Commun. Math. Res.
doi: 10.4208/cmr.2026-0014

Vol. **42**, No. 1, pp. 97-120
March 2026

Observation-Driven INAR(1) Models with Novel and Flexible Links

Yiling Dong and Fukang Zhu*

School of Mathematics, Jilin University, Changchun 130012, P.R. China.

Received 25 January 2026; Accepted 3 March 2026

Abstract. Observation-driven integer-valued autoregressive models are widely used for modeling count time series exhibiting dynamic dependence, yet their performance critically depends on the way that thinning probabilities are linked to past observations. Most existing specifications rely on the logit link and may respond excessively to large counts. In this paper, we introduce a class of new observation-driven integer-valued autoregressive models using logarithmic and soft-clipping links that attenuate the influence of large observations. The proposed framework allows for stochastic covariates. Estimation is carried out using conditional maximum likelihood and conditional least squares methods. Simulation studies and two real data applications are used to illustrate the proposed models.

AMS subject classifications: 62M10, 62M20

Key words: Observation-driven model, covariate, logarithmic link, soft-clipping link, conditional maximum likelihood.

1 Introduction

To model integer-valued time series with values in $\mathbb{N}_0 = \{0,1,2,\ldots\}$, traditional autoregressive (AR) frameworks developed for continuous-valued data are often inadequate because they rely on Gaussian assumptions. To address this limitation, McKenzie [16] and Al-Osh and Alzaid [1] introduced the integer-valued autoregressive (INAR) model, in which the multiplicative structure of AR models is replaced by a binomial thinning operator. Numerous extensions of the INAR framework have been proposed. With respect to statistical characteristics of count

*Corresponding author. *Email address:* `zfk8010@163.com` (F. Zhu)

time series, Kang *et al.* [12, 13] developed models that accommodated under-, equi-, and overdispersion; Yang *et al.* [30] considered a bivariate threshold Poisson INAR process; Kang *et al.* [14] proposed a zero-modified geometric specification; and Kang *et al.* [10] introduced a parsimonious specification designed to capture multiple empirical features. Serially dependent innovations were also considered; see [11,26]. From a Bayesian perspective, Miao and Wang [17] developed Bayesian procedures for estimating the order of INAR(q) models.

An important extension was introduced by Zheng *et al.* [35], who proposed the random-coefficient INAR model with time-varying thinning probabilities. Zheng and Basawa [34] further introduced observation-driven INAR models in which the dependence structure evolves with past observations. Several subsequent studies incorporated covariates into this framework, including the empirical likelihood approach of Ding and Wang [6], the minification-based INAR process of Qian and Zhu [21], and the random-coefficient threshold INAR model of Yang *et al.* [29]. Other extensions have focused on alternative thinning operators. Yu *et al.* [32] introduced a class of observation-driven random-coefficient INAR processes based on negative binomial thinning, and Yu and Tao [31] further generalized observation-driven parameters under Poisson thinning. For recent comprehensive overviews of integer-valued time series models, see [3, 18, 20].

Most existing works on observation-driven INAR models treat covariates as fixed design sequences or conditionally exogenous inputs. When covariates are modeled as stochastic processes, such models can be viewed as Markov chains in random environments, as formalized by Kifer [15]. Fokianos and Truquet [9] consider finite-state Markov chains in random environments, whereas Truquet [24] develops a framework for general state spaces; in both cases, the observed dynamics are conditionally time-inhomogeneous given an exogenous process. A common strategy is to enlarge the state space by incorporating the shifted environment, thereby obtaining a time-homogeneous representation for the joint process and a convenient route to stationarity and ergodicity. Complementary to these results, Doukhan *et al.* [7] provide a coupling-based treatment of observat ion-driven models in random environments via backward iterations in Wasserstein distance, without relying on small-set or uniform contraction assumptions. In observation-driven INAR models with stochastic covariates, the thinning mechanism is specified through a conditional transition kernel rather than an explicit recursion. This makes drift or Lyapunov verification on the count state space nontrivial.

The choice of link plays a crucial role in observation-driven INAR models. In most existing formulations, the thinning probability is modeled as a logit function of past observations. While this specification ensures that the thinning probability remains within the unit interval, it directly couples an unbounded count process with a probability scale. As a result, large counts may induce overly sensitive probabilistic responses. Similar concerns have motivated the use of alternative links in related count models. For instance, Fokianos and Tjøstheim [8] proposed a log-linear Poisson autoregressive model in which past counts enter through a logarithmic transformation, thereby tempering the influence of large observations. More recently, Weiß and Jahn [27] introduced the soft-clipping link in integer-valued generalized autoregressive conditional heteroskedasticity (INGARCH) models to bound the effect of extreme predictors, while Weiß *et al.* [28] considered the softplus link as smooth approximations to linear dynamics that preserve positivity. Despite these developments, the impact of links on the stability of observation-driven INAR models has not been examined.

To address these two issues, we introduce a class of observation-driven INAR models with stochastic covariates, employing logarithmic and soft-clipping links to mitigate scale-sensitivity induced by large observations. Treating the covariate process as a random environment, we construct an augmented state representation that yields a time-homogeneous Markov structure. Within this framework, we establish the existence of an invariant distribution via drift and minorization conditions, and prove ergodicity through an invariant-set argument.

The remainder of the paper is organized as follows. Section 2 introduces a general observation-driven INAR(1) framework and presents the link specifications. Section 3 presents conditional maximum likelihood (CML) and conditional least squares (CLS) estimation methods; in particular, since the likelihood involves an integral representation, we employ an expectation-maximization (EM) algorithm for its computation. Section 4 reports the numerical results of the above two estimators. Section 5 presents real-world data applications. Section 6 concludes with a discussion.

2 Proposed observation-driven INAR(1) models

We consider an observation-driven INAR(1) process $\{X_t\}$ defined by

$$X_t = \phi_t \circ X_{t-1} + \varepsilon_t, \quad \phi_t = g\big(\alpha + h(X_{t-1}) + \boldsymbol{\gamma}^\top \boldsymbol{Z}_t + U_t\big) \tag{2.1}$$

for $t \geq 1$, where

(i) $\alpha \in \mathbb{R}$, $\phi_t \in (0,1)$, and the binomial thinning operator is defined as $\phi_t \circ X_{t-1} = \sum_{k=1}^{X_{t-1}} B_{t,k}$, with $\{B_{t,k}\}$ being independent and identically distributed (i.i.d.) Bernoulli random variables with success probability ϕ_t.

(ii) $g:\mathbb{R}\to(0,1)$ is a known link, and $h:\mathbb{N}_0\to\mathbb{R}$ is a measurable function of the lagged observation X_{t-1}.

(iii) $\{\varepsilon_t\}$ is a sequence of i.i.d. nonnegative integer-valued innovations with strictly positive probability mass function f_ε, finite mean $E(\varepsilon_t)=\lambda$, and finite variance $\mathrm{Var}(\varepsilon_t)=\sigma^2$.

(iv) $\boldsymbol{Z}_t=(Z_{1t},\ldots,Z_{pt})^\top$ is a p-dimensional vector of exogenous covariates taking values in $\mathbb{R}^p$, with associated coefficient vector $\boldsymbol{\gamma}=(\gamma_1,\ldots,\gamma_p)^\top\in\mathbb{R}^p$.

(v) $\{U_t\}$ is a sequence of i.i.d. latent disturbances with $U_t\sim N(0,\eta^2)$, capturing unobserved stochastic effects; X_0, $\{\boldsymbol{Z}_t\}$, $\{\varepsilon_t\}$ and $\{U_t\}$ are mutually independent.

Conditioned on $\{\boldsymbol{Z}_t\}$, the process $\{X_t\}$ forms a time-inhomogeneous Markov chain on the state space $\mathbb{N}_0$. The one-step transition probabilities can be expressed as a mixture with respect to the thinning probability ϕ_t as

$$P(X_t=i|X_{t-1}=j,\boldsymbol{Z}_t)=\int_0^1 P(X_t=i|X_{t-1}=j,\phi_t)\,p(\phi_t|X_{t-1}=j,\boldsymbol{Z}_t)\,d\phi_t, \tag{2.2}$$

where

$$P(X_t=i|X_{t-1}=j,\phi_t)=\sum_{k=0}^{\min(i,j)}\binom{j}{k}\phi_t^k(1-\phi_t)^{j-k}f_\varepsilon(i-k) \tag{2.3}$$

for all $i,j\in\mathbb{N}_0$, and $p(\phi_t|X_{t-1}=j,\boldsymbol{Z}_t)$ denotes the conditional density of the thinning probability ϕ_t. Given this, we can derive the following conditional moments for the process.

Proposition 2.1. *Let $m_t=E(\phi_t|X_{t-1},\boldsymbol{Z}_t)$, $v_t=\mathrm{Var}(\phi_t|X_{t-1},\boldsymbol{Z}_t)$. Then, for $t\geq 1$,*

(i) $E(X_t|X_{t-1},\boldsymbol{Z}_t)=m_tX_{t-1}+\lambda$,

(ii) $\mathrm{Var}(X_t|X_{t-1},\boldsymbol{Z}_t)=\big(m_t(1-m_t)-v_t\big)X_{t-1}+v_tX_{t-1}^2+\sigma^2$.

The result follows directly from the mixture representation above and standard properties of binomial thinning, thus the proof is omitted.

To reveal how the link g and the state transformation h jointly shape the response to past counts x, we study the conditional expectation of the thinning probability m_t and derive a general bound on $|\partial m_t/\partial x|$.

Proposition 2.2. *Consider the observation-driven INAR(1) process defined in* (2.1). *Assume that the link $g:\mathbb{R}\to(0,1)$ is continuously differentiable with a bounded derivative*

$$L_g \equiv \sup_{u\in\mathbb{R}} |g'(u)| < \infty,$$

and that the state transformation $h:\mathbb{N}_0\to\mathbb{R}$ is differentiable on $(0,\infty)$. Then, for any $\boldsymbol{z}\in\mathbb{R}^p$ and any $x>0$, the mapping

$$x\mapsto m_t(x,\boldsymbol{z}) = E(\phi_t|X_{t-1}=x,\boldsymbol{Z}_t=\boldsymbol{z})$$

is differentiable and satisfies

$$\left|\frac{\partial m_t(x,\boldsymbol{z})}{\partial x}\right| \le L_g|h'(x)|.$$

The result follows from differentiation under the integral sign, and the proof is omitted for brevity. This proposition is primarily a technical result and serves as a tool for comparing the sensitivity properties of different link specifications considered below.

(i) Logit-linear link. The logit-linear specification, introduced by Zheng and Basawa [34], is given by

$$\phi_t = \text{logit}^{-1}\left(\alpha+\beta X_{t-1}+\gamma^\top \boldsymbol{Z}_t+U_t\right), \quad \beta\le 0, \tag{2.4}$$

where $\text{logit}^{-1}(x)=\exp(x)/[1+\exp(x)]$; $\beta\le 0$ ensures that the ϕ_t does not increase with lagged counts. This formulation directly links the thinning probability to lagged counts on the original scale and has been widely used in observation-driven INAR models.

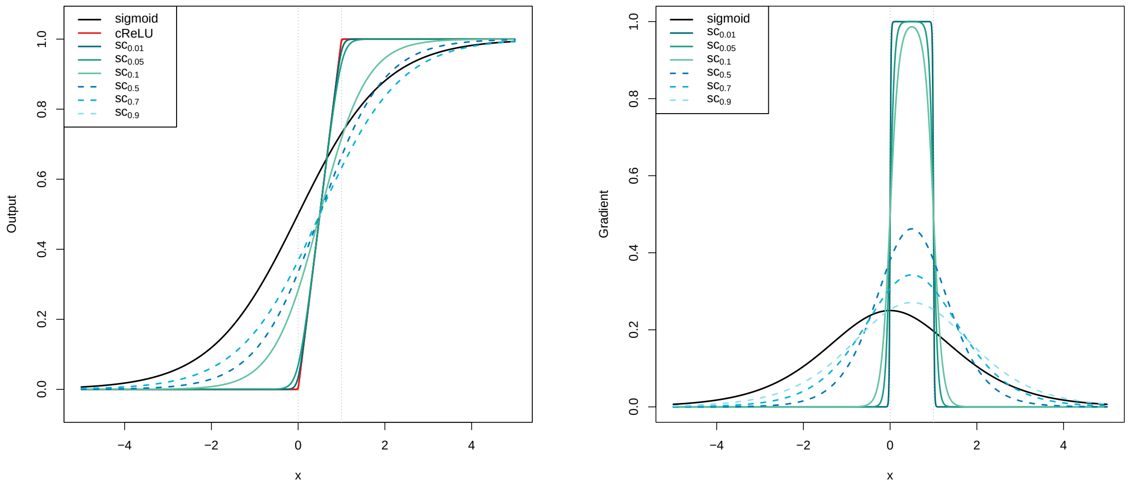
(ii) Logit-log link. We propose a logit-log specification of the form

$$\phi_t = \text{logit}^{-1}\left(\alpha+\beta\log(X_{t-1}+1)+\gamma^\top \boldsymbol{Z}_t+U_t\right), \quad \beta\le 0, \tag{2.5}$$

which introduces a logarithmic transformation of lagged counts. This transformation moderates the influence of large observations while naturally accommodating zero counts. To the best of our knowledge, this specification is new in the literature.

(iii) Soft-clipping link. Specifically, the soft-clipping function is defined as

$$\text{sc}_c(x) = c\log\left\{\frac{\exp(x/c)+1}{\exp((x-1)/c)+1}\right\}. \tag{2.6}$$



(a) (b)

Figure 1: Plots (against x) of (a) the inverse logit (sigmoid) and soft-clipping functions, and (b) their gradients.

Inspired by the soft-clipping INGARCH model of Weiß and Jahn [27], an alternative link specification is defined as

$$\phi_t = \mathrm{sc}_c\left(\alpha + \beta X_{t-1} + \gamma^\top \mathbf{Z}_t + U_t\right), \quad c > 0, \quad \beta \in \mathbb{R}, \tag{2.7}$$

where $\mathrm{sc}_c : \mathbb{R} \to (0,1)$ denotes the soft-clipping function with smoothing parameter $c > 0$. As shown in Fig. 1, compared with the inverse logit, which approaches its bounds gradually, the soft-clipping link saturates over a finite input range, thereby limiting the effect of large past counts. Smaller values of c induce earlier saturation, while larger values of c yield a smoother transition closer to the inverse logit.

We apply Proposition 2.2 to the link specifications considered above.

Corollary 2.1. *Under the conditions of Proposition* 2.2*, the local sensitivity bound for* $m_t(x,z)$ *specializes as follows:*

(i) (Logit-linear link) If $g(u) = \mathrm{logit}^{-1}(u)$ *and* $h(x) = \beta x$*, then*

$$\sup_{x \geq 0} \left| \frac{\partial m_t(x,z)}{\partial x} \right| \leq \frac{|\beta|}{4},$$

which does not depend on the magnitude of x*.*

(ii) (Logit-log link) If $g(u) = \text{logit}^{-1}(u)$ *and* $h(x) = \beta\log(x+1)$*, then*

$$\left|\frac{\partial m_t(x,z)}{\partial x}\right| \le \frac{|\beta|}{4(x+1)},$$

so the marginal effect of lagged counts attenuates as $x \to \infty$.

(iii) (Soft-clipping link) If $g(u) = \text{sc}_c(u)$ *and* $h(x) = \beta x$*, then*

$$\sup_{x\ge 0}\left|\frac{\partial m_t(x,z)}{\partial x}\right| \le |\beta|\kappa_c,$$

where $\kappa_c = \sup_{u\in\mathbb{R}} g'(u) \in (0,1)$.

Remark 2.1. Due to the absence of closed-form expressions for the autocovariance function, we include a brief illustrative simulation of process (2.1) without covariates. We generate realizations of length 12,000, discarding the first 4,000 observations as burn-in. Fig. 2 shows representative sample paths and the corresponding sample autocorrelation functions (ACFs) for the parameter setting

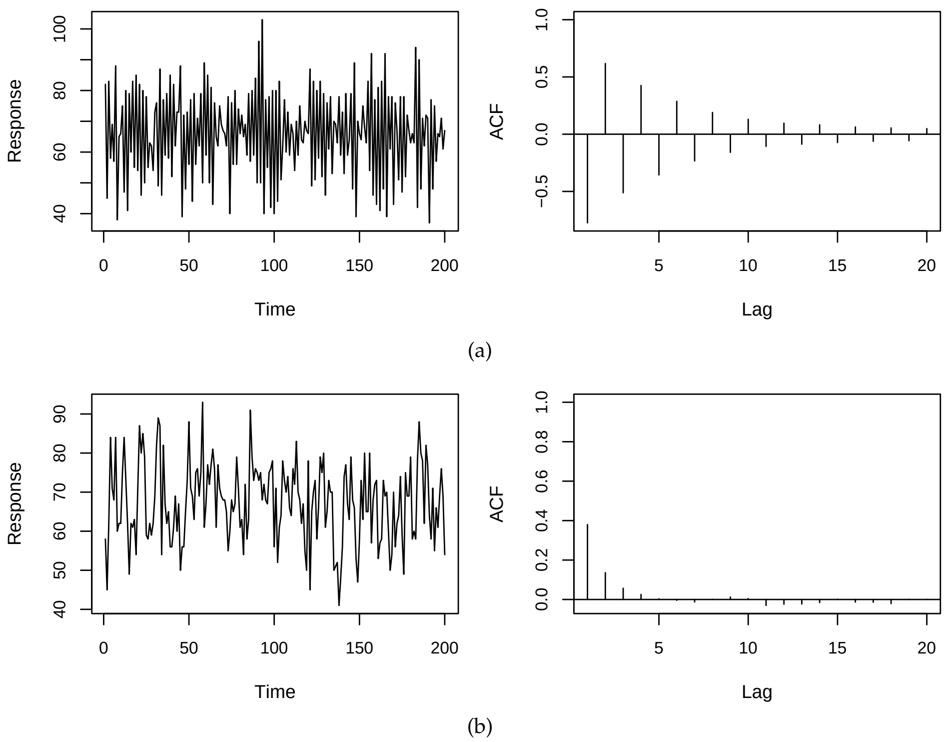


Figure 2: Simulated sample paths and ACFs under (a) logit-linear and (b) logit-log links.

($\lambda=40$, $\beta=-0.10$, $\eta=0.25$). The value $\lambda=40$ is chosen solely for illustration, as it produces higher count levels that make the differences between link functions easier to visualize. In contrast, the simulation study in Section 4 focuses on parameter configurations calibrated to empirically relevant count levels. Under this illustrative parameter setting, the logit-linear link tends to produce more volatile trajectories, whereas the logit-log link yields smoother paths with rapidly decaying autocorrelations.

We turn to the long-run behavior of the INAR(1) process defined in (2.1) and make the following assumption.

Assumption 2.1. (i) The innovation distribution satisfies $f_\varepsilon(i)>0$ for all $i\in\mathbb{N}_0$.

(ii) The covariate process $\mathcal{Z}=\{\mathbf{Z}_t\}_{t\in\mathbb{Z}}$ is stationary and shift-ergodic on the measurable space $(G^{\mathbb{Z}},\mathcal{B}(G^{\mathbb{Z}}))$: For any $D\in\mathcal{B}(G^{\mathbb{Z}})$ such that

$$1_D\big((\mathbf{Z}_t)_{t\in\mathbb{Z}}\big)=1_D\big((\mathbf{Z}_{t+1})_{t\in\mathbb{Z}}\big) \quad \text{almost surely,}$$

we have $P(\mathcal{Z}\in D)\in\{0,1\}$, where 1_D denotes the indicator function of the set D.

(iii) There exists a constant $\delta\in(0,1)$ such that

$$\sup_{t\geq 1}\sup_{x\in\mathbb{N}_0}\sup_{z\in G}E(\phi_t|X_{t-1}=x,\,\mathbf{Z}_t=z)\leq 1-\delta.$$

Assumption 2.1(iii) requires the conditional expectation to be uniformly bounded away from one, preventing near unit root behaviour; see Liu *et al.* [19, Remark 5]. Corollary 2.1 shows that the logit-log and soft-clipping links dampen the impact of large lagged counts on ϕ_t, making Assumption 2.1(iii) easier to satisfy than under the logit-linear link.

The following theorem establishes stationarity and ergodicity by considering an augmented state that includes the shifted covariate environment, so that the transition depends only on the current state, which renders the joint process time-homogeneous and permits standard Markov chain arguments.

Theorem 2.1. *Suppose that Assumption* 2.1 *holds. Then model* (2.1) *admits a stationary and ergodic joint solution* $\{(X_t,\mathbf{Z}_t)\}_{t\in\mathbb{Z}}$.

3 Estimation

We consider CML and CLS estimation approaches for the parameters of interest. Without the assumption of stationarity regarding the covariates, which is more common in reality and will be abandoned in the following, the stationarity of the considered process cannot be guaranteed. This makes the related asymptotic analysis for parameter estimation technically challenging. We therefore defer asymptotics and refer to Qian and Zhu [21] and the references therein for related results.

3.1 Conditional maximum likelihood estimation

Let $\boldsymbol{\theta}=(\alpha,\beta,\lambda,\sigma^2,\eta^2,\boldsymbol{\gamma}^\top)^\top$ denote the parameter vector. When the innovations follow a Poisson distribution, $\sigma^2=\lambda$ and is therefore not treated as a separate parameter. Although parameter estimation is based on maximization of the conditional log-likelihood, direct maximization of

$$\ell_n(\boldsymbol{\theta})=\sum_{t=2}^{n}\log P(X_t|X_{t-1},\boldsymbol{Z}_t) \tag{3.1}$$

is computationally infeasible due to non-closed form transition probabilities (2.2). To address this issue, we adopt an EM algorithm by treating $\{\phi_t\}$ as latent variables (see Yu and Zheng [33]). For a fixed t, the complete-data log-likelihood contribution is

$$\log f(X_t,\phi_t|X_{t-1},\boldsymbol{Z}_t)=\log f(X_t|X_{t-1},\phi_t)+\log f(\phi_t|X_{t-1},\boldsymbol{Z}_t),$$

where $f(\cdot)$ denotes the corresponding probability density or mass function.

Let $\boldsymbol{\theta}^{(s-1)}$ denote the parameter estimate obtained at iteration $s-1$. In the E-step, we evaluate the expected complete-data log-likelihood,

$$S\big(\boldsymbol{\theta}|\boldsymbol{\theta}^{(s-1)}\big)=\sum_{t=2}^{n}S_t\big(\boldsymbol{\theta}|\boldsymbol{\theta}^{(s-1)}\big), \tag{3.2}$$

where, for each t,

$$\begin{aligned}S_t\big(\boldsymbol{\theta}|\boldsymbol{\theta}^{(s-1)}\big)&=E\big[\log f(X_t,\phi_t|X_{t-1},\boldsymbol{Z}_t;\boldsymbol{\theta})|X_t,X_{t-1},\boldsymbol{Z}_t;\boldsymbol{\theta}^{(s-1)}\big]\\&=\int_0^1\log f\big(X_t|X_{t-1},\phi_t;\boldsymbol{\theta}\big)f_{\text{post}}\big(\phi_t|X_t,X_{t-1},\boldsymbol{Z}_t;\boldsymbol{\theta}^{(s-1)}\big)\,d\phi_t\\&\quad+\int_0^1\log f\big(\phi_t|X_{t-1},\boldsymbol{Z}_t;\boldsymbol{\theta}\big)f_{\text{post}}\big(\phi_t|X_t,X_{t-1},\boldsymbol{Z}_t;\boldsymbol{\theta}^{(s-1)}\big)\,d\phi_t.\end{aligned} \tag{3.3}$$

Here $f_{\text{post}}(\phi_t|X_t, X_{t-1}, \boldsymbol{Z}_t; \boldsymbol{\theta}^{(s-1)})$ denotes the posterior density of the latent variable ϕ_t, which is given by

$$f_{\text{post}}\big(\phi_t|X_t,X_{t-1},\boldsymbol{Z}_t;\boldsymbol{\theta}^{(s-1)}\big) = \frac{f(X_t|X_{t-1},\phi_t;\boldsymbol{\theta}^{(s-1)})f(\phi_t|X_{t-1},\boldsymbol{Z}_t;\boldsymbol{\theta}^{(s-1)})}{\int_0^1 f(X_t|X_{t-1},\phi_t;\boldsymbol{\theta}^{(s-1)})f(\phi_t|X_{t-1},\boldsymbol{Z}_t;\boldsymbol{\theta}^{(s-1)})d\phi_t}. \tag{3.4}$$

The integrals are evaluated numerically using the Gauss-Hermite quadrature.

In the M-step, we update

$$\hat{\boldsymbol{\theta}}^{(s)} = \arg\max_{\boldsymbol{\theta}} S\big(\boldsymbol{\theta}|\boldsymbol{\theta}^{(s-1)}\big), \tag{3.5}$$

where $S(\boldsymbol{\theta}|\boldsymbol{\theta}^{(s-1)})$ can be decomposed into two components: (i) a term involving λ through the conditional distribution $f(X_t|X_{t-1},\phi_t)$, and (ii) a term involving $(\alpha,\beta,\gamma,\eta^2)$ through $f(\phi_t|X_{t-1},\boldsymbol{Z}_t)$. Accordingly, λ is updated numerically using the Newton-Raphson method, while the remaining parameters are updated analytically as described below.

Recall that in (2.1) the link $g:\mathbb{R}\to(0,1)$ maps the input to ϕ_t, we can obtain that, conditional on $(X_{t-1},\boldsymbol{Z}_t)$,

$$g^{-1}(\phi_t) \sim N(\nu_t,\eta^2), \quad \nu_t = y_t^\top \boldsymbol{\xi},$$

where $y_t = (1, h(X_{t-1}), \boldsymbol{Z}_t^\top)^\top$, $\boldsymbol{\xi} = (\alpha,\beta,\boldsymbol{\gamma}^\top)^\top$. The conditional log-density of ϕ_t is given by

$$\log f(\phi_t|X_{t-1},\boldsymbol{Z}_t;\boldsymbol{\theta}) = -\frac{1}{2}\log(2\pi\eta^2) - \log|(g^{-1})'(\phi_t)| - \frac{(g^{-1}(\phi_t)-\nu_t)^2}{2\eta^2}. \tag{3.6}$$

Further, up to a constant independent of $(\boldsymbol{\xi},\eta^2)$, the relevant part of $S(\boldsymbol{\theta}|\boldsymbol{\theta}^{(s-1)})$ takes the form

$$S\big(\boldsymbol{\xi},\eta^2|\boldsymbol{\theta}^{(s-1)}\big) \propto -(n-1)\log\eta - \frac{1}{2\eta^2}\sum_{t=2}^{n} E\big[(g^{-1}(\phi_t)-\boldsymbol{\xi}^\top y_t)^2|X_t,X_{t-1},\boldsymbol{Z}_t;\boldsymbol{\theta}^{(s-1)}\big]. \tag{3.7}$$

The M-step update for $\boldsymbol{\xi}$ is therefore obtained by solving the least squares problem

$$\boldsymbol{\xi}^{(s)} = \arg\min_{\boldsymbol{\xi}} \sum_{t=2}^{n}\Big(E\big[g^{-1}(\phi_t)|X_t,X_{t-1},\boldsymbol{Z}_t;\boldsymbol{\theta}^{(s-1)}\big] - y_t^\top\boldsymbol{\xi}\Big)^2, \tag{3.8}$$

while η^2 is updated as

$$\eta^{2(s)}=\frac{1}{n-1}\sum_{t=2}^{n}E\left[(g^{-1}(\phi_t)-y_t^\top\boldsymbol{\xi}^{(s)})^2|X_t,X_{t-1},\boldsymbol{Z}_t;\boldsymbol{\theta}^{(s-1)}\right]. \tag{3.9}$$

The E-step and M-step are alternated until $\|\boldsymbol{\theta}^{(s)}-\boldsymbol{\theta}^{(s-1)}\|/\|\boldsymbol{\theta}^{(s-1)}\|<10^{-5}$. Based on the general EM algorithm described above, we detail the corresponding updates under each of the link specifications.

(i) Logit-linear link. With (2.4),

$$y_t=\left(1,X_{t-1},\boldsymbol{Z}_t^\top\right)^\top,\quad \nu_t=\alpha+\beta X_{t-1}+\boldsymbol{\gamma}^\top\boldsymbol{Z}_t,$$

and $g^{-1}(\cdot)=\text{logit}(\cdot)$, with the constraint $\beta\le 0$.

(ii) Logit-log link. With (2.5),

$$y_t=\left(1,\log(X_{t-1}+1),\boldsymbol{Z}_t^\top\right)^\top,\quad \nu_t=\alpha+\beta\log(X_{t-1}+1)+\boldsymbol{\gamma}^\top\boldsymbol{Z}_t,$$

and $g^{-1}(\cdot)=\text{logit}(\cdot)$, subject to $\beta\le 0$.

(iii) Soft-clipping link. With (2.7),

$$y_t=\left(1,X_{t-1},\boldsymbol{Z}_t^\top\right)^\top,\quad \nu_t=\alpha+\beta X_{t-1}+\boldsymbol{\gamma}^\top\boldsymbol{Z}_t,$$

and $g^{-1}(\cdot)=\text{sc}_c^{-1}(\cdot)$.

3.2 Conditional least squares estimation

Let $\boldsymbol{\psi}=(\alpha,\lambda,\beta,\eta^2,\boldsymbol{\gamma}^\top)^\top$ denote the parameters entering the conditional mean $E(X_t|X_{t-1},\boldsymbol{Z}_t)$, and let $\boldsymbol{\theta}=(\boldsymbol{\psi}^\top,\sigma^2)^\top$ denote the full parameter vector. The CLS estimator $\tilde{\boldsymbol{\psi}}_{\text{CLS}}$ is defined as

$$\tilde{\boldsymbol{\psi}}_{\text{CLS}}=\arg\min_{\boldsymbol{\psi}}\sum_{t=2}^{n}\left(X_t-E(X_t|X_{t-1},\boldsymbol{Z}_t)\right)^2. \tag{3.10}$$

In practice, $E(X_t|X_{t-1},\boldsymbol{Z}_t)$ is evaluated numerically via the Gauss-Hermite quadrature, with the parameter constraints imposed according to the chosen link case.

Given $\tilde{\boldsymbol{\psi}}_{\text{CLS}}$, the innovation variance σ^2 is then estimated using a two-step CLS procedure (see Chen *et al.* [4]) by minimizing

$$S_n(\sigma^2)=\sum_{t=2}^{n}\left[(X_t-E(X_t|X_{t-1},\boldsymbol{Z}_t;\tilde{\boldsymbol{\psi}}_{\text{CLS}}))^2-\text{Var}(X_t|X_{t-1},\boldsymbol{Z}_t;\tilde{\boldsymbol{\psi}}_{\text{CLS}})\right]^2. \tag{3.11}$$

4 Simulations

Assuming the innovation term ε_t follows a Poisson distribution with mean parameter $\lambda>0$, we present simulation results using the CML and CLS methods for estimating the parameter vector $\boldsymbol{\theta}=(\alpha,\beta,\lambda,\eta^2,\gamma)^\top$. In the simulation study we consider a single covariate ($p=1$); accordingly, γ is scalar. We consider sample sizes $n=100,300,500$ and perform 1000 replications using R under the following scenarios:

logit-log link:

(A1) $\boldsymbol{\theta}=(1,-0.05,1,1,-0.2)^\top$, with $Z_t=\sin\left(\frac{2\pi t}{12}\right)+\zeta_t$, $\zeta_t\overset{\text{i.i.d.}}{\sim}N(0,1.1^2)$;

(A2) $\boldsymbol{\theta}=(1,-0.4,1,1,-0.2)^\top$, with $Z_t=0.3Z_{t-1}+\zeta_t$, $\zeta_t\overset{\text{i.i.d.}}{\sim}N(0,1.5^2)$, $Z_1=\zeta_1$;

(A3) $\boldsymbol{\theta}=(1,-0.15,1,1,0.2)^\top$, with $Z_t=\zeta_t$, $\zeta_t\overset{\text{i.i.d.}}{\sim}\Gamma(2,1)$.

soft-clipping link:

(B1) $\boldsymbol{\theta}=(1,-0.05,1,1,-0.2)^\top$, with $Z_t=\sin\left(\frac{2\pi t}{12}\right)+\zeta_t$, $\zeta_t\overset{\text{i.i.d.}}{\sim}N(0,1.1^2)$;

(B2) $\boldsymbol{\theta}=(0.5,0.05,1,1,-0.2)^\top$, with $Z_t=0.3Z_{t-1}+\zeta_t$, $\zeta_t\overset{\text{i.i.d.}}{\sim}N(0,1.5^2)$, $Z_1=\zeta_1$;

(B3) $\boldsymbol{\theta}=(1,-0.15,1,1,0.2)^\top$, with $Z_t=\zeta_t$, $\zeta_t\overset{\text{i.i.d.}}{\sim}\Gamma(2,1)$.

Since the soft-clipping specification requires choosing the parameter c, we report a sensitivity analysis using the CML estimator for illustration. Table 1 considers scenarios B2 and B3 at $n=500$, using 200 replications for each $c\in\{0.005,0.01,0.05,0.1\}$. Bias, mean squared error (MSE), and mean absolute deviation error (MADE) are qualitatively unchanged across this grid, so we can fix $c=0.01$ in the subsequent simulations. Tables 2 and 3 show that as the sample size increases, the estimates in all scenarios tend to converge to the true parameter values. Overall, the CML estimator exhibits lower bias, MSE, and MADE than the CLS estimator. This superior performance is expected, since the CLS approach is moment-based, whereas the CML explicitly accounts for the latent thinning probabilities and the full conditional distribution. Some finite-sample bias is observed in small samples, which gradually diminishes as the sample size increases.

Table 1: Sensitivity analysis over the soft-clipping parameter c.

Scenario	Para.	c	Bias	MSE	MADE	c	Bias	MSE	MADE
B2	$\alpha=0.5$	0.005	0.0369	0.0289	0.1308	0.010	0.0201	0.0224	0.1164
	$\beta=0.05$		-0.0133	0.0016	0.0291		-0.0130	0.0012	0.0270
	$\lambda=1$		0.0050	0.0048	0.0548		0.0039	0.0040	0.0492
	$\eta=1$		-0.0582	0.0314	0.1460		-0.0611	0.0402	0.1638
	$\gamma=-0.2$		0.0080	0.0037	0.0482		0.0084	0.0041	0.0508
	$\alpha=0.5$	0.050	0.0354	0.0252	0.1222	0.100	0.0322	0.0274	0.1296
	$\beta=0.05$		-0.0152	0.0014	0.0283		-0.0148	0.0015	0.0298
	$\lambda=1$		0.0001	0.0042	0.0529		0.0060	0.0041	0.0517
	$\eta=1$		-0.0513	0.0347	0.1543		-0.0332	0.0360	0.1597
	$\gamma=-0.2$		0.0056	0.0037	0.0486		0.0070	0.0038	0.0494
B3	$\alpha=1$	0.005	-0.0066	0.0464	0.1686	0.010	-0.0313	0.0501	0.1759
	$\beta=-0.15$		-0.0007	0.0023	0.0379		0.0076	0.0020	0.0365
	$\lambda=1$		0.0080	0.0037	0.0468		0.0047	0.0039	0.0496
	$\eta=1$		-0.0638	0.0298	0.1412		-0.0770	0.0274	0.1339
	$\gamma=0.2$		-0.0109	0.0052	0.0571		-0.0130	0.0034	0.0468
	$\alpha=1$	0.050	-0.0333	0.0499	0.1814	0.100	-0.0106	0.0507	0.1759
	$\beta=-0.15$		0.0059	0.0021	0.0364		0.0020	0.0026	0.0392
	$\lambda=1$		-0.0004	0.0037	0.0494		0.0053	0.0035	0.0460
	$\eta=1$		-0.0605	0.0285	0.1411		-0.0351	0.0324	0.1496
	$\gamma=0.2$		-0.0055	0.0040	0.0485		-0.0064	0.0051	0.0584

5 Real data examples

We fit the observation-driven INAR(1) model (2.1) to two real-world datasets and compare the following specifications with assuming $\varepsilon_t \sim \text{Poisson}(\lambda)$:

$$
\begin{aligned}
&S0 \quad \phi_t \equiv \phi,\\
&S1 \quad \phi_t = \text{logit}^{-1}\big(\alpha+\beta X_{t-1}+\gamma^\top \mathbf{Z}_t+U_t\big),\\
&S2 \quad \phi_t = \text{logit}^{-1}\big(\alpha+\beta\log(X_{t-1}+1)+\gamma^\top \mathbf{Z}_t+U_t\big),\\
&S3 \quad \phi_t = \text{sc}_c\big(\alpha+\beta X_{t-1}+\gamma^\top \mathbf{Z}_t+U_t\big), \quad c=0.01,\\
&S4 \quad \phi_t = \text{sc}_c\big(\alpha+\beta X_{t-1}+\gamma^\top \mathbf{Z}_t+U_t\big), \quad c=0.1.
\end{aligned}
$$

Table 2: Biases of estimates, MSEs, and MADEs for logit-log specification. Part 1.

Scenario	Sample size	Para.	CML			CLS		
			Bias	MSE	MADE	Bias	MSE	MADE
A1	100	$\alpha=1$	0.1465	0.2080	0.3580	0.3373	0.3426	0.4642
		$\beta=-0.05$	-0.0959	0.0360	0.1391	-0.1812	0.0503	0.1829
		$\lambda=1$	0.0213	0.0270	0.1303	0.0024	0.0350	0.1478
		$\eta=1$	0.0995	0.1227	0.2641	0.1476	0.0465	0.1504
		$\gamma=-0.2$	0.0879	0.0288	0.1377	0.1221	0.0338	0.1569
	300	$\alpha=1$	0.1650	0.1351	0.2861	0.3664	0.2826	0.4269
		$\beta=-0.05$	-0.0997	0.0340	0.1380	-0.1746	0.0491	0.1792
		$\lambda=1$	-0.0024	0.0085	0.0725	-0.0261	0.0163	0.1020
		$\eta=1$	0.0066	0.0503	0.1794	0.1196	0.0321	0.1234
		$\gamma=-0.2$	0.0586	0.0119	0.0890	0.0717	0.0159	0.1053
	500	$\alpha=1$	0.1458	0.1130	0.2577	0.3468	0.2405	0.3914
		$\beta=-0.05$	-0.0968	0.0327	0.1355	-0.1730	0.0487	0.1777
		$\lambda=1$	0.0017	0.0053	0.0574	-0.0218	0.0102	0.0808
		$\eta=1$	-0.0016	0.0342	0.1485	0.1145	0.0314	0.1171
		$\gamma=-0.2$	0.0383	0.0072	0.0688	0.0445	0.0097	0.0799
A2	100	$\alpha=1$	0.0116	0.1222	0.2836	0.0363	0.2815	0.4258
		$\beta=-0.4$	0.0887	0.0581	0.2018	0.0497	0.0478	0.1866
		$\lambda=1$	-0.0197	0.0194	0.1123	-0.0137	0.0320	0.1437
		$\eta=1$	0.2867	0.2702	0.3615	0.1085	0.0529	0.1145
		$\gamma=-0.2$	0.0667	0.0246	0.1259	0.1045	0.0255	0.1376
	300	$\alpha=1$	0.0097	0.0951	0.2588	-0.0286	0.2220	0.3865
		$\beta=-0.4$	0.0182	0.0362	0.1586	0.0319	0.0471	0.1865
		$\lambda=1$	-0.0051	0.0084	0.0745	0.0073	0.0145	0.0960
		$\eta=1$	0.0933	0.0689	0.1962	0.1084	0.0527	0.1111
		$\gamma=-0.2$	0.0325	0.0079	0.0715	0.0511	0.0113	0.0882
	500	$\alpha=1$	0.0101	0.0837	0.2408	-0.0226	0.1935	0.3638
		$\beta=-0.4$	0.0011	0.0300	0.1441	0.0128	0.0448	0.1803
		$\lambda=1$	-0.0020	0.0052	0.0575	0.0108	0.0095	0.0784
		$\eta=1$	0.0456	0.0394	0.1530	0.1071	0.0520	0.1102
		$\gamma=-0.2$	0.0219	0.0046	0.0544	0.0286	0.0058	0.0613

Table 2: Biases of estimates, MSEs, and MADEs for logit-log specification. Part 2.

Scenario	Sample size	Para.	CML			CLS		
			Bias	MSE	MADE	Bias	MSE	MADE
A3	100	$\alpha=1$	0.1730	0.2728	0.4108	0.4218	0.4105	0.5295
		$\beta=-0.15$	0.0087	0.0268	0.1462	−0.0818	0.0243	0.1262
		$\lambda=1$	0.0148	0.0249	0.1259	−0.0200	0.0375	0.1578
		$\eta=1$	0.1098	0.0982	0.2286	−0.0495	0.0247	0.0985
		$\gamma=0.2$	−0.0896	0.0268	0.1350	−0.1314	0.0367	0.1679
	300	$\alpha=1$	0.1504	0.1652	0.3268	0.2874	0.2403	0.3889
		$\beta=-0.15$	−0.0216	0.0255	0.1381	−0.0727	0.0236	0.1243
		$\lambda=1$	0.0065	0.0083	0.0727	−0.0123	0.0164	0.1028
		$\eta=1$	0.0275	0.0431	0.1617	−0.0660	0.0128	0.0803
		$\gamma=0.2$	−0.0575	0.0120	0.0905	−0.0825	0.0168	0.1109
	500	$\alpha=1$	0.1256	0.1329	0.2931	0.2130	0.1780	0.3427
		$\beta=-0.15$	−0.0376	0.0248	0.1331	−0.0715	0.0229	0.1237
		$\lambda=1$	0.0033	0.0054	0.0581	−0.0069	0.0125	0.0889
		$\eta=1$	0.0058	0.0281	0.1343	−0.0636	0.0111	0.0739
		$\gamma=0.2$	−0.0358	0.0070	0.0677	−0.0538	0.0107	0.0871

Table 3: Biases of estimates, MSEs, and MADEs for soft-clipping specification. Part 1.

Scenario	Sample size	Para.	CML			CLS		
			Bias	MSE	MADE	Bias	MSE	MADE
B1	100	$\alpha=1$	0.0310	0.1474	0.2931	0.2829	0.6774	0.6130
		$\beta=-0.05$	−0.0374	0.0085	0.0678	−0.0875	0.0296	0.1286
		$\lambda=1$	−0.0067	0.0172	0.1048	0.0123	0.0477	0.1790
		$\eta=1$	−0.2175	0.1244	0.2954	0.0580	0.0464	0.1329
		$\gamma=-0.2$	0.0280	0.0190	0.1057	−0.0387	0.0560	0.1784
	300	$\alpha=1$	−0.0032	0.0486	0.1734	0.2289	0.2809	0.4049
		$\beta=-0.05$	−0.0141	0.0023	0.0369	−0.0488	0.0090	0.0718
		$\lambda=1$	−0.0029	0.0060	0.0618	−0.0307	0.0224	0.1238
		$\eta=1$	−0.1035	0.0472	0.1768	0.0703	0.0310	0.1251
		$\gamma=-0.2$	0.0132	0.0070	0.0671	−0.0239	0.0172	0.1024
	500	$\alpha=1$	−0.0046	0.0274	0.1327	0.2214	0.2347	0.3786
		$\beta=-0.05$	−0.0085	0.0012	0.0270	−0.0402	0.0060	0.0604
		$\lambda=1$	−0.0002	0.0034	0.0472	−0.0379	0.0205	0.1139
		$\eta=1$	−0.0737	0.0291	0.1398	0.0633	0.0218	0.1173
		$\gamma=-0.2$	0.0099	0.0044	0.0527	−0.0223	0.0112	0.0797

Table 3: Biases of estimates, MSEs, and MADEs for soft-clipping specification. Part 2.

Scenario	Sample size	Para.	CML			CLS		
			Bias	MSE	MADE	Bias	MSE	MADE
B2	100	$\alpha=0.5$	0.1263	0.1465	0.2959	0.3045	0.5764	0.5727
		$\beta=0.05$	−0.0409	0.0084	0.0691	−0.0708	0.0217	0.1076
		$\lambda=1$	0.0040	0.0153	0.0984	−0.0129	0.0380	0.1569
		$\eta=1$	−0.0544	0.0418	0.1727	0.0386	0.0276	0.1097
		$\gamma=-0.2$	0.0085	0.0162	0.0993	−0.0242	0.0444	0.1564
	300	$\alpha=0.5$	0.0462	0.0443	0.1638	0.2675	0.2844	0.4025
		$\beta=0.05$	−0.0198	0.0025	0.0371	−0.0465	0.0080	0.0660
		$\lambda=1$	0.0036	0.0061	0.0629	−0.0498	0.0245	0.1251
		$\eta=1$	−0.0709	0.0291	0.1455	0.0562	0.0216	0.1075
		$\gamma=-0.2$	0.0140	0.0051	0.0569	−0.0250	0.0149	0.0918
	500	$\alpha=0.5$	0.0298	0.0251	0.1240	0.2309	0.2029	0.3411
		$\beta=0.05$	−0.0141	0.0013	0.0276	−0.0358	0.0049	0.0521
		$\lambda=1$	0.0046	0.0038	0.0491	−0.0516	0.0186	0.1096
		$\eta=1$	−0.0645	0.0244	0.1314	0.0625	0.0212	0.1072
		$\gamma=-0.2$	0.0077	0.0033	0.0454	−0.0303	0.0104	0.0778
B3	100	$\alpha=1$	−0.0054	0.2082	0.3602	0.1755	0.7863	0.6509
		$\beta=-0.15$	−0.0126	0.0083	0.0714	−0.0640	0.0296	0.1254
		$\lambda=1$	0.0060	0.0166	0.1016	0.0125	0.0489	0.1799
		$\eta=1$	−0.0726	0.0351	0.1590	0.0261	0.0250	0.1081
		$\gamma=0.2$	0.0050	0.0191	0.1094	0.0484	0.0496	0.1652
	300	$\alpha=1$	−0.0310	0.0685	0.2083	0.0980	0.2908	0.4221
		$\beta=-0.15$	0.0039	0.0029	0.0429	−0.0289	0.0097	0.0751
		$\lambda=1$	0.0041	0.0053	0.0583	−0.0053	0.0241	0.1255
		$\eta=1$	−0.0808	0.0265	0.1364	0.0339	0.0202	0.1063
		$\gamma=0.2$	−0.0090	0.0057	0.0599	0.0215	0.0160	0.0936
	500	$\alpha=1$	−0.0218	0.0436	0.1669	0.1155	0.2270	0.3781
		$\beta=-0.15$	0.0047	0.0019	0.0346	−0.0259	0.0072	0.0657
		$\lambda=1$	0.0028	0.0038	0.0490	−0.0175	0.0207	0.1160
		$\eta=1$	−0.0699	0.0223	0.1246	0.0352	0.0166	0.1024
		$\gamma=0.2$	−0.0113	0.0040	0.0506	0.0152	0.0093	0.0746

5.1 Possession of drugs offences in Southwark

We analyze monthly observations of reported possession of drugs offences in the Faraday ward of Southwark, London, available at `https://data.london.gov.uk/dataset/mps-recorded-crime-geographic-breakdown-exy3m/`. The sample covers the period from April 2010 to September 2023 and consists of 162 observations. The series has sample mean of 10.3148 and variance of 22.3661, indicating substantial overdispersion. Given the potential impact of seasonality, we selected $\sin(2\pi t/12)+\zeta_t$ as a sinusoidal covariate, where $\zeta_t \overset{\text{i.i.d.}}{\sim} N(0,0.3^2)$.

Fig. 3 presents the histogram, time series plot, and the corresponding ACF and partial autocorrelation function (PACF) plots of the data, which are consistent with an INAR(1) dependence structure. As shown in Table 4, the parameter estimates, negative log-likelihood values ($-\ell$), Akaike information criterion (AIC), and Bayesian information criterion (BIC) are reported for specifications S0-S4. The observation-driven specifications S1-S4 achieve very similar log-likelihood values and all outperform the constant-thinning specification S0. Among them, the soft-clipping specification S3 attains slightly lower AIC and BIC values. Moreover, the similar performance of S3 and S4 indicates that, for this dataset, the choice of c within this range has little influence on the overall model fit.

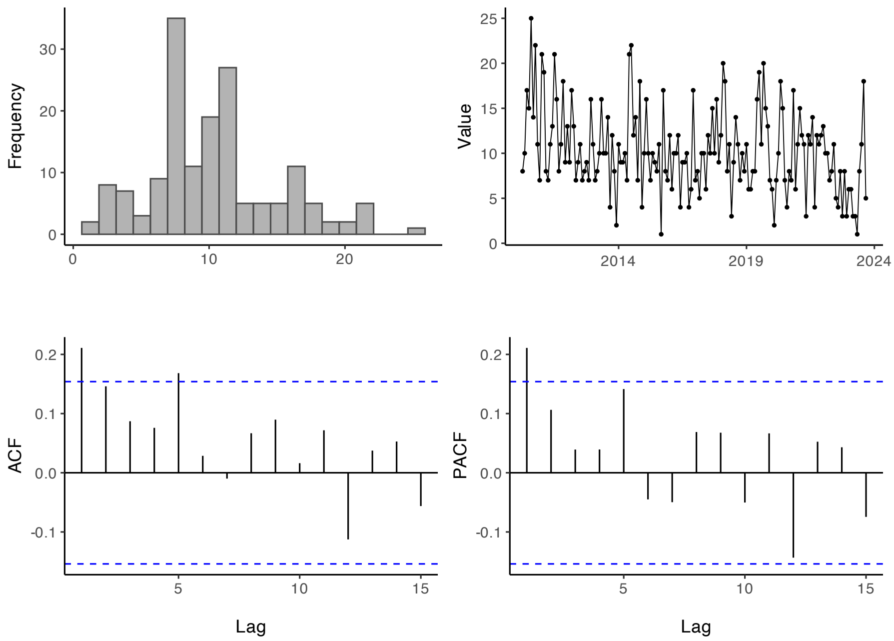


Figure 3: The histogram and time series plot of the monthly drugs offences data, along with the corresponding ACF and PACF plots.

Table 4: Parameter estimates, $-\ell$, AIC and BIC values for the drugs offences data.

	ϕ	α	β	λ	η	γ	$-\ell$	AIC	BIC
S0	0.139			8.882			501.537	1007.073	1013.248
S1		-1.754	-0.033	7.761	2.665	0.468	476.605	963.210	978.647
S2		-1.003	-0.443	7.734	2.647	0.463	476.577	963.155	978.593
S3		0.178	-0.005	7.798	0.478	0.092	476.273	962.546	977.984
S4		0.139	-0.006	7.767	0.548	0.098	476.304	962.608	978.046

5.2 Motor vehicle theft offences in Ballina

As a second example, we consider the monthly motor vehicle theft offence records from Ballina, New South Wales (`https://www.bocsar.nsw.gov.au/`), spanning from June 1995 to September 2018 and comprising a total of 280 observations. The empirical mean and variance of the series are 5.8464 and 11.8366, respectively. The potential influence of climate and temperature on criminal activity has been documented in the existing literature; see [2,22]. Motivated by this evidence, we select the monthly average maximum temperature (°C) as a covariate and apply seasonal differencing of order 12; see Fig. 4. The temperature data are obtained from the Australian Government Bureau of Meteorology (station 058198; `http://www.bom.gov.au/`). In addition, the monthly number of fraud offences in Ballina is included as a second covariate.

Fig. 5 illustrates key empirical features of the motor vehicle theft data in Ballina, which exhibit clear overdispersion and right skewness. Table 5 reports the

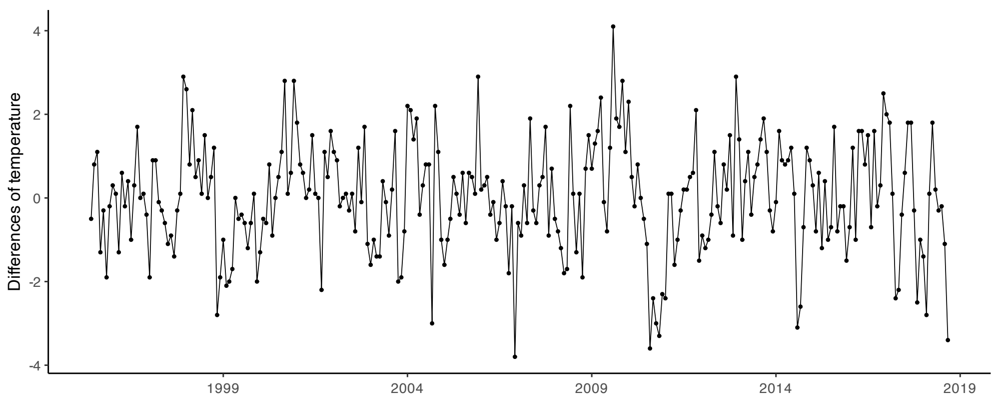


Figure 4: The monthly average maximum temperature after 12-order difference.

Table 5: Parameter estimates, $-\ell$, AIC and BIC values for the motor vehicle theft data.

	ϕ	α	β	λ	η	γ_1	γ_2	$-\ell$	AIC	BIC
S0	0.137			5.045				757.722	1519.444	1526.713
S1		-2.362	-0.060	4.551	3.507	-0.196	-0.013	728.955	1469.910	1491.719
S2		-2.205	-0.345	4.571	3.581	-0.204	-0.013	729.210	1470.420	1492.229
S3		0.063	-0.010	4.539	0.647	-0.036	-0.003	728.463	1468.926	1490.735
S4		0.017	-0.011	4.547	0.720	-0.039	-0.003	728.701	1469.403	1491.211

estimation results for specifications S0-S4. Introducing observation-driven thinning yields a substantial improvement over the constant-thinning specification S0, as evidenced by the markedly lower negative log-likelihood values. Differences among specifications S1-S4 are relatively small. Among them, the soft-clipping specification S3 achieves the lowest AIC and BIC values, while further increasing the clipping parameter c (S4) does not lead to additional gains. Although the information criteria are similar across S1-S4, the logit-log and soft-clipping links offer a more robust and interpretable response to unusually large counts than the logit-linear link.

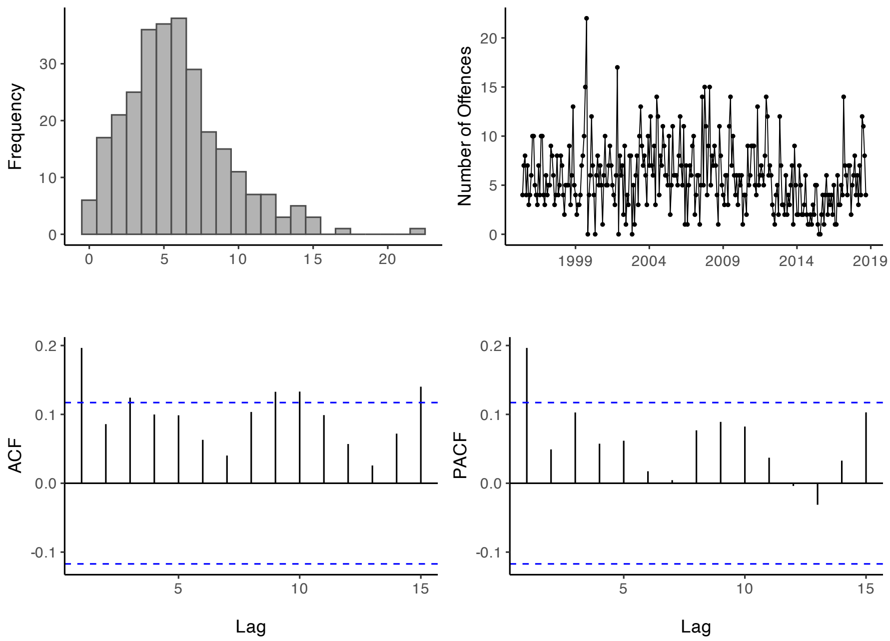


Figure 5: The histogram, time series plot of the monthly motor vehicle theft data in Ballina, along with the corresponding ACF and PACF plots.

6 Conclusion

In this paper, we propose observation-driven INAR(1) models with covariates under two novel link specifications. The proposed models are shown to be stationary and ergodic under the stationarity regarding the covariates. Parameter estimation is conducted using CML and CLS, and simulation results indicate superior finite-sample performance of the former. Two real-data applications are analyzed. As a practical recommendation, if the logit-linear specification frequently yields near-zero fitted thinning probabilities, the logit-log or soft-clipping link provides a competitive alternative.

Several directions merit further investigation. Related work by Wang *et al.* [25] demonstrates that alternative links may lead to distinct dynamic behaviors in binary INGARCH models. Extending the proposed framework to accommodate more flexible and possibly asymmetric links is therefore a natural topic. Another promising avenue is to explore more flexible distributions with richer shape properties, as in the binomial-Poisson Lindley regression framework of Chesneau *et al.* [5]. It would also be of interest to develop conditional quantile inference and distributional forecasting for INAR models, building on recent advances for Poisson autoregressive processes in Sheng and Wang [23]. In addition, Bayesian estimation methods, including Markov chain Monte Carlo algorithms, may provide a useful alternative for inference in more complex model specifications.

Acknowledgments

We thank the editor and reviewers for their insightful and constructive comments that greatly improve the overall presentation.

Zhu's work is supported by the Scientific Research Project Funding from the Education Department of Jilin Province (Grant No. JJKH20261608KJ) and by the National Natural Science Foundation of China (Grant No. 12271206).

Appendix A

Proof of Theorem 2.1. We consider an enlarged state space that incorporates the shifted environment and establish ergodicity via a 0-1 invariant-set argument, as in Fokianos and Truquet [9, Theorem 2] and Truquet [24, Theorem 1]. Existence and uniqueness of the stationary law are obtained by verifying a Lyapunov drift inequality and a uniform one-step minorization condition, in the spirit of Truquet [24].

Let $W_t := (\boldsymbol{Z}_{t+s})_{s\in\mathbb{Z}} \in G^{\mathbb{Z}}$, so W_{t+1} is the left shift of W_t and $(W_t)_0 = \boldsymbol{Z}_t$, $(W_{t-1})_1 = \boldsymbol{Z}_t$. Define $H_t := (X_t, W_t) \in \mathbb{N}_0 \times G^{\mathbb{Z}}$. Fix $t \in \mathbb{Z}$ and condition on $H_{t-1} = (x,w)$. Then $\boldsymbol{Z}_t = (w)_1$ and W_t is obtained deterministically from w by one shift. Given H_{t-1}, (ϕ_t, X_t) is generated from (U_t, ε_t), which is independent of the past. Hence, the conditional distribution of H_t given the past depends only on H_{t-1}, so $\{H_t\}$ is time-homogeneous Markov.

Consider the conditional transition kernel of the count component given the environment. For $w \in G^{\mathbb{Z}}$, define

$$Q_w(x,A) := P\big(X_t \in A \,|\, H_{t-1} = (x,w)\big) = P\big(X_t \in A \,|\, X_{t-1} = x,\, \boldsymbol{Z}_t = (w)_1\big), \quad A \subset \mathbb{N}_0.$$

Let $V : \mathbb{N}_0 \times G^{\mathbb{Z}} \to [0,\infty)$ be given by $V(x,w) = x$. Conditioning on $H_{t-1} = (x,w)$ fixes $X_{t-1} = x$ and $\boldsymbol{Z}_t = (w)_1$, and Assumption 2.1(iii) yields

$$E\big[V(H_t) \,|\, H_{t-1} = (x,w)\big] = E\big(X_t \,|\, X_{t-1} = x,\, \boldsymbol{Z}_t = (w)_1\big) \le (1-\delta)x + \lambda, \tag{A.1}$$

which is a Lyapunov drift inequality.

Fix $m \in \mathbb{N}$ and let $x \in \{0,1,\ldots,m\}$. By Assumption 2.1(i), conditional on $H_{t-1} = (x,w)$ the distribution of X_t has full support on $\mathbb{N}_0$. For any $i \in \mathbb{N}_0$,

$$Q_w(x,\{i\}) = P(X_t = i \,|\, H_{t-1} = (x,w)) \ge E[(1-\phi_t)^x \,|\, X_{t-1} = x,\, \boldsymbol{Z}_t = (w)_1] f_\varepsilon(i).$$

If $x \ge 1$, since $u \mapsto (1-u)^x$ is convex on $[0,1]$, Jensen's inequality and Assumption 2.1(iii) give

$$E[(1-\phi_t)^x \,|\, X_{t-1} = x,\, \boldsymbol{Z}_t = (w)_1] \ge (1 - E(\phi_t \,|\, X_{t-1} = x,\, \boldsymbol{Z}_t = (w)_1))^x \ge \delta^x \ge \delta^m,$$

and the case $x = 0$ is trivial since $(1-\phi_t)^0 \equiv 1$. Define φ on $\mathbb{N}_0$ by $\varphi(\{i\}) = f_\varepsilon(i)$. Then for any $A \subset \mathbb{N}_0$,

$$Q_w(x,A) \ge \delta^m \varphi(A), \quad x \in \{0,1,\ldots,m\}, \quad \forall w \in G^{\mathbb{Z}}. \tag{A.2}$$

That is, $\{0,1,\ldots,m\}$ satisfies a uniform one-step minorization for Q_w, uniformly over w. The drift inequality (A.1) and the minorization (A.2) satisfy the drift-minorizat ion assumptions of Truquet [24, Theorem 1]. In particular, there exists a unique invariant probability measure $\nu \in \mathcal{M}_P$, where

$$\mathcal{M}_P := \big\{\nu \;\text{ on } \mathbb{N}_0 \times G^{\mathbb{Z}} :\, \nu(\mathbb{N}_0 \times \cdot) = P(\cdot)\big\}.$$

We then prove ergodicity of $\{H_t\}$ under ν using an invariant-set argument. Let $C \subset \mathbb{N}_0 \times G^{\mathbb{Z}}$ be measurable with $\nu(C) > 0$ and invariant for $\{H_t\}$. Then for

ν-almost every $(x,w)\in C$, $P(H_1\in C|H_0=(x,w))=1$. Given $H_0=(x,w)$, the environment component evolves deterministically by the shift, and, by Assumption 2.1(i), conditional on $W_0=w$ the count component X_1 has full support on $\mathbb{N}_0$. Consequently,

$$P\big((i,W_1)\in C|H_0=(x,w)\big)=1 \quad \text{for all} \quad i\in\mathbb{N}_0,$$

where W_1 is the shift of w. This implies that for a given shifted environment w', either $(i,w')\in C$ for all $i\in N_0$, or $(i,w')\notin C$ for all $i\in N_0$. Hence, there exists a measurable set $D\subset G^{\mathbb{Z}}$ such that $C=\mathbb{N}_0\times D$ ν-almost surely. Invariance of C implies that D is invariant under the shift on $G^{\mathbb{Z}}$. Since the $G^{\mathbb{Z}}$-marginal of ν equals P and Z is shift-ergodic by Assumption 2.1(ii), we have $P(W_0\in D)\in\{0,1\}$ and thus $\nu(C)=\nu(\mathbb{N}_0\times D)=P(W_0\in D)\in\{0,1\}$. Therefore, $\{H_t\}$ is ergodic under ν; consequently, the marginal law of ν on $(X_t,\boldsymbol{Z}_t)$ is stationary and ergodic for the joint process $\{(X_t,\boldsymbol{Z}_t)\}_{t\in\mathbb{Z}}$. $\square$

References

[1] M. A. Al-Osh and A. A. Alzaid, *First-order integer-valued autoregressive (INAR(1)) process*, J. Time Ser. Anal. 8 (1987), 261–275.

[2] D. S. Blakeslee and R. Fishman, *Weather shocks, agriculture, and crime: Evidence from India*, J. Hum. Resour. 53 (2018), 750–782.

[3] H. Chen, Q. Li, and F. Zhu, *Two classes of dynamic binomial integer-valued ARCH models*, Braz. J. Probab. Stat. 34 (2020), 685–711.

[4] H. Chen, F. Zhu, and X. Liu, *Two-step conditional least squares estimation for the bivariate $\mathbb{Z}$-valued INAR(1) model with bivariate Skellam innovations*, Commun. Stat.-Theory Methods 53 (2024), 4085–4106.

[5] C. Chesneau, H. S. Bakouch, Y. Akdoğan, and K. Karakaya, *The binomial-discrete Poisson-Lindley model: Modeling and applications to count regression*, Commun. Math. Res. 38 (2022), 28–51.

[6] X. Ding and D. Wang, *Empirical likelihood inference for INAR(1) model with explanatory variables*, J. Korean Stat. Soc. 45 (2016), 623–632.

[7] P. Doukhan, M. H. Neumann, and L. Truquet, *Stationarity and ergodic properties for some observation-driven models in random environments*, Ann. Appl. Probab. 33 (2023), 5145–5170.

[8] K. Fokianos and D. Tjøstheim, *Log-linear Poisson autoregression*, J. Multivar. Anal. 102 (2011), 563–578.

[9] K. Fokianos and L. Truquet, *On categorical time series models with covariates*, Stoch. Process. Their Appl. 129 (2019), 3446–3462.

[10] Y. Kang, D. Sheng, and F. Lu, *A simple INAR(1) model for analyzing count time series with multiple features*, Commun. Stat.-Theory Methods 54 (2025), 457–475.

[11] Y. Kang, D. Sheng, and J. Yue, *Threshold integer-valued autoregressive model with serially dependent innovation*, J. Stat. Comput. Simul. 94 (2024), 3826–3863.

[12] Y. Kang, D. Wang, F. Lu, and S. Wang, *Flexible INAR(1) models for equidispersed, underdispersed or overdispersed counts*, J. Korean Stat. Soc. 51 (2022), 1268–1301.

[13] Y. Kang, D. Wang, K. Yang, and Y. Zhang, *A new thinning-based INAR(1) process for underdispersed or overdispersed counts*, J. Korean Stat. Soc. 49 (2020), 324–349.

[14] Y. Kang, F. Zhu, D. Wang, and S. Wang, *A zero-modified geometric INAR(1) model for analyzing count time series with multiple features*, Can. J. Stat. 52 (2024), 873–899.

[15] Y. Kifer, *Perron-Frobenius theorem, large deviations, and random perturbations in random environments*, Math. Z. 222 (1996), 677–698.

[16] E. McKenzie, *Some simple models for discrete variate time series*, J. Am. Water Resour. Assoc. 21 (1985), 645–650.

[17] G. Miao and D. Wang, *Bayesian estimation for the order of INAR(q) model*, Commun. Math. Res. 32 (2016), 325–331.

[18] Q. Li, H. Chen, and F. Zhu, *$\mathbb{Z}$-valued time series: Models, properties and comparison*, J. Stat. Plan. Inference 230 (2024), Paper No. 106099.

[19] C. Liu, H. Lv, and D. Wang, *Observation-driven random coefficient threshold INAR models for count time series*, Stat 14 (2025), Paper No. e70105.

[20] M. Liu, F. Zhu, J. Li, and C. Sun, *A systematic review of INGARCH models for integer-valued time series*, Entropy 25 (2023), Paper No. 922.

[21] L. Qian and F. Zhu, *A new minification integer-valued autoregressive process driven by explanatory variables*, Aust. N. Z. J. Stat. 64 (2022), 478–494.

[22] M. Ranson, *Crime, weather, and climate change*, J. Environ. Econ. Manage. 67 (2014), 274–302.

[23] D. Sheng and D. Wang, *Quantile regression estimation for Poisson autoregressive models*, J. Time Ser. Anal. 47 (2026), 378–413.

[24] L. Truquet, *Ergodic properties of some Markov chains models in random environments*, arXiv:2108.06211, 2021.

[25] C. Wang, Q. Li, and F. Zhu, *Binary INGARCH models with different link functions*, J. Stat. Comput. Simul. 95 (2025), 2848–2868.

[26] C. H. Weiß, *A Poisson INAR(1) model with serially dependent innovations*, Metrika 78 (2015), 829–851.

[27] C. H. Weiß and M. Jahn, *Soft-clipping INGARCH models for time series of bounded counts*, Stat. Model. 24 (2024), 295–319.

[28] C. H. Weiß, F. Zhu, and A. Hoshiyar, *Softplus INGARCH models*, Stat. Sin. 32 (2022), 1099–1120.

[29] K. Yang, H. Li, D. Wang, and C. Zhang, *Random coefficients integer-valued threshold*

autoregressive processes driven by logistic regression, AStA Adv. Stat. Anal. 105 (2021), 533–557.

[30] K. Yang, Y. Zhao, H. Li, and D. Wang, *On bivariate threshold Poisson integer-valued autoregressive processes*, Metrika 86 (2023), 931–963.

[31] K. Yu and T. Tao, *An observation-driven random parameter INAR(1) model based on the Poisson thinning operator*, Entropy 25 (2023), Paper No. 859.

[32] M. Yu, D. Wang, and K. Yang, *A class of observation-driven random coefficient INAR(1) processes based on negative binomial thinning*, J. Korean Stat. Soc. 48 (2019), 248–264.

[33] W. Yu and H. Zheng, *First-order multivariate integer-valued autoregressive model with multivariate mixture distributions*, J. Stat. Comput. Simul. 94 (2024), 3015–3041.

[34] H. Zheng and I. V. Basawa, *First-order observation-driven integer-valued autoregressive processes*, Stat. Probab. Lett. 78 (2008), 1–9.

[35] H. Zheng, I. V. Basawa, and S. Datta, *First-order random coefficient integer-valued autoregressive processes*, J. Stat. Plan. Inference 137 (2007), 212–229.