Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2026, 2026:56
https://doi.org/10.28919/cmbn/9839
ISSN: 2052-2541

A NON-STATIONARY NB–INAR(1) MODEL WITH TIME-VARYING BETA–BINOMIAL THINNING: PROPERTIES AND ROLLING ESTIMATION

MOISE SALEH WEMBO[1,*], BONFACE MIYA MALENJE[2], MICHAEL MARKO SESAY[1]

[1]Department of Mathematics, Pan African University Institute for Basic Sciences, Technology and Innovation (PAUSTI), JKUAT Campus, P.O. Box 62000-00200, Nairobi, Kenya

[2]Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology (JKUAT), P.O. Box 62000-00200, Nairobi, Kenya

Abstract. Many count time series exhibit non-stationary behavior such as drifting intensity, regime switching, structural breaks, evolving serial dependence, and changing innovation levels, which can invalidate classical stationary INAR models. We propose a non-stationary Negative Binomial INAR(1) model based on Beta–Binomial thinning, where the thinning variable is randomized through a Beta mixing distribution. This construction yields a transparent decomposition of dynamic persistence and dispersion: the time-varying mean persistence $\tau_t = \mathbb{E}(\Theta_t)$ governs dependence, while the time-varying precision $\phi_t = a_t + b_t$ controls survivor-variance inflation induced by mixing. We derive closed-form conditional moment identities and time-varying recursions for the unconditional mean, variance, and autocovariance structure, highlighting how dependence propagates through products of τ_t. For inference, we develop practical moment/CLS-based procedures under local stationarity, including rolling-window CLS and smooth parametric evolution for (τ_t, m_t), together with feasibility constraints that preserve the probabilistic interpretation. Monte Carlo experiments under multiple non-stationary regimes demonstrate accurate tracking under gradual evolution, quantify the RMSE–adaptation-delay trade-off induced by window length under breaks and regime switching, and illustrate the volatility impact of precision variation. An application to weekly accident

*Corresponding author
E-mail address: wembo.saleh@students.jkuat.ac.ke
Received February 21, 2026

counts shows that a smooth time-varying innovation mean improves one-step-ahead forecasting, residual lag dependence becomes negligible after mean correction, and negative binomial predictive intervals provide improved coverage under mild overdispersion.

Keywords: beta–binomial thinning; count time series; INAR model; local stationarity; negative binomial; non-stationarity; rolling-window estimation; structural break; time-varying persistence; uncertainty quantification.

2020 AMS Subject Classification: 62M10, 60G10.

1. Introduction

Integer-valued (count) time series occur across epidemiology, economics, finance, environmental sciences, transportation, and public safety. Because the observations are discrete and nonnegative, Gaussian ARMA models are often inappropriate, motivating models that preserve integer support while retaining an interpretable dependence structure. A leading approach is the *thinning-based* integer-valued autoregressive (INAR) family, which replaces scalar multiplication by a stochastic "survival" operator so that past counts generate integer-valued offspring at the next time point [1, 2, 3, 4].

1.1. Motivation: when stationarity fails for count series.

Most classical INAR specifications are *time-homogeneous*: the thinning mechanism (persistence) and the innovation law (new arrivals) are constant over time [4, 5]. This can be reasonable over short horizons, but many real count series are *structurally non-stationary*. Typical departures include gradual mean drift driven by changing exposure or reporting, abrupt breaks due to interventions or system changes, and evolving dependence or volatility as persistence strengthens or weakens over time [6, 7, 8]. In these settings, a stationary fit may be locally acceptable yet globally misleading, yielding unstable forecasts and distorted uncertainty quantification. At the same time, allowing all components to vary freely can be ill-posed; non-stationarity must therefore be introduced in a controlled, estimable, and interpretable way.

1.2. Related work: INAR(1), random-coefficient thinning, and time-varying count dynamics.

The INAR(1) process was introduced via binomial thinning, yielding an AR(1)-like dependence structure for discrete data [1, 2]. Higher-order extensions and core existence/ergodicity results for INAR(p) were established by Du and Li [3], while thinning operators and modeling variants were reviewed by Weiß [4] (see also [5]).

To accommodate overdispersion and latent heterogeneity, *generalized* and *random-coefficient* thinning mechanisms have been proposed. Early constructions with negative-binomial or geometric marginals appear in McKenzie [9], and convolution-closed infinitely divisible (CCID) frameworks were developed by Joe [10]. Generalized thinning and the GINAR(p) class were studied by Latour [11], while explicit random-coefficient INAR models with stochastic thinning parameters and inference theory were developed by Zheng et al. [12]. Beta–Binomial-type mechanisms and extra-binomial variation in thinning contexts have also been discussed by Weiß and Kim [13].

Non-stationary count dynamics have been treated via models with covariates or trends, time-varying parameters, and state-space formulations [6, 7, 8]. Nonetheless, non-stationary INAR-type models that preserve a clear survival or innovation decomposition while allowing time variation in both persistence and dispersion, and that remain tractable through usable moment recursions with implementable feasibility constraints, are still comparatively limited.

1.3. Contributions.

We develop a *non-stationary* Negative Binomial INAR(1) model based on *Beta–Binomial thinning* and provide a practical inferential toolkit. The main contributions are:

- **Model.** A non-stationary NB–INAR(1) process in which thinning follows a Beta–Binomial random-environment mechanism and innovations are Negative Binomial, with key parameters allowed to evolve over time.
- **Probabilistic properties.** Closed-form conditional moments and iterated (unconditional) recursions for the mean and variance, including a variance decomposition that isolates survivor-variance inflation from Beta mixing and a time-varying dependence structure driven by the evolving persistence sequence.
- **Estimation under feasibility constraints.** Rolling or local conditional least squares (CLS) and complementary moment-based procedures for tracking time-varying persistence and innovation intensity, together with explicit constraints ensuring admissible Beta and Negative Binomial parameterizations.

- **Monte Carlo validation.** Simulation studies under multiple non-stationary regimes (smooth drift, regime switching, and breaks) quantifying bias and RMSE, break adaptation delay, and the effect of thinning precision on volatility and estimator noise.

2. Related Literature and Stationary Benchmark

This section provides a compact literature map and a stationary benchmark that anchors the non-stationary extension.

2.1. Compact literature map.

Table 1 positions the proposed non-stationary NB–INAR(1) with Beta–Binomial thinning relative to the core INAR lineage (binomial thinning), stationary NB-type count autoregressions, random-coefficient thinning, and recent time-varying parameter developments.

2.2. Thinning operators recap: binomial vs. Beta–Binomial.

Thinning operators transfer AR-type dependence to the integer domain while preserving non-negativity and discreteness [14, 4, 5].

Binomial thinning. Let $X \in \mathbb{N}_0$ and $\alpha \in (0,1)$. The classical binomial thinning operator is

$$\alpha \circ X = \sum_{i=1}^{X} B_i, \qquad B_i \sim \text{Bernoulli}(\alpha), \tag{1}$$

so that $\alpha \circ X \mid X = n \sim \text{Bin}(n, \alpha)$. In particular,

$$\mathbb{E}[\alpha \circ X \mid X = n] = \alpha n, \qquad \text{Var}(\alpha \circ X \mid X = n) = \alpha(1-\alpha)n. \tag{2}$$

This operator underpins the original INAR(1) construction [15, 1].

Beta–Binomial thinning. To incorporate extra-binomial variability, randomize the survival probability through a Beta law. Let $\Theta \sim \text{Beta}(a, b)$ be independent of X, and define

$$\Theta \circ X = \sum_{i=1}^{X} B_i, \qquad B_i \mid \Theta = \theta \sim \text{Bernoulli}(\theta). \tag{3}$$

Then $(\Theta \circ X) \mid (X = n, \Theta = \theta) \sim \text{Bin}(n, \theta)$ and, marginally, $(\Theta \circ X) \mid X = n$ follows a Beta–Binomial distribution with pmf

$$\mathbb{P}(\Theta \circ X = k \mid X = n) = \binom{n}{k} \frac{B(k+a, n-k+b)}{B(a,b)}, \qquad k = 0, \ldots, n, \tag{4}$$

where $B(\cdot,\cdot)$ denotes the Beta function. Write the mean persistence and precision as

$$\tau = \mathbb{E}(\Theta) = \frac{a}{a+b}, \qquad \kappa = a+b. \tag{5}$$

The conditional moments are

$$\mathbb{E}[\Theta \circ X \mid X = n] = \tau n, \tag{6}$$

and

$$\mathrm{Var}(\Theta \circ X \mid X = n) = n\,\tau(1-\tau)\,\frac{\kappa+n}{\kappa+1}. \tag{7}$$

Relative to (2), the factor $(\kappa+n)/(\kappa+1)$ inflates the conditional variance when κ is small, reflecting extra-binomial variation [13]. As $\kappa \to \infty$ with τ fixed, Θ concentrates at τ and (3) reduces to binomial thinning.

2.3. Stationary NB–INAR(1) with Beta–Binomial thinning (benchmark). As a stationary baseline (used later as a local approximation), consider

$$Y_t = \eta_t + \xi_t, \qquad \eta_t = \Theta_t \circ Y_{t-1}, \qquad t \geq 1, \tag{8}$$

where $\{\Theta_t\}$ are i.i.d. Beta(a,b) and independent of the innovation sequence $\{\xi_t\}$. To represent overdispersion in new arrivals, let ξ_t be i.i.d. Negative Binomial (NB), as in stationary NB-type count autoregressions [16, 17, 18]. The mean persistence is $\tau = a/(a+b) \in (0,1)$ and the precision is $\kappa = a+b$.

This benchmark extends the classical Poisson-INAR(1) model [15, 1] by (i) randomizing the thinning probability through Θ_t and (ii) allowing NB innovations to flexibly capture overdispersion.

2.4. Key stationary properties (used later). Let $\mu_\xi = \mathbb{E}(\xi_t)$ and $\sigma_\xi^2 = \mathrm{Var}(\xi_t)$, assumed finite.

Mean stationarity. From $\mathbb{E}(Y_t \mid Y_{t-1}) = \tau Y_{t-1} + \mu_\xi$ by (6), stationarity implies

$$\mu := \mathbb{E}(Y_t) = \frac{\mu_\xi}{1-\tau}, \tag{9}$$

provided $0 < \tau < 1$ [4, 5].

Variance stationarity and the role of precision. By the law of total variance and (7),

$$\sigma^2 := \operatorname{Var}(Y_t) = \sigma_\xi^2 + \tau^2\sigma^2 + \mathbb{E}[\operatorname{Var}(\eta_t \mid Y_{t-1})], \tag{10}$$

and a closed-form expression can be written in terms of (τ,κ), μ, and (μ_ξ,σ_ξ^2). One convenient form is

$$\sigma^2 = \frac{\sigma_\xi^2 + \dfrac{\tau(1-\tau)}{\kappa+1}\left(\kappa\mu+\mu^2\right)}{(1-\tau)\left[(1+\tau) - \dfrac{\tau}{\kappa+1}\right]}. \tag{11}$$

Equation (11) makes explicit that smaller κ inflates variability through increased randomness in Θ_t; as $\kappa \to \infty$, the binomial-thinning variance formula is recovered.

Geometric autocorrelation decay. Under (8), the lag-k autocorrelation decays geometrically:

$$\rho(k) = \operatorname{Corr}(Y_t, Y_{t-k}) = \tau^k, \qquad k \geq 1, \tag{12}$$

mirroring an AR(1) structure in the count setting [1, 4, 5]. This property is used later to motivate local dependence estimation when parameters are approximately constant within short windows.

TABLE 1. Summary of closely related work and the present contribution.

Reference	Model class	Thinning / dependence	Main focus / limitation
[14]	Foundational thinning theory	Discrete self-decomposability; basis for thinning operators	Establishes core discrete-time ideas; not an applied INAR model.
[15, 1]	INAR(1) / Poisson-INAR	Binomial thinning with fixed α	Canonical INAR(1); limited for overdispersed counts.
[16, 17]	Stationary NB-type count AR/INAR	Fixed-coefficient dependence (NB constructions)	Handles overdispersion; typically time-homogeneous.
[10, 12]	Random-coefficient thinning INAR	Random α_t (heterogeneity)	Random persistence; often treated in stationary settings.
[13]	Extra-binomial variation	Beta–Binomial-type variability	Motivates Beta mixing; limited non-stationary toolkit.
[8]	Time-varying parameter count time series	Time-varying parameters (local methods)	Time variation addressed; model components differ from the Beta–Binomial plus NB setup considered here.
This work	Non-stationary NB–INAR(1)	**Beta–Binomial thinning plus NB innovations**, time-varying components	Moment recursions plus time-varying dependence; rolling or local CLS and moment estimators with feasibility constraints; Monte Carlo under drift, switching, and breaks.

3. Model: Non-stationary NB–INAR(1) with Beta–Binomial Thinning

We generalize the classical INAR(1) construction by randomizing the thinning probability through Beta mixing and by allowing persistence and innovation intensity to vary over time. The resulting model preserves the INAR decomposition into survivors and new arrivals while accommodating evolving dependence and overdispersion [1, 4, 5].

3.1. Definition and hierarchical construction.

Let $\{X_t\}_{t\geq 0}$ be a process on $\mathbb{N}_0$ with X_0 given. For $t \geq 1$,

$$X_t = \eta_t + \xi_t, \qquad \eta_t = \Theta_t \circ X_{t-1}, \tag{13}$$

where $\Theta_t \in (0,1)$ is a time-varying random thinning probability and $\circ$ denotes binomial thinning conditional on Θ_t. The innovations $\{\xi_t\}$ are independent of (Θ_t, X_{t-1}).

A convenient hierarchical specification is

$$\Theta_t \sim \mathrm{Beta}(a_t, b_t), \tag{14}$$

$$\eta_t \mid (X_{t-1} = x, \Theta_t = \theta) \sim \mathrm{Bin}(x, \theta), \tag{15}$$

$$\xi_t \sim \mathrm{NB}(r, p_t), \qquad p_t = \frac{r}{r + m_t}, \tag{16}$$

and

$$X_t = \eta_t + \xi_t, \tag{17}$$

where $m_t = \mathbb{E}(\xi_t) > 0$ is time-varying and $r > 0$ is a dispersion parameter, taken to be fixed or piecewise fixed; see Section 3.4. This defines a time-inhomogeneous first-order Markov count process.

3.2. Beta–Binomial thinning with time-varying (a_t, b_t): persistence and precision.

Marginally, conditional on $X_{t-1} = x$, the survivor term follows a Beta–Binomial law:

$$\eta_t \mid (X_{t-1} = x) \sim \text{Beta–Binomial}(x; a_t, b_t), \tag{18}$$

capturing extra-binomial variability [13].

We reparameterize (a_t, b_t) by separating mean persistence and precision:

$$\tau_t = \mathbb{E}(\Theta_t) = \frac{a_t}{a_t + b_t} \in (0,1), \qquad \phi_t = a_t + b_t > 0, \tag{19}$$

so that

$$a_t = \tau_t \phi_t, \qquad b_t = (1 - \tau_t)\phi_t, \tag{20}$$

and

$$\mathrm{Var}(\Theta_t) = \frac{\tau_t(1-\tau_t)}{\phi_t + 1}. \tag{21}$$

Thus, smaller ϕ_t implies a more variable random environment for survival.

The conditional moments of η_t are

$$\mathbb{E}(\eta_t \mid X_{t-1} = x) = \tau_t x, \tag{22}$$

and

$$\mathrm{Var}(\eta_t \mid X_{t-1} = x) = x\tau_t(1-\tau_t)\frac{\phi_t + x}{\phi_t + 1}. \tag{23}$$

The factor $(\phi_t + x)/(\phi_t + 1)$ inflates the conditional variance relative to binomial thinning; as $\phi_t \to \infty$ with τ_t fixed, Beta–Binomial thinning reduces to standard binomial thinning.

3.3. Negative Binomial innovations with time variation. We model arrivals by a Negative Binomial law, suitable when variance exceeds the mean [16, 17, 18]. Using the mean-dispersion parametrization,

$$\xi_t \sim \mathrm{NB}(r, p_t), \qquad m_t = \mathbb{E}(\xi_t) = \frac{r(1-p_t)}{p_t}, \qquad \mathrm{Var}(\xi_t) = m_t + \frac{m_t^2}{r}. \tag{24}$$

Equivalently, $p_t = r/(r + m_t)$ as in (16). This isolates non-stationarity in the innovation mean through m_t while keeping dispersion parsimonious via r.

From (22) and (24), the conditional mean is

$$\mathbb{E}(X_t \mid X_{t-1}) = \tau_t X_{t-1} + m_t, \tag{25}$$

and, using conditional independence of η_t and ξ_t given X_{t-1},

$$\mathrm{Var}(X_t \mid X_{t-1} = x) = x\tau_t(1-\tau_t)\frac{\phi_t + x}{\phi_t + 1} + m_t + \frac{m_t^2}{r}. \tag{26}$$

The first term is survivor uncertainty, amplified when ϕ_t is small, and the second is innovation overdispersion, controlled by r.

3.4. Notation and identifiability convention. Throughout the paper:

- $\Theta_t \sim \mathrm{Beta}(a_t, b_t)$ is the random thinning probability;
- $\tau_t = \mathbb{E}(\Theta_t)$ is the mean persistence;
- $\phi_t = a_t + b_t$ is the precision of Θ_t;
- $\xi_t \sim \mathrm{NB}(r, p_t)$ has time-varying mean $m_t = \mathbb{E}(\xi_t)$ and dispersion r.

The stationary benchmark is recovered when $(\tau_t, \phi_t, m_t) \equiv (\tau, \phi, m)$.

We do not let (r_t, p_t) vary freely over time. Allowing both components of the innovation law to evolve simultaneously is weakly identified from a single trajectory, especially within short rolling windows, because distinct pairs can reproduce similar local mean-variance patterns. This confounds changes in intensity with changes in overdispersion and can destabilize local estimation.

We therefore let the innovation mean m_t vary over time and treat r as either (i) globally constant or (ii) piecewise constant across detected regimes. With r fixed or piecewise fixed, m_t uniquely determines

$$p_t = \frac{r}{r + m_t}, \tag{27}$$

and rolling or local estimators can target (τ_t, m_t) without over-parameterization.

The kpic in Figure 1 summarizes the modeling logic and makes explicit where non-stationarity enters. The next count X_t is decomposed into a *survivor* component $\eta_t = \Theta_t \circ X_{t-1}$ and an *innovation* component ξ_t, preserving the standard INAR interpretation. Time variation is introduced through two interpretable pathways: the persistence $\tau_t = \mathbb{E}(\Theta_t)$, and optionally the precision $\phi_t = a_t + b_t$, governs how much of X_{t-1} is expected to carry over and how variable that survival mechanism is, while the innovation mean $m_t = \mathbb{E}(\xi_t)$ controls the evolving arrival intensity, with dispersion regulated by r. This separation clarifies identification and estimation: τ_t primarily drives dependence through products $\prod \tau_j$, whereas ϕ_t affects volatility via survivor-variance inflation, allowing changes in dependence and changes in dispersion to be discussed and modeled distinctly.

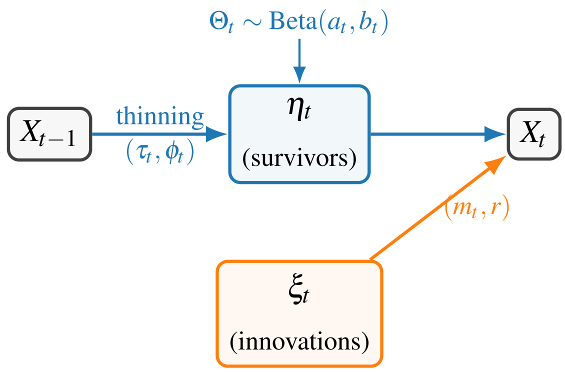

FIGURE 1. **kpic (model backbone).** The INAR decomposition $X_t = \Theta_t \circ X_{t-1} + \xi_t$ is preserved. Non-stationarity enters through time-varying persistence $\tau_t = \mathbb{E}(\Theta_t)$ and/or innovation mean $m_t = \mathbb{E}(\xi_t)$, and optionally through precision $\phi_t = a_t + b_t$.

4. PROBABILISTIC PROPERTIES

We collect the main distributional components and moment recursions used for interpretation and estimation: Markov structure, conditional moments, unconditional recursions under time variation, and the induced time-varying dependence. Related thinning-based derivations can be found in [1, 4, 5].

4.1. Markov structure and conditional distributional components.

The process $\{X_t\}$ is a time-inhomogeneous first-order Markov chain:

$$\mathbb{P}(X_t \in A \mid X_{t-1}, X_{t-2}, \dots) = \mathbb{P}(X_t \in A \mid X_{t-1}), \qquad A \subseteq \mathbb{N}_0. \tag{28}$$

Given $X_{t-1} = x$, $X_t = \eta_t + \xi_t$ is a convolution of a Beta–Binomial survivor term and a Negative Binomial innovation:

$$\mathbb{P}(X_t = y \mid X_{t-1} = x) = \sum_{k=0}^{\min(x,y)} \mathbb{P}(\eta_t = k \mid X_{t-1} = x)\,\mathbb{P}(\xi_t = y - k). \tag{29}$$

The survivor distribution is

$$\mathbb{P}(\eta_t = k \mid X_{t-1} = x) = \binom{x}{k} \frac{B(k + a_t, x - k + b_t)}{B(a_t, b_t)}, \qquad k = 0, 1, \dots, x, \tag{30}$$

and, for $\xi_t \sim \mathrm{NB}(r, p_t)$,

$$\mathbb{P}(\xi_t = j) = \frac{\Gamma(j+r)}{\Gamma(r)\, j!}\, p_t^r (1-p_t)^j, \qquad j = 0,1,2,\ldots, \tag{31}$$

with the standard Gamma-function extension when r is non-integer.

4.2. Conditional mean (core estimating equation).

Let $\mathscr{F}_{t-1} = \sigma(X_{t-1}, X_{t-2}, \ldots)$. Then

$$\mathbb{E}(X_t \mid \mathscr{F}_{t-1}) = \mathbb{E}(X_t \mid X_{t-1}) = \tau_t X_{t-1} + m_t, \tag{32}$$

which preserves an AR(1)-type linear conditional mean while allowing (τ_t, m_t) to evolve over time.

4.3. Conditional variance decomposition.

By conditional independence of η_t and ξ_t given X_{t-1},

$$\mathrm{Var}(X_t \mid X_{t-1}) = \mathrm{Var}(\eta_t \mid X_{t-1}) + \mathrm{Var}(\xi_t). \tag{33}$$

The Beta–Binomial survivor variance is

$$\mathrm{Var}(\eta_t \mid X_{t-1} = x) = x\tau_t(1-\tau_t)\frac{\phi_t + x}{\phi_t + 1}, \tag{34}$$

exhibiting survivor-variance inflation through $(\phi_t + x)/(\phi_t + 1)$ when ϕ_t is finite [13]. The NB innovation variance is

$$\mathrm{Var}(\xi_t) = m_t + \frac{m_t^2}{r}. \tag{35}$$

Hence,

$$\mathrm{Var}(X_t \mid X_{t-1} = x) = x\tau_t(1-\tau_t)\frac{\phi_t + x}{\phi_t + 1} + m_t + \frac{m_t^2}{r}. \tag{36}$$

4.4. Unconditional mean recursion and explicit solution.

Let $\mu_t = \mathbb{E}(X_t)$ with $\mu_0 = \mathbb{E}(X_0)$. Taking expectations in (32) yields

$$\mu_t = \tau_t \mu_{t-1} + m_t, \qquad t \geq 1. \tag{37}$$

Iterating gives

$$\mu_t = \left(\prod_{j=1}^{t} \tau_j\right) \mu_0 + \sum_{i=1}^{t} \left(m_i \prod_{j=i+1}^{t} \tau_j\right), \tag{38}$$

with the convention that an empty product equals 1.

4.5. Unconditional variance recursion (role of precision). Let $\sigma_t^2 = \mathrm{Var}(X_t)$. By the law of total variance,

$$\sigma_t^2 = \mathbb{E}[\mathrm{Var}(X_t \mid X_{t-1})] + \mathrm{Var}(\mathbb{E}(X_t \mid X_{t-1})). \tag{39}$$

Using (32),

$$\mathrm{Var}(\mathbb{E}(X_t \mid X_{t-1})) = \tau_t^2 \sigma_{t-1}^2. \tag{40}$$

Moreover,

$$\mathbb{E}[\mathrm{Var}(\eta_t \mid X_{t-1})] = \frac{\tau_t(1-\tau_t)}{\phi_t + 1} \mathbb{E}[X_{t-1}(\phi_t + X_{t-1})], \tag{41}$$

and since $\mathbb{E}(X_{t-1}^2) = \sigma_{t-1}^2 + \mu_{t-1}^2$, we obtain

$$\sigma_t^2 = \tau_t^2 \sigma_{t-1}^2 + \frac{\tau_t(1-\tau_t)}{\phi_t + 1} \left(\phi_t \mu_{t-1} + \sigma_{t-1}^2 + \mu_{t-1}^2\right) + m_t + \frac{m_t^2}{r}. \tag{42}$$

Thus, smaller ϕ_t increases the survivor-variance contribution and inflates σ_t^2 even when (τ_t, m_t) are unchanged.

4.6. Time-varying autocovariance and autocorrelation. Define $\gamma(t,s) = \mathrm{Cov}(X_t, X_s)$ for $s < t$. Using the orthogonality of conditional expectations and (32),

$$\gamma(t,s) = \tau_t \, \gamma(t-1,s), \qquad s \leq t-1. \tag{43}$$

Iterating yields

$$\gamma(t,s) = \left(\prod_{j=s+1}^{t} \tau_j \right) \sigma_s^2, \qquad 0 \leq s < t, \tag{44}$$

and, for lag $k \geq 1$,

$$\gamma_t(k) := \mathrm{Cov}(X_t, X_{t-k}) = \left(\prod_{j=t-k+1}^{t} \tau_j \right) \sigma_{t-k}^2. \tag{45}$$

Therefore,

$$\rho_t(k) := \mathrm{Corr}(X_t, X_{t-k}) = \left(\prod_{j=t-k+1}^{t} \tau_j \right) \sqrt{\frac{\sigma_{t-k}^2}{\sigma_t^2}}. \tag{46}$$

Dependence is driven by products of the persistence path $\{\tau_j\}$, while the normalization reflects time variation in marginal variance.

4.7. Special and limiting cases.

- **Stationary benchmark.** If $(\tau_t, \phi_t, m_t) \equiv (\tau, \phi, m)$, then (37)–(42) reduce to stationary recursions and (46) collapses to geometric decay $\rho(k) = \tau^k$ [1, 4, 5].
- **Binomial thinning limit.** If $\phi_t \to \infty$ with τ_t fixed, then $\Theta_t \to \tau_t$ and the survivor-variance inflation vanishes, since the factor $(\phi_t + x)/(\phi_t + 1) \to 1$ in (34).
- **No persistence.** If $\tau_t \equiv 0$, then $X_t \equiv \xi_t$ is an independent, possibly non-stationary in mean, NB sequence.
- **Poisson innovation limit.** If $r \to \infty$ with m_t fixed, then ξ_t approaches Poisson(m_t) and $\mathrm{Var}(\xi_t) \to m_t$.

4.8. Stability, feasibility, and interpretation. Because the model is time-inhomogeneous, stability is understood as non-explosiveness, that is, bounded moments and well-behaved time-varying recursions, rather than strict stationarity [1, 4, 5].

4.8.1. *Mean stability and non-explosiveness.* Recall the mean recursion

$$\mu_t = \tau_t \mu_{t-1} + m_t, \qquad t \geq 1. \tag{47}$$

A convenient sufficient condition for mean non-explosiveness is

$$\sup_{t \geq 1} \tau_t \leq \bar{\tau} < 1, \qquad \sup_{t \geq 1} m_t \leq \bar{m} < \infty. \tag{48}$$

Then

$$\mu_t \leq \bar{\tau}^t \mu_0 + \sum_{i=1}^{t} \bar{\tau}^{t-i} \bar{m} \leq \bar{\tau}^t \mu_0 + \bar{m} \frac{1 - \bar{\tau}^t}{1 - \bar{\tau}} \leq \mu_0 + \frac{\bar{m}}{1 - \bar{\tau}}, \tag{49}$$

so $\sup_t \mu_t < \infty$. In particular, $(1 - \tau_t)^{-1}$ acts as an effective memory length: values of τ_t close to 1 imply slow forgetting and persistent shocks.

4.8.2. *Near-nonstationary behavior when $\tau_t \approx 1$.* If τ_t remains close to 1 over extended intervals, mean reversion is weak and small changes in m_t can accumulate. Writing $\tau_t = 1 - \varepsilon_t$ with $\varepsilon_t > 0$ small, two practical consequences follow:

- **Window-lag trade-off.** Local estimators typically require larger windows to reduce variance when persistence is high, but larger windows increase break or drift detection delay.

- **Amplified variability.** High persistence propagates both innovation noise and survivor variability, widening uncertainty even if ϕ_t and r are unchanged.

Sustained $\hat{\tau}_t$ near 1 should therefore be interpreted as strong inertia, not as a Gaussian unit-root analogue.

4.8.3. *Periodic or seasonal parameter sequences.* Seasonality can be represented by periodic parameter paths of period S:

$$\tau_{t+S} = \tau_t, \qquad \phi_{t+S} = \phi_t, \qquad m_{t+S} = m_t, \qquad t \geq 1. \tag{50}$$

A natural stability condition over one cycle is

$$\prod_{j=1}^{S} \tau_j < 1, \tag{51}$$

which ensures contraction per season and convergence to a unique periodic mean profile $\{\mu_t^*\}$ satisfying $\mu_{t+S}^* = \mu_t^*$.

4.8.4. *Practical parameter constraints for estimation.* Rolling or local procedures can yield infeasible raw estimates. We enforce the core region

$$\tau_t \in (0,1), \qquad \phi_t > 0, \qquad m_t > 0, \qquad r > 0, \tag{52}$$

so that

$$a_t = \tau_t \phi_t > 0, \qquad b_t = (1-\tau_t)\phi_t > 0, \tag{53}$$

and $p_t = r/(r+m_t) \in (0,1)$ is well defined.

Link functions (recommended). For parametric paths, map unconstrained predictors into the feasible region:

$$\tau_t = \text{logit}^{-1}(u_t), \qquad \phi_t = \exp(v_t), \qquad m_t = \exp(w_t), \qquad r = \exp(\psi). \tag{54}$$

Clipping (rolling fallback). For direct rolling estimates, use

$$\hat{\tau}_t \leftarrow \min\{1-\varepsilon_\tau, \max(\varepsilon_\tau, \hat{\tau}_t)\}, \qquad \hat{\phi}_t \leftarrow \max(\varepsilon_\phi, \hat{\phi}_t), \qquad \hat{m}_t \leftarrow \max(\varepsilon_m, \hat{m}_t), \tag{55}$$

to avoid boundary pathologies and degenerate Beta or NB parameterizations.

Second-moment stability (practical sufficient condition). From (42), the coefficient multiplying σ^2_{t-1} is

$$A_t = \tau_t^2 + \frac{\tau_t(1-\tau_t)}{\phi_t+1}. \tag{56}$$

A convenient sufficient condition for controlling variance growth is: there exist $\bar{\tau} < 1$ and $\underline{\phi} > 0$ such that

$$\tau_t \le \bar{\tau} < 1, \qquad \phi_t \ge \underline{\phi} > 0 \quad \Longrightarrow \quad \sup_t A_t \le \bar{\tau}^2 + \frac{\bar{\tau}(1-\bar{\tau})}{\underline{\phi}+1} < 1, \tag{57}$$

together with bounded $\{m_t\}$ and fixed, or piecewise fixed, r. This formalizes an important point: very small ϕ_t can inflate survivor variability even when τ_t is moderate.

Interpretation summary.

- τ_t (*mean persistence*): expected fraction of X_{t-1} carried to time t; dependence is driven by products $\prod \tau_j$.
- ϕ_t (*precision*): concentration of Θ_t around τ_t; smaller values imply stronger random-environment effects and higher survivor volatility.
- m_t (*innovation mean*): expected new arrivals at time t.
- r (*innovation dispersion*): controls innovation overdispersion via $\mathrm{Var}(\xi_t) = m_t + m_t^2/r$; fixed, or piecewise fixed, for identifiability.

4.9. Estimation and inference (non-stationary setting). Inference is anchored on the conditional-mean identity

$$\mathbb{E}(X_t \mid X_{t-1}) = \tau_t X_{t-1} + m_t, \tag{58}$$

which motivates conditional least squares (CLS), rolling or local regression, and local moment estimators [1, 4, 5].

4.9.1. *General strategy: local stationarity, smooth evolution, and piecewise regimes.* Estimating a free parameter at every time point is ill-posed without additional structure. We therefore assume that (τ_t, m_t), and if modeled also ϕ_t, are either smooth, locally constant, or piecewise constant. This leads to three practical estimation strategies:

- **Smooth parametric evolution:** $\tau_t = g(t;\theta)$ and $m_t = h(t;\beta)$ with low-dimensional parameter vectors (θ,β).
- **Rolling or local estimation:** within a window of length W, treat (τ_t, m_t) as approximately constant.
- **Piecewise regimes:** combine local estimation with break detection and segment-wise re-initialization.

State-space approaches are possible but are not pursued here; the focus is on interpretable moment and CLS procedures.

4.9.2. *Time-varying CLS with parametric paths.* Assume

$$\tau_t = g(t;\theta) \in (0,1), \qquad m_t = h(t;\beta) > 0, \tag{59}$$

with feasibility enforced through link functions, for example

$$\tau_t = \operatorname{logit}^{-1}(u_t(\theta)), \qquad m_t = \exp(w_t(\beta)). \tag{60}$$

Define the CLS criterion

$$Q(\theta,\beta) = \sum_{t=1}^{T} \left(X_t - g(t;\theta)X_{t-1} - h(t;\beta)\right)^2, \tag{61}$$

and obtain $(\widehat{\theta},\widehat{\beta})$ by numerical minimization. The fitted paths

$$\widehat{\tau}_t = g(t;\widehat{\theta}), \qquad \widehat{m}_t = h(t;\widehat{\beta}), \tag{62}$$

can then be inserted into the mean, variance, and dependence recursions developed earlier.

4.9.3. *Local or rolling-window CLS (window length W).* For each target time t, let I_t denote a window of size W (trailing or centered). Estimate (τ_t, m_t) by

$$(\widehat{\tau}_t, \widehat{m}_t) = \arg\min_{\tau\in(0,1),\, m>0} \sum_{s\in I_t} \left(X_s - \tau X_{s-1} - m\right)^2. \tag{63}$$

This is ordinary least squares with an intercept, yielding

$$\widehat{\tau}_t = \frac{\sum_{s\in I_t}(X_{s-1} - \bar{X}_{-1,t})(X_s - \bar{X}_t)}{\sum_{s\in I_t}(X_{s-1} - \bar{X}_{-1,t})^2}, \tag{64}$$

and

$$\widehat{m}_t = \bar{X}_t - \widehat{\tau}_t \bar{X}_{-1,t}, \tag{65}$$

where

$$\bar{X}_t = \frac{1}{W}\sum_{s\in I_t} X_s, \qquad \bar{X}_{-1,t} = \frac{1}{W}\sum_{s\in I_t} X_{s-1}. \tag{66}$$

Bias–variance and detection delay. Smaller W adapts quickly but is noisy; larger W reduces variance but mixes regimes and delays break detection. This trade-off is unavoidable and is quantified later in the simulation study.

4.9.4. *Local moment-based estimators.* From the lag-one covariance identity,

$$\mathrm{Cov}(X_t, X_{t-1}) = \tau_t \,\mathrm{Var}(X_{t-1}), \tag{67}$$

a local moment estimator is

$$\widetilde{\tau}_t = \frac{\widehat{\mathrm{Cov}}_{I_t}(X_s, X_{s-1})}{\widehat{\mathrm{Var}}_{I_t}(X_{s-1})}, \qquad \widetilde{m}_t = \bar{X}_t - \widetilde{\tau}_t \bar{X}_{-1,t}, \tag{68}$$

with sample moments computed over the window I_t.

Equivalence. With an intercept, (68) coincides numerically with the local CLS estimator in (64)–(65). We keep both formulations because one emphasizes dependence-based identification, while the other emphasizes conditional-mean fitting.

Estimating ϕ_t (optional). Time-varying ϕ_t is difficult to identify without strong regularization. A practical recommendation is therefore to assume either $\phi_t \equiv \phi$ with sensitivity analysis, or piecewise-constant ϕ across detected regimes. When ϕ is constant within a segment or window, it can be inferred from second-moment relations using the conditional variance formula.

4.9.5. *Innovation component estimation under identifiable NB parametrization.* We adopt the identifiable convention

$$\xi_t \sim \mathrm{NB}(r, p_t), \qquad m_t = \mathbb{E}(\xi_t), \qquad r > 0 \text{ fixed, or piecewise fixed.} \tag{69}$$

Given $\widehat{\tau}_t$, local estimates of m_t follow directly from (65) or (68).

To estimate r globally, or by segments, define residuals

$$\widehat{u}_t = X_t - \widehat{\tau}_t X_{t-1} - \widehat{m}_t. \tag{70}$$

Approximating $\mathbb{E}(\widehat{u}_t^2 \mid X_{t-1})$ by the conditional variance, set

$$v_t(r,\phi) = X_{t-1}\widehat{\tau}_t(1-\widehat{\tau}_t)\frac{\phi + X_{t-1}}{\phi+1} + \widehat{m}_t + \frac{\widehat{m}_t^2}{r}, \tag{71}$$

and estimate (r,ϕ) by

$$(\widehat{r},\widehat{\phi}) = \arg\min_{r>0,\,\phi>0}\sum_{t=1}^{T}\left(\widehat{u}_t^2 - v_t(r,\phi)\right)^2. \tag{72}$$

This keeps dispersion low-dimensional while allowing (τ_t, m_t) to capture the main non-stationarity.

4.9.6. *Implementation details: constraints, boundaries, and break diagnostics.*

Feasibility enforcement. We enforce the admissible region through either link functions in the parametric case or clipping in the rolling case, for example

$$\widehat{\tau}_t \leftarrow \min\{1-\varepsilon_\tau, \max(\varepsilon_\tau, \widehat{\tau}_t)\}, \qquad \widehat{m}_t \leftarrow \max(\varepsilon_m, \widehat{m}_t), \qquad \widehat{\phi} \leftarrow \max(\varepsilon_\phi, \widehat{\phi}), \tag{73}$$

to avoid degenerate Beta or NB parameterizations.

Boundary handling. Near the beginning or end of the series, use expanding windows until size W is reached, or use trailing windows, to avoid discarding early observations.

Break diagnostics and piecewise regimes. Abrupt changes can be flagged by

- jumps in $\widehat{\tau}_t$ or $\widehat{m}_t$,
- sustained increases in $\widehat{u}_t^2$,
- standard change-point methods applied to $\{\widehat{\tau}_t\}$ or $\{\widehat{m}_t\}$.

After a detected break, re-initialize estimation on the new segment and optionally allow piecewise-constant r and ϕ.

Downstream use. Estimated paths $(\widehat{\tau}_t, \widehat{m}_t)$ feed directly into the unconditional mean and variance recursions and into the time-varying dependence formulas, enabling dynamic interpretation and uncertainty tracking.

5. Results

We report (i) simulation evidence on tracking time variation in persistence τ_t and innovation mean m_t, including window-length trade-offs and the role of Beta precision ϕ, and (ii) an empirical application to weekly accident counts illustrating model selection and forecast and uncertainty behavior.

5.1. Simulation study: tracking time-varying dependence and mean.

5.1.1. *Scenario design and evaluation criteria.* We study five regimes: **S1** smooth drift in τ_t, **S2** alternating low/high τ_t, **S3** a single break in τ_t, **S4** smooth drift in m_t, and **S5** varying Beta precision ϕ (volatility and noise) with fixed mean persistence τ. Rolling CLS produces $(\hat{\tau}_t, \hat{m}_t)$; accuracy is summarized by time-averaged Bias and RMSE. For the break setting (S3), we also report an **adaptation delay**: the first post-break time when the Monte Carlo mean path enters a $\pm\delta$ band around the post-break persistence.

5.1.2. *Tracking τ_t under drift, regime switching, and breaks (S1–S3).* Figure 2 compares the true τ_t with the Monte Carlo mean of $\hat{\tau}_t$ and its empirical 95% band for rolling CLS with $W = 40$. As expected for a local-constant estimator, rolling CLS tracks gradual evolution (S1) but smooths abrupt changes (S2–S3) because the window mixes heterogeneous regimes.

S1 (smooth drift). The tracking curve follows the increasing trend with a modest negative bias (Table 2), consistent with trailing-window lag: older, smaller persistence values remain in the window and pull $\hat{\tau}_t$ downward.

S2 (alternating regimes). Frequent switching induces regime-mixing bias: near transitions, $\hat{\tau}_t$ is pulled toward an intermediate value, yielding larger RMSE and a positive average bias. Simple thresholding of the mean tracking path gives a regime-classification accuracy of about 0.68 for $W = 40$ (Table 2).

S3 (single break). After a break, rolling CLS adapts with delay: for $W = 40$, the mean tracking path enters the ± 0.05 band around $\tau_+ = 0.8$ about 6 steps after $t_b = 200$ (Table 2). This delay is the practical cost of smoothing and is quantified further in the W-sensitivity results below.

TABLE 2. Time-averaged tracking accuracy for τ_t across Scenarios S1–S3 using rolling CLS ($W = 40$).

			Performance		**Data**		**Regimes**			**Break**		
Scen.	**Estimator**	W	**Bias**	**RMSE**	T	B	$\tau_{L/\min}$	$\tau_{H/\max}$	**Acc.**	**Delay**	δ	t_b
S1	Rolling CLS	40	-0.078	0.165	400	200	0.20	0.80	–	–	–	–
S2	Rolling CLS	40	0.169	0.283	400	200	0.25	0.75	0.681	–	–	–
S3	Rolling CLS	40	-0.060	0.160	400	200	0.30	0.80	–	6	0.05	200

Note: All scenarios use $\phi = 50$, $m = 8$, and $r = 20$; entries are rounded.

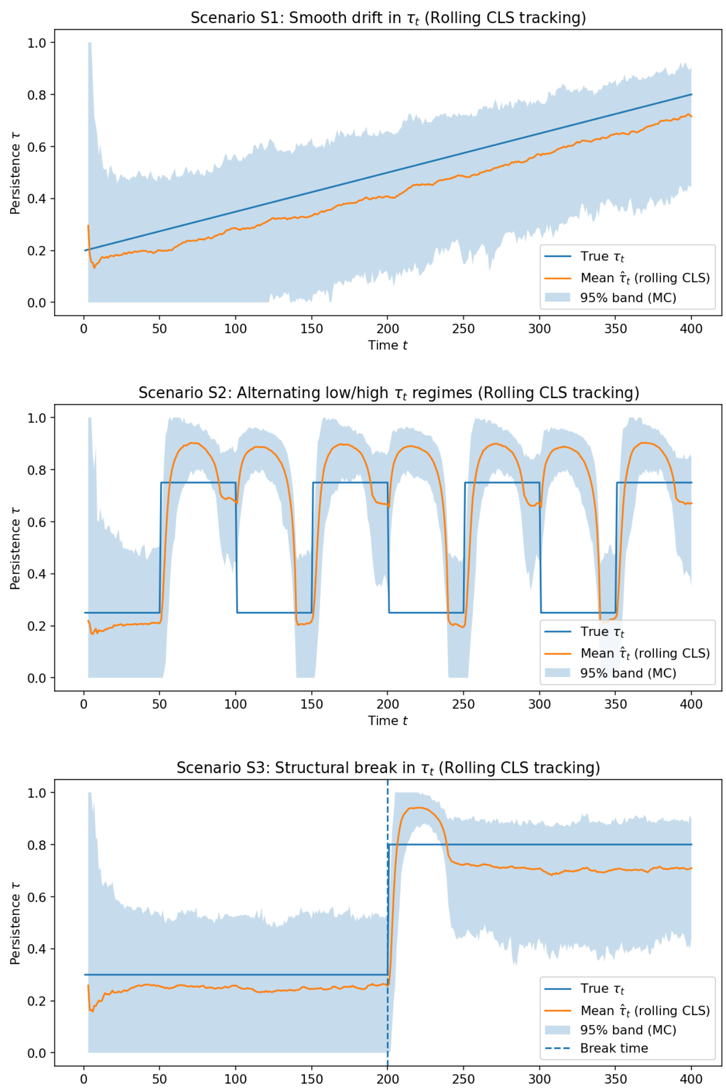

FIGURE 2. **Tracking time-varying persistence across Scenarios S1–S3.** Top: S1, smooth drift in τ_t ($W = 40$). Middle: S2, alternating regimes in τ_t ($W = 40$). Bottom: S3, break in τ_t ($W = 40$). Blue: true τ_t; orange: Monte Carlo mean of $\hat{\tau}_t$ (rolling CLS); shaded: empirical 95% band; dashed in S3: break time.

5.1.3. *Window-length sensitivity and RMSE–delay trade-off.* Window length W controls a bias–variance trade-off and, under breaks, an RMSE–delay trade-off. Figure 3 summarizes RMSE sensitivity for $\hat{\tau}_t$ (S1–S3) and break adaptation delay (S3); Table 3 reports the RMSE-minimizing $W^{\star}$.

Regime dependence of $W^{\star}$. Under smooth drift (S1) and a single break (S3), RMSE decreases as W increases, reflecting variance reduction when the target evolves slowly relative to the window. In contrast, under alternating regimes (S2), large windows amplify regime-mixing bias; RMSE is minimized at a smaller window (here $W^{\star} = 20$).

Break adaptation versus RMSE (S3). Larger W improves time-averaged RMSE but delays adaptation: with $\delta = 0.05$ and $\tau_+ = 0.8$, the delay increases from 4 steps at $W \in \{10, 20\}$ to 8 steps at $W = 80$ (Table 4). Thus, window choice should reflect whether the priority is overall tracking accuracy or rapid break response.

TABLE 3. Recommended rolling window length $W^{\star}$ minimizing time-averaged RMSE by scenario and target.

Scenario	Target	$W^{\star}$	RMSE($W^{\star}$)
S1	τ	80	0.122946
S2	τ	20	0.257692
S3	τ	80	0.123395
S4	m	80	2.517600

TABLE 4. Scenario S3: break adaptation delay versus window length W.

W	δ	t_{adapt}	Delay	RMSE($\hat{\tau}$)	T	B	t_b	τ_-	τ_+
10	0.05	204	4	0.322191	400	120	200	0.30	0.80
20	0.05	204	4	0.222691	400	120	200	0.30	0.80
40	0.05	206	6	0.154739	400	120	200	0.30	0.80
60	0.05	207	7	0.129795	400	120	200	0.30	0.80
80	0.05	208	8	0.121914	400	120	200	0.30	0.80

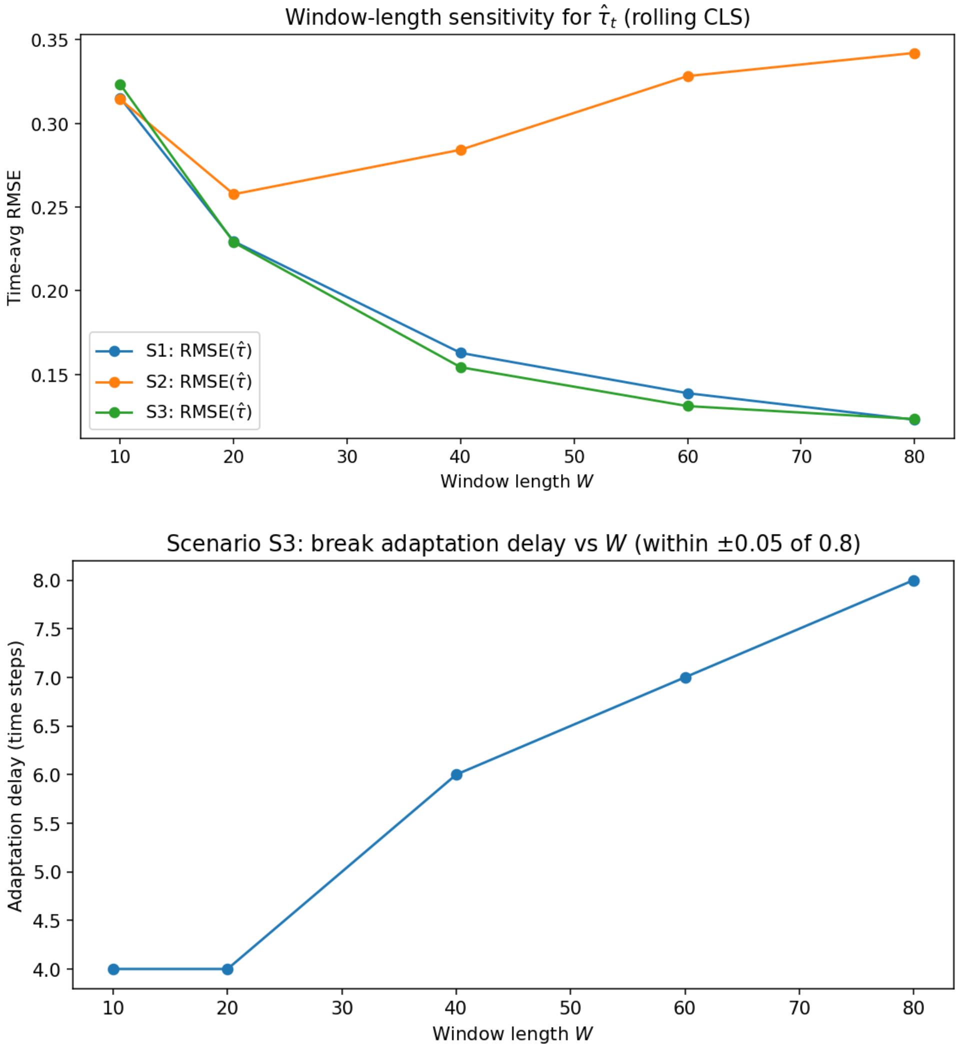

FIGURE 3. **Window-length trade-offs.** Top: RMSE of $\hat{\tau}_t$ versus W (S1–S3). Bottom: S3 adaptation delay versus W within ± 0.05 of $\tau_+ = 0.8$. Larger W reduces variance under slow evolution but increases break delay and can worsen performance under frequent switching due to regime mixing.

5.1.4. *Tracking m_t under smooth evolution (S4).* Scenario S4 isolates mean non-stationarity by letting m_t drift smoothly while holding (τ, ϕ, r) fixed. Figure 4 shows that rolling CLS recovers the upward trend but exhibits window-induced lag; for $W = 40$ the estimator has positive average bias and moderate RMSE (Table 5). Because the signal evolves smoothly, increasing W primarily reduces variance: RMSE$(\hat{m}_t)$ drops sharply as W grows, motivating larger windows or smooth parametric m_t in gradual-evolution settings.

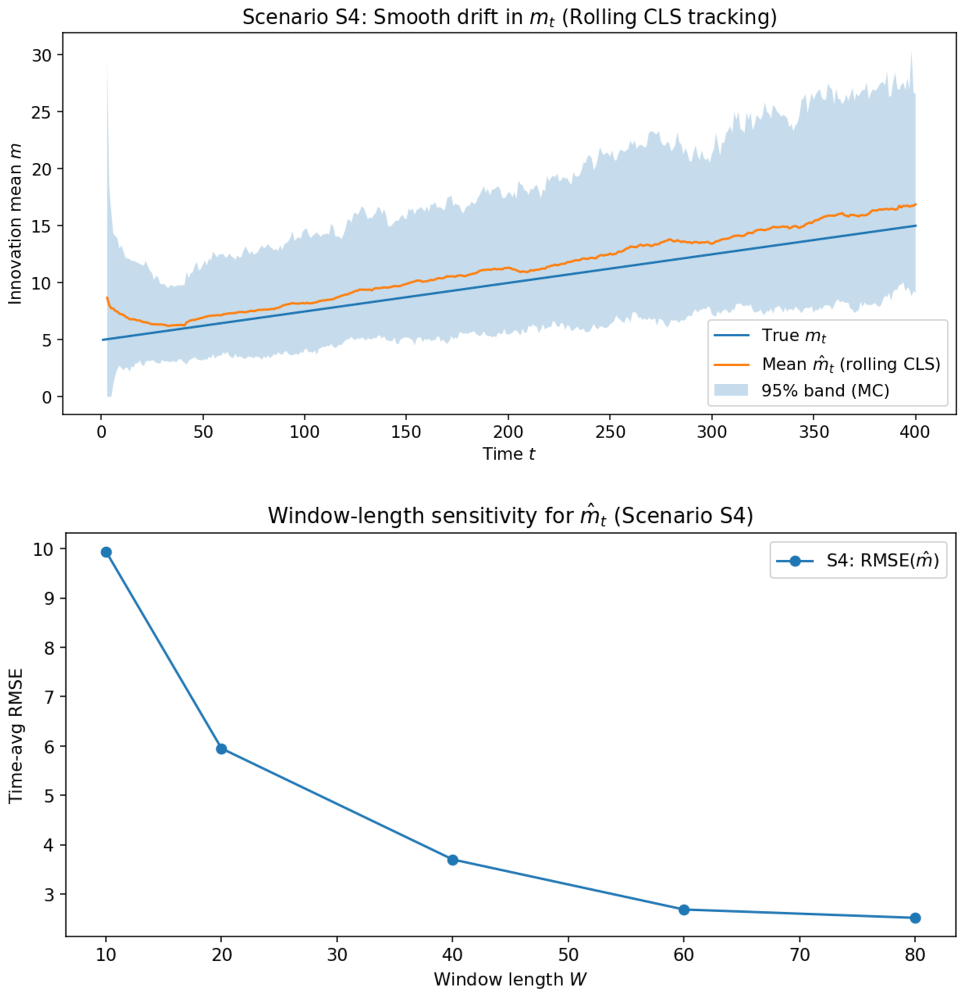

FIGURE 4. **Tracking the time-varying innovation mean in S4.** Top: smooth drift in m_t ($W = 40$). Bottom: RMSE of $\hat{m}_t$ versus W. Rolling CLS captures the smooth drift, and larger windows substantially reduce tracking RMSE.

TABLE 5. Innovation-mean tracking in S4 using rolling CLS ($W = 40$).

Scenario	Estimator	W	Bias	RMSE	T	B	τ	ϕ	$m_{\min}$	$m_{\max}$	r
S4	Rolling CLS	40	1.220753	3.624852	400	200	0.60	50	5	15	20

5.1.5. *Precision ϕ governs volatility and estimator noise (S5).* Scenario S5 varies Beta precision $\phi = a + b$ while holding (τ, m, r) fixed, isolating the survivor-variance inflation implied by the conditional variance decomposition in Section 4.3. Figure 5 shows markedly higher marginal volatility under low precision ($\phi = 5$) than under high precision ($\phi = 200$). Time-averaged marginal variance nearly doubles (Table 6), and tracking RMSE increases slightly under low precision, consistent with noise amplification rather than systematic mean shifts.

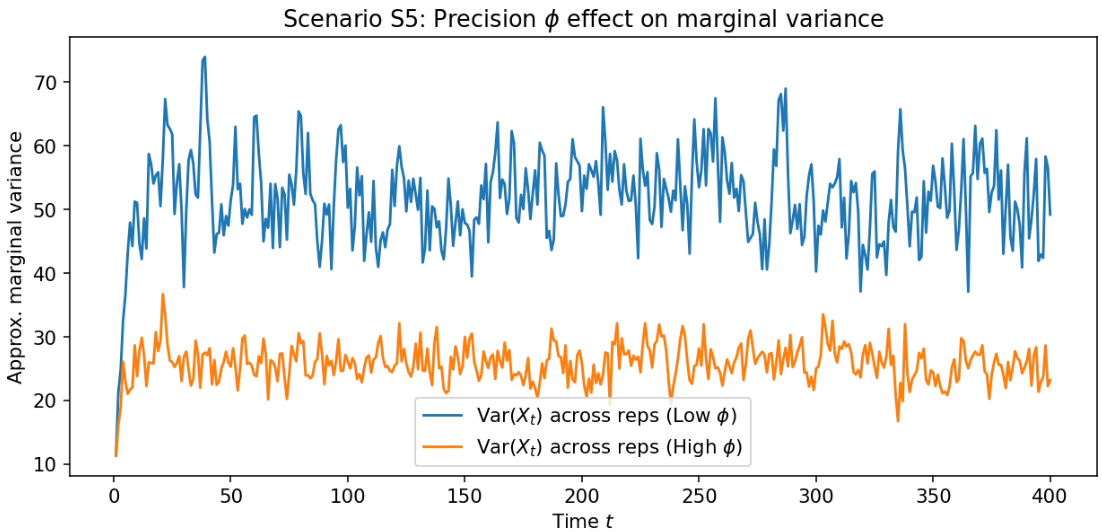

FIGURE 5. **S5: precision effect on volatility.** Lower ϕ increases Beta mixing variability and inflates volatility even when τ is unchanged.

TABLE 6. Precision effect on volatility and tracking accuracy in S5.

ϕ level	ϕ	mean Var(X_t)	RMSE$(\hat{\tau})$	RMSE$(\hat{m})$	T	B	W	τ	m	r
Low	5	51.746	0.1725	3.2710	400	200	40	0.60	8	20
High	200	25.820	0.1580	3.2211	400	200	40	0.60	8	20

5.2. Empirical application: weekly total accidents.

5.2.1. *Data description and basic diagnostics.* The application uses $n = 262$ consecutive weekly counts from 2009–01–05 to 2014–01–06 with no missing weeks (Table 7). Counts range from 3 to 55 with mean 31.90 and variance 56.33 (variance-to-mean ratio ≈ 1.63), indicating mild overdispersion. Figure 6 suggests a slowly varying level and seasonal structure; the raw ACF shows weak short-lag dependence, which can partly reflect unmodeled mean and seasonality rather than genuine survivor persistence.

Data source: UK Road Safety Data (Department for Transport), available via the Road Safety Data portal: https://ckan.publishing.service.gov.uk/dataset/road-safety-data.

TABLE 7. Weekly total accidents: dataset summary.

n_{weeks}	start	end	min	max	mean	variance	% zeros	missing
262	2009-01-05	2014-01-06	3	55	31.900763	56.334942	0.000000	0

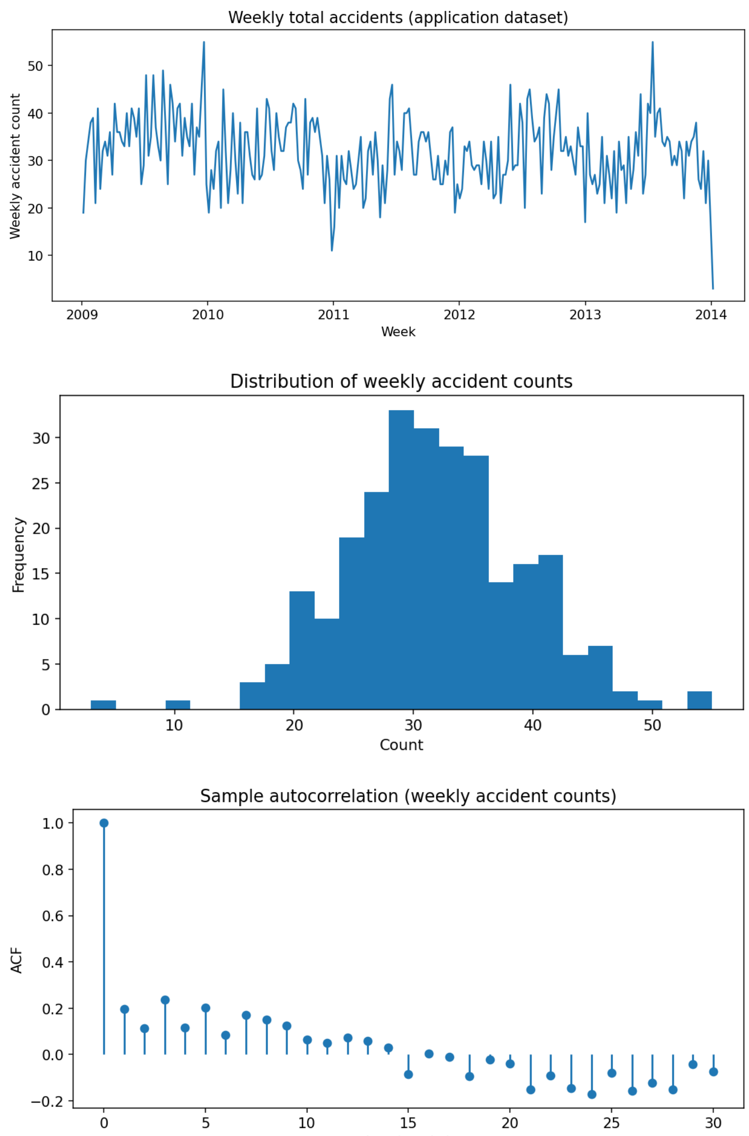

FIGURE 6. **Diagnostics for weekly accidents.** Top: time series. Middle: marginal distribution. Bottom: sample ACF. The series shows substantial week-to-week variability around a slowly evolving and seasonal level; the distribution is unimodal; and the raw ACF exhibits weak short-lag dependence.

5.2.2. *Point forecasting: stationary versus non-stationary mean models.* We assess one-step-ahead forecasts using the last 52 weeks as test data. Table 8 shows that stationary CLS outperforms rolling CLS ($W = 12$) and simple baselines, indicating that unregularized local-constant tracking can be too noisy when the dominant non-stationarity is smooth.

Motivated by the diagnostics, we fit a parsimonious smooth mean

$$m_t = \beta_0 + \beta_1 \tilde{t} + \beta_2 \sin\left(\frac{2\pi t}{52}\right) + \beta_3 \cos\left(\frac{2\pi t}{52}\right), \tag{74}$$

and estimate the conditional-mean equation by CLS. This parametric model achieves the best test accuracy (RMSE 6.887, MAE 5.036; Table 9). Under this specification, $\hat{\tau} \approx 0.0068$, and removing the lag term yields virtually identical performance (Table 10), implying that once m_t is modeled, residual persistence is negligible for this series.

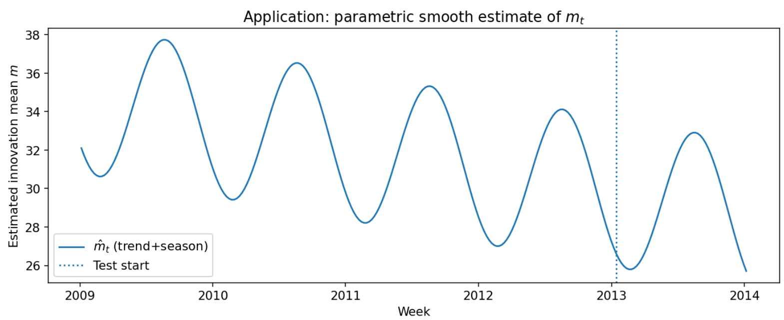

FIGURE 7. **Application:** fitted smooth innovation-mean path $\hat{m}_t$ (trend and annual seasonality). The dotted line marks the test split.

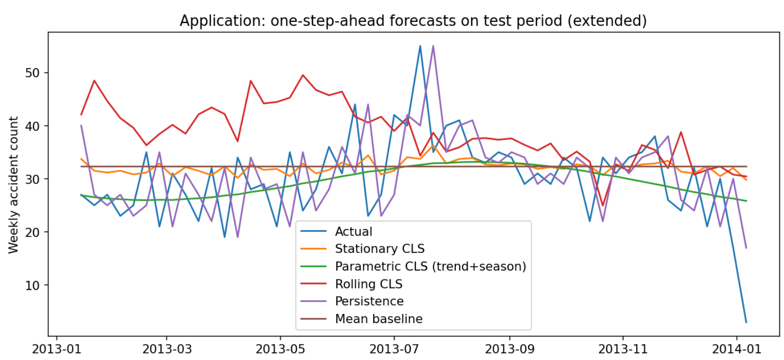

FIGURE 8. **Application:** one-step-ahead forecasts over the test period. The smooth mean model tracks the level more consistently than rolling CLS, which exhibits higher variability.

TABLE 8. Application: one-step-ahead test accuracy (weekly accidents).

Model	RMSE	MAE	n_{pred}
Stationary CLS (NB–INAR)	7.991146	5.985241	52
Mean baseline	8.283571	6.137912	52
Persistence baseline	9.119927	7.403846	52
Non-stationary rolling CLS ($W = 12$)	12.895038	10.382410	52

TABLE 9. Application: extended comparison including smooth parametric m_t.

Model	RMSE	MAE	n_{pred}
Parametric CLS (trend+season m_t)	6.887319	5.035609	52
Stationary CLS (NB–INAR)	7.991146	5.985241	52
Mean baseline	8.283571	6.137912	52
Persistence baseline	9.119927	7.403846	52
Non-stationary rolling CLS	12.895038	10.382410	52

TABLE 10. Application: smooth parametric CLS with versus without lag dependence.

Model	RMSE	MAE
Parametric CLS (with lag)	6.887319	5.035609
Parametric CLS (no lag)	6.889625	5.039238

5.2.3. *Overdispersion and predictive-uncertainty calibration.* On the training set, var/mean $\approx$ 1.63 (Table 11), and a moment-based estimate gives $\hat{r} \approx 88$, indicating moderate extra-Poisson variability. Using $\hat{m}_t$, nominal 90% predictive intervals on the test set slightly over-cover under both Poisson and NB assumptions; NB intervals attain higher empirical coverage (0.942 versus 0.923) at the cost of wider average length (Table 12, Figure 9). This supports NB innovations as a practical mechanism for uncertainty quantification under mild overdispersion.

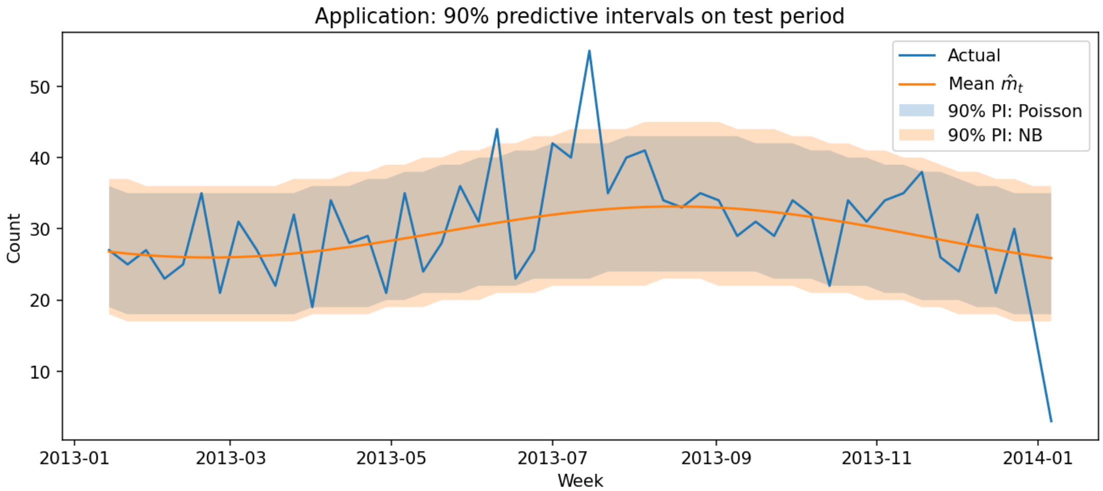

FIGURE 9. **Application:** nominal 90% one-step predictive intervals under Poisson and NB assumptions using $\hat{m}_t$.

TABLE 11. Application (train): overdispersion check and moment-based estimate of r under the smooth mean model.

$\bar{y}_{\text{train}}$	$\text{Var}(y_{\text{train}})$	var/mean	$\overline{\hat{m}}$	$\text{Var}(\hat{\xi})$	$\hat{r}$
32.411483	52.926021	1.632940	32.411483	44.348761	88.001997

TABLE 12. Application (test): empirical coverage and average length of nominal 90% predictive intervals.

Model	$\text{Coverage}_{0.90}$	Avg. interval length
$\text{Poisson}(\hat{m}_t)$	0.923077	17.846154
$\text{NB}(\hat{m}_t, \hat{r} = 88.00)$	0.942308	20.538462

5.3. Summary of findings.

- Rolling CLS tracks smooth evolution (S1, S4), but its performance depends critically on window length W; larger W reduces variance when parameters evolve slowly.
- Under frequent switching (S2), large windows induce regime-mixing bias; smaller W is preferred, whereas breaks (S3) create a clear RMSE–delay trade-off.
- Lower precision ϕ inflates volatility via Beta mixing and slightly increases tracking RMSE (S5), consistent with the conditional variance decomposition.

- In the accident data, predictive gains come primarily from modeling smooth non-stationarity in m_t; once m_t is corrected, estimated persistence is negligible. NB innovations improve interval calibration under mild overdispersion.

6. Discussion

We developed a non-stationary NB–INAR(1) model with Beta–Binomial thinning in which time variation enters through an interpretable persistence–precision decomposition: $\tau_t = \mathbb{E}(\Theta_t)$ governs dynamic dependence, while $\phi_t = a_t + b_t$ controls survivor-variance inflation. We derived conditional moments, iterated unconditional recursions, and an explicit time-varying dependence structure driven by products of τ_t, and we proposed rolling/local CLS and smooth parametric variants for estimation under feasibility constraints.

6.1. Implications for modeling non-stationary count series. The simulations underline that *non-stationarity is regime-dependent*. Smooth drift, frequent switching, and structural breaks impose different requirements on estimation. Rolling CLS performs well when parameters evolve slowly relative to the window, but it becomes biased when windows span heterogeneous regimes (S2) and it cannot avoid adaptation delay after breaks (S3). Consequently, the window length W is not merely a tuning parameter; it encodes an assumption about the fastest plausible rate of structural change.

Separating τ_t from ϕ_t also yields a practical interpretive advantage. While τ_t primarily determines dependence, ϕ_t acts through variance inflation: even with identical mean persistence, lower precision increases survivor randomness and raises $\mathrm{Var}(X_t)$ (S5). This clarifies how series with similar short-lag dependence can exhibit materially different volatility.

6.2. Estimation trade-offs and practical guidance. The results suggest three practical recommendations:

(1) **Match the estimator to the change regime.** Under gradual evolution, larger windows or smooth parametric paths stabilize estimation (S1, S4). Under rapid switching, smaller windows mitigate regime-mixing bias (S2).

(2) **Report delay metrics when breaks are plausible.** Under structural breaks, an RMSE-optimal window can be undesirable if timely adaptation is required. Reporting both

RMSE and an adaptation-delay measure (S3) provides a more informative summary than RMSE alone.

(3) **Enforce feasibility to preserve interpretation.** Noise and boundary effects can yield infeasible raw estimates, for example $\hat{\tau}_t \notin (0,1)$ or $\hat{m}_t < 0$. Constraints via links, clipping, or constrained optimization are essential for maintaining the probabilistic meaning of the model components.

6.3. Insights from the accident-count application. The weekly accident series illustrates a setting where non-stationarity is driven mainly by a smooth intensity component rather than persistent autoregressive dependence. Once a trend-plus-seasonal $\hat{m}_t$ is introduced, estimated persistence becomes negligible and with-lag versus no-lag specifications perform essentially identically for one-step-ahead prediction. Rather than being a limitation, this is diagnostically useful: the framework can indicate when apparent dependence in raw counts is largely explained by non-stationary mean structure.

The application also highlights the role of the Negative Binomial layer for uncertainty quantification. Mild overdispersion leads NB predictive intervals to achieve higher empirical coverage than Poisson intervals, with the expected trade-off of wider intervals.

6.4. Limitations and extensions. Several directions deserve further work. First, our main estimators are CLS- and moment-based and target one-step conditional moments; likelihood-based or Bayesian approaches could leverage the full transition structure for improved efficiency, especially for short series. Second, time variation was modeled either locally (rolling windows) or via simple parametric forms; spline regularization, state-space evolution, or change-point formulations could better accommodate mixed regimes. Third, extending the model to multivariate counts, for example through matrix thinning with Beta mixing, would broaden applicability to coupled count processes. Finally, developing formal inferential uncertainty for $(\hat{\tau}_t, \hat{m}_t)$ under non-stationarity, beyond Monte Carlo bands, remains an important open problem.

6.5. Concluding remarks. Overall, the proposed non-stationary NB–INAR(1) model provides an interpretable mechanism for evolving dependence and overdispersion in count time series. The results emphasize that estimation should be aligned with the underlying change regime and that separating persistence from precision clarifies how time variation impacts both dependence and volatility.

7. Conclusion

We introduced a non-stationary NB–INAR(1) model with Beta–Binomial thinning for count time series whose dependence and dispersion evolve over time. A central feature is the persistence–precision decomposition: $\tau_t = \mathbb{E}(\Theta_t)$ controls time-varying dependence, while $\phi_t = a_t + b_t$ governs survivor-variance inflation induced by Beta mixing. We derived closed-form conditional moments and time-varying recursions for the mean, variance, and dependence, including explicit expressions showing that autocovariances evolve through products of the persistence path.

For inference, we proposed locally stationary and smoothly evolving estimation strategies, namely rolling/local CLS and parametric time-varying specifications, together with feasibility constraints required for valid Beta and NB components. Simulations demonstrated that rolling CLS tracks gradual change effectively, while the window length W induces a fundamental trade-off: larger windows reduce variance but increase regime-mixing bias and post-break adaptation delay when changes are abrupt. We also confirmed that lower precision ϕ inflates volatility and modestly increases tracking error, consistent with the model's conditional-variance decomposition.

In the weekly accident application, the dominant non-stationarity is captured by a smooth trend-plus-seasonal m_t, and residual persistence becomes negligible once the mean is modeled, illustrating how the framework can distinguish mean-driven non-stationarity from genuine survivor dependence. Finally, mild overdispersion motivates NB predictive distributions, which improve empirical interval coverage relative to Poisson intervals.

In summary, the model offers an interpretable and flexible toolkit for non-stationary count processes, clarifying the distinct roles of evolving persistence, precision-driven volatility, and innovation intensity. Future work may develop likelihood-based or Bayesian inference, richer evolution schemes such as splines, state-space models, or change-point methods, multivariate extensions with cross-dependence, and formal uncertainty quantification for time-varying estimators.

DATA AVAILABILITY STATEMENT

We have deposited the derived weekly dataset and cleaning scripts on Zenodo and cite the DOI in the manuscript.

The accident-level data analyzed in this study are publicly available from the UK Road Safety Data portal: https://surl.li/awxcxd.

The specific CSV resource used to construct the weekly accident-count series is available online at: https://surl.li/eivjor.

The derived weekly aggregated dataset used in the analyses (`weekly_total_accidents_clean.csv`) and the data-cleaning scripts have been deposited on Zenodo: https://doi.org/10.5281/zenodo.18762302.

Note: The openly shared analysis files and results derived from the weekly dataset have been deposited on Zenodo: https://doi.org/10.5281/zenodo.18389818.

ACKNOWLEDGMENTS

The authors express their gratitude to the Pan African University Institute for Basic Sciences, Technology, and Innovation (PAUSTI) for its valuable support in facilitating this study.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] M.A. Al-Osh, A.A. Alzaid, First-order Integer-valued Autoregressive (INAR(1)) Process, J. Time Ser. Anal. 8 (1987), 261–275. https://doi.org/10.1111/j.1467-9892.1987.tb00438.x.

[2] E. Mckenzie, Some ARMA Models for Dependent Sequences of Poisson Counts, Adv. Appl. Probab. 20 (1988), 822–835. https://doi.org/10.2307/1427362.

[3] D. Jin-Guan, L. Yuan, The Integer-valued Autoregressive (INAR(p)) Model, J. Time Ser. Anal. 12 (1991), 129–142. https://doi.org/10.1111/j.1467-9892.1991.tb00073.x.

[4] C.H. Weiß, Thinning Operations for Modeling Time Series of Counts–A Survey, AStA Adv. Stat. Anal. 92 (2008), 319–341. https://doi.org/10.1007/s10182-008-0072-3.

[5] C. H. Weiß, Introduction to Discrete-Valued Time Series, Wiley, 2018.

[6] R. Zhu, H. Joe, Modelling Count Data Time Series with Markov Processes Based on Binomial Thinning, J. Time Ser. Anal. 27 (2006), 725–738. https://doi.org/10.1111/j.1467-9892.2006.00485.x.

[7] R.A. Davis, K. Fokianos, S.H. Holan, H. Joe, J. Livsey, et al., Count Time Series: A Methodological Review, J. Am. Stat. Assoc. 116 (2021), 1533–1547. https://doi.org/10.1080/01621459.2021.1904957.

[8] Y. Pang, D. Wang, M. Goh, A Two-Step Estimation Method for a Time-Varying INAR Model, Axioms 13 (2023), 19. https://doi.org/10.3390/axioms13010019.

[9] E. McKenzie, Autoregressive Moving-Average Processes with Negative-Binomial and Geometric Marginal Distributions, Adv. Appl. Probab. 18 (1986), 679–705. https://doi.org/10.2307/1427183.

[10] H. Joe, Time Series Models with Univariate Margins in the Convolution-Closed Infinitely Divisible Class, J. Appl. Probab. 33 (1996), 664–677. https://doi.org/10.2307/3215348.

[11] A. Latour, Existence and Stochastic Structure of a Non-negative Integer-valued Autoregressive Process, J. Time Ser. Anal. 19 (1998), 439–455. https://doi.org/10.1111/1467-9892.00102.

[12] H. Zheng, I.V. Basawa, S. Datta, First-Order Random Coefficient Integer-Valued Autoregressive Processes, J. Stat. Plan. Inference 137 (2007), 212–229. https://doi.org/10.1016/j.jspi.2005.12.003.

[13] C.H. Weiß, H. Kim, Diagnosing and Modeling Extra-binomial Variation for Time-dependent Counts, Appl. Stoch. Model. Bus. Ind. 30 (2013), 588–608. https://doi.org/10.1002/asmb.2005.

[14] F. W. Steutel, K. van Harn, Discrete analogues of self-decomposability and stability, The Annals of Probability 7 (5) (1979), 893–899. https://doi.org/10.1214/aop/1176994782.

[15] E. McKenzie, Some Simple Models for Discrete Variate Time Series, J. Am. Water Resour. Assoc. 21 (1985), 645–650. https://doi.org/10.1111/j.1752-1688.1985.tb05379.x.

[16] M.A. Al-Osh, E.E.A. Aly, First Order Autoregressive Time Series with Negative Binomial and Geometric Marginals, Commun. Stat. - Theory Methods 21 (1992), 2483–2492. https://doi.org/10.1080/03610929208830925.

[17] M.M. Ristić, A.S. Nastić, H.S. Bakouch, Estimation in an Integer-Valued Autoregressive Process with Negative Binomial Marginals (NBINAR(1)), Commun. Stat. - Theory Methods 41 (2012), 606–618. https://doi.org/10.1080/03610926.2010.529528.

[18] R. Zhu, H. Joe, Negative Binomial Time Series Models Based on Expectation Thinning Operators, J. Stat. Plan. Inference 140 (2010), 1874–1888. https://doi.org/10.1016/j.jspi.2010.01.031.