics (2026) 22:19
https://doi.org/10.1007/s41060-025-01008-2

REGULAR PAPER

Forecasting refugee migration with high-dimensional covariate space

Haodong Qi[1,2] · Alina Sîrbu[3] · Rahman Momeni[4] · Enes Hisam[5] · Carlos Arcila-Calderón[6] · Tuba Bircan[7] · Stefano Iacus[8]

Received: 8 May 2025 / Accepted: 19 December 2025

Abstract

Forecasting refugee migration is challenging, exacerbated by the high dimensional and dynamic nature of its drivers, such as climatic, economic, and political stressors. This article introduces a novel forecasting framework based on the Dynamic Elastic Net (DynENet) algorithm, which incorporates a time-varying regularization and a new model selection criterion: the penalized deviance ratio (PDR). Unlike conventional metrics such as the deviance ratio (DR), which emphasize in-sample fit, PDR explicitly penalizes model complexity, enhancing generalization in high-dimensional covariate setting. We apply this framework to forecast asylum-seeker rates (ASR) from Somalia to EU member states, leveraging a comprehensive set of district-level predictors. Extensive validation demonstrates that PDR-tuned models consistently outperform DR-based benchmarks in out-of-sample accuracy, reducing average point prediction errors by 40% and improving interval forecasts by 79%. Furthermore, we demonstrate how the DynENet framework supports explanatory insights at multiple levels—origin district, destination, and temporal—revealing both persistent and transient nature of migration drivers. The proposed methodology not only advances forecasting accuracy under high-dimensional covariate conditions but also enhances the interpretability of complex and evolving migration systems.

Keywords Migration forecasting · Feature selection · Model complexity · Penalized deviance ratio

1 Introduction

Forecasting refugee migration is a critical task, with profound implications for anticipatory humanitarian response as well as for ensuring safe and orderly migration. However, building accurate forecasting systems for forced migration remains a challenge. The drivers of refugee migration are inherently complex, spanning environmental, political, and socioeconomic factors, each operating across multiple temporal and spatial scales. Traditional approaches, particularly those based on the gravity equation, are often fall short in this context [1]. While gravity models offer broad insights into aggregate migration patterns, they fail to capture the crisis-driven, dynamic nature of forced displacement, resulting in limited predictive utility. In this article, we propose a novel machine learning framework designed to model asylum-related migration as a complex yet explainable system.

Recent methodological advances have attempted to address these limitations. In particular, the flow-specific temporal gravity (FTG) model has been shown to better capture temporal variation in refugee flows [2]. By exploiting within-dyad temporal structure, the FTG model improves predictive performance relative to standard gravity-based models. Building on this foundation, we extend the FTG framework to incorporate not only temporal but also spatial heterogeneity in migration responses. Specifically, we enrich the FTG model with sub-national, high-resolution covariates, enabling it to reflect localized shocks and conditions that drive refugee movements.

Incorporating highly granular spatiotemporal data introduces substantial methodological challenges. It leads to a

✉ Haodong Qi
haodong.qi@mau.se

1 Malmö University, Nördenskildsgatan 1, Malmö 205 06, Sweden

2 Stockholm University Demography Unit, Stockholm, Sweden

3 Department of Computer Science, University of Pisa, Pisa, Italy

4 European Space Agency, Harwell, United Kingdom

5 GMV, Oxfordshire, UK

6 University of Salamanca, Salamanca, Spain

7 Vrije Universiteit Brussels, Brussels, Belgium

8 Harvard University, Cambridge, USA

Published online: 10 January 2026

high-dimensional covariate space, composed of numerous time series potentially associated with refugee flows at varying lags. This presents a particularly difficult problem in migration forecasting, where outcome data are typically low frequency and short in duration. The central challenge, therefore, is how to reduce dimensionality effectively without compromising predictive accuracy.

To address this challenge, we propose a novel feature selection metric: the penalized deviance ratio (PDR). PDR extends the traditional deviance ratio (DR) by introducing a penalty term that discourages overly complex models, thus enhancing generalization performance in high-dimensional settings. We apply this framework to the case of refugee migration from Somalia to the European Union (EU), using time- and space-varying covariates that capture conflict intensity, climate variability, and socio-political instability. While the PDR-based model shows inferior fit on training data relative to DR, it achieves superior predictive performance on unseen data. Crucially, beyond predictive gains, the model structure selected by PDR offers interpretable insights into both the empirical signals—such as localized droughts or violent events—and the conceptual mechanisms, including delayed responses to shocks and regional spillover effects, that underpin forced migration. By revealing which features, locations, and lag structures contribute most to migration flows, our approach offers not only a forecasting tool, but also an explanatory lens into the drivers of forced migration.

2 Related work

Traditional migration models grounded in the gravity equation are foundational for macro-analyses of how bilateral flows relate to economic and sociopolitical indicators in both origin and destination countries. More recently, these models have been extended to incorporate climate and environmental variables (see, e.g., [3–7]). While they remain foundational for understanding broad patterns of migration [8, 9], gravity-based models struggle to capture the complexity of forced migration and often perform poorly in forecasting contexts.

Rooted in the random utility framework, gravity models typically assume that migrants are forward-looking, utility-maximizing agents who move voluntarily in pursuit of better opportunities or improved living conditions. However, this assumption has been challenged by the literature on involuntary immobility, which highlights that many individuals lack the capability or resources to migrate, even when faced with existential threats [10–16]. These models also tend to impose homogeneous parameter estimates across dyads, leading to exaggerated or overly generalized conclusions about migration responses to economic changes [8, 9, 17–20], as well as to variations in climatic conditions [3–7].

Recent studies have demonstrated that when model parameters are allowed to vary across origin–destination dyads, the estimated migration responses reveal pervasive heterogeneity [2, 21]. This heterogeneity reflects contextual factors such as geographic variation, capability to migrate, and the influence of migration-related policies [11, 16, 22]. Moreover, the drivers of migration are not static and they evolve over time. Abrupt events like conflict or natural disasters can generate sudden displacement, while slower-onset phenomena such as environmental degradation may lead to gradual and persistent mobility trends. Current models largely ignore these temporal distinctions, assuming stable parameter estimates and overlooking the transient or enduring nature of migration responses.

In addition to these limitations in explanation, traditional models are also weak in prediction. Many introduce fixed effects to adjust for time-invariant flow characteristics (e.g., cultural or geographic proximity), but these so-called fixed-effects gravity models are not designed for forecasting [1, 2]. Although some studies have used such models for forward-looking projections, their predictive validity is questionable [3, 16]. These practices risk reinforcing deterministic narratives of 'mass' migration, which may be misinterpreted in policy and media discourse [23, 24].

To address the shortcomings of the migration models discussed above, we identify three key avenues for methodological advancement. First, models need to be parameterized in a less restrictive fashion such that they can account for non-uniform and non-spontaneous mobility responses. Second, models need to be more adaptive such that they can capture temporal dynamics of migration drivers—time-varying relationships between predictors and migration outcomes. Third, implementing rigorous validation to ensure models are robust and reliable in out-of-sample forecasting. Our work contributes to each of these areas.

Our approach builds on the dynamic elastic net (DynENet) algorithm, recently proposed for forecasting asylum-related migration [25], and extends it in three important ways. First, we introduce a new regularization-based model selection criterion—the Penalized Deviance Ratio (PDR)—to guide hyperparameter tuning in DynENet. This method addresses the challenge posed by high-dimensional time-series data with short outcome histories, helping reduce overfitting while preserving predictive performance. Second, we expand the DynENet model estimated in [25] by adding a layer of climate indicators derived from satellite imagery. This addition allows us to further investigate how refugee migration may be driven by environmental stressors. Third, we aggregate predictors at the sub-national instead of country-level. This increased spatial resolution allows us to test DynENet's ability to handle high-dimensional covariates; it also provides additional insights into how the magnitude of migration drivers may vary within a country of origin—source of het-

erogeneity that has been largely overlooked in the existing literature on international migration.

3 A framework for migration forecasting with high-dimensional covariate space

As discussed at the outset, traditional gravity models are inadequate in forecasting future migration flows due to restrictive assumptions posed on model parameters. Specifically, those models assume migration drivers to be homogeneous across contexts (countries). Under such assumptions, models cannot effectively explain or predict the temporal dynamics of migration outcomes [1, 2].

To address this shortcoming, we impose fewer restrictions on model parameters and specify our model as:

$$Y_{i,t} = b_{0,i} + \sum_{k=1}^{K} \sum_{\theta_i=1}^{T-1} b_{k,i}^{\theta_i} L^{\theta_i} X_{k,i,t} + \epsilon_{i,t} \tag{1}$$

where $Y_{i,t}$ denotes the migration flow between an origin–destination dyad i at time t; $L^{\theta_i} X_{k,i,t}$ represents the k-th predictor at lag θ_i; T is the length of the time series; and $\epsilon_{i,t}$ is a normally distributed error term with zero mean and constant variance.

Equation (1) comprises multiple time-series models stratified by dyad and is known as the flow-specific temporal gravity (FTG) model—a new class of migration models designed to capture temporal migration patterns more accurately [2]. In FTG, model parameters vary across flows i, isolating spatial correlation and relying on temporal variation for identification.

Since the covariate space spans $K \times (T-1)$ dimensions, estimating Eq. (1) can be challenging, particularly for large K and small T. We address this challenge through a six-step framework (Fig. 1):

Step 1: Data preparation

Develop time-series indicators from structured and unstructured data (e.g., satellite imagery, global news), convert them into tabular format, and align with migration flow time series.

Step 2: Lead–Lag analysis

To reduce the dimensionality of the covariate space, we perform an initial selection of predictors before mapping them into the DynENet algorithm. Specifically, following [25], we use the lead–lag analysis to (i) select the predictors that have significant correlations with the $Y_{i,t}$ and (ii) find the optimal lag length for each selected predictor.

The lead–lag approach is considered a robust alternative to the Granger causality test [26]. It uses the Hayashi–Yoshida estimator to evaluate the correlations between x_t and $y_{t+\theta}$ and to find the optimal lag $\widehat{\theta}$ by maximizing the contrast function with respect to θ [27–30]. Let $\theta \in (-\delta, \delta)$ be time lags between predictor x_t and outcome y_t, and let $U_s(\theta)$ be a contrast function. The optimal lag $\widehat{\theta}$ is estimated as:

$$\widehat{\theta} = \text{argmax}_{-\delta<\theta<\delta} |U_s(\theta)| \tag{2}$$

If $\widehat{\theta} > 0$, x_t leads y_t by an amount of $\widehat{\theta}$, and conversely if $\widehat{\theta} < 0$, x_t lags y_t by an amount of $\widehat{\theta}$. A statistical test is performed to assess whether $\widehat{\theta}$ is significantly different from zero. Moreover, when $\widehat{\theta} > 0$ and $\widehat{\theta} < 0$ are both significant, the lead–lag-ratio (LLR) test is performed to determine whether the lead or lag effect dominates.

In our framework, we use the `yuima` R package [31] to select predictors to be mapped into the DynENet model and to identify the optimal lag of each selected predictor. Specifically, for a given predictor $x_{t-\widehat{\theta}}$ to be selected, it needs to satisfy: (i) x_t leads y_t; (ii) the corresponding lead–lag ratio (LLR) for $\widehat{\theta} > 0$ relative to $\widehat{\theta} < 0$ is greater than 1; and (iii) $\widehat{\theta} > 0$ is significant at 95% confidence level.

Given the lead–lag analysis, Eq.(1) is reduced to K dimensions in the covariate space:

$$Y_{i,t} = b_{0,i} + \sum_{k=1}^{K} b_{k,i} L^{\widehat{\theta_{k,i}}} X_{k,i,t} + \epsilon_{i,t} \tag{3}$$

where $\widehat{\theta_{k,i}}$ is estimated by the contrast function Eq. (2) for each dyad flow i and predictor k.

Step 3: DynENet hyperparameter tuning

We use the Dynamic Elastic Net (DynENet) [25], which regularizes regression over rolling windows. Its objective function is:

$$\min_{\beta_i} \left\{ \frac{1}{T} \sum_{t=1}^{T} L(Y_{i,t}, X_{i,t}\beta_i) + \frac{\lambda_i}{2} \left[(1-\alpha_i)\beta_i^2 + 2\alpha|\beta_i| \right] \right\} \tag{4}$$

where $X_{i,t}\beta_i = b_{0,i} + \sum_{k=1}^{K} b_{k,i} L^{\widehat{\theta_{k,i}}} X_{k,i,t}$. T is the length of time-series data, $L(.)$ is a loss function, λ_i determines the magnitude of penalty on β_i, and α_i is a mixing factor determining the fraction of penalty applied to β_i^2 and to $|\beta_i|$, respectively.

Equation (4) combines two types of penalized regression: Ridge and LASSO (Least Absolute Shrinkage and Selection Operator). For $\alpha_i = 0$, Eq. (4) is a Ridge regression which will shrink the coefficients through the penalty factor $\frac{\lambda_i}{2}\beta_i^2$.

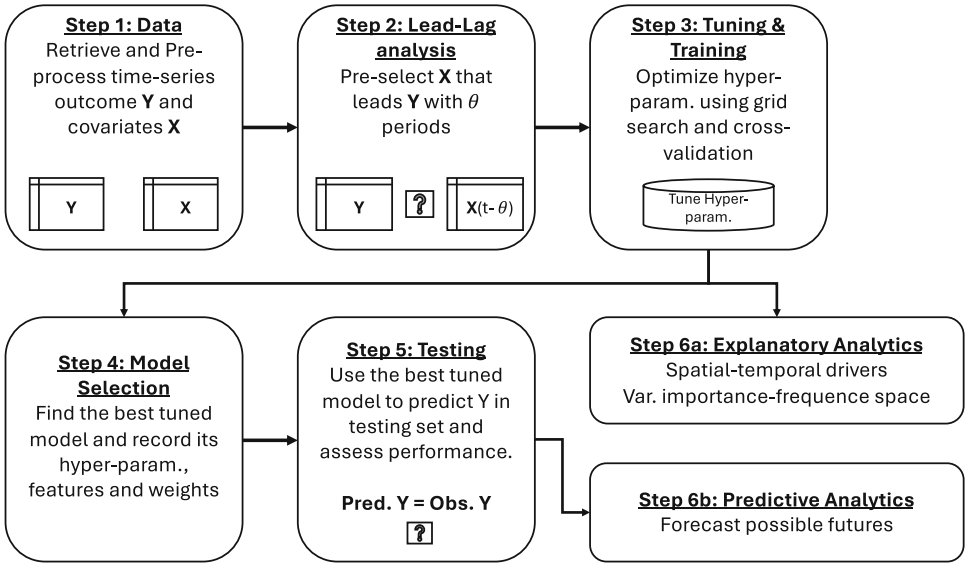

Fig. 1 Methodological framework

For $\alpha_i = 1$, Eq.(4) becomes a LASSO regression, which will zero out the coefficients through the penalty factor $\lambda_i|\beta_i|$. When $\alpha_i = 0.5$, the model becomes the DynENet with half Ridge and half LASSO regression. This mix is considered a good compromise in terms of prediction and interpretation [25].

We set $\alpha_i = 0.5$ and tune λ_i over a predefined grid for each rolling window using a new metric, penalized deviance ratio (PDR):

$$\text{PDR}_{i,w} = \left[1 - \frac{\sum_{t_w=1}^{T_w}\left(Y_{i,t_w} - \widehat{Y_{i,t_w}}\right)^2}{\sum_{t_w=1}^{T_w}\left(Y_{i,t_w} - \frac{\sum_{t_w=1}^{T_w} Y_{i,t_w}}{T_w}\right)^2}\right] \times \frac{T_w - 1 - k_{i,w}}{T_w - 1} \tag{5}$$

where T_w is the length of the training data within each fold, $k_{i,w}$ is the number of selected features for each flow in each fold.

In essence, $\text{PDR}_{i,w}$ measures the fraction of null deviance in Y_{i,t_w} (i.e., the sum of squared deviations from the unconditional mean of Y_{i,t_w}) explained by the model after adjusting for the number of selected covariates $k_{i,w}$. The metric is a decreasing function of $k_{i,w}$, and hence, the deviance ratio is penalized when model complexity increases. The optimal $\lambda_{i,w}$ is chosen from the tuning regression that produces the highest value of $\text{PDR}_{i,w}$.

The tuning procedure described above aims to prevent the migration model from overfitting, enhancing its predictive accuracy. Specifically, the optimized $\lambda_{i,w}$ balances the trade-off between model complexity and training fit, which is critical for our model to make stable and robust extrapolation into possible futures. Moreover, the rolling window design for tuning $\lambda_{i,w}$ is valuable for uncovering how the complexity of asylum-related migration systems may evolve overtime. In particular, when $\lambda_{i,w}$ changes, the number of predictors (i.e., migration drivers) retained as well as their weights may vary accordingly.

Step 4: Model selection

Once the optimal λ_i for each flow is found, we store its corresponding model features and weights, i.e., the estimates of $b_{k,i}$ and $b_{0,i}$ in Eq.(3).

Step 5: DynENet forecasting model testing

The performance of the best-tuned model is evaluated by comparing its predictions against the observed values in the testing set. We use two metrics: sum of squared errors (SSE) and prediction interval score (IS).

SSE assesses how well the point predictions made by Eq. (3) may resemble the actual outcomes:

$$\text{SSE}_i = \sum_{t=1}^{T}(Y_{i,t} - \widehat{Y_{i,t}})^2 \tag{6}$$

Prediction IS assesses the quality of probabilistic forecasts, by assigning a numerical score based on the predictive distribution and on the actual data points of the outcome [32]. Considering the prediction interval of $\widehat{Y_{i,t}}$ to be at $(1-\alpha)\times 100\%$ confidence level, the upper and lower bounds of predictive quantiles are $Q_{1-\alpha/2}(\widehat{Y_{i,t}})$ and $Q_{\alpha/2}(\widehat{Y_{i,t}})$, respectively, and the IS for the predictive interval is defined as:

$$\text{IS}_\alpha\left(Q_{1-\alpha/2}(\widehat{Y_{i,t}}),\ Q_{\alpha/2}(\widehat{Y_{i,t}});\ Y_{i,t}\right)$$

$$= \left(Q_{1-\alpha/2}(\widehat{Y_{i,t}}) - Q_{\alpha/2}(\widehat{Y_{i,t}})\right) + 2/\alpha \left(Q_{\alpha/2}(\widehat{Y_{i,t}}) - Y_{i,t}\right) \mathbb{I}\left(Y_{i,t} < Q_{\alpha/2}(\widehat{Y_{i,t}})\right) + 2/\alpha \left(Y_{i,t} - Q_{1-\alpha/2}(\widehat{Y_{i,t}})\right) \mathbb{I}\left(Y_{i,t} > Q_{1-\alpha/2}(\widehat{Y_{i,t}})\right) \quad (7)$$

where $\mathbb{I}(\cdot)$ is an indicator function equals to 1 if the condition is true, 0 otherwise.

Step 6: Explanatory and predictive analytics

A unique feature of our proposed framework is that it can not only provide a tool for predictive analytics based on a rigorously tested model in Step 5, but also offer insights into how the importance of different migration drivers may vary across time and space based on Step 3.

For predictive analytics, our approach may provide both point and interval forecasts. The probabilistic interval of predictions can be obtained through a bootstrapping procedure. Specifically, for each flow i and for S forecasting steps, we can draw S residuals from the training set, assuming the residuals are independent over time. These residuals are then added to $\widehat{Y_{i,t}}$.

$$\widetilde{Y_{i,t+s}} = \widehat{Y_{i,t+s}} + \widetilde{\epsilon_{i,t+s}},\ \widetilde{\epsilon_{i,t+s}} \overset{\text{i.i.d.}}{\sim} \text{Uniform}\left(\{\widehat{\epsilon_{i,1}}, \widehat{\epsilon_{i,2}}, \ldots, \widehat{\epsilon_{i,T}}\}\right) \quad (8)$$

where $\hat{\epsilon}$'s are residuals from Eq. (3), and T is the length of each flow's time-series in the training set.

Repeating this procedure for many times yields a sample of $\widetilde{Y_{i,t+s}}$, and the $Q_{\alpha/2}(\widetilde{Y_{i,t+s}})$ and $Q_{1-\alpha/2}(\widetilde{Y_{i,t+s}})$ of the sampled $\widetilde{Y_{i,t+s}}$ constitute the $(1-\alpha) \times 100\%$ prediction interval.

The rolling window cross-validation procedure in Step 3 provides a basis for explanatory analytics. Specifically, the flow-specific and time-varying weights allow us to investigate how migration responses to different factors varied across origins and destinations as well as over time. We also construct an importance-frequency (IF) space to analyze the relationship between the relative importance of predictors and how often different covariates are selected by the DynENet model.

In our framework, we define the relative importance of predictors as:

$$\text{Rank}_{i,k} = \frac{1}{W} \sum_{w \in W}^{W} \frac{\text{vip}_{w,k,i}}{\sum_k \text{vip}_{w,k,i}} \quad (9)$$

where w is an index for each rolling window or fold, and W is the total number of folds. $\text{vip}_{w,k,i}$ is the variable importance of a given predictor $L^{\widehat{\theta_{k,i}}} X_{k,i,t}$ in Eq. (3). It quantifies how much a model's performance would change if the effect of a selected predictor is removed through permutation of the predictor values [33].

The frequency of a predictor being selected is defined as:

$$\text{Freq}_{i,k} = \frac{1}{W} \sum_{w \in W}^{W} \mathbb{I}(b_{w,k,i} \neq 0) \quad (10)$$

where w is an index for each rolling window or fold, and W is the total number of folds. $\mathbb{I}(\cdot)$ is an indicator function equal to 1 if the parameter $b_{w,k,i}$ in Eq. (3) is nonzero, 0 otherwise.

4 Case study: Somalia-EU refugee migration

In this section, we demonstrate the application of our proposed framework using a case study on refugee migration from Somalia to the European Union (EU). While the framework is broadly applicable across migration contexts, we focus here on this specific case to manage the computational demands of processing high-resolution spatiotemporal satellite data. Somalia faces severe development challenges, with approximately 72% of the population living below the international poverty line of $2.50 per day [34]. More urgently, the country has been increasingly affected by extreme weather events, which undermine fragile, agriculture-based livelihoods and contribute to widespread food insecurity and human displacement.

4.1 Data preparation

This case study integrates multiple data sources. To capture climatic conditions, we process gridded remote sensing data and derive two key indicators:

1. Standardized Precipitation and Evapotranspiration Index (SPEI): a normalized index measuring the intensity of extreme climate conditions [35]. A value of $+1$ or -1 denotes wet or dry conditions one standard deviation from the mean. Unlike single metrics (e.g., temperature or rainfall), SPEI captures the net water balance between precipitation and evapotranspiration.
2. Soil Moisture Index (SMI): derived from satellite observations from the European Space Agency Climate Change Initiative Soil Moisture Climate Copernicus. It measures the water content in soil layers, critical for understanding agricultural stress.

Both indices are aggregated at the district level across Somalia. Due to resource constraints, we process historical data from January 2016 to December 2020. Thus, our analysis is limited to this 48-month window, and all other data described below are aligned to this time frame.

To assess conflict, economic, and sociopolitical conditions, we use the Global Database of Events, Language, and Tone (GDELT).[1] GDELT collects global news data and classifies events into 316 categories based on the CAMEO codebook [36]. Following [25], we group these into five macro-categories:

- GD:Political
- GD:Social (unrest)
- GD:Conflict
- GD:Economic
- GD:Governance

Event intensity is quantified using the *Goldstein scale*, which assigns numerical scores to events ranging from -10 (highly conflictual) to $+10$ (highly cooperative). Positive values reflect peaceful actions (e.g., treaty signing), while negative values indicate hostile actions (e.g., military attack).

To capture the stringency of immigration policies in different EU member states, we compute the recognitions rate defined as the number of positive decisions as a fraction of total asylum applications. The data for computing these statistics are obtained from EUROSTAT. Moreover, following [25], we also include the number of illegal border crossings (ibc) from Somalia to the EU external borders, from the Frontex Migratory Map.

The dependent variable is the monthly asylum-seeking rate (ASR) of Somali nationals to EU countries, measured as the number of first-time asylum seekers per 1,000 people remaining in Somalia. This is calculated using monthly asylum application data from EUROSTAT, normalized by Somalia's population from the World Bank. To match the satellite data time frame, we restrict analysis to January 2016–December 2020. Some EU countries are excluded due to excessive missing values or insufficient ASR variability.

We focus on asylum applications (rather than general migration statistics) for two reasons: (i) Higher temporal resolution (monthly) enables the use of time-series forecasting via DynENet and (ii) Refugee migration is central to ongoing EU asylum policy debates, particularly since the 2015 refugee crisis [3]. Figure 2 illustrates ASR trends for Somali nationals across EU destinations. Between January 2016 and December 2019, Germany (DEU) received the highest number of applications, followed by France (FRA) and Italy (ITA). Notably, the peaks of asylum requests vary by destination, suggesting that the drivers of asylum flows to different EU member states are not homogeneous. As noted earlier, this justifies the use flexible models like DynENet that accommodate flow-specific parameters.

[1] https://www.gdeltproject.org

4.2 Preprocessing for modeling

Before training, all data are preprocessed as follows:

- ASR Transformation: Monthly ASR is first transformed as $\text{ASR} = \dfrac{N_{asylum}+1}{N_{origin}+1}$ to ensure $\text{ASR} \in (0,1]$. It is then log-transformed using the natural logarithm.
- Covariate Aggregation: All predictors are aggregated to monthly frequencies to match ASR.
- Normalization of Predictors: Each time series x_t is normalized via:
  1. Translation: $x'_t = x_t - \min_t(x_t)$ (ensures all values are positive)
  2. Normalized growth trends: $x''_t = \dfrac{(x'_t + \max(x'_t) \times 0.01)}{(x'_1 + \max(x'_t) \times 0.01)} - 1$

Finally, to evaluate forecasting performance, the dataset is partitioned into training and testing sets, separated by a vertical line in Fig. 2. In what follows, we demonstrate how our DynENet model can identify temporal and spatial patterns in climate, conflict, economic, and socio-political drivers of Somali refugee migration to various EU destinations, as well as the model's ability to forecast possible futures.

4.3 Cross-validation

To search for the optimal hyperparameter λ_i in Eq. (4), we apply cross-validation by partitioning the data as shown in Fig. 3. The total data span is 48 months, of which the last 6 months are reserved for testing (see Fig. 2). The training set contains 42 months of data, which are further divided into 10 folds (sub-periods) with varying length.

4.4 Lead–lag analysis

For each training fold, the algorithm starts with a lead–lag analysis using the contrast function in Eq. (2). This step effectively reduces the dimensionality of covariate space in Eq. (1) by:

- Pre-select predictors that exhibit statistically significant correlations with the asylum-seeking rate (ASR); and
- Determine the optimal lag length for each selected predictor.

Figure 4 depicts the correlations between pre-selected predictors and ASR for each EU destination. These correlations are averaged across all cross-validation folds (as shown in Fig. 3) and aggregated over all administrative districts within Somalia. The heatmap is structured by variable

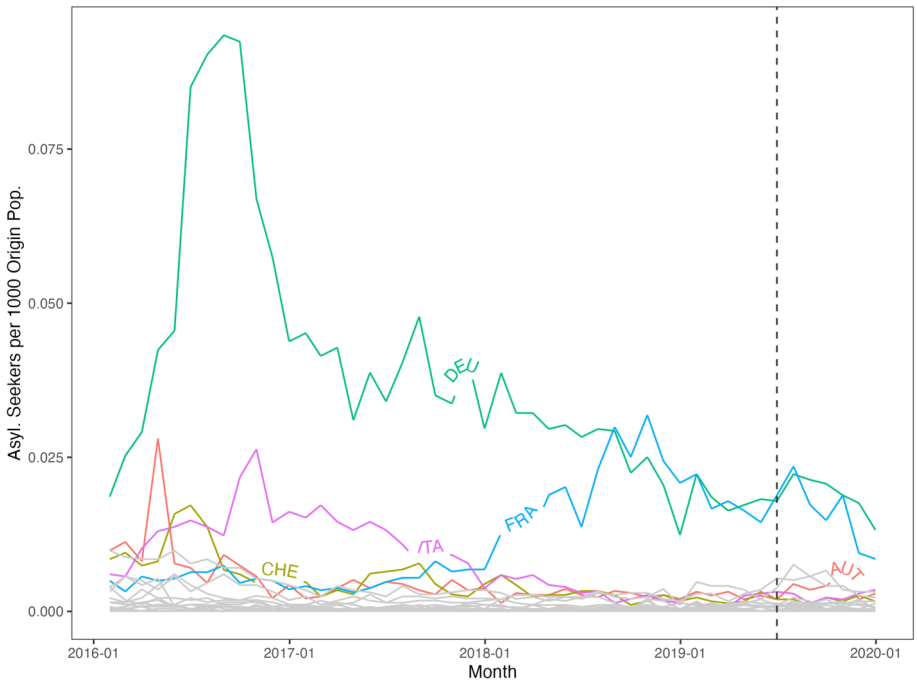

Fig. 2 Trends in monthly asylum-seeking rate in the EU for Somali nationals, Jan. 2016–Dec. 2019

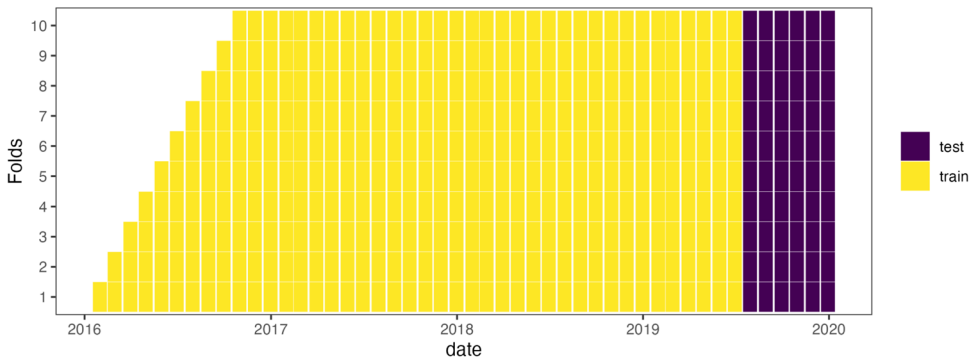

Fig. 3 Rolling origin cross-validation scheme

categories, including climatic (e.g., SPEI and soil moisture), socio-economic, political, governance, and conflict indicators, each assessed across multiple temporal lags (L1–L6).

Notably, the figure reveals that correlations between climatic predictors and ASR are relatively weak, suggesting that climate fluctuations exert limited direct influence on refugee flows in this context. In contrast, socioeconomic and political indicators derived from GDELT exhibit stronger associations, with several variables maintaining moderate positive or negative correlations across multiple lags. This pattern implies that governance quality, political stability, and social dynamics may play a more immediate and sustained role in shaping refugee migration than environmental stressors alone.

The heterogeneity of correlations across destination countries also highlights that the drivers of asylum-seeking behavior are context-dependent: while some destinations exhibit stronger associations with governance, socio-political, economic and/or conflict indicators, others display weaker or even inverse relationships. These findings underscore the importance of integrating adaptive modeling approaches to account for diverse nature and temporal dynamics of human mobility.

4.5 Hyperparameter λ

We tune λ_i via a grid search, which is defined as $\lambda_s = \exp(-5/59 \times s)$, where $s \in [0, 1, 2, ..., 49]$. This results in 50×10 Elastic Net regressions for each flow, yielding a total of 500 λ_i values and corresponding model specifications.

The relationship between the training performance—measured by $\text{PDR}_{i,w}$—and the 500 λ_i values can be nonlinear. From Eq. (5), $\text{PDR}_{i,w}$ is a decreasing function of the number of selected features $k_{i,w}$, *ceteris paribus*. This implies a positive relationship between $\text{PDR}_{i,w}$ and λ_i, as higher λ_i values lead to fewer selected predictors, thereby

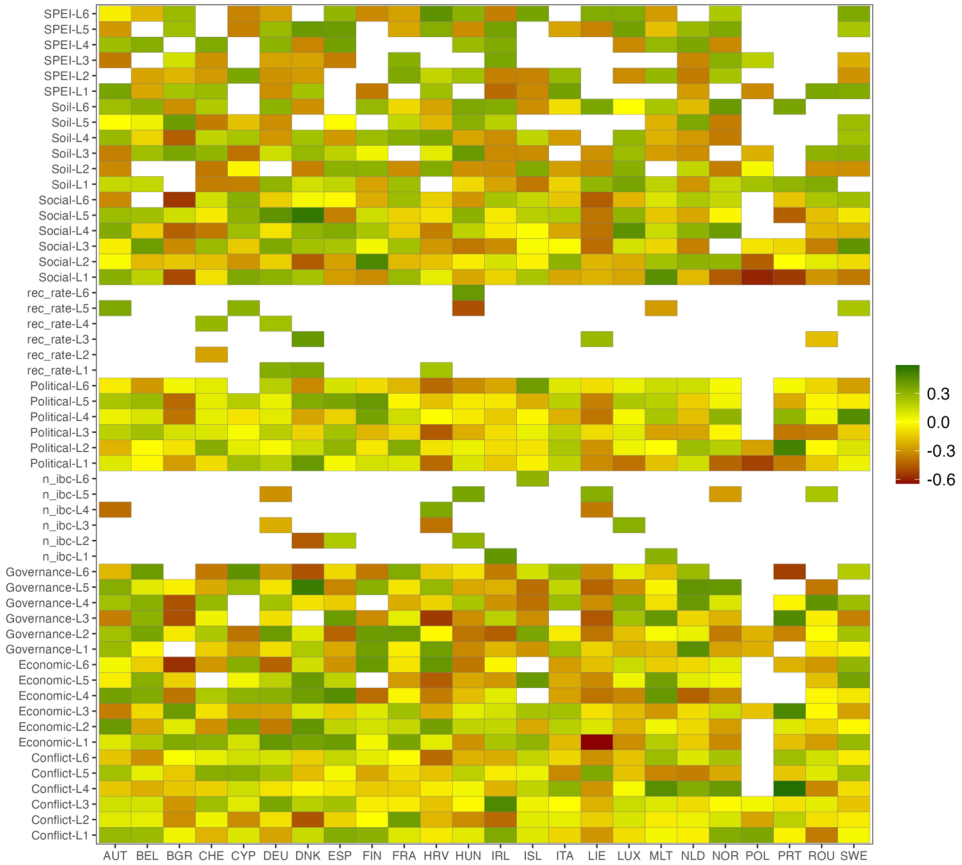

Fig. 4 Correlations between pre-selected predictors and ASR (averaged across training folds and districts in Somali)

reducing $k_{i,w}$. However, the standard deviance ratio (DR)[2] tends to decrease monotonically with increasing λ_i in the training set, because stronger L1 penalization (with fixed L2 penalty) enforces greater sparsity and reduces model fit. These properties suggest a potentially concave relationship between $\text{PDR}_{i,w}$ and λ_i.

Figure 5 illustrates how both the standard DR and $\text{PDR}_{i,w}$ vary with respect to λ_i. Each point represents an Elastic Net model from the grid search. As anticipated, the standard DR (shown as red circles) declines monotonically with λ_i, while $\text{PDR}_{i,w}$ (shown as blue circles) exhibits a concave shape.

Notably, the optimal λ_i values selected by maximizing $\text{PDR}_{i,w}$ are systematically larger than those chosen by the standard DR. This implies that our proposed metric favors sparser, more parsimonious models relative to traditional DR-based selection.

[2] The standard deviance ratio is given by:

$$\text{DR}_{i,w} = 1 - \frac{\sum_{t_w=1}^{T_w} \left(Y_{i,t_w} - \widehat{Y}_{i,t_w}\right)^2}{\sum_{t_w=1}^{T_w} \left(Y_{i,t_w} - \frac{1}{T_w}\sum_{t_w=1}^{T_w} Y_{i,t_w}\right)^2}$$

4.6 Training and testing performance

To examine the effect of our proposed PDR metric on model performance, we compare its prediction errors with a benchmark model tuned using the standard deviance ratio (DR) metric. Both models are trained on data up to June 2019, with the subsequent six months reserved for evaluating forecasting performance. Figure 6 presents the performance of the PDR model relative to the DR benchmark. A negative value indicates superior performance of the PDR model.

Panel (a) reports the percentage differences in the Sum of Squared Errors (SSE) for point predictions, computed according to Eq. (6). Across all EU destinations, the PDR model exhibits a significantly worse fit on the training data. This outcome is expected, as shown in Figure 5, since the PDR-tuned λ tends to be higher, resulting in sparser models. However, in the testing set, the PDR model mostly outperforms the DR benchmark, delivering lower point prediction errors. These contrasting performances between training and testing sets suggest that models tuned with the standard DR are more prone to overfitting.

Panel (b) reports the percentage differences in the prediction interval score (IS), calculated using Eq. (7). Similar to the pattern observed with SSE, the PDR model yields higher IS values in the training set, again indicating inferior in-sample

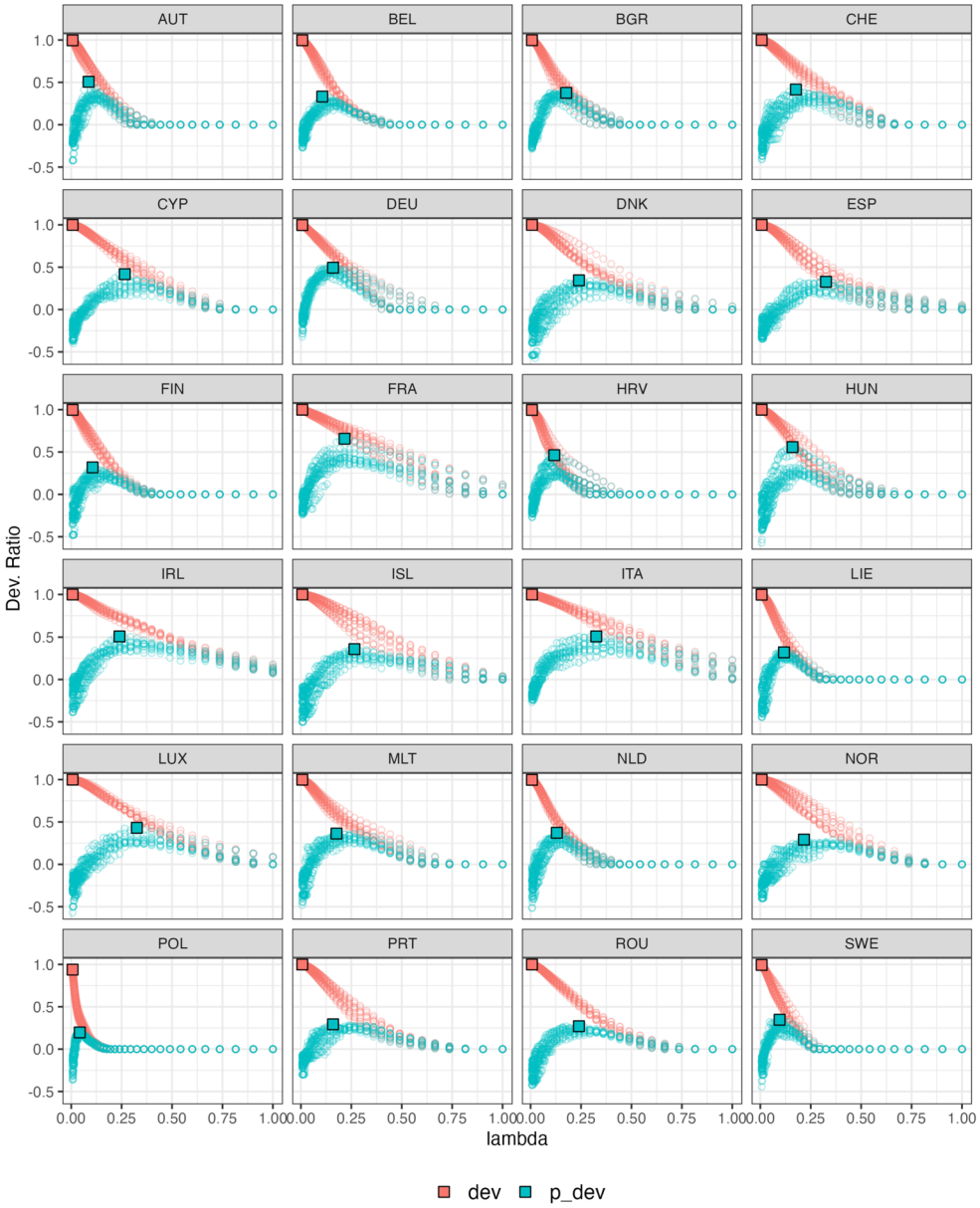

Fig. 5 Optimal λ_i

fit. However, in the testing set, the PDR model achieves superior performance in most cases, except for Austria. These results further reinforce the view that DR-tuned models tend to overfit, which undermines their generalizability.

In terms of aggregate testing performance, PDR is clearly favored. As illustrated by the blue bars in Fig. 6, the PDR-tuned model achieves an average reduction of 40% in point prediction error and a 79% reduction in interval score, compared to the DR benchmark. This demonstrates the predictive advantage of the PDR criterion, despite some isolated cases of under-performance.

4.7 Predictive analytics

We construct 95% prediction intervals using Eq. (8). Figure 7 illustrates the models' ability to reproduce the trends in asylum-seeker rates (ASR) from Somalia to the EU and to individual member states. All values are presented on the original scale (asylum seekers per 1000 Somali population), having been exponentiated from the log-transformed model estimates.

The top panel displays the aggregate ASR to the EU, while subsequent panels disaggregate the trends by destination country. The vertical dashed line marks the boundary between the training and testing sets, with the forecasting

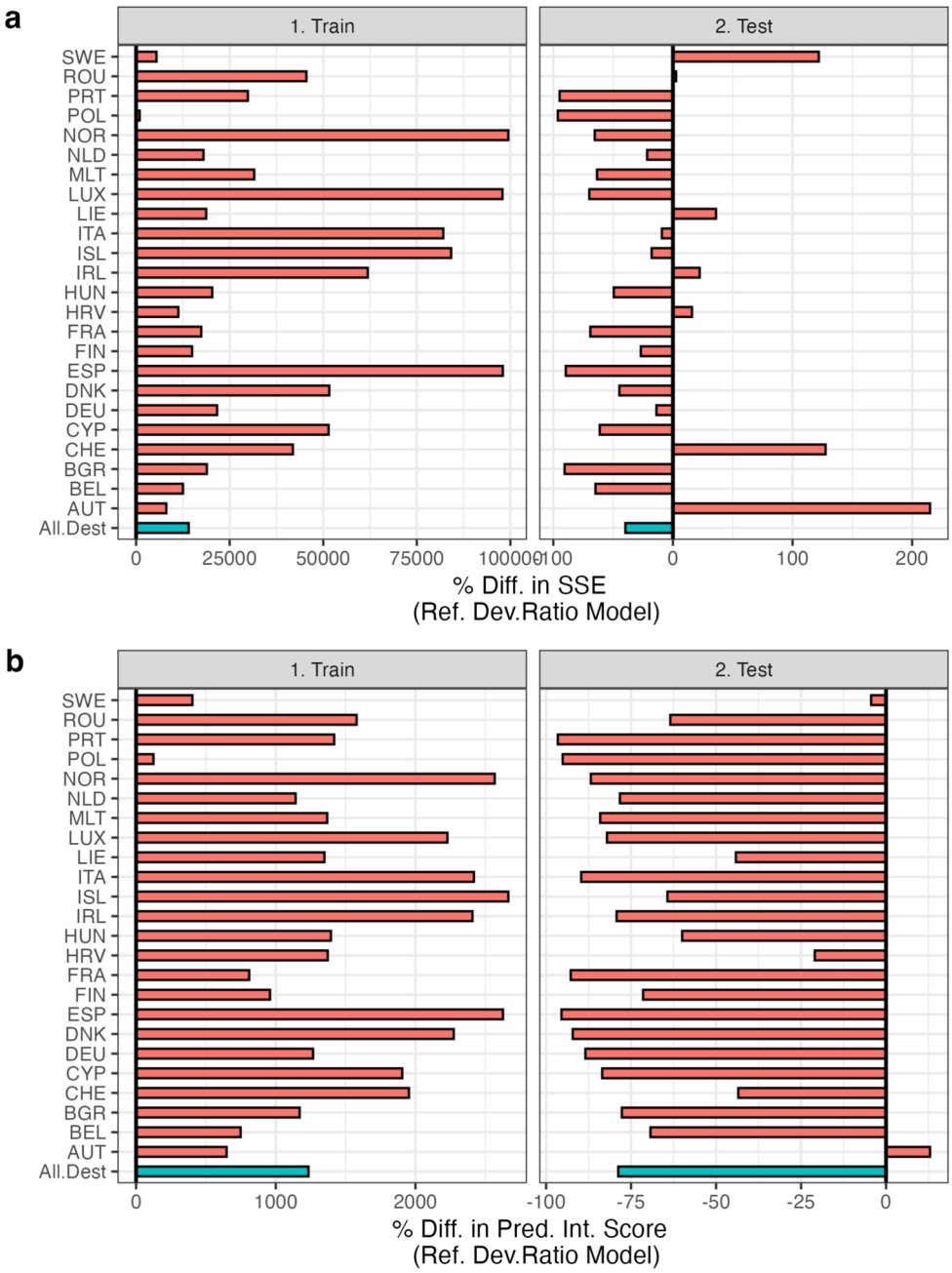

Fig. 6 Difference in training and testing errors between DynENet and Benchmark AR(1) models

target corresponding to the final six months of the sample period (July–December 2019).

The PDR-selected models yield point forecasts that are more stable and exhibit reduced variability relative to the benchmark. This reflects stronger L1 penalization and greater sparsity in the DynENet model under the PDR criterion. As expected, this leads to a poorer fit during the training period, but substantially enhances predictive accuracy out-of-sample. These results underscore the PDR metric's effectiveness in managing the bias–variance trade-off.

Additionally, the wider prediction intervals produced by the PDR model reflect a higher level of epistemic uncertainty—driven by larger residuals—and result in improved empirical coverage. While this comes at the cost of increased interval width, the net effect is a reduction in the prediction interval score (IS), as defined in Eq. (7). The improved coverage outweighs the width penalty, leading to more accurate and reliable probabilistic forecasts.

4.8 Explanatory analytics

The cross-validation scheme illustrated in Fig. 3 implies that parameter estimates may vary over time. Moreover, since models are estimated for individual flows—each representing

Fig. 7 Predicted vs. observed asylum-seeking rate

migration from Somalia to a specific EU country, parameter values for the same predictor may differ across destinations. This flexible and adaptive structure offers nuanced insights into the spatiotemporal dynamics of migration drivers.

To guide interpretation, we distinguish three levels of model parameters:

- **Level I**: A destination-district-specific and time-varying parameter captures how migration intensity to a specific EU destination responds to changes in a district-level predictor (e.g., how ASR from Somalia to Germany changes in response to soil moisture conditions in the district of Eyl).
- **Level II**: A district-specific and time-invariant parameter reflects how overall migration to the EU responds, on average, to changes in a district-level predictor (e.g., how ASR from Somalia to the EU is affected by changes in soil moisture in Eyl).
- **Level III**: A destination-specific and time-varying parameter captures how migration to a specific EU destination

responds, on average, to changes in a predictor across all Somali districts (e.g., how ASR from Somalia to Germany is affected by soil moisture changes across Somalia).

Unlike standard least squares regression, DynENet does not yield conventional statistical significance levels. Instead, it uses LASSO-based shrinkage to retain only the most predictive covariates. We thus interpret nonzero (retained) coefficients as indicators of statistically important migration drivers.

Level II: district-specific and time-invariant parameters

By aggregating Level I parameters across destinations and time, we derive Level II parameters that characterize how migration responses vary across districts in Somalia. These metrics enable the identification of potential "hotspots" of forced migration sensitivity. Figure 8 displays the spatial distribution of average coefficients describing the relationship between local stressors and Somalia–EU refugee migration. The maps visualize how changes in conflict, economic, governance, political, social, soil, and climatic (SPEI) conditions relate to asylum flows to the EU. Red/green shades, respectively, denote positive and negative migration responses to adverse local conditions, while white or pale yellow areas indicate weak or negligible relationships (i.e., near-zero coefficients).

A key pattern is that political and conflict-related variables yield a higher number of nonzero coefficients (*N.Dist*), suggesting that these factors are more frequently associated with migration dynamics compared to other covariate groups. However, their average magnitude of these coefficients (*Avg.Coef*) remain modest, indicating that although such factors are spatially widespread, their overall influence on migration intensity is limited. This points to a complex interplay between local instability and migration potential rather than a simple causal relationship.

Figure 8 further reveals pronounced spatial heterogeneity. Coefficients for the same predictor often differ in both sign and magnitude across districts. This implies that similar stressors may operate as either "push" or "trapping" forces depending on local context. For example, intensified conflict in Burao appears positively associated with asylum flows, whereas in areas such as Bosaso or Hobyo, conflict correlates negatively with refugee migration, possibly reflecting mobility constraints, differing livelihood strategies, or the protective effects of social and family networks.

Overall, these patterns suggest that in many Somali districts, deteriorating conditions do not automatically translate into increased refugee movements. Instead, they align with the concept of resource-constrained immobility, whereby worsening sociopolitical or environmental circumstances limit individuals' ability to migrate rather than motivating displacement [12, 14, 16]. This underscores the importance of considering not only the presence of adverse conditions but also the underlying capacities and constraints that shape migration responses across heterogeneous local contexts.

Level III: Destination-specific and time-varying parameters

By employing a rolling-origin cross-validation design (as illustrated in Fig. 3), the estimated parameters explicitly account for temporal variation. This dynamic modeling framework allows for a fine-grained examination of how migration drivers evolve over time rather than assuming static relationships. Figure 9 displays the resulting Level III coefficients, capturing time-dependent associations between predictor variables and asylum-seeker rates (ASR) for each EU destination.

The results reveal pronounced temporal and spatial heterogeneity even though the analysis focuses on a single-origin country (Somalia) and a single migration stream (refugee movements). Such variation underscores the complex and context-contingent nature of forced migration systems, where the relevance and magnitude of different drivers fluctuate over time and across receiving countries. For instance, while conflict indicators frequently emerge as significant and positively associated with refugee outflows, their intensity and persistence vary across destinations, suggesting that geopolitical shocks and instability in Somalia are consistently influential but mediated by country-specific reception dynamics.

This temporal heterogeneity allows us to make a distinction between persistent and transitory migration drivers. Persistent drivers are those that recur across multiple time windows, maintaining stable predictive power throughout the observation period. In this case, conflict intensity and governance instability fall into this category, highlighting their role as structural determinants of Somalia–EU refugee flows. These drivers reflect deep-rooted vulnerabilities under enduring adverse conditions that systematically shape migration pressures.

In contrast, transitory drivers are those exhibit intermittent or episodic importance. For example, fluctuations in the recognition rate (used here as a proxy for destination-country asylum policy) appear episodic, possibly reflecting shifts in administrative practices, political climates, or international protection norms. These results align with prior findings emphasizing the responsiveness of refugee flows to policy volatility and the externalization of asylum policy regimes within the EU context.

The temporal instability and spatial heterogeneity of coefficients across destinations indicate that migration drivers cannot be generalized uniformly across the migration systems. Instead, they interact with destination-specific con-

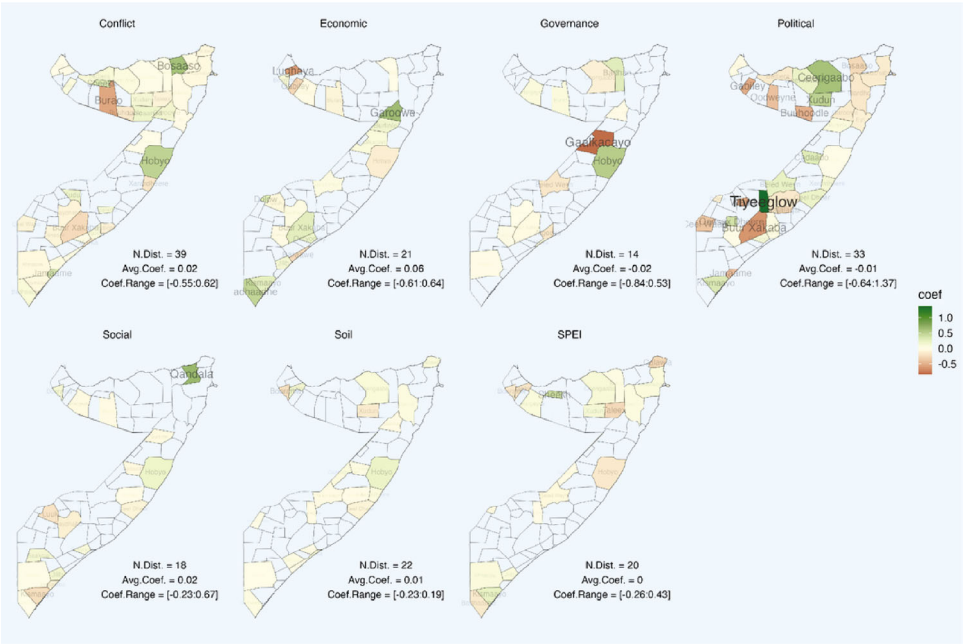

Fig. 8 Average parameter estimates by Somali districts

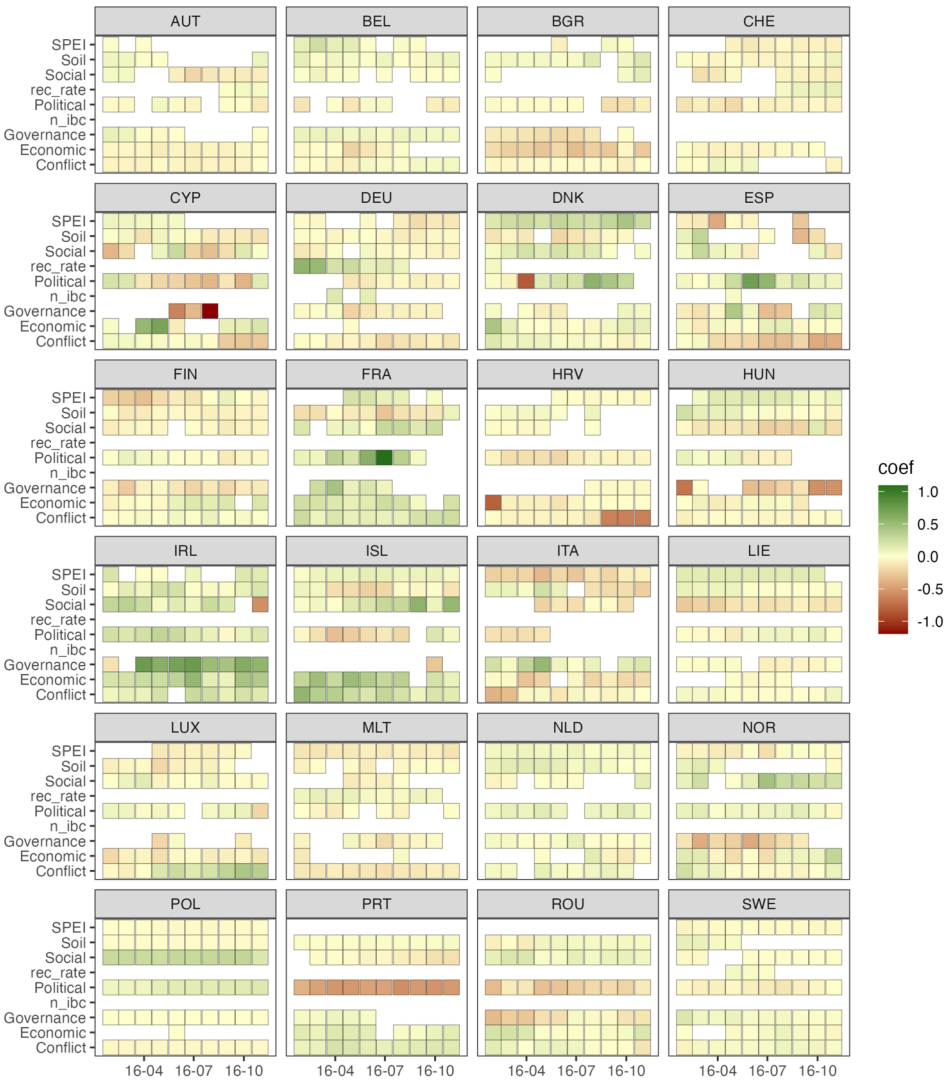

Fig. 9 Average parameter estimates by sub-periods/folds and EU destinations

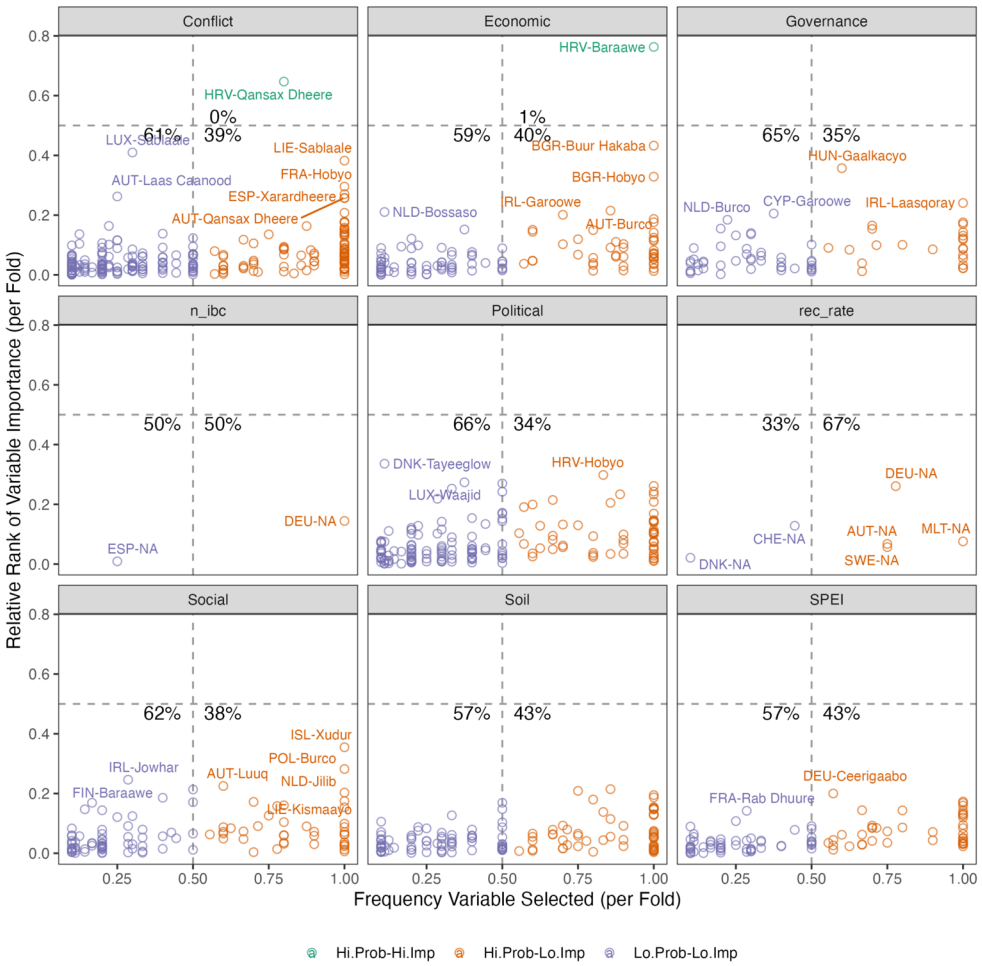

Fig. 10 Importance–frequency (IF) space

texts, such as reception capacity, policy frameworks, diaspora networks, as well as the overall costs associated with the movement. Such temporal fluidity and multi-scalar nature of migration drivers underscore the need for more adaptive models that can capture non-stationary and context-specific migration processes. DynENet fills this void by offering a path toward disentangling the structural from the contingent dimensions of refugee mobility, which has critical implications for both predictive modeling and policy planning.

Variable importance and frequency (IF) space

To synthesize the spatiotemporal dynamics shown in Figs. 8 and 9 , we construct the importance-frequency (IF) space, depicted in Fig. 10.

The y-axis shows the average relative importance rank of each predictor across folds (Eq. (9)), indicating its predictive power. The x-axis displays the average frequency with which each predictor is selected (Eq. (10)), indicating its likelihood of being retained. A frequency of 1 implies consistent selection; a value of 0 means it was never selected.

This IF space allows classification of migration drivers into four quadrants:

- *High-probability, high-impact* Frequently selected and highly predictive.
- *High-probability, low-impact* Frequently selected but weakly predictive.
- *Low-probability, high-impact* Rarely selected but strongly predictive when included.
- *Low-probability, low-impact* Infrequently selected and weakly predictive.

As most predictors fall in the bottom two quadrants in Fig. 10, it is evident that the majority of environmental and societal factors act as low-impact drivers. Furthermore, no covariates appear in the top-left quadrant, implying an absence of "black swan" events—rare triggers with large impacts—during the study period. This finding reinforces the view that Somali–EU refugee migration is shaped more by persistent, structural drivers than by episodic shocks.

5 Conclusions

This paper introduces a novel framework for forecasting refugee migration that directly addresses the challenges posed by high-dimensional covariate spaces. Our key innovation is the penalized deviance ratio (PDR), a new metric for hyperparameter tuning that extends traditional model selection criteria to better suit forecasting tasks with sparse, noisy, and multi-collinear predictors.

Unlike the standard deviance ratio (DR), which tends to prioritize in-sample fit and can lead to overfitting in high-dimensional settings, the PDR explicitly balances model sparsity with predictive accuracy by penalizing over-parameterization. This makes it particularly well-suited to dynamic elastic net (DynENet) models, where controlling the bias–variance trade-off is critical for stable out-of-sample predictions.

Through extensive cross-validation and application to Somali–EU asylum flows, we demonstrate that PDR-tuned models outperform DR-based benchmarks in both point and probabilistic forecasts. Improvements in the prediction interval score are observed across 23 of 24 EU destinations, underscoring the robustness of the PDR approach even within a single-origin case study. The method produces parsimonious and interpretable models that remain stable across time and destination contexts, effectively mitigating the challenges of high dimensionality without sacrificing generalization.

Beyond forecasting, our framework provides new insights into the drivers of forced migration by leveraging dynamic regularization to uncover persistent and transitory migration determinants at multiple spatial levels. This dual focus on predictive performance and interpretability marks a substantive advancement in the methodological approaches for migration modeling.

We acknowledge, however, that the empirical scope of this study is limited to a single-origin country. This constraint reflects the considerable computational demands of processing and harmonizing gridded remote sensing datasets at high temporal and spatial resolution. Expanding the analysis to include multiple origin contexts would require substantial data-processing capacity that exceeds the resources currently available and is planned for future work. Nevertheless, even within this constrained design, the strong and consistent predictive improvements achieved across nearly all EU destinations suggest that the advantages of the PDR are not case-specific but rather indicative of its general applicability to other migration and forecasting domains.

In sum, the penalized deviance ratio and the accompanying high-dimensional forecasting framework represent a robust, interpretable, and generalizable approach for analyzing and anticipating migration flows in an increasingly complex world, where human mobility is shaped by a web of interrelated and often opaque factors.

Acknowledgements Financial support by European Union's Horizon 2020 Programme (grant agreement 870661, 871042, and 101004535), U.S. National Science Foundation (Award Number: 2310908), and the Swedish Research Council Vetenskapsrådet (grant agreement 2022-06012-3) are gratefully noted.

Author Contributions H.Q., A.S., and S.I. conceptualized and developed the machine learning algorithms. H.Q., A.S., R.M., E.H., and C.A. processed and prepared the data for analyses. All authors drafted, reviewed, and revised the manuscript.

Funding Open access funding provided by Malmö University. This work has received support from the European Union's Horizon 2020 Programme (grant agreement 870661). In addition, Haodong Qi has received support from the Swedish Research Council Vetenskapsrådet (grant agreement 2022–06012-3) and from the European Union's Horizon 2020 Programme (grant agreement 101004535). Alina Sîrbu has received support from the European Union's Horizon 2020 Programme (grant agreement 871042). Stefano Iacus has received support from U.S. National Science Foundation (Award Number: 2310908).

Data Availability Data that support the findings of this study have been deposited at https://github.com/HaodongQi/HumMingBird_MigForecast.git.

Declarations

Conflict of interest The authors declare no Conflict of interest.

Code availability Code that support the findings of this study have been deposited at https://github.com/HaodongQi/HumMingBird_MigForecast.git.

References

1. Beyer, R., Schewe, J., Lotze-Campen, H.: Gravity models do not explain, and cannot predict, international migration dynamics. Human. Soc. Sci. Commun. **9**(1), 56 (2022). https://doi.org/10.1057/s41599-022-01067-x
2. Qi, H., Bircan, T.: Modelling and predicting forced migration. PLoS ONE **18**(4), 0284416 (2023). https://doi.org/10.1371/journal.pone.0284416
3. Missirian, A., Schlenker, W.: Asylum applications respond to temperature fluctuations. Science **358**(6370), 1610–1614 (2017)

4. Rigaud, K., De Sherbinin, A., Jones, B., Casals Fernandez, A., Adamo, S.: Groundswell: Preparing for internal climate. Migration (2018). https://doi.org/10.1596/36447
5. Abel, G.J., Brottrager, M., Cuaresma, J.C., Muttarak, R.: Climate, conflict and forced migration. Glob. Environ. Chang. **54**, 239–249 (2019). https://doi.org/10.1016/j.gloenvcha.2018.12.003
6. Cattaneo, C., Beine, M., Fröhlich, C.J., Kniveton, D., Martinez-Zarzoso, I., Mastrorillo, M., Millock, K., Piguet, E., Schraven, B.: Human migration in the era of climate change. Rev. Environ. Econ. Policy **13**(2), 189–206 (2019). https://doi.org/10.1093/reep/rez008
7. Clement, V., Rigaud, K., De Sherbinin, A., Jones, B., Adamo, S., Schewe, J., Sadiq, N., Shabahat, E.: Groundswell Part 2: acting on internal climate. Migration (2021). https://doi.org/10.1596/36248
8. Bertoli, S., Moraga, J.F.-H.: Multilateral resistance to migration. J. Dev. Econ. **102**, 79–100 (2013). https://doi.org/10.1016/j.jdeveco.2012.12.001
9. Beine, M., Bertoli, S., Moraga, J.F.-H.: A practitioners' guide to gravity models of international migration. World Econ. **39**(4), 496–512 (2015). https://doi.org/10.1111/twec.12265
10. Carling, J.: Migration in the age of involuntary immobility: Theoretical reflections and Cape Verdean experiences. J. Ethn. Migr. Stud. **28**(1), 5–42 (2002)
11. Black, R., Bennett, S.R., Thomas, S.M., Beddington, J.R.: Migration as adaptation. Nature **478**(7370), 447–449 (2011)
12. Carling, J., Schewel, K.: Revisiting aspiration and ability in international migration. J. Ethn. Migr. Stud. **44**(6), 945–963 (2018)
13. Collins, F.L.: Desire as a theory for migration studies: temporality, assemblage and becoming in the narratives of migrants. J. Ethn. Migr. Stud. **44**(6), 964–980 (2018)
14. De Haas, H.: A theory of migration: the aspirations-capabilities framework. Comp. Migr. Stud. **9**(1), 1–35 (2021)
15. Schewel, K.: Understanding immobility: Moving beyond the mobility bias in migration studies. Int. Migr. Rev. **54**(2), 328–355 (2020)
16. Benveniste, H., Oppenheimer, M., Fleurbaey, M.: Climate change increases resource-constrained international immobility. Nat. Clim. Chang. **12**(7), 634–641 (2022)
17. Clark, X., Hatton, T.J., Williamson, J.G.: Explaining US immigration, 1971–1998. Rev. Econ. Stat. **89**(2), 359–373 (2007)
18. Pedersen, P.J., Pytlikova, M., Smith, N.: Selection and network effects—migration flows into OECD countries 1990–2000. Eur. Econ. Rev. **52**(7), 1160–1186 (2008). https://doi.org/10.1016/j.euroecorev.2007.12.002
19. Mayda, A.M.: International migration: a panel data analysis of the determinants of bilateral flows. J. Popul. Econ. **23**(4), 1249–1274 (2010)
20. Grogger, J., Hanson, G.H.: Income maximization and the selection and sorting of international migrants. J. Dev. Econ. **95**(1), 42–57 (2011). https://doi.org/10.1016/j.jdeveco.2010.06.003
21. Qi, H., Bircan, T.: Understanding migration drivers: Methodology, sensitivity, and heterogeneity. In: HumMingBird Project 870661 H2020, Deliverable 3.2, (2021)
22. Hoffmann, R., Dimitrova, A., Muttarak, R., Crespo Cuaresma, J., Peisker, J.: A meta-analysis of country-level studies on environmental change and migration. Nat. Clim. Chang. **10**(10), 904–912 (2020)
23. Boas, I., Farbotko, C., Adams, H., Sterly, H., Bush, S., Geest, K., Wiegel, H., Ashraf, H., Baldwin, A., Bettini, G., et al.: Climate migration myths. Nat. Clim. Chang. **9**(12), 901–903 (2019)
24. Wiegel, H., Boas, I., Warner, J.: A mobilities perspective on migration in the context of environmental change. Wiley Interdiscip. Rev.: Clim. Change **10**(6), 610 (2019)
25. Carammia, M., Iacus, S.M., Wilkin, T.: Forecasting asylum-related migration flows with machine learning and data at scale. Sci. Rep. **12**(1), 1–16 (2022)
26. Comte, F., Renault, E.: Long memory continuous time models. J. Econ. **73**(1), 101–149 (1996)
27. Hoffmann, M., Rosenbaum, M., Yoshida, N.: Estimation of the lead-lag parameter from non-synchronous data (2013)
28. Hayashi, T., Yoshida, N.: On covariance estimation of non-synchronously observed diffusion processes. Bernoulli **11**(2), 359–379 (2005)
29. Hayashi, T., Yoshida, N.: Asymptotic normality of a covariance estimator for nonsynchronously observed diffusion processes. Ann. Inst. Stat. Math. **60**, 367–406 (2008)
30. Hayashi, T., Yoshida, N.: Nonsynchronous covariation process and limit theorems. Stochastic Process. Appl. **121**(10), 2416–2454 (2011)
31. Iacus, S.M., Yoshida, N.: Simulation and inference for stochastic processes with yuima. A comprehensive R framework for SDEs and other stochastic processes, Use R (2018)
32. Gneiting, T., Raftery, A.E.: Strictly proper scoring rules, prediction, and estimation. J. Am. Stat. Assoc. **102**(477), 359–378 (2007). https://doi.org/10.1198/016214506000001437
33. Fisher, A., Rudin, C., Dominici, F.: All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. J. Mach. Learn. Res. **20**(177), 1–81 (2019)
34. World Bank: Sub-Saharan Africa-Macro Poverty Outlook: Country-by-Country Analysis and Projections for the Developing World, April 2022. World Bank (2022)
35. Stagge, J.H., Tallaksen, L.M., Xu, C.Y., Van Lanen, H.A.: Standardized precipitation-evapotranspiration index (SPEI): Sensitivity to potential evapotranspiration model and parameters. Hydrol. Chang. World **363**, 367–373 (2014)
36. Schrodt, P.: Conflict and Mediation Event Observations event and actor codebook V. 1.1 b3 (2012)