

DYNAMIC COUNT MODELS WITH FLEXIBLE INNOVATION PROCESSES FOR IRREGULAR MARITIME MIGRATION

BY GREGOR ZENS^{1,a} AND JAKUB BIJAK^{2,b}

¹*Population and Just Societies Program, International Institute for Applied Systems Analysis (IIASA), zens@iiasa.ac.at*
²*Leverhulme Centre for Demographic Science, Nuffield Department of Population Health and Reuben College, University of Oxford, jakub.bijak@demography.ox.ac.uk*

Motivated by the challenge of analyzing the dynamics of weekly sea border crossings in the Mediterranean (2015–2025) and the English Channel (2018–2025), we develop a Bayesian dynamic framework for modeling heteroskedastic count time series. Building on theoretical considerations and empirical stylized facts, our approach utilizes a Poisson random walk model that allows for heavy-tailed innovations or stochastic volatility dynamics, while incorporating an explicit mechanism to separate structural from sampling zeros. Posterior inference is carried out via a straightforward Markov chain Monte Carlo algorithm. Applying this methodology to Mediterranean and English Channel data, we compare alternative model specifications through a comprehensive out-of-sample forecasting exercise. Using log predictive scores and empirical coverage at predictive quantiles to evaluate each model, we find strong evidence for stochastic volatility in migration innovations. These models deliver the strongest out-of-sample forecasts with empirical coverage close to nominal levels up to the 99th percentile. Our framework can be used to develop risk indicators with direct policy implications for improving governance and preparedness for migration surges. More broadly, the methodology extends to other zero-inflated nonstationary count time series applications, including epidemiological surveillance and public safety incident monitoring.

1. Introduction. Irregular maritime migration and sea border crossings are major issues in contemporary political discourse and come with severe humanitarian consequences due to the high mortality risks associated with sea migration routes. At the same time, such “small boat” passages constitute one of the few irregular migration contexts where arrivals can be tracked with relative precision, based on systematic search-and-rescue logs and immigration or asylum registration systems in arrival countries. The resulting availability of high-frequency data should, in principle, enable a rigorous statistical investigation.

However, the literature on modeling the dynamics of irregular maritime migration and border crossings remains surprisingly thin. This gap may stem from two intertwined obstacles. First, data on irregular arrivals by sea present unique challenges, as high-frequency counts of migrant arrivals are noisy, nonstationary, volatile, and often zero-inflated—a combination that overwhelms many off-the-shelf modeling techniques. Second, these characteristics make point forecasts notoriously unreliable, potentially creating the impression that any attempt to effectively model irregular maritime border crossings is futile. This leaves a significant research gap both in terms of forecasting capabilities, which could enhance preparedness and humanitarian response, and in terms of developing structural insights into the underlying drivers and dynamics of irregular maritime migration.

In this article we demonstrate that both meaningful structural insights and well-calibrated distributional forecasts of sea border crossing activity can be obtained using a simple context-specific probabilistic model. Based on theoretical considerations and stylized empirical facts

Received August 2025; revised March 2026.

Key words and phrases. Zero-inflated Poisson, Bayesian dynamic models, migration time series, distributional forecasting, heavy-tailed innovations.

of irregular maritime border crossing data, we develop a Bayesian dynamic modeling framework for count data time series. This framework is designed for discrete time series jointly characterized by: (i) highly persistent trending patterns, (ii) volatility clustering or sudden shifts in the log-scale intensity, and (iii) zero-inflation. We develop a simple Markov chain Monte Carlo (MCMC) algorithm for posterior simulation. The proposed methods and algorithms are flexible, easily extensible, and applicable to any nonnegative integer time series. These features suggest their broader utility across various fields beyond modeling irregular sea border crossings.

We apply the model to two case study datasets on weekly maritime border crossings in the Mediterranean Sea and in the English Channel. We benchmark various versions of the model to gain insight into irregular migration dynamics, focusing particularly on the problem of producing and evaluating distributional forecasts. Specific emphasis is placed on empirical coverage in the tails of the predictive distribution of maritime border crossings. This focus is especially relevant, as accurate tail probability forecasts directly enhance preparedness for extreme events. For instance, well-calibrated predictive distribution tails may enable more effective contingency planning in the context of irregular migration, which is often characterized by rare events of high magnitude (see, e.g., [Bijak \(2024\)](#)). Such approaches are already entering the mainstream debate and policy in other areas, such as preparedness for extreme climate events (e.g., [IPCC \(2012\)](#)), or civil contingencies more generally ([HM Government \(2025\)](#)). The emerging frameworks for responding to high-impact migration events (e.g., the [EU Migration Preparedness and Crisis Blueprint, EC \(2020\)](#)) open the possibility of extending such approaches to migration in a formal and rigorous way.

Our results indicate that the dynamics of irregular sea border crossings are effectively captured by a Poisson state-space model with log intensity expressed as a random walk with stochastic volatility. In both our case studies, this model delivers the strongest one-week-ahead probabilistic forecasting performance for irregular sea border crossing activity among the specifications considered, with empirical coverage close to nominal levels at upper-tail quantiles such as q_{95} and q_{99} . We connect these findings with the underlying theoretical considerations and discuss implications for policy and future empirical research under the assumption that sea border crossings are indeed generated by the proposed model.

This article, therefore, offers two primary contributions. First, from a statistical perspective, we introduce a flexible modeling framework for analyzing zero-inflated and potentially nonstationary count time series, accompanied by a straightforward computational algorithm for posterior simulation. This advances the literature on Bayesian approaches to dynamic modeling of complex integer-valued time series. Second, in terms of substantive application, we conduct the first comprehensive predictive evaluation to determine effective approaches for probabilistic modeling of irregular maritime border crossing data through two high-profile and policy-relevant case studies. This yields a new benchmark model for examining such processes and generates evidence-based policy insights regarding the underlying structural dynamics.

The remainder of the article is structured as follows. Section 2 presents the motivating case study datasets, empirical stylized facts, theoretical considerations, and related literature. Section 3 presents our model, prior specification, and the computational approach. Section 4 applies the methodology to the case study datasets. Section 5 discusses the policy implications of the results and concludes with directions for future research. A simulation study and additional results are presented in the Supplementary Material ([Zens and Bijak \(2026\)](#)).

2. Data, stylized facts, and theoretical considerations. We analyze weekly time series data on irregular sea border crossings across two distinct geographical regions. The first case study examines irregular arrivals in Italy across the Mediterranean Sea, encompassing $T =$

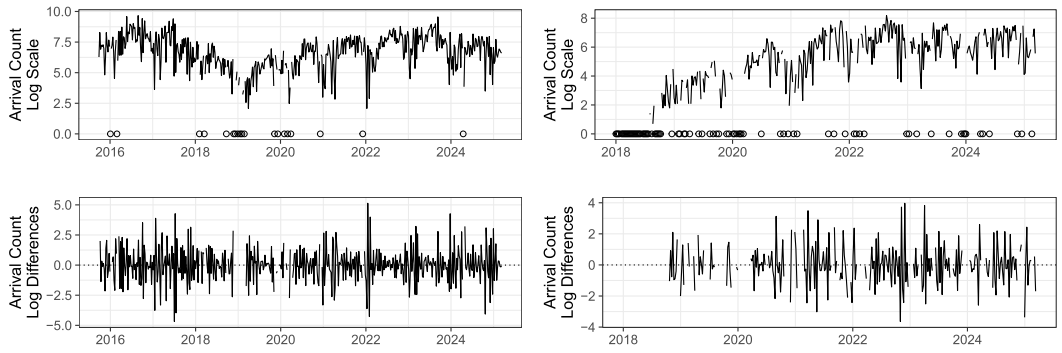


FIG. 1. Irregular maritime border crossing case study datasets. Left: Mediterranean Sea border crossings to Italy by sample week from 2015W40 to 2025W11. Right: English Channel crossings by sample week from 2018W01 to 2025W11. Top panels display log-transformed counts, with zero counts highlighted separately as dots on the zero line. Bottom panels show log differences of counts for observations where both $y_t > 0$ and $y_{t-1} > 0$.

494 weeks from 2015W40 to 2025W11. These data are sourced from the *Operational Data Portal* of the United Nations High Commissioner for Refugees (UNHCR).¹ On average, we observe 1,607 arrivals per week, with a weekly median of 851 arrivals. The series ranges from zero arrivals up to a maximum of 15,694 arrivals in the week of August 29, 2016.

For our second case study, we investigate irregular arrivals by sea to the United Kingdom across the English Channel, covering $T = 376$ weeks from 2018W01 to 2025W11. This dataset is provided by the Border Force branch of the U.K. Home Office.² On average, we observe 413 arrivals per week, with a weekly median of 152 arrivals. The series ranges from zero arrivals up to a maximum of 3,564 arrivals in the week of August 22, 2022.

2.1. Empirical stylized facts. Figure 1 presents the data after a logarithmic transformation (top panels) and after taking first differences of the log-transformed counts (bottom panels). The untransformed count time series are shown in Figure S2 in the Supplementary Material. Figure 1 highlights the three main challenges inherent in capturing the dynamics of maritime migration time series.

First, as is common in many migration processes (Bijak (2010)), the count-valued data exhibit high persistence with respect to mean intensity. While crossing conditions typically lead to somewhat higher activity in summer months than in winter, there is no clear deterministic trend over the full sample period. To capture this behavior, we specify a benchmark count time series model that allows for random walk behavior of the log intensity parameter. The simple random walk assumption is further supported by the log differences in the bottom panels of Figure 1 being almost exclusively centered around the zero line, suggesting the absence of a deterministic drift.

Second, as becomes evident from the time series of log differences, the innovation process on the log scale is heteroskedastic. A priori, it is unclear whether this heteroskedasticity is the result of an overall heavy-tailed (but constant in time) process or whether volatility itself follows a dynamic process that produces persistent time-varying volatilities that may be clustered in time. To investigate this, we will compare various volatility models empirically by benchmarking them against each other in a large-scale pseudo-out-of-sample forecasting exercise.

¹Source: <https://data.unhcr.org/en/situations/europe-sea-arrivals/location/24521> (as of 7 August 2025).

²Source: <https://www.gov.uk/government/publications/migrants-detected-crossing-the-english-channel-in-small-boats/migrants-detected-crossing-the-english-channel-in-small-boats-last-7-days> (as of 7 August 2025).

Third, observed zero counts may reflect one of two distinct phenomena. On the one hand, zero arrivals may be the result of weeks where external circumstances—such as coast guard operations, adverse weather conditions, or low smuggler availability—temporarily prevent crossings despite relatively high migration intensity in surrounding weeks (compare the zero counts after 2022 in the U.K. time series). In other cases, zero counts may stem from genuinely low migration intensity during those periods (such as in the early period of the U.K. data). There is no clear a priori way to determine whether a zero should be treated as an unexpected outlier—effectively missing for the purpose of modeling migration intensity—or as an informative observation indicating genuinely low migration activity. This motivates an approach that treats the probability of crossing and the intensity of the underlying migration process as two separate processes that jointly generate the observed arrival counts. This conceptualization naturally leads to a parameter-driven time series model for zero-inflated counts.

2.2. Theoretical considerations. The empirical patterns observed in Figure 1—namely, high persistence, heteroskedasticity, and zero-inflation—align with what one would expect from irregular maritime migration processes, which emerge from the complex interaction of individual decisions, weather conditions, smuggling operations, border enforcement measures (including pushbacks and deportations), political and diplomatic developments, and numerous other factors (see, e.g., [Rodríguez Sánchez et al. \(2023\)](#), [Frontex \(2024\)](#), [Migration Observatory \(2025\)](#)). Many of these migration determinants can be conceptualized as evolving approximately as first-order random walks with time-varying innovation variances. Such dynamics represent plausible assumptions for classical migration drivers, including conflict pressures, environmental disasters, macroeconomic developments, smuggler capacity, and border enforcement efforts.³ Adopting a simplified perspective, one may reasonably assume that the dynamic properties of these migration drivers are inherited by the migration process itself. This process is then punctuated by occasional events—such as coast guard operations or exceptionally bad weather conditions—that explain the zero-inflation by temporarily halting all movement. While these considerations connect our three stylized facts, and ultimately our statistical model, to basic theoretical intuition, a fully formalized and consistent micro-theoretical treatment of irregular sea migration remains largely absent from the literature. The development of such a framework is left for future research.

Note that many of the underlying migration determinants and drivers are shared between irregular maritime migration and irregular land border crossings. Consequently, a random walk with stochastic volatility dynamics could be expected to provide a reasonable baseline for modeling the latter as well. A formal investigation of this matter is left for future research. The primary reasons we study maritime migration as a distinct subject in the current manuscript are: (i) the availability of relatively high-frequency weekly data, which is seldom the case for land border crossings—at least not in publicly available data sources, (ii) the zero-inflation mechanism, whereby physical barriers, such as the sea, render crossings impossible during certain periods, a feature that is presumably less relevant for land border crossings, and (iii) the relatively distinct public perception and impact of maritime vs. land migration dynamics ([Bhatiya and Kadam \(2025\)](#)).

2.3. Existing statistical approaches. Conceptual statistical work on high-frequency irregular sea migration data (and high-frequency migration patterns more generally) remains

³Note that in many cases, the direct effects of these drivers may be difficult to model and quantify explicitly. In such instances, random walk models have been proposed as atheoretical alternatives for analyzing migration when precise knowledge of drivers is unavailable ([Bijak \(2010\)](#)).

relatively scarce. The majority of quantitative literature in this domain focuses on causal inference studies that evaluate the impact of factors such as search-and-rescue operations, coast guard activity, or smuggler network behavior on migration patterns (Camarena et al. (2020), Hoffmann Pham and Komiyama (2024), Friebe et al. (2024), Deiana, Maheshri and Mastrobuoni (2024), Zambiasi and Albarosa (2025)). These studies typically employ relatively simplistic regression frameworks for count data or Gaussian models for log-transformed counts, based on a few observable covariates or gravity-type models. Predictive evaluations are infrequent and, if provided, focus exclusively on point prediction accuracy.

In parallel with recent efforts to apply machine learning techniques for predicting asylum-related migration activity (Carammia, Iacus and Wilkin (2022)), few studies have attempted to forecast boat migration across various frequencies and case studies—often focusing on Mediterranean border crossings—using an array of machine learning and standard time series analytical tools, sometimes combined in ensembles (Georgiou (2016), Bosco et al. (2024, 2025)). While these studies demonstrate some degree of point predictability in sea border crossing data, the fitted models are largely black-box approaches that provide limited evidence-based structural insights into underlying patterns. In addition, distributional forecasts are typically not considered. Bridging the gap between causal inference-oriented and prediction-oriented work, Rodríguez Sánchez et al. (2023) use Bayesian time series models to generate counterfactuals for Mediterranean small boat migration activity, investigating the role of search-and-rescue operations.

In substantive terms we identify three key limitations in existing statistical approaches to irregular maritime migration. First, irregular migration is typically modeled using frameworks similar to those used for regular migration flows (e.g., labor mobility), despite fundamentally different underlying dynamics. Even for regular migration dynamics, these conventional models can demonstrate poor predictive performance (Beyer, Schewe and Lotze-Campen (2022)). Second, model evaluation—if any—relies almost exclusively on point forecast metrics, neglecting probabilistic accuracy measures related to the calibration of predictive distributions. As we discuss in more detail later, such calibration is essential for uncertainty quantification and risk assessment. Third, maritime migration patterns are almost exclusively analyzed in the context of the Mediterranean Sea,⁴ limiting our understanding of whether observed dynamics are region-specific or reflect more universal patterns in irregular migration dynamics. In general, little is known about the dynamics and statistical properties of irregular maritime migration.

Our work addresses these gaps and hence makes several contributions to the literature on migration modeling. First, we construct a parsimonious probabilistic model tailored specifically to high-frequency irregular maritime migration time series, grounded in both theoretical reasoning and observed empirical patterns. Despite requiring no covariates, this framework generates accurate distributional forecasts and yields insights relevant to policy design. Second, we pioneer the evaluation of distributional forecast accuracy in this domain, with special emphasis on tail behavior, which is crucial for evaluating uncertainty and providing accurate risk assessments. Third, we provide the first predictive analysis of English Channel crossings alongside Mediterranean routes, broadening the empirical evidence base. More broadly, the dynamic modeling framework we develop extends naturally to applications beyond migration, including epidemiological analyses, accident modeling, or reliability contexts.

⁴An exception is Wood (2025), which appeared after earlier versions of this manuscript were circulated as a working paper.

3. Dynamic zero-inflated heteroskedastic count models.

3.1. *Model specification.* Let $y_t \in \mathbb{N}_0$ denote a univariate count observed at equally-spaced time points $t = 1, \dots, T$. Each count can be a *structural* zero or arise from a *sampling* mechanism. Introduce the latent indicator $s_t \in \{0, 1\}$ with interpretation $s_t = 0$ (structural zero) and $s_t = 1$ (sampling path active). Conditionally on s_t and a latent log-intensity $z_t \in \mathbb{R}$ the data-generating process is

$$(1) \quad y_t | s_t, z_t \sim \begin{cases} \delta_0 & s_t = 0, \\ \mathcal{P}(\lambda_t) & s_t = 1, \end{cases} \quad \lambda_t = \exp(z_t), t = 1, \dots, T,$$

where δ_0 denotes a point mass at zero. The binary indicators are taken to be i.i.d. $s_t \sim \text{Bernoulli}(\pi)$, $\pi \in (0, 1)$. More elaborate dynamic or covariate-driven choices are discussed in Section 5 but are not pursued here. Marginalizing over the indicator yields the zero-inflated Poisson mixture

$$(2) \quad p(y_t | z_t, \pi) = \begin{cases} (1 - \pi) + \pi \exp(-\lambda_t) & y_t = 0, \\ \pi \frac{\lambda_t^{y_t}}{y_t!} \exp(-\lambda_t) & y_t \geq 1, \end{cases} \quad \lambda_t = \exp(z_t).$$

The latent state evolves as a random walk

$$(3) \quad z_t = z_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim f_\varepsilon(\cdot; \boldsymbol{\vartheta}), t = 1, \dots, T,$$

with $p(z_0) \propto 1$ as prior on the initial value and where the density f_ε is parametrized using hyperparameters $\boldsymbol{\vartheta}$.⁵ For f_ε , we consider four distinct specifications. The homoskedastic Gaussian case $f_\varepsilon = \mathcal{N}(0, \sigma^2)$ is used as simple benchmark. For constant heavy-tailed densities, we consider a scaled t -density with ν degrees of freedom $f_\varepsilon = t_\nu(0, \sigma^2)$ and a scale mixture of two Gaussian densities $f_\varepsilon = \sum_{h=1}^2 \eta_h \mathcal{N}(0, \sigma_h^2 \sigma^2)$ with $\eta_1 + \eta_2 = 1$. Finally, following Bijak (2010), to model persistent time-varying volatility patterns, we consider a stochastic volatility model (Kastner (2016)) $f_\varepsilon = \mathcal{N}(0, \exp(h_t))$, where $h_t = \mu + \phi(h_{t-1} - \mu) + \xi_t$ and $\xi_t \sim \mathcal{N}(0, \sigma_\xi^2)$.

The considered model is essentially a heteroskedastic non-Gaussian⁶ state-space model. Poisson state-space formulations and their extensions are well known in the Bayesian literature (West, Harrison and Migon (1985), Frühwirth-Schnatter and Wagner (2006), Berry and West (2020)), and state-space models have been used in migration-related applications in a handful of studies (e.g., Rodríguez Sánchez et al. (2023), Zens and Thalheimer (2025)). Poisson state-space models with a random-walk log intensity augmented with locally adaptive shrinkage or dynamic horseshoe priors have been applied to trend filtering in count series (Kowal, Matteson and Ruppert (2019), Schafer and Matteson (2025)); differenced zero-inflated negative binomial specifications have been used to capture time-varying dispersion in high-frequency financial counts (Barra, Borowska and Koopman (2018)), and spatially varying-dispersion frameworks have been considered in epidemiological settings (e.g.,

⁵Note that the underlying factors driving sea migration are extremely challenging to model explicitly. Consider, for instance, only conflict as a single potential driver of irregular migration. Data on conflict are available in a timely, geocoded format, which is already a major benefit compared to measurements of other potential drivers, enabling detailed modeling of conflict-related migration patterns (see, e.g., Zens and Thalheimer (2025)). At the same time, accounting for the activity of the smuggling gangs, time lag due to travel from the origin region to the point of embarkation, and the waiting time there is essentially a black box. As a result, a latent variable model is particularly useful in this context.

⁶For large counts, a zero-inflated log-normal model may offer a computationally more efficient alternative to a Poisson specification. We do not pursue this approach for two reasons. First, counts are not necessarily large in several applications, including the English Channel data considered here. Second, from a modeling perspective, migration data inherently consists of counts and is, therefore, most naturally treated as such.

and

$$\mathbf{K}(\boldsymbol{\vartheta}) = \text{diag}(k_1, \dots, k_{T+1})$$

collects the increment precisions. The exact form of $\mathbf{K}(\boldsymbol{\vartheta})$ depends on the chosen specification for f_ε and will be detailed below. Combining this prior kernel with the Poisson likelihood contributions for those observations that are on the sampling path ($s_t = 1$) yields

$$p(z_0, \mathbf{z}, z_{T+1} | \mathbf{y}, \mathbf{s}, \boldsymbol{\vartheta}) \propto \exp\left(-\frac{1}{2} \mathbf{z}^{\star\top} \mathbf{P}(\boldsymbol{\vartheta}) \mathbf{z}^{\star}\right) \prod_{t:s_t=1} \mathcal{P}(y_t; \exp(z_t)).$$

Direct sampling from this joint distribution is difficult, so we proceed by single-site updates $p(z_t | \mathbf{z}_{-t}, s_t, y_t, \boldsymbol{\vartheta})$ for $t = 0, \dots, T + 1$, where \mathbf{z}_{-t} collects all latent log-intensities except z_t . For the boundary states z_0 and z_{T+1} , there is no likelihood contribution, so they are updated from their Gaussian full conditionals alone. For $t = 1, \dots, T$, the full conditionals are

$$p(z_t | \mathbf{z}_{-t}, s_t, y_t, \boldsymbol{\vartheta}) \propto \begin{cases} \mathcal{N}(z_t; m_t, v_t) & s_t = 0, \\ \mathcal{N}(z_t; m_t, v_t) \mathcal{P}(y_t; \lambda_t) & s_t = 1, \end{cases} \quad \lambda_t = \exp(z_t),$$

while for $t \in \{0, T + 1\}$,

$$p(z_t | \mathbf{z}_{-t}, \boldsymbol{\vartheta}) \propto \mathcal{N}(z_t; m_t, v_t).$$

The conditional moments follow from standard Gaussian conditioning:

$$m_t = -\frac{1}{P_{tt}} \sum_{j \neq t} P_{tj} z_j, \quad v_t = \frac{1}{P_{tt}},$$

where P_{tj} denotes entry (t, j) of $\mathbf{P}(\boldsymbol{\vartheta})$. Each univariate full conditional is proper, and we sample z_t using an adaptive Metropolis–Hastings step (Roberts and Rosenthal (2009)) that gradually diminishes adaptation of the proposal variance during MCMC and targets an acceptance rate of 0.234. The Poisson term $\mathcal{P}(y_t; \lambda_t)$ dominates whenever y_t is large, effectively collapsing the conditional toward a near-point mass. In the empirical examples considered below, we, therefore, did not encounter notable MCMC efficiency problems despite employing single-site updates.

ii. *Updating s.* To update the indicators s_t , fix $s_t = 1$ for all t where $y_t > 0$. For observations where $y_t = 0$, sample s_t from the conditional posterior $p(s_t | y_t, z_t, \pi)$. Recall that the prior probability of the sampling path being active is π and that $p(y_t = 0 | z_t) = \mathcal{P}(0; \lambda_t) = \exp(-\lambda_t)$. Using Bayes’ theorem, the conditional posterior $p(s_t | y_t, z_t, \pi)$ is hence Bernoulli with

$$p(s_t = 1 | y_t, z_t, \pi) = \frac{\pi \exp(-\lambda_t)}{(1 - \pi) + \pi \exp(-\lambda_t)} \quad \lambda_t = \exp(z_t).$$

iii. *Updating π.* The conditional posterior density $p(\pi | \mathbf{s})$ is Beta, $\pi \sim \mathcal{B}(1 + \sum_{t=1}^T s_t, 1 + T - \sum_{t=1}^T s_t)$.

iv. *Updating ϑ.* Let $\Delta z_t = z_t - z_{t-1}$. The hyperparameter updates depend on the choice of the innovation density f_ε . We present the posterior updates for four distributional assumptions below:

Option 1: Gaussian innovations. When $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$, the variance has a conjugate inverse-gamma posterior,

$$\sigma^2 | \mathbf{z} \sim \text{IG}\left(2.5 + \frac{T + 1}{2}, 0.5 + \frac{1}{2} \sum_{t=1}^{T+1} (\Delta z_t)^2\right).$$

For the latent precision updates, set $k_t = \sigma^{-2}$ for all t .

Option 2: Student- t innovations. We employ the scale mixture representation of the Student- t distribution,

$$\Delta z_t | \omega_t, \sigma^2 \sim \mathcal{N}\left(0, \frac{\sigma^2}{\omega_t}\right), \quad \omega_t | \nu \sim \mathcal{G}\left(\frac{\nu}{2}, \frac{\nu}{2}\right),$$

with prior $\nu - 3 \sim \text{Exp}(1/6)$ ensuring the first three moments of the innovation density exist. The resulting conditional posteriors are

$$\begin{aligned} \sigma^2 | \boldsymbol{\omega}, \mathbf{z} &\sim \mathcal{IG}\left(2.5 + \frac{T+1}{2}, 0.5 + \frac{1}{2} \sum_{t=1}^{T+1} \omega_t \Delta z_t^2\right), \\ \omega_t | \nu, \sigma^2, \mathbf{z} &\sim \mathcal{G}\left(\frac{1+\nu}{2}, \frac{\nu}{2} + \frac{\Delta z_t^2}{2\sigma^2}\right), \quad t = 1, \dots, T+1. \end{aligned}$$

The degrees of freedom ν are updated via adaptive Metropolis–Hastings steps on the log scale. Propose $\log \nu^* \sim \mathcal{N}(\log \nu, \tau_\nu)$, and accept with probability,

$$\alpha = \min\left\{1, \frac{\nu^*}{\nu} \frac{p(\boldsymbol{\omega} | \nu^*)}{p(\boldsymbol{\omega} | \nu)} \frac{p(\nu^*)}{p(\nu)}\right\},$$

where the adaptation of τ_ν targets an acceptance rate of 0.234. For the latent precision updates, set $k_t = \omega_t / \sigma^2$.

Option 3: Finite Gaussian scale mixture. Consider H mixture components with weights $\boldsymbol{\eta} = (\eta_1, \dots, \eta_H)$, component variances σ_h^2 , and allocation indicators $\rho_t \in \{1, \dots, H\}$,

$$\begin{aligned} \Delta z_t | \rho_t = h, \sigma^2, \sigma_h^2 &\sim \mathcal{N}(0, \sigma^2 \sigma_h^2), \quad \sigma_h^2 \sim \mathcal{IG}(2.5, 0.5), \\ \rho_t | \boldsymbol{\eta} &\sim \text{Categorical}(\boldsymbol{\eta}), \quad \boldsymbol{\eta} \sim \text{Dirichlet}(1, \dots, 1). \end{aligned}$$

Define $R_h = \sum_{t=1}^{T+1} \mathbb{1}(\rho_t = h)$ as the number of allocations to component h . The resulting conditional posteriors are

$$\begin{aligned} \sigma^2 | \sigma_{\rho_t}^2, \mathbf{z} &\sim \mathcal{IG}\left(2.5 + \frac{T+1}{2}, 0.5 + \frac{1}{2} \sum_{t=1}^{T+1} \sigma_{\rho_t}^{-2} \Delta z_t^2\right), \\ \text{Pr}(\rho_t = h | \cdot) &\propto \eta_h \cdot \varphi(\Delta z_t; 0, \sigma^2 \sigma_h^2), \\ \boldsymbol{\eta} | \boldsymbol{\rho} &\sim \text{Dirichlet}(1 + R_1, \dots, 1 + R_H), \\ \sigma_h^2 | \cdot &\sim \mathcal{IG}\left(2.5 + \frac{R_h}{2}, 0.5 + \frac{1}{2\sigma^2} \sum_{t:\rho_t=h} \Delta z_t^2\right), \end{aligned}$$

where $\varphi(\cdot; \mu, \sigma^2)$ denotes the Gaussian probability density function with mean μ and variance σ^2 . For latent precision updates, set $k_t = (\sigma^2 \sigma_{\rho_t}^2)^{-1}$. We use $H = 2$ mixture components in the applications below.

Option 4: Stochastic volatility. For the SV case, we employ the auxiliary-mixture sampler of Kim, Shephard and Chib (1998), combined with the multivariate Gaussian joint density representation for state-space formulations (McCausland, Miller and Pelletier (2011)); see Kastner (2016) and the `stochvol` R package for an efficient software implementation. The estimated variance pathways $\sigma_t^2 = \exp(h_t)$ can then be used to update the latent precisions by setting $k_t = \sigma_t^{-2}$.

3.4. *Forecasting and log predictive score evaluation.* To sample from the posterior predictive distribution of y_{T+1} , for each MCMC draw m sample $s_{T+1}^{(m)} \sim \text{Bernoulli}(\pi^{(m)})$. If $s_{T+1}^{(m)} = 0$, set $y_{T+1}^{(m)} = 0$. If $s_{T+1}^{(m)} = 1$, sample z_{T+1} from the state evolution, which under the random-walk prior satisfies $z_{T+1}|z_T, k_{T+1} \sim \mathcal{N}(z_T, k_{T+1}^{-1})$, and then sample $y_{T+1}^{(m)}|z_{T+1}^{(m)} \sim \mathcal{P}(\exp(z_{T+1}^{(m)}))$. Forecasts conditional on $s_{T+1} = 1$ are obtained by skipping the Bernoulli step.

For log predictive distribution evaluation, we integrate out z_{T+1} numerically. For each posterior draw m , define

$$g^{(m)}(y) = \int \mathcal{P}(y; \exp(z)) p(z_{T+1} = z | z_T^{(m)}, k_{T+1}^{(m)}) dz,$$

which we approximate using Gauss–Hermite quadrature. The corresponding full zero-inflated predictive pmf is

$$p^{(m)}(y|\mathbf{y}) = \begin{cases} (1 - \pi^{(m)}) + \pi^{(m)} g^{(m)}(0) & y = 0, \\ \pi^{(m)} g^{(m)}(y) & y \geq 1. \end{cases}$$

Given M posterior draws, the estimated log predictive distribution at y_{T+1} is

$$\widehat{\log p}(y_{T+1}|\mathbf{y}) = \log \left\{ \frac{1}{M} \sum_{m=1}^M p^{(m)}(y_{T+1}|\mathbf{y}) \right\},$$

while the score conditional on $s_{T+1} = 1$ is

$$\widehat{\log p}(y_{T+1}|s_{T+1} = 1, \mathbf{y}) = \log \left\{ \frac{1}{M} \sum_{m=1}^M g^{(m)}(y_{T+1}) \right\}.$$

3.5. *Simulation exercises.* We evaluate the proposed mechanisms and posterior simulation algorithms in a controlled setting using a simulation study. Full details are provided in Supplementary Material Section S1; here we summarize only the main insights for brevity. We compare zero-inflated homoskedastic and stochastic volatility Poisson random walk models against a naive homoskedastic Poisson approach that treats all zeros as arising from the sampling count process. In the first setting, the random walk innovation variance is simulated as constant. The homoskedastic Poisson random walk recovers parameters well, while the stochastic volatility model approximates the constant variance accurately, though, as expected, with somewhat lower efficiency. The naive model severely overestimates variance by misattributing occasional drops to zero as part of the sampling process. In a second setting, the innovation variance is assumed to oscillate between low and high volatility periods. The stochastic volatility model accurately recovers the underlying volatility dynamics. The zero-inflated homoskedastic model produces averaged estimates that overestimate volatility during tranquil periods and underestimate it during volatile periods. The naive Poisson random walk performs worst, severely overestimating innovation variance by ignoring zero-inflation.

Overall, the stochastic volatility specification proves sufficiently flexible to accommodate both constant and time-varying volatility regimes, nesting the simpler homoskedastic model as a special case. This flexibility translates to strong empirical performance in the application presented in Section 4. Regarding zero-inflation, we find no substantial differences between the stochastic volatility and homoskedastic Poisson specification in the settings considered. While the stochastic volatility model performs similarly or marginally better in classifying zeros as structural, both models classify truly structural zeros with very high accuracy.

4. Modeling irregular migration in the Mediterranean Sea and the English Channel.

4.1. *Predictive performance and benchmarking.* Before discussing inferential results under any of the proposed specifications, we first assess distributional forecasting performance under alternative choices of f_{ϵ} . The purpose of this exercise is not to claim that any single specification dominates all conceivable competitors, but rather to identify which model components (e.g., heavy tails or stochastic volatility) are required for reasonable distributional forecasting performance in the context of irregular maritime migration.

We conduct a benchmarking exercise in which we repeatedly split the data into a training sample of length $T - 250$ and a one-step-ahead hold-out observation, whose distribution we aim to predict. We repeat this procedure for 250 hold-out periods using a rolling-window scheme. For each model we collect 75,000 posterior samples after a burn-in period of 7,500 iterations. We then compute several performance metrics, including cumulative log-predictive scores, root mean square error, and correlations between posterior mean forecasts and realized values. To evaluate tail calibration, we additionally report empirical coverage with respect to the predictive tail quantiles q_{01} , q_{05} , q_{10} , q_{90} , q_{95} , and q_{99} .

4.1.1. *Operational target and the role of structural zeros.* As our ultimate aim is to provide operationalizable forecasts *conditional on the sampling regime being active* and to accurately capture the upper tail of the predictive distribution, we report baseline results based on predictive distributions conditional on $s_{T+1} = 1$ and compute out-of-sample metrics using only observations with $y_t > 0$, thereby focusing evaluation on weeks with realized crossing activity. This evaluates how well a model predicts the magnitude of arrivals when they occur. In the context of policy preparedness, this conditional forecasting target is arguably the most operationally relevant. For completeness and as an additional robustness check, we also report results that integrate over uncertainty in s_{T+1} and include all hold-out observations in the Supplementary Material; these are referenced and briefly discussed below.

4.1.2. *Baseline predictive results.* Table 1 reports the baseline out-of-sample results. They confirm our initial assessment that accurate point predictions for irregular sea border crossings are difficult to obtain. The generally low correlation between point forecasts and realized values highlights the limited point predictability of the irregular maritime border-crossing process, reinforcing our decision to emphasize distributional forecasts. For the Mediterranean data, specifications with heavy-tailed innovations—implemented either via mixture models or t -distributions—outperform the homoskedastic Gaussian benchmark in terms of log predictive scores and tail coverage.

The strongest performance in both case studies is achieved by specifications that incorporate stochastic volatility, providing compelling evidence of *persistent* time variation in volatility in the underlying migration intensity process. This suggests that the process alternates between relatively stable periods and episodes of heightened fluctuation. Figure 2 further illustrates this point by tracking cumulative log predictive scores relative to the stochastic volatility model over the hold-out period. For the Mediterranean data, the gains from stochastic volatility are consistent throughout and are particularly pronounced during high-volatility episodes (e.g., early 2022), when the stochastic volatility model is able to adapt more flexibly. For the U.K. data, the gains of stochastic volatility relative to static heavy-tailed specifications are weaker and less uniform over time. Nonetheless, over the considered hold-out period the improvements remain substantial, and we view stochastic volatility as a strong benchmark also for this case study. Future work may further validate or revise this conclusion as longer time series become available.

Turning to tail calibration, empirical tail coverage closely tracks nominal coverage under the stochastic volatility model. It is noteworthy that a highly parsimonious specification

TABLE 1
Results of the predictive exercise and hold-out evaluation

		LPS	RMSE	Corr.	q_{01}	q_{05}	q_{10}	q_{90}	q_{95}	q_{99}	n
ITA	Gaussian (excl. zeros)	-2,144.436	7.563	0.366	0.020	0.057	0.081	0.899	0.931	0.980	247
ITA	Gaussian (incl. zeros)	-2,180.422	7.564	0.367	0.004	0.016	0.040	0.931	0.955	0.984	247
ITA	Gaussian	-2,144.371	7.565	0.365	0.020	0.057	0.081	0.899	0.931	0.980	247
ITA	Mixture	-2,135.893	7.565	0.364	0.012	0.057	0.081	0.895	0.935	0.984	247
ITA	Student- t	-2,138.267	7.565	0.366	0.012	0.057	0.117	0.895	0.931	0.984	247
ITA	Stoch. Vol	-2,121.401	7.562	0.366	0.008	0.057	0.101	0.907	0.955	0.992	247
U.K.	Gaussian (excl. zeros)	-1,726.586	6.546	0.367	0.023	0.068	0.108	0.869	0.919	0.973	222
U.K.	Gaussian (incl. zeros)	-1,854.025	6.594	0.335	0.000	0.000	0.005	0.883	0.905	0.955	222
U.K.	Gaussian	-1,726.057	6.545	0.367	0.023	0.068	0.108	0.869	0.914	0.973	222
U.K.	Mixture	-1,724.793	6.545	0.367	0.023	0.068	0.108	0.869	0.919	0.973	222
U.K.	Student's t	-1,725.238	6.546	0.366	0.014	0.072	0.113	0.860	0.914	0.977	222
U.K.	Stoch. Vol	-1,721.395	6.547	0.368	0.009	0.059	0.117	0.878	0.932	0.986	222

Note: Columns report the cumulative log predictive score (LPS), root mean square error (RMSE) on the logarithmic scale, correlation of forecasts and true values (Corr.), and empirical coverage with respect to the predictive first, fifth, 10th, 90th, 95th, and 99th quantiles ($q_{01} - q_{99}$), with n denoting the number of nonzero hold-out observations. Results are summaries across out-of-sample periods where $y_{T+1} > 0$ and conditional on $s_{T+1} = 1$. Second column refers to the innovation model for the Poisson random walk. ‘‘Gaussian (excl. zeros)’’ is a homoskedastic model where all zeros are treated as missing. ‘‘Gaussian (incl. zeros)’’ is a homoskedastic model where all zeros are considered part of the sampling process. All other models use an explicit zero-inflation mechanism to estimate s_t .

achieves good tail calibration for a highly complex and volatile social process. Accurate tail quantile estimates and forecasts are particularly useful in this context, as they can be used for contingency planning and as operational risk indicators. From the perspective of statistical decision theory, an important consideration is that quantiles solve optimal decision problems under (asymmetric) linear loss functions. If such loss functions could be elicited from forecast users (e.g., migration practitioners or policymakers), even approximately, policy and

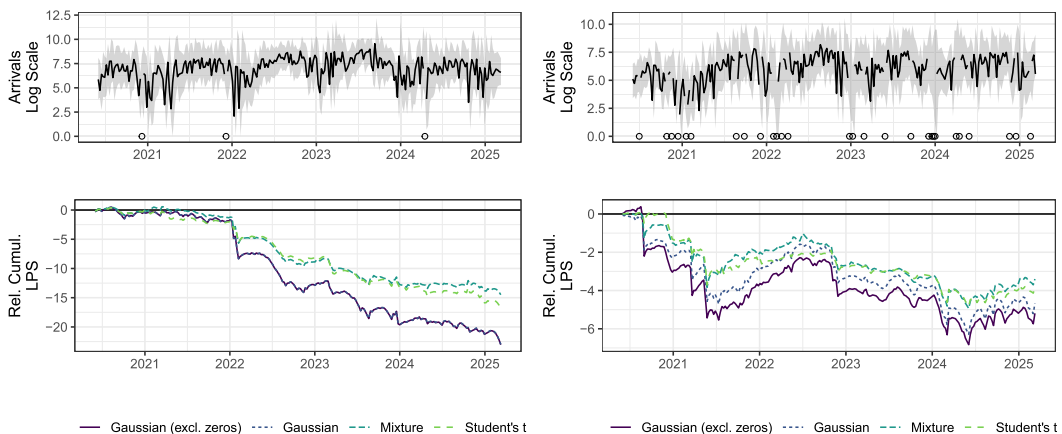


FIG. 2. Top row: One-step-ahead predictive distributions and realized observations (black line) over the 250 hold-out periods. Zero observations are highlighted separately on the x-axis. Grey shaded areas are 90% credible intervals. Bottom row: Cumulative log predictive scores relative to the stochastic volatility model. The Poisson model without zero-inflation is not shown due to noncompetitive performance. Left column: Mediterranean crossing data. Right column: English Channel crossing data. Hold-out period runs from 2020W23 to 2025W11. Results are computed across out-of-sample periods where $y_{T+1} > 0$ and conditional on $s_{T+1} = 1$.

operational preparedness could be substantially improved based on distributional forecasts (see, e.g., [Bijak \(2010\)](#)). So far, such elicitation has not been carried out in the context of migration, although, especially given the high policy salience, there are strong arguments for implementing it as a part of broader preparedness efforts ([EC \(2020\)](#)). Additionally, an explicit elicitation of potential impacts of different migration scenarios would be in line with existing examples of preparedness and contingency planning (e.g., *National Risk Register*, [HM Government \(2025\)](#)) and recent work on infectious disease response ([Mills et al. \(2025\)](#)), which use probabilistic assessment for risk management purposes.

In addition, quantiles are invariant under monotonic transformations. This implies that if the policy objective functions of different decision makers (such as the cost of running reception centers or time of processing asylum applications) can be expressed as monotonic transformations $g(\cdot)$ of the number of irregular migrants y_t , then under linear loss functions, the solutions to such decision problems are simply the values of $g(\cdot)$ evaluated at the respective quantiles of the predictive distribution of y_t . This feature additionally underscores the versatility of the approach, where different users, with different objectives and loss functions, could use the same underlying probability distributions to support their decisions as long as these distributions are well calibrated, including in the tails.

Finally, quantile estimates and forecasts also admit straightforward operational interpretations in our context. For example, the 90th percentile corresponds to an approximate 10% weekly probability of exceeding the indicated threshold, implying an expected exceedance about once every 10 weeks (roughly every two to three months). Likewise, the 95th percentile corresponds to about a 5% weekly probability, or one exceedance approximately every 20 weeks (around every five months). The 99th percentile implies roughly a 1% weekly probability, corresponding to an extreme event about once every 100 weeks, that is, close to once every two years. While these interpretations rely on an independence approximation and should, therefore, be viewed as heuristic, they nonetheless yield transparent and interpretable indicators that can serve as starting points for contingency planning. Developing more comprehensive operational risk indicators for this domain remains an avenue for future research.

4.1.3. Further predictive results and insights. Table S1 in the Supplementary Material reports log predictive scores and tail coverage when marginalizing over s_t and evaluating performance on all hold-out observations (including weeks with $y_t = 0$). Supplementary Material Figure S4 presents the corresponding posterior predictive distributions and cumulative log predictive score plots. Overall, these results lead to very similar conclusions but indicate slightly worse empirical coverage in the lower tail.

This difference primarily reflects our convenient, but simplifying, assumption of a constant probability π over the sample. A constant π has two implications for forecasting. First, when the estimated probability of a structural zero ($1 - \pi$) is sufficiently large, the lowest predictive quantiles become exactly zero (compare Figure S4). This also affects in-sample fits; see Supplementary Material Figure S5. The effect is modest for the Mediterranean series, where zeros are rare and π is estimated close to one. The quantiles are, therefore, similar under conditioning on $s_t = 1$ and marginalizing over s_t . In contrast, for the U.K. series with a larger fraction of zeros, the differences are more substantial. For example, q_{05} is equal to zero throughout the sample period when marginalizing over s_t . Second, a constant π typically overstates the prior probability of structural zeros during periods in which they are uncommon and vice versa. In our application this does not appear to materially affect overall predictive performance, as the likelihood is typically informative about the separation between structural zeros and nonzero observations. Nevertheless, for forecasting it implies that the predictive distribution may be biased, with size and direction of the bias depending on the period under consideration. Model extensions and generalizations that relax the constant- π assumption are discussed in Section 5.

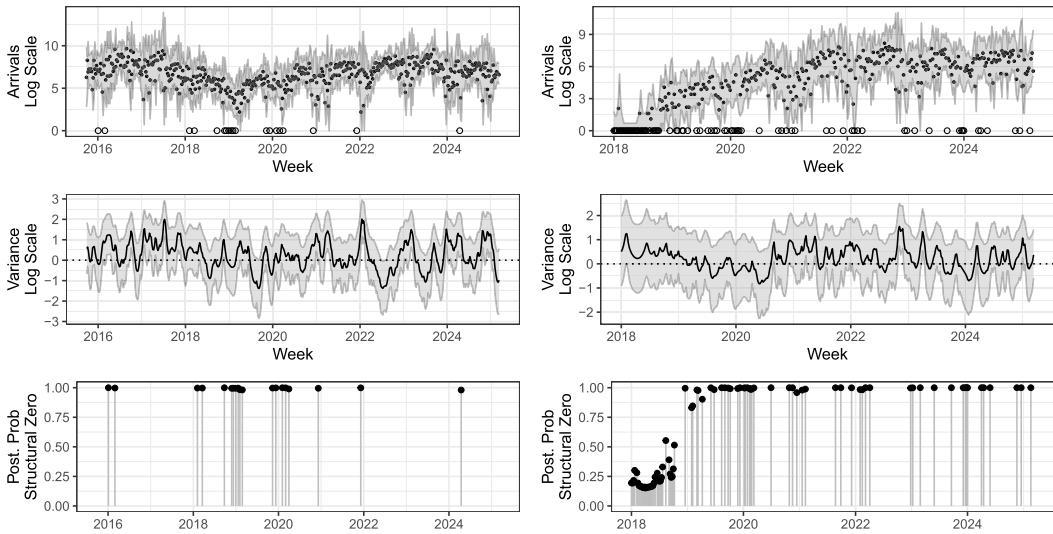


FIG. 3. *Top row: Crossing counts and distributional fit by sample week on a logarithmic scale. Zero counts are highlighted separately on the zero line. Middle row: Estimated posterior density of h_t . Bottom row: Estimated posterior mean probability of y_t being a structural zero. Left column: Mediterranean crossings by sample week from 2015W40 to 2025W11. Right column: English Channel crossings by sample week from 2018W01 to 2025W11. Solid lines are posterior means. Grey shaded areas indicate 90% credible intervals.*

4.2. *Model fit, volatility estimates, zero-inflation patterns.* Figure 3 displays *smoothed one-step-ahead predictive distributions* for the count data models. At each time point t , these distributions are obtained by propagating draws from the full-sample (smoothed) posterior of the latent intensity one step forward through the state-space model and then generating the corresponding count outcome. Repeating this procedure yields a Monte Carlo approximation to the model-implied distribution of y_t conditional on $s_t = 1$ and on the full information in \mathbf{y} .⁷ The middle row of Figure 3 shows posterior density estimates of the log variances $h_t = \log(\sigma_t^2)$, illustrating the time-varying uncertainty captured by the stochastic volatility model. All reported summaries are based on 75,000 posterior draws after a burn-in of 7,500 iterations.

For both the U.K. and Mediterranean crossings data, Figure 3 helps explain why the stochastic volatility specification outperforms competing models in predictive terms. In the Mediterranean case, the estimated variance exhibits clear alternation between high-volatility episodes (e.g., early 2022) and more tranquil periods (e.g., mid-2022). Similar, albeit less pronounced, patterns emerge in the U.K. case study, consistent with the smaller performance gap between the stochastic volatility model and the competing specifications. In both applications, the stochastic volatility component captures shifts between tranquil and volatile regimes, allowing the predictive distribution to widen or tighten adaptively and thereby improving predictive distribution estimates.

In terms of the zero-inflation component, the bottom row of Figure 3 reports the posterior probabilities that an observed zero corresponds to a structural zero. For the Mediterranean data, the posterior mean $\hat{\pi} = 0.96$ indicates that the sampling regime is active most of the time. When zeros do occur, the model assigns them to the structural-zero state ($s_t = 0$) with near-unit probability. As a result, the implied structural-zero probability closely matches the empirical frequency of zeros in the data (3.8%). For the U.K. data, we obtain $\hat{\pi} = 0.85$,

⁷We use this quantity as an illustrative in-sample distributional fit diagnostic. It differs from the real-time one-step-ahead predictive distributions in Figure 2, which condition only on information available up to $t - 1$.

compared to an empirical zero frequency of 22.3%, indicating that not all zeros are classified as structural. In particular, most zeros prior to 2019 are attributed to the sampling process ($s_t = 1$), whereas zeros after 2019 are predominantly classified as structural ($s_t = 0$). This pattern reflects that migration intensity is comparatively low early in the sample. Zeros in that period are interpreted as arising from this low intensity (i.e., few or no attempts) rather than from temporary conditions that render crossings effectively impossible.

4.3. Further exploratory exercises. To gain deeper insight into the drivers of time-varying volatility patterns and structural zeros in sea migration data, we conduct additional empirical analyses. Specifically, we aim to explain the posterior mean of: (i) log variances and (ii) the logit-transformed a posteriori structural-zero probabilities for each t in both case studies using linear regression. For completeness, we also consider: (iii) an exploratory linear regression model with the logarithm of weekly border crossings as the outcome variable.⁸ As explanatory variables these models include monthly and annual effects to capture seasonal patterns and long-term trends as well as potential policy shifts. Additionally, we incorporate a variable measuring wave height in the broader crossing regions (see Supplementary Material Figure S3) to approximate prevailing weather and sea conditions.⁹

The results of these exploratory models, which are detailed in Supplementary Material Table S2 reveal several distinct dynamics. We find that both seasonal and annual patterns predict variance patterns in the Mediterranean case study, with the summer months typically being the most tranquil. A similar pattern is visible in the English Channel example, although it is less pronounced. In both case studies, the wave height variable is positively related to log variance, though the association is relatively weak once seasonality is controlled for in the same model. This likely reflects that wave height and broader meteorological conditions are themselves strongly seasonal.

For both case studies, structural zeros do not exhibit clear seasonal patterns. Wave height, however, is a strong predictor of structural zeros in both settings, underscoring the relevance of weather and sea conditions in maritime migration; compare also the findings in Wood (2025). Given the small number of zero weeks in the Mediterranean data, we find no strong annual effects, whereas for the U.K. the later years show an increase in structural zeros. Note that structural zeros may reflect interception by coast guards, weather effects, accidents, or lack of smuggler availability. We can only explore one of these mechanisms explicitly, since reliable data on interceptions or smuggler activity are typically not publicly available. Finally, wave height also has strong effects on overall arrival intensity on the log scale in both case studies, suggesting increased accident risk and a greater reluctance to initiate a journey in adverse weather conditions as potential underlying channels.

Overall, the explanatory power of the covariates—as measured by the coefficient of determination—remains modest, suggesting that a broader constellation of factors likely drives the volatility, feasibility, and magnitude of crossings. Our findings, therefore, indicate that at least half of the variation in volatility and feasibility cannot be attributed to long-run trends, annual effects, seasonal patterns, or weather and maritime conditions. This substantial unexplained component likely reflects unmeasured factors such as fluctuations in smuggling operations, coast guard activities, and upstream determinants of the potential migrant population (e.g., conflict events and political instability). Importantly, both the reported R^2 values

⁸Note that this procedure is of course quite imperfect and exploratory in nature, given that it ignores various aspects such as correlations across time or uncertainty in the latent outcome variables, among others. Nevertheless, it provides some first-order insights into process dynamics and potential drivers.

⁹Data on significant wave height are collected by the European Centre for Medium-Range Weather Forecasts (ECMWF) and are available at <https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels> (as of 7 August 2025).

and the covariate importance assessments should be interpreted cautiously. They are overly confident, as the regression models do not account for uncertainty in the estimated volatility paths and in the structural-zero probabilities, and due to the presence of autocorrelation.

Future work could incorporate covariate-driven volatility or structural-zero dynamics directly into the model to strengthen latent volatility estimation rather than using covariates only in post-hoc exploratory regressions. This would directly connect to related zero-inflated models to analyze migration dynamics (Wood (2025), Zhang et al. (2026)). For real-world deployment, however, such an approach would face practical challenges, since weather data and other relevant covariates often become available only ex post and would, therefore, also need to be forecast, introducing additional modeling complexity.

5. Discussion and concluding remarks.

5.1. Policy implications. The empirical evidence in favor of a random walk with persistent time-varying volatility in border crossing activity has at least three direct policy implications. First, our finding that irregular migration is well approximated by a random walk confirms earlier intuitions about the inherent difficulty of predicting these flows (Bijak (2010)). Because random walks imply limited forecastability and uncertainty that expands rapidly with the forecast horizon, turning points may easily be missed, even by elaborate point forecasting tools. Policymakers should, therefore, avoid relying on medium-run forecasts and rigid annual quotas in reception infrastructure. Instead, the focus should be on real-time monitoring tools and on maintaining flexible surge capacity (e.g., modular reception centers, standby search-and-rescue assets, or dynamic solidarity mechanisms) that can scale up or down as the migration process unfolds.

Second, the evidence for persistence in volatility patterns suggests clustering of calm and volatile periods, with long stretches of low variance alternating with bursts of high variance. This mirrors field observations that Mediterranean crossings can remain subdued for months and then increase sharply within weeks when, for example, a civil-war front shifts, a smuggling route reopens, or the business model of smuggling networks changes. From a policy perspective, this points to treating migration management as a *risk management* problem, not only as flow management. Risk management explicitly requires taking into account both the *uncertainty* and potential *impact* of unforeseen events (e.g., HM Government (2025); for a migration example, see Bijak et al. (2019)). Our results suggest that an important part of the uncertainty considerations may be related to the *volatility* of flows. During low-volatility windows, policymakers may invest in contingency planning and insurance-type funding (e.g., an EU-wide migration contingency fund) so that cash, assets, and staff can be prepositioned for the next high-volatility spell and scaled up rapidly as demand increases.

Third, a random walk with stochastic volatility implies a heavy-tailed distribution of outcomes, in the sense that extreme events occur more frequently than under standard homoskedastic models. Conceptually, occasional extremes should, therefore, not be treated as “unexpected outliers” or “black swan” events but rather as a structural feature of irregular migration processes. Policy should accordingly prioritize systems that are robust to tail risk. Examples include contingency plans that ensure sufficient search-and-rescue capacity for “worst-case” scenarios, fast-track legal pathways that can relieve pressure once a surge begins, the negotiation of standby humanitarian corridors with coastal states, or trigger-based burden-sharing clauses to reduce pressure on countries of first arrival (e.g., automatic relocation once arrivals exceed a predefined high percentile).

Note that replacing our simpler unit-root assumption with a less stringent but still highly persistent mean-reverting process (e.g., an AR(1) model with persistence parameter close to one) would only slightly alter these conclusions. However, an AR(1) regime would suggest

two additional policy levers. First, interventions that effectively reduce persistence—for instance, through diplomacy in origin states, legal work-migration channels, or humanitarian cash transfers—could speed up the decay of shocks and spikes. Second, because the resulting predictive intervals are large but not unbounded, risk-based insurance and reinsurance instruments (e.g., bonds that pay out when arrivals exceed a high percentile) may be priced more realistically, especially when coupled with statistical decision theory. In short, high persistence implies that shocks linger, yet even weak mean reversion leaves scope for proactive measures that accelerate a return toward long-run “normality.”

In summary, viewing irregular maritime migration through the lens of a random walk with stochastic volatility shifts the policy focus from forecasting and flow management toward risk management and preparedness, and from average-case planning toward governance informed by volatility and tail risk. While a fully fledged real-world operationalization of these ideas is beyond the scope of this article, the well-calibrated distributional forecasts produced by our model may provide a useful starting point in that direction.

5.2. Model extensions and future research. Several avenues for future research remain. First, alternative and more flexible specifications for f_ε could be considered, including asymmetric innovations or combinations of stochastic volatility and heavy tails. Second, a dynamic specification for the sampling probability π_t —for example, via dynamic probit models—would be a natural extension and could improve forecasting performance. Our simplifying assumption of constant π is convenient but does not reflect real-world observations, such as the clustering of zeros over time. Moreover, both the π_t process and the volatility process could be allowed to co-move with the latent count intensity, which may further strengthen out-of-sample predictive performance, particularly for the zero-inflation component. Finally, the conditional mean dynamics could be explored further by assessing whether a highly persistent, but mean-reverting, specification outperforms the RW(1) benchmark. Overall, in the irregular sea border crossings applications studied here, several modeling choices could be refined to improve log predictive scores or achieve more accurate tail coverage. Nevertheless, based on our empirical exercises, we are confident that a random walk with persistent, time-varying volatility provides a useful, simple, and well-performing benchmark framework for irregular migration processes.

Moreover, our analysis has focused on a univariate setting, but extensions to multivariate count data frameworks could be highly valuable. Potential applications in migration and border crossing data are numerous. For example, jointly modeling crossings at different European entry points could enable borrowing strength across borders, border types, and origin countries when modeling sparse and noisy migration time series. This may also help to provide new, granular evidence on the impact of differentially restrictive visa and migration enforcement regimes on changes in migration routes and preferred destinations and on switching from legal to illegal modes of crossing borders (Czaika and Hobolth (2016)). Conceptually, we expect that a straightforward multivariate extension of our current model could provide a useful baseline for multivariate irregular migration dynamics. At the same time, many applications likely exhibit complex cross-series dependency structures, as we have observed, for instance, in ongoing work on probabilistic modeling of asylum dynamics in the EU-27.

On a final note, ethical considerations are paramount when developing models in this domain. Any predictive approach can be problematic given the risk of dual use or misuse—for instance, being used to advocate more restrictive border-closure policies or repressive measures, or exploited by smuggling networks to inform their operations. In our setting, this risk is relatively limited because we model weekly arrivals at the country level rather than, for example, daily route-specific departures. Moreover, as discussed above, border crossing

series are barely point-predictable, which further constrains operational misuse. Ultimately, these concerns intersect with political choices that reflect societal values and are negotiated through democratic processes. At the same time, they must be weighed against the potential benefits of improved preparedness and more effective humanitarian resource allocation, particularly for sea border crossings. Risks of misuse can be further mitigated through ethical oversight, strict data protection protocols, and inclusive stakeholder engagement to ensure that such models serve public safety and humanitarian objectives without undermining individual rights.

Significance statement. Irregular maritime migration—or “small boat” crossings—poses major humanitarian and policy challenges, yet statistical understanding of these flows remains limited. We develop a probabilistic model that captures three key features of weekly sea arrival data: persistent migration intensity, periods of calm alternating with volatile times, and occasional weeks with seemingly unexpected zero crossings. Applied to Mediterranean and English Channel data, we find strong evidence that volatility itself changes over time in persistent patterns. Our model produces well-calibrated forecasts, including accurate predictions of rare tail events. This offers a foundation for policy approaches centered on risk management and preparedness rather than point predictions. The methodology extends to other nonstationary count time series with zero-inflation.

Acknowledgments. The authors gratefully acknowledge Maximilian Böck for insightful discussions that helped shape the early development of this work. We further thank the Associate Editor and three anonymous referees for constructive feedback, and the participants of the *Wittgenstein Centre Conference 2025: Demographic Perspectives on Migration in the 21st Century* for helpful comments.

Funding. Gregor Zens gratefully acknowledges financial support from the European Research Council (ERC) through the *2C-RISK* grant (Grant Agreement No. 101162653).

Jakub Bijak’s work was supported by the Leverhulme Trust (Grant RC-2018-003) for the Leverhulme Centre for Demographic Science, based on earlier work from the Horizon 2020 project *QuantMig* funded by the European Commission (Grant No. 870299).

Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union, the European Research Council Executive Agency, or the Leverhulme Trust. Neither the European Union nor the granting authorities can be held responsible for them.

SUPPLEMENTARY MATERIAL

Supplement to “Dynamic count models with flexible innovation processes for irregular maritime migration” (DOI: [10.1214/26-AOAS2171SUPP](https://doi.org/10.1214/26-AOAS2171SUPP); .pdf). The Supplementary Material contains a simulation study evaluating the proposed zero-inflated models and posterior simulation algorithms, as well as additional tables and figures including full-sample predictive evaluations, exploratory regression results, and further data plots.

REFERENCES

- AKTEKIN, T., SOYER, R. and ZHANG, D. (2026). Bayesian forecasting of zero-inflated time-series of counts. *Int. J. Forecast.*
- BARRA, I., BOROWSKA, A. and KOOPMAN, S. J. (2018). Bayesian dynamic modeling of high-frequency integer price changes. *J. Financ. Econom.* **16** 384–424.
- BERRY, L. R. and WEST, M. (2020). Bayesian forecasting of many count-valued time series. *J. Bus. Econom. Statist.* **38** 872–887. [MR4154894 https://doi.org/10.1080/07350015.2019.1604372](https://doi.org/10.1080/07350015.2019.1604372)

- BEYER, R. M., SCHEWE, J. and LOTZE-CAMPEN, H. (2022). Gravity models do not explain, and cannot predict, international migration dynamics. *Humanit. Soc. Sci. Commun.* **9** 1–10.
- BHATIYA, A. and KADAM, S. (2025). Small Boats, Big Impacts: the Ripple Effects of Irregular Migration Technical Report, Competitive Advantage in the Global Economy (CAGE).
- BIJAK, J. (2010). *Forecasting International Migration in Europe: A Bayesian View*. Springer, Berlin.
- BIJAK, J., ed. (2024). *From Uncertainty to Policy: A Guide to Migration Scenarios* Edward Elgar, Cheltenham Glos.
- BIJAK, J., DISNEY, G., FINDLAY, A. M., FORSTER, J. J., SMITH, P. W. F. and WIŚNIOWSKI, A. (2019). Assessing time series models for forecasting international migration: Lessons from the United Kingdom. *J. Forecast.* **38** 470–487. MR4002373 <https://doi.org/10.1002/for.2576>
- BOSCO, C., MINORA, U., DE RIGO, D., PINGSDORF, J. and CORTINOVIS, R. (2025). Supporting migration policies with forecasts: Illegal border crossings in Europe through a mixed approach. arXiv preprint. Available at [arXiv:2512.10633](https://arxiv.org/abs/2512.10633).
- BOSCO, C., MINORA, U., ROSIŃSKA, A., TEOBALDELLI, M. and BELMONTE, M. (2024). A machine learning architecture to forecast irregular border crossings and asylum requests for policy support in Europe: A case study. *Data Policy* **6** e81.
- CAMARENA, K. R., CLAUDY, S., WANG, J. and WRIGHT, A. L. (2020). Political and environmental risks influence migration and human smuggling across the Mediterranean Sea. *PLoS ONE* **15** e0236646.
- CARAMMIA, M., IACUS, S. M. and WILKIN, T. (2022). Forecasting asylum-related migration flows with machine learning and data at scale. *Sci. Rep.* **12** 1457.
- CHAN, J. C. and JELIAZKOV, I. (2009). Efficient simulation and integrated likelihood estimation in state space models. *Int. J. Math. Model. Numer. Optim.* **1** 101–120.
- CZAIKA, M. and HOBOLTH, M. (2016). Do restrictive asylum and visa policies increase irregular migration into Europe? *Eur. Union Polit.* **17** 345–365. <https://doi.org/10.1177/1465116516633299>
- DAVIS, R. A., HOLAN, S. H., LUND, R. and RAVISHANKER, N., eds. (2016). *Handbook of Discrete-Valued Time Series. Chapman & Hall/CRC Handbooks of Modern Statistical Methods*. CRC Press, Boca Raton, FL. MR3642975
- DEIANA, C., MAHESHRI, V. and MASTROBUONI, G. (2024). Migrants at sea: Unintended consequences of search and rescue operations. *Amer. Econ. J., Econ. Policy* **16** 335–365.
- EC (2020). Commission Recommendation (EU) 2020/1366 of 23 September 2020 on an EU mechanism for preparedness and management of crises related to migration (Migration Preparedness and Crisis Blueprint). *OJ L* **317** 26–38. 1.10.2020. Available at <https://eur-lex.europa.eu/eli/reco/2020/1366/oj/eng>.
- FRIEBEL, G., MANCHIN, M., MENDOLA, M. and PRAROLO, G. (2024). International migration and illegal costs: Evidence from Africa-to-Europe smuggling routes. *J. Int. Econ.* **148** 103878.
- FRONTEX (2024). Migratory Routes. Available at <https://www.frontex.europa.eu/what-we-do/monitoring-and-risk-analysis/migratory-routes/migratory-routes/>.
- FRÜHWIRTH-SCHNATTER, S. and WAGNER, H. (2006). Auxiliary mixture sampling for parameter-driven models of time series of counts with applications to state space modelling. *Biometrika* **93** 827–841. MR2285074 <https://doi.org/10.1093/biomet/93.4.827>
- GEORGIU, H. V. (2016). Identification of refugee influx patterns in Greece via model-theoretic analysis of daily arrivals. arXiv preprint. Available at [arXiv:1605.02784](https://arxiv.org/abs/1605.02784).
- HM GOVERNMENT (2025). National Risk Register: 2025 Edition. Available at <https://bit.ly/uk-nrr-2025>.
- HOFFMANN PHAM, K. and KOMIYAMA, J. (2024). Strategic choices of migrants and smugglers in the central Mediterranean Sea. *PLoS ONE* **19** e0300553.
- IPCC (2012). Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. Special Report of the Intergovernmental Panel on Climate Change. Cambridge Univ. Press, Cambridge, MA. Available at <https://www.cambridge.org/9781107607804>.
- JUNG, R. C., KUKUK, M. and LIESENFELD, R. (2006). Time series of count data: Modeling, estimation and diagnostics. *Comput. Statist. Data Anal.* **51** 2350–2364. MR2307505 <https://doi.org/10.1016/j.csda.2006.08.001>
- KASTNER, G. (2016). Dealing with stochastic volatility in time series using the R package stochvol. *J. Stat. Softw.* **69** 1–30.
- KIM, S. and ALBERT, P. S. (2018). Latent variable Poisson models for assessing the regularity of circadian patterns over time. *J. Amer. Statist. Assoc.* **113** 992–1002. MR3862334 <https://doi.org/10.1080/01621459.2017.1379402>
- KIM, S., SHEPHARD, N. and CHIB, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Rev. Econ. Stud.* **65** 361–393.
- KOWAL, D. R., MATTESON, D. S. and RUPPERT, D. (2019). Dynamic shrinkage processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 781–804. MR3997101 <https://doi.org/10.1111/rssb.12325>

- MCCAUSLAND, W. J., MILLER, S. and PELLETIER, D. (2011). Simulation smoothing for state-space models: A computational efficiency analysis. *Comput. Statist. Data Anal.* **55** 199–212. MR2736547 <https://doi.org/10.1016/j.csda.2010.07.009>
- MIGRATION OBSERVATORY (2025). People crossing the English Channel in small boats. Available at <https://migrationobservatory.ox.ac.uk/resources/briefings/people-crossing-the-english-channel-in-small-boats>.
- MILLS, C., IRONS, N. J., TSUI, J. L. H., SPARROW, S., CARVALHO, L. M., KUCHARSKI, A. J., RATMANN, O., LAMBERT, B., DONNELLY, C. A. et al. (2025). From metric to action: An evaluation framework to translate infectious disease forecasts into policy decisions. medRxiv. <https://doi.org/10.1101/2025.07.20.25331802>
- MUTISO, F., PEARCE, J. L., BENJAMIN-NEELON, S. E., MUELLER, N. T., LI, H. and NEELON, B. (2022). Bayesian negative binomial regression with spatially varying dispersion: Modeling COVID-19 incidence in Georgia. *Spat. Stat.* **52** Paper No. 100703, 19. MR4496717 <https://doi.org/10.1016/j.spasta.2022.100703>
- NEELON, B. (2019). Bayesian zero-inflated negative binomial regression based on Pólya-Gamma mixtures. *Bayesian Anal.* **14** 829–855. MR3960773 <https://doi.org/10.1214/18-BA1132>
- ROBERTS, G. O. and ROSENTHAL, J. S. (2009). Examples of adaptive MCMC. *J. Comput. Graph. Statist.* **18** 349–367. MR2749836 <https://doi.org/10.1198/jcgs.2009.06134>
- RODRÍGUEZ SÁNCHEZ, A., WUCHERPFENNIG, J., RISCHKE, R. and IACUS, S. M. (2023). Search-and-rescue in the central Mediterranean route does not induce migration: Predictive modeling to answer causal queries in migration research. *Sci. Rep.* **13** 11014.
- SCHAFFER, T. L. J. and MATTESON, D. S. (2025). Locally adaptive shrinkage priors for trends and breaks in count time series. *Technometrics* **67** 157–167. MR4866660 <https://doi.org/10.1080/00401706.2024.2407316>
- WEST, M., HARRISON, P. J. and MIGON, H. S. (1985). Dynamic generalized linear models and Bayesian forecasting. *J. Amer. Statist. Assoc.* **80** 73–83. With discussion. MR0786598 <https://doi.org/10.1080/01621459.1985.10477131>
- WOOD, R. M. (2025). Factors affecting illegal migration to the UK by small-boat crossing of the English channel: Descriptive, inferential and predictive analyses. *Int. Migr. Rev.* <https://doi.org/10.1177/0197918325139400>
- ZAMBIASI, D. and ALBAROSA, E. (2025). Externalizing rescue operations at sea: The migration deal between Italy and Libya. *J. Econ. Geogr.* **25** 41–58.
- ZENS, G. and BIJAK, J. (2026). Supplement to “Dynamic count models with flexible innovation processes for irregular maritime migration.” <https://doi.org/10.1214/26-AOAS2171SUPP>
- ZENS, G. and THALHEIMER, L. (2025). The short-term dynamics of conflict-driven displacement: Bayesian modeling of disaggregated data from Somalia. *Ann. Appl. Stat.* **19** 286–301. MR4888111 <https://doi.org/10.1214/24-aos1959>
- ZHANG, L., ZHOU, M., YU, K. and TIAN, M. (2026). A spatiotemporal marginalized zero-inflated Conway–Maxwell–Poisson regression model: Application to international population outmigration within Asia. *J. Roy. Statist. Soc., Ser. A, Statist. Soc.* <https://doi.org/10.1093/jrssa/qnag009>