Statistica Sinica Preprint No: SS-2025-0390	
Title	Modelling Time Series of Counts with Hysteresis
Manuscript ID	SS-2025-0390
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0390
Complete List of Authors	Xintong Ma, Dong Li and Howell Tong
Corresponding Authors	Xintong Ma
E-mails	mxt22@mails.tsinghua.edu.cn
Notice: Accepted author version.	

Modelling time series of counts with hysteresis

Xintong Ma[1], Dong Li[1] and Howell Tong[1,2,3]

[1] *Tsinghua University,* [2] *Xiamen University*

and [3] *The London School of Economics and Political Science*

Abstract: In this article, we propose a novel model for time series of counts called the hysteretic Poisson autoregressive model with thresholds (HPART) by extending the linear Poisson autoregressive model into a nonlinear model. Unlike other approaches that bear the adjective "hysteretic", our model incorporates a scientifically relevant and essential controlling factor that produces genuine hysteresis. Further, we re-analyse the buffered Poisson autoregressive model with thresholds (BPART). Although the two models share the convenient piecewise linear structure, the HPART model probes deeper into the intricate dynamics that governs regime switching. We study the maximum likelihood estimation of the parameters of both models and their asymptotic properties in a unified manner, establish tests of separate families of hypotheses for the non-nested case involving a BPART model and a HPART model, and demonstrate the finite-sample efficacy of parameter estimation and tests with Monte Carlo simulation in the Supplementary Material. We showcase advantages of the HPART model with two real time series, including plausible interpretations and improved out-of-sample predictions.

Key words and phrases: Buffered Poisson autoregression, Hysteresis, Hysteretic Poisson autoregression, Non-nested models, Separate family of hypotheses, Thresholds.

1. Introduction

1.1 What is hysteresis in science and technology?

Let us start by stating that our use of the term *hysteresis* is in line with the phenomenon defined in physics, engineering, biology and many other scientific and engineering disciplines. It is associated with a hysteresis curve that is a *multivalued* function with two branches. Which branch a trajectory follows depends critically on a *controlling factor*. For example, the action of loading/unloading a weight to/from a spring is a controlling factor: unloading a weight returns the spring to (possibly) its original state but along a different branch than the one initially traversed by the loading of a weight on the same spring. Fig. 1 illustrates a hysteresis mechanism intuitively. The output is unambiguous (state 1) for input less than or equal to r or (state 0) for greater than s. However, for input between r and s, the output depends on whether the input is increasing or decreasing. Here the difference of two consecutive inputs is the controlling factor. A controlling factor is essential for the generation of hysteresis. For details, see,

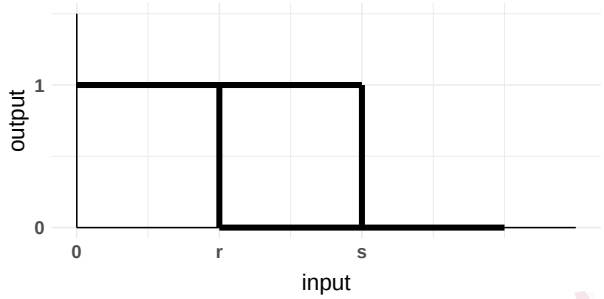

Figure 1: The relay figure illustrating the hysteresis mechanism.

e.g., Morris (2012). Hysteresis was first discovered in magnetism by Ewing (1885) and is commonly observed in mechanical gears and smart materials, see, e.g., Kennedy and Chua (1991), Brokate and Sprekels (1996), Smith (2005), etc.

1.2 Background

As early as in 1980, Tong and Lim (1980, p.280) recognised the importance of hysteresis in the above sense; specifically, they referred to the cusp catastrophe that generates hysteresis among other features (*op cit.*, p.292 that cited Zeeman (1977)). In fact, they even fitted model (A1) (*op cit.*, p.283) to the Canadian lynx data. Model (A1) uses the sign of the first difference of the observations as the controlling factor. As far as we know, model (A1) is the first hysteretic autoregressive model for real-valued time series in the literature.

Later, for *real-valued (i.e. non-count)* time series data, Zhu et al. (2014) and Li et al. (2015) studied another interesting mechanism by introducing a buffer zone. Further development of this approach followed, e.g. Lo et al. (2016), Zhu et al. (2017), Li et al. (2020), Liu et al. (2020), Wang and Li (2020), and others.

Time series of counts is an important subclass of time series, ubiquitous in the real world. The statistical analysis of such time series has received much attention in recent years; see, e.g., Liu et al. (2019), Armillotta and Fokianos (2023, 2024), Kong and Lund (2023), Jia et al. (2023), Weiß and Zhu (2024), and the references therein. For comprehensive reviews, see Tjøstheim (2012), Davis et al. (2016), Weiß (2018), Davis et al. (2021), Fokianos et al. (2022), Karlis and Mamode Khan (2023), Liu et al. (2023), and Sellers (2023).

A well-known model for time series of counts is the Poisson autoregressive (PAR) model (Fokianos et al., 2009), also known as the BIN model (Rydberg and Shephard, 2000), the ACP model (Heinen, 2003), and the INGARCH model (Ferland et al., 2006). It has been further studied extensively by Fokianos and Tjøstheim (2011, 2012), Neumann (2011), Wang et al. (2014), Christou and Fokianos (2015), Ahmad and Francq (2016), Davis and Liu (2016), Douc et al. (2017), Alzahrani et al. (2018), Fokianos

et al. (2020), Diop and Kengne (2021), Doukhan et al. (2021, 2022), Weiß et al. (2022), Huang and Khabou (2023), and others.

To address shortcomings of the PAR model, by adopting the threshold approach of Tong (1978), Wang et al. (2014) proposed a new self-excited threshold PAR (SETPAR) model, which provides a more flexible intensity process. Subsequently, Truong et al. (2017) introduced the buffer mechanism into the PAR model and proposed a new count time series model with the adjective *hysteretic*, although its use differs from the sense explained at the opening paragraph. Therefore, to distinguish this model from the one that we are about to develop, we assign the acronym BPART to their buffered PAR model with thresholds. Related studies on buffered count models include works by Liu et al. (2020), Chen et al. (2021), Yang et al. (2024, 2025), Zhang and Dong (2026), etc.

Later in this article, we propose a hysteretic PAR model with thresholds (HPART), in which a hysteretic regime is introduced into the PAR. We also explore the BPART model further. On the one hand, the HPART model and the BPART model are similar: besides being piecewise linear, they initially divide the state space into three regimes: an upper regime, a lower regime, and an 'intermediary' hysteretic/buffer regime before ending up with just the first two regimes. On the other hand, the regime-switching mechanisms

of the Poisson intensity process for data within the hysteretic/buffer regime are quite different as we shall see in Subsection 1.3. Briefly, in the HPART model, the mechanism involves an explicitly defined controlling factor that is based on the first difference of the two immediately past observations as the control factor, as in model (A1) mentioned above. (The controlling factor can be a much more general function, including, e.g., a function of a covariate time series of either count data or non-count data.) However, the BPART model does not involve such a controlling factor.

To distinguish between these two mechanisms, we develop tests for non-nested models to detect departure from a BPART/HPART model in the direction of a HPART/BPART model, in the context of separate families of hypotheses initiated by Cox (1960, 1962). See also Cox (2013).

1.3 Two nonlinear mechanisms

A count time series $\{y_t : t \in \mathbb{Z}\}$ is said to follow a parametrically stochastic dynamic Poisson model, if y_t conditionally follows a Poisson distribution, namely

$$\mathcal{L}(y_t \mid \mathcal{F}_{t-1}) = \text{Poisson}(\lambda_t), \quad t \in \mathbb{Z},$$

where $\mathcal{F}_{t-1}$ is a sequence of σ-fields generated by $(y_j : j \leq t-1)$, and the intensity process $\lambda_t = g_\theta(y_{t-1}, \lambda_{t-1}, y_{t-2}, \lambda_{t-2}, ...)$ is a positive function

depending on parameter θ and is measurable with respect to $\mathcal{F}_{t-1}$ for all $t \in \mathbb{Z}$, i.e., $\lambda_t \in \mathcal{F}_{t-1}$.

(a) A count time series $\{y_t : t \in \mathbb{Z}\}$ is said to follow a first-order BPART model, if the intensity process λ_t satisfies the stochastic recurrence

$$\lambda_t = \begin{cases} \omega_1 + \alpha_1 y_{t-1} + \beta_1 \lambda_{t-1}, & \text{if } R_t = 1, \\ \omega_2 + \alpha_2 y_{t-1} + \beta_2 \lambda_{t-1}, & \text{if } R_t = 0, \end{cases} \qquad R_t = \begin{cases} 1, & \text{if } y_{t-1} \le r, \\ 0, & \text{if } y_{t-1} > s, \\ R_{t-1}, & \text{otherwise,} \end{cases} \tag{1.1}$$

where $\omega_i > 0, \alpha_i \ge 0, \beta_i \ge 0$, $i = 1, 2$, are the coefficients, r, s are non-negative integers called thresholds with $r \le s$, and $\{R_t\}$ is a $\{0, 1\}$-valued stochastic sequence with $R_t \in \mathcal{F}_{t-1}$. Clearly, from (1.1), it follows the recurrence relation

$$R_t = I(y_{t-1} \le r) + I(r < y_{t-1} \le s) R_{t-1}, \quad t \in \mathbb{Z},$$

where $I(\cdot)$ is the indicator function.

(b) A count time series $\{y_t : t \in \mathbb{Z}\}$ is said to follow a first-order HPART model, if the intensity process λ_t satisfies the stochastic recurrence

$$\lambda_t = \begin{cases} \omega_1 + \alpha_1 y_{t-1} + \beta_1 \lambda_{t-1}, & \text{if } \{\Delta y_{t-1} \ge c, y_{t-1} \le s\} \text{ or } \{\Delta y_{t-1} < c, y_{t-1} \le r\}, \\ \omega_2 + \alpha_2 y_{t-1} + \beta_2 \lambda_{t-1}, & \text{if } \{\Delta y_{t-1} \ge c, y_{t-1} > s\} \text{ or } \{\Delta y_{t-1} < c, y_{t-1} > r\}, \end{cases} \tag{1.2}$$

where $\Delta y_{t-1} = y_{t-1} - y_{t-2}$, $\omega_i > 0$, $\alpha_i \geq 0$, $\beta_i \geq 0$, $i = 1, 2$, are the coefficients, r, s, c are integers called thresholds with $0 \leq r \leq s$ and $c \in \mathbb{Z}$. For the HPART model, we have the following recurrence-free relations

$$\begin{aligned} I_t :=& I(\Delta y_{t-1} \geq c, y_{t-1} \leq s) + I(\Delta y_{t-1} < c, y_{t-1} \leq r) \\ =& I(y_{t-1} \leq r) + I(r < y_{t-1} \leq s) I(\Delta y_{t-1} \geq c). \end{aligned}$$

It is interesting to compare the role of $I(\Delta y_{t-1} \geq c)$ in the HPART model to that of R_{t-1} in the BPART model. (i) The HPART model switches data in the hysteretic regime $\{r < y_{t-1} \leq s\}$ at different thresholds: at r if $\Delta y_{t-1} < c$, but at s if $\Delta y_{t-1} \geq c$. (ii) The BPART model does not switch regime for data in the buffer regime $\{r < y_{t-1} \leq s\}$. (iii) While Δy_{t-1} for the HPART only depends on $\{y_{t-1}, y_{t-2}\}$, i.e. a 2-finite past, R_t for the BPART depends on $\{R_{t-i}, i > 1\}$, i.e. $\{y_{t-i}, i > 1\}$, an infinite past.

1.4 Main contributions

There are five main contributions.

(1) Unlike existing practice in the literature, our approach is motivated by the genuinely and physically meaningful notion of hysteresis in which a controlling factor plays an essential role; we propose a new nonlinear mechanism, the HPART model, for the intensity process of the parametric dynamic stochastic Poisson models.

(2) We have probed deeper into the dynamics that underpin the hysteretic model and the buffered model.

(3) For these two models, we discuss, in a unified manner and with weaker conditions than in the existing literature. Specifically, we weaken the condition $\alpha_1 < 1$ to α_1 being bounded, a condition in connection with proving consistency and asymptotic normality of maximum likelihood estimators.

(4) As the two models are separate in the sense of Cox (1960, 1962), we develop two separate tests for a BPART/HPART model vs. a HPART/BPART model and derive their respective limiting distributions. Specifically, we indicate that the test statistic under the BPART model can be represented as a function of multiple chi-square distributions, while that under the HPART model approximately follows a chi-square distribution. We validate our results through Monte Carlo simulations, which are presented in the Supplementary Material.

(5) In the real data examples we have introduced a novel idea based on identity cards which allows us to construct binary sequences to help decide if the two fitted models have the same mechanism.

1.5 Outline

The remainder of the article is organized as follows. For the BPART model and the HPART model, Section 2 considers, in a unified manner, the maximum likelihood estimation of the first-order case and establishes their asymptotic properties under general conditions. Section 3 studies separate families of hypotheses and develops two tests for a BPART/HPART model vs. a HPART/BPART model. Section 4 summarizes the simulation results. Section 5 showcases two real applications. Section 6 concludes. Simulation study results and proofs of all theorems with some technical lemmas are relegated to the Supplementary Material.

2. Maximum Likelihood Estimation

Let $\vartheta = (\omega_1, \alpha_1, \beta_1, \omega_2, \alpha_2, \beta_2)^T$ and $\theta^{\mathrm{b}} = (\vartheta^T, r, s)^T$ and $\theta^{\mathrm{h}} = (\vartheta^T, r, s, c)^T$, $\tau_{\mathrm{b}} = (r, s)^T \in \mathbb{Z}^2$ and $\tau_{\mathrm{h}} = (r, s, c)^T \in \mathbb{Z}^3$. Supposing that the observations $\{y_1, \ldots, y_n\}$ are available with the true parameter θ_0. The conditional log-likelihood function is given by

$$\widetilde{L}_n(\theta) = \sum_{t=1}^{n} \tilde{\ell}_t(\theta) \quad \text{with} \quad \tilde{\ell}_t(\theta) = -\widetilde{\lambda}_t(\theta) + y_t \log \widetilde{\lambda}_t(\theta) - \log(y_t!).$$

The maximum likelihood estimator (MLE) of θ_0 is defined as

$$\widehat{\theta}_n = \underset{\theta \in \Theta_\vartheta \times \Theta_\tau}{\arg\max}\, \widetilde{L}_n(\theta),$$

where Θ_ϑ is the coefficient parameter space, and Θ_τ is the threshold parameter space.

For the BPART model,

$$\widetilde{\lambda}_t(\theta) = \{\omega_1 + \alpha_1 y_{t-1} + \beta_1 \widetilde{\lambda}_{t-1}(\theta)\}\widetilde{R}_t(\theta) + \{\omega_2 + \alpha_2 y_{t-1} + \beta_2 \widetilde{\lambda}_{t-1}(\theta)\}[1 - \widetilde{R}_t(\theta)],$$

$$\widetilde{R}_t(\theta) = I(y_{t-1} \le r) + I(r < y_{t-1} \le s)\widetilde{R}_{t-1}(\theta), \qquad t \ge 1,$$

where the initial value $(y_0, \widetilde{\lambda}_0(\theta), \widetilde{R}_0(\theta))^T$ is fixed.

For the HPART model,

$$\widetilde{\lambda}_t(\theta) = \{\omega_1 + \alpha_1 y_{t-1} + \beta_1 \widetilde{\lambda}_{t-1}(\theta)\}I_t(\theta) + \{\omega_2 + \alpha_2 y_{t-1} + \beta_2 \widetilde{\lambda}_{t-1}(\theta)\}[1 - I_t(\theta)],$$

$$I_t(\theta) = I(\Delta y_{t-1} \ge c)I(y_{t-1} \le s) + I(\Delta y_{t-1} < c)I(y_{t-1} \le r), \qquad t \ge 1,$$

where the initial value $(y_0, \widetilde{\lambda}_0(\theta), \Delta y_0)^T$ is fixed.

To facilitate the study on the asymptotic properties of $\widehat{\theta}_n$, we define the theoretical log-likelihood function as

$$L_n(\theta) = \sum_{t=1}^{n} \ell_t(\theta) \quad \text{with} \quad \ell_t(\theta) = -\lambda_t(\theta) + y_t \log \lambda_t(\theta) - \log(y_t!),$$

where, for the BPART model,

$$\lambda_t(\theta) = \{\omega_1 + \alpha_1 y_{t-1} + \beta_1 \lambda_{t-1}(\theta)\}R_t(\theta) + \{\omega_2 + \alpha_2 y_{t-1} + \beta_2 \lambda_{t-1}(\theta)\}[1 - R_t(\theta)],$$

$$R_t(\theta) = I(y_{t-1} \le r) + I(r < y_{t-1} \le s)R_{t-1}(\theta), \qquad t \in \mathbb{Z}.$$

and for the HPART model,

$$\lambda_t(\theta) = \{\omega_1 + \alpha_1 y_{t-1} + \beta_1 \lambda_{t-1}(\theta)\} I_t(\theta) + \{\omega_2 + \alpha_2 y_{t-1} + \beta_2 \lambda_{t-1}(\theta)\}[1 - I_t(\theta)],$$

$$I_t(\theta) = I(\Delta y_{t-1} \geq c) I(y_{t-1} \leq s) + I(\Delta y_{t-1} < c) I(y_{t-1} \leq r), \qquad t \in \mathbb{Z}.$$

To study the asymptotic properties of $\widehat{\theta}_n$, we need some assumptions, which are standard in the study of count time series.

Assumption 1. $\Theta_\vartheta \subseteq \{\vartheta \in \mathbb{R}^6 | \omega_i > 0, \alpha_i > 0, 0 < \beta_i < 1, \alpha_2 + \beta_2 < 1, i = 1, 2\}$ is compact. Θ_τ is bounded, where $\Theta_\tau := \Theta_{\tau_{\mathrm{b}}} = \{(r, s)^T \in \mathbb{Z}^2 |\ 0 \leq r < s\}$ for the BPART model, and $\Theta_\tau := \Theta_{\tau_{\mathrm{h}}} = \{(r, s, c)^T \in \mathbb{Z}^3 |\ 0 \leq r < s\}$ for the HPART model.

Assumption 2. ϑ_0 satisfies $(\omega_{10}, \alpha_{10}, \beta_{10})^T \neq (\omega_{20}, \alpha_{20}, \beta_{20})^T$.

Assumption 3. $\{y_t\}$ is strictly stationary and ergodic.

Assumption 4. ϑ_0 is an interior point of Θ_ϑ.

Remark 1. Assumption 3 is adopted immediately here. Indeed, it can be proved under some mild conditions via weakly Feller and e-chain techniques. More detailed and sophisticated analyses remain necessary. We leave this topic for future research.

Theorem 1. *Under Assumptions 1-3, then* $\widehat{\theta}_n \to \theta_0$ *a.s. as* $n \to \infty$.

Since thresholds r_0, s_0 and c_0 are integer-valued, the consistency of $\widehat{\tau}_n$ implies that $\widehat{r}_n = r_0$, $\widehat{s}_n = s_0$ and $\widehat{c}_n = c_0$ eventually. Thus, without loss of generality, we can assume that the parameter τ is known when we study the limiting distribution of the MLE $\widehat{\vartheta}_n$.

Theorem 2. *Under Assumptions 1-4, then* $\sqrt{n}\,(\widehat{\vartheta}_n - \vartheta_0) \xrightarrow{d} N(0, G^{-1})$, *where* $\xrightarrow{d}$ *stands for convergence in distribution and*

$$G = E\Big\{\frac{1}{\lambda_t(\theta_0)}\frac{\partial\lambda_t(\theta_0)}{\partial\vartheta}\frac{\partial\lambda_t(\theta_0)}{\partial\vartheta^T}\Big\}. \tag{2.3}$$

In order to construct a confidence region of the parameter ϑ_0, it is necessary to estimate the matrix G. A weakly consistent estimate is

$$\widehat{G} = \frac{1}{n}\sum_{t=1}^{n}\frac{1}{\widetilde{\lambda}_t(\theta)}\frac{\partial\widetilde{\lambda}_t(\theta)}{\partial\vartheta}\frac{\partial\widetilde{\lambda}_t(\theta)}{\partial\vartheta^T}\Big|_{\theta=\widehat{\theta}_n}.$$

3. Tests

From models (1.1) and (1.2), we consider a coupled Poisson model with a compound intensity process, namely $\mathcal{L}(y_t|\mathcal{F}_{t-1}) = \text{Poisson}(\lambda_t)$, where λ_t is a function of a weight factor $\delta \in [0,1]$ and the parameter $\theta = (\vartheta^T, r, s, c)^T$, defined as

$$\lambda_t = (1-\delta)\lambda_t^{\text{b}} + \delta\lambda_t^{\text{h}}, \quad t \in \mathbb{Z}. \tag{3.4}$$

3.1 Test under a BPART model

The test "$H_0 : \delta = 0$ against $H_a : \delta \neq 0$ " tests the departure of a BPART model in the direction of model (3.4).

The conditional log-likelihood function is given by

$$\widetilde{L}_n(\delta,\theta) = \sum_{t=1}^{n} \tilde{\ell}_t(\delta,\theta) \quad \text{with} \quad \tilde{\ell}_t(\delta,\theta) = -\widetilde{\lambda}_t(\delta,\theta) + y_t \log \widetilde{\lambda}_t(\delta,\theta) - \log(y_t!),$$

$$\widetilde{\lambda}_t(\delta,\theta) = (1-\delta)\widetilde{\lambda}_t^{\mathrm{b}}(\theta) + \delta\widetilde{\lambda}_t^{\mathrm{h}}(\theta).$$

Then, we get the first and second derivatives for δ as below

$$\frac{\partial \widetilde{L}_n(\delta,\theta)}{\partial \delta} = \sum_{t=1}^{n} \Big(\frac{y_t}{\widetilde{\lambda}_t(\delta,\theta)} - 1 \Big) \{\widetilde{\lambda}_t^{\mathrm{h}}(\theta) - \widetilde{\lambda}_t^{\mathrm{b}}(\theta)\},$$

$$\frac{\partial^2 \widetilde{L}_n(\delta,\theta)}{\partial \delta^2} = -\sum_{t=1}^{n} \frac{y_t}{\widetilde{\lambda}_t^2(\delta,\theta)} \{\widetilde{\lambda}_t^{\mathrm{h}}(\theta) - \widetilde{\lambda}_t^{\mathrm{b}}(\theta)\}^2.$$

In model (1.1), the threshold c is a nuisance parameter. Under H_0, we fix $c \in \mathbb{Z}$ and the score function and the information matrix are respectively

$$\frac{\partial \widetilde{L}_n(0,\theta)}{\partial \delta} = \sum_{t=1}^{n} \Big(\frac{y_t}{\widetilde{\lambda}_t^{\mathrm{b}}(\theta)} - 1 \Big) \{\widetilde{\lambda}_t^{\mathrm{h}}(\theta) - \widetilde{\lambda}_t^{\mathrm{b}}(\theta)\},$$

$$\frac{\partial^2 \widetilde{L}_n(0,\theta)}{\partial \delta^2} = -\sum_{t=1}^{n} \frac{y_t}{\{\widetilde{\lambda}_t^{\mathrm{b}}(\theta)\}^2} \big\{\widetilde{\lambda}_t^{\mathrm{h}}(\theta) - \widetilde{\lambda}_t^{\mathrm{b}}(\theta)\big\}^2.$$

For a given c, let

$$T_n^{\mathrm{b}}(c) = \Big\{ -\frac{\partial^2 \widetilde{L}_n(0,\theta^{\mathrm{b}},c)}{\partial \delta^2} \Big\}^{-1} \Big\{ \frac{\partial \widetilde{L}_n(0,\theta^{\mathrm{b}},c)}{\partial \delta} \Big\}^2 \Big|_{\theta^{\mathrm{b}}=\widehat{\theta}_n^{\mathrm{b}}}.$$

Assume the range of c is $\{c_1, c_2, ..., c_k\}$. Denote $[k] = \{1, ..., k\}$ and

$$\sigma_1(c_i, c_j) = E\Big[\frac{1}{\lambda_t^{\rm b}(\theta_0)}\{\lambda_t^{\rm h}(\vartheta_0, r_0, s_0, c_i) - \lambda_t^{\rm b}(\theta_0)\}\{\lambda_t^{\rm h}(\vartheta_0, r_0, s_0, c_j) - \lambda_t^{\rm b}(\theta_0)\}\Big],$$

$$\sigma_2(c_i, c_j) = \sigma_1(c_i, c_j) - E\Big[\frac{1}{\lambda_t^{\rm b}(\theta_0)}\{\lambda_t^{\rm h}(\vartheta_0, r_0, s_0, c_i) - \lambda_t^{\rm b}(\theta_0)\}\frac{\partial \lambda_t^{\rm b}(\theta_0)}{\partial \vartheta^T}\Big]$$
$$\times \Sigma_1^{-1} E\Big[\frac{1}{\lambda_t^{\rm b}(\theta_0)}\{\lambda_t^{\rm h}(\vartheta_0, r_0, s_0, c_j) - \lambda_t^{\rm b}(\theta_0)\}\frac{\partial \lambda_t^{\rm b}(\theta_0)}{\partial \vartheta}\Big], \quad i, j \in [k],$$

where $\Sigma_1 = G\big|_{\lambda_t = \lambda_t^{\rm b}}$ and G is defined in (2.3).

The score-based test statistic for testing BPART model is defined as $S_n = \max_{i\in[k]} T_n^{\rm b}(c_i)$.

Theorem 3. *Under H_0, if Assumptions 1-4 hold, then*

$$S_n \xrightarrow{d} \max_{i\in[k]} \frac{Z^2(c_i)}{\sigma_1(c_i, c_i)},$$

where $(Z(c_1), ..., Z(c_k))^T \sim N(0, \Sigma)$ with $\Sigma = (\sigma_2(c_i, c_j))_{k\times k}$.

3.2 Test under a HPART model

The test "$\widetilde{H}_0 : \delta = 1$ against $\widetilde{H}_a : \delta \neq 1$" tests the departure of a HPART model in the direction of model (3.4). Under $\widetilde{H}_0$, we obtain the score function and information matrix as follows

$$\frac{\partial \widetilde{L}_n(1, \theta)}{\partial \delta} = \sum_{t=1}^{n}\Big(\frac{y_t}{\widetilde{\lambda}_t^{\rm h}(\theta)} - 1\Big)\{\widetilde{\lambda}_t^{\rm h}(\theta) - \widetilde{\lambda}_t^{\rm b}(\theta)\},$$

$$\frac{\partial^2 \widetilde{L}_n(1, \theta)}{\partial \delta^2} = -\sum_{t=1}^{n}\frac{y_t}{\{\widetilde{\lambda}_t^{\rm h}(\theta)\}^2}\{\widetilde{\lambda}_t^{\rm h}(\theta) - \widetilde{\lambda}_t^{\rm b}(\theta)\}^2.$$

The score-based test statistic under $\widetilde{H}_0$ is

$$T_n^{\mathrm{h}} = \Big\{ -\frac{\partial^2 \widetilde{L}_n(1,\theta)}{\partial \delta^2} \Big\}^{-1} \Big\{ \frac{\partial \widetilde{L}_n(1,\theta)}{\partial \delta} \Big\}^2 \Big|_{\theta=\widehat{\theta}_n^{\mathrm{h}}}$$

Similarly, denote

$$\sigma_1' = E\Big\{ -\frac{1}{n}\frac{\partial^2 L\,(1,\theta_0)}{\partial \delta^2} \Big\} = E\Big[\frac{1}{\lambda_t^{\mathrm{h}}(\theta_0)} \{\lambda_t^{\mathrm{h}}(\theta_0) - \lambda_t^{\mathrm{b}}(\theta_0)\}^2 \Big],$$

$$\sigma_2' = \sigma_1' - E\Big[\frac{1}{\lambda_t^{\mathrm{h}}(\theta_0)} \{\lambda_t^{\mathrm{h}}(\theta_0) - \lambda_t^{\mathrm{b}}(\theta_0)\} \frac{\partial \lambda_t^{\mathrm{h}}\,(\theta_0)}{\partial \vartheta^T} \Big] \Sigma_2^{-1} E\Big[\frac{1}{\lambda_t^{\mathrm{h}}(\theta_0)} \{\lambda_t^{\mathrm{h}}(\theta_0) - \lambda_t^{\mathrm{b}}(\theta_0)\} \frac{\partial \lambda_t^{\mathrm{h}}\,(\theta_0)}{\partial \vartheta} \Big],$$

where $\Sigma_2 = G\big|_{\lambda_t=\lambda_t^{\mathrm{h}}}$ and G is defined in (2.3).

Theorem 4. *Under $\widetilde{H}_0$, if Assumptions 1-4 hold, then $(\hat{\sigma}_1'/\hat{\sigma}_2')T_n^{\mathrm{h}} \xrightarrow{d} \chi_1^2$.*

4. Simulation Studies

The detailed simulation results and analysis are reported in the Supplementary Material. A brief summary is as follows.

4.1 Performance of estimators

To evaluate the finite-sample performance of $\widehat{\theta}_n$, we conducted Monte Carlo simulations with $n = 500$, 1000, and 2000. The threshold bounds are determined using the γ_i-quantiles of $\{y_t\}_{t=1}^n$ and $\{\Delta y_t\}_{t=2}^n$ with $\gamma_1 = 0.2$ and $\gamma_2 = 0.8$. The results show that $\widehat{\tau}_n$ generally converges to τ_0, but the speed of convergence depends on the parameter values. In some cases, $\widehat{\tau}_n$

hits τ_0 exactly for small n, while in other cases, it fails to achieve perfect estimation even for large n.

The asymptotic behavior of the estimators aligns with expectations, but the intercept parameter ω_2 shows a larger variance-to-mean ratio compared to other parameters, a phenomenon observed in similar models like PAR and SETPAR. So far, no explanation has been given in the literature.

4.2 Performance of tests

We evaluate the performance of the tests, where observations are generated using the same true parameters as above, except for setting $c_0 = -1, 0, 1$ for the HPART model. We take n = 500, 1000, 2000, and 4000, and the significance levels are 0.1, 0.05, and 0.01. Critical values for testing H_0 are simulated with 20,000 replications.

The results summarize the size and power based on 1000 replications. As n increases, the size and power of the test approach the desired values. For example, Table S.6 in the Supplementary Material shows satisfactory results in terms of power. However, Table S.5 exhibits less satisfactory power behavior before n = 4000, even though the size is well-matched. Additionally, the value of c_0 affects the power of the test. The results in Table S.7 and Table S.8 for testing $\widetilde{H}_0$ generally perform better than those

for testing H_0.

5. Applications to Real Data

5.1 Escape custody Data

To illustrate the efficacy of the two models and the tests, as well as to gain further insights, we analyse the monthly number of escape custody in the New South Wales during the period from 2010 to 2024. The data, available at https://bocsar.nsw.gov.au, consist of 180 observations and are plotted in Fig. 2. The sample mean and variance are 15.17 and 25.93, respectively, suggesting some degree of over-dispersion in the data. The data also exhibit positive serial dependence, as illustrated in Fig. 3.

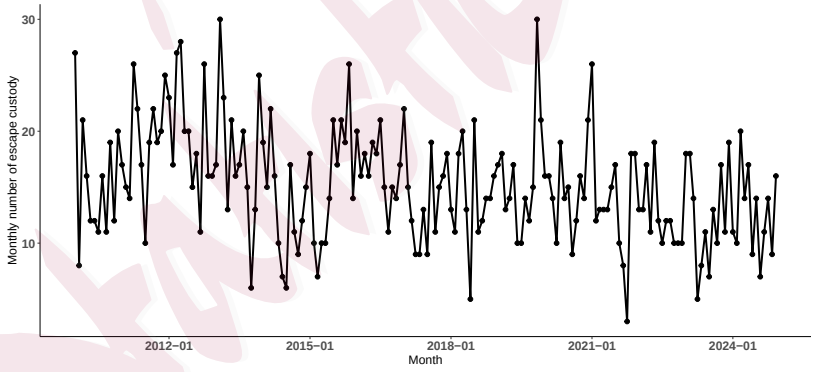

Figure 2: The monthly number of escape custody during 2010-2024.

We use the first 160 data to fit models and the last 20 data to perform

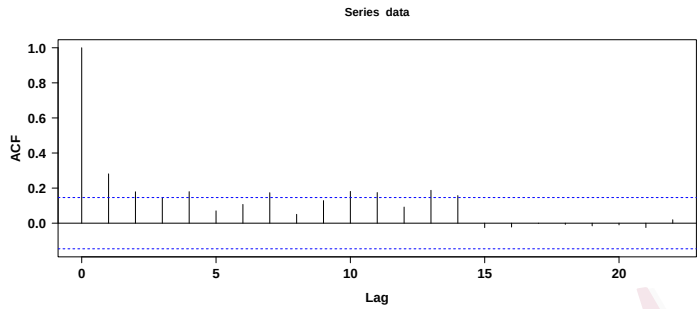

Figure 3: ACF of the escape custody data.

rolling out-of-sample one-step-ahead predictions. The fitted PAR model is

$$\mathcal{L}(y_t \mid \mathcal{F}_{t-1}) = \text{Poisson}(\lambda_t),$$

$$\lambda_t = 2.42_{(1.42)} + 0.16_{(0.05)} y_{t-1} + 0.68_{(0.12)} \lambda_{t-1},$$

where numbers in parentheses are standard deviations. On noticing that the upward spikes tend to be steeper than the downward spikes, a SETPAR model might be worthy of consideration. The fitted SETPAR model is

$$\mathcal{L}(y_t \mid \mathcal{F}_{t-1}) = \text{Poisson}(\lambda_t),$$

$$\lambda_t = \begin{cases} \underset{(6.35)}{3.45} \; \underset{(0.26)}{+ 0.04 y_{t-1}} \; \underset{(0.39)}{+ 0.70 \lambda_{t-1}}, & \text{if } \; y_{t-1} \leq 11, \\ \underset{(1.66)}{1.61} \; \underset{(0.08)}{+ 0.20 y_{t-1}} \; \underset{(0.14)}{+ 0.69 \lambda_{t-1}}, & \text{if } \; y_{t-1} > 11. \end{cases}$$

Noticing that much more data lie within the middle range, we investigate the possibility of a BPART/HPART model. First, consider tests

between the BPART model and the HPART model.

$$H_0: \quad \mathcal{L}(y_t \mid \mathcal{F}_{t-1}) = \text{Poisson}(\lambda_t),$$

$$\lambda_t = \begin{cases} \underset{(4.40)}{0.03} + \underset{(0.11)}{0.24} y_{t-1} + \underset{(0.32)}{0.82} \lambda_{t-1}, & \text{if } R_t = 1, \\ \underset{(1.66)}{0.007} + \underset{(0.08)}{0.27} y_{t-1} + \underset{(0.12)}{0.69} \lambda_{t-1}, & \text{if } R_t = 0. \end{cases} \qquad R_t = \begin{cases} 1, & \text{if } y_{t-1} \le 11, \\ 0, & \text{if } y_{t-1} > 16, \\ R_{t-1}, & \text{otherwise.} \end{cases} \tag{5.5}$$

$$\widetilde{H}_0: \quad \mathcal{L}(y_t \mid \mathcal{F}_{t-1}) = \text{Poisson}(\lambda_t),$$

$$\lambda_t = \begin{cases} \underset{(3.46)}{1.02} + \underset{(0.12)}{0.24} y_{t-1} + \underset{(0.26)}{0.75} \lambda_{t-1}, & \text{if } \{\Delta y_{t-1} \ge -2, y_{t-1} \le 16\} \\ & \quad \text{or } \{\Delta y_{t-1} < -2, y_{t-1} \le 11\}, \\ \underset{(2.63)}{0.02} + \underset{(0.10)}{0.39} y_{t-1} + \underset{(0.17)}{0.53} \lambda_{t-1}, & \text{if } \{\Delta y_{t-1} \ge -2, y_{t-1} > 16\} \\ & \quad \text{or } \{\Delta y_{t-1} < -2, y_{t-1} > 11\}. \end{cases} \tag{5.6}$$

The AIC values and Pearson residual analysis indicate that all fitted models are adequate (See Subsection S2.1 in the Supplementary Material).

Before doing tests, a preliminary examination of the four models reveals the following features: (i) the four models share a more or less similar coefficient for λ_{t-1}, implying that the difference of their λ_t values is determined to a large extent by the combined effect of the intercept and y_{t-1}; (ii) on

setting y_{t-1} at its sample mean, 15.17, the combined effect is roughly 4.8 for the PAR model; (iii) for the SETPAR model, similar calculations suggest that the combined effect (now regime dependent) is roughly 4.0 for the lower regime with sample mean 13.64, and 4.8 for the upper regime with sample mean 15.71. Since there are much fewer data points in the lower regime, the combined effect is roughly the same as for the PAR model; (iv) for the BPART model, the intercepts are negligible, implying that the key influencer is y_{t-1} with a similar coefficient in the two regimes, leading to combined effects of roughly 3.4 for the lower regime with sample mean 14.19, and 4.3 for the upper regime with sample mean 15.82, but the number 3.4 cannot be dismissed because the buffer regime recruits more data points into the lower regime than it is the case for the SETPAR model; (v) for the HPART model, similar calculations lead to combined effects of 4.5 for the lower regime with the sample mean 14.35, and 6.2 for the upper regime with the sample mean 15.93; (vi) overall, the combined effects on the PAR model and the SETPAR model are similar, but for the BPART model and the HPART model, the combined effects are clearly regime dependent, affecting the two models quite differently.

Now, from Table 1, when testing H_0, we do not reject model (5.5) at each of the three levels. When testing $\widetilde{H}_0$, again we do not reject the model

(5.6) at each of the three levels. These test results mean that we are not in a position to prefer one model to the other as far as model fitting is concerned.

Table 1: Results on the tests for models (5.5) and (5.6).

	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
Test (5.5)	Not rejected	Not rejected	Not rejected
Test (5.6)	Not rejected	Not rejected	Not rejected

In the BPART model, the intercepts ($\omega_1 \approx 0$, $\omega_2 \approx 0$) represent the baseline escape rate when past counts and past conditional means are negligible. The higher value of $\beta_1 = 0.82$ in the lower regime suggests that the process remains at a low level for a longer period, reflecting a more stable and persistent pattern. In contrast, a smaller β_2 than β_1 implies that once a large number is observed, the conditional mean of the process decreases more quickly compared to the lower regime. This suggests that the monthly number of escape custody cases is unlikely to increase unduly reflecting a relatively stable situation. The number of escape custody cases reached its minimum of 3 only in one month, namely October 2021, and the maximum value of 30 was observed only in two months, namely February 2013 and November 2021. In fact, although the maximum minus the minimum is as large as 27, the 80th percentile (at 19) minus the 20th percentile (at 11)

is only 8. Therefore, the monthly counts exhibit a relatively stable and remain mostly within a mid-range over time. The HPART model leads generally to a similar conclusion, with minor variations: (i) $\omega_1 = 1.02$ indicates that a small number of escape custody cases in a given month tends to lead to an increased number of cases in the subsequent month. (ii) The relatively small magnitude of c suggests that within the range $(11, 16)$, the fluctuations in y_t are mild.

For further comparison, we turn to out-of-sample prediction. The results are given in Table 2, which includes additionally the PAR model and the SETPAR model for the sake of comparison. According to both the mean square error (MSE) and the mean absolute error (MAE), the HPART model outperforms the PAR model by about 3.7% and 2% respectively and the BPART model less so. This result reveals that incorporating a hysteretic/buffer regime is potentially beneficial, to a more or less extent, within the context of nonlinearity and prediction performance.

Table 2: Summary of prediction error

	PAR	SETPAR	BPART	HPART
MSE	16.89	17.39	16.80	16.27
MAE	3.55	3.60	3.50	3.47

Since the fitted HPART model and the fitted BPART model are very

close in terms of model parameters and prediction performance, it is relevant to probe deeper. Notice that both models allocate each datum to one of two regimes. Let us label them regime 1 (the upper regime) and regime 0 (the lower regime). So, each datum carries an identity (ID) card that bears either a 1 marking or an 0 marking. Of the first 160 data, 60 of them are in the same buffer/hysteretic regime and the ID cards of 46 of them bear the same 1 or 0 marking. That is to say a high proportion of the data in the buffer regime of the BPART model is assigned to the same upper or lower regime as data in the hysteretic regime of the HPART model. The high overlap hints at the possibility that the controlling factor of the HPART model could effectively be the mechanism of the BPART model governing the latter's buffer regime. This hint is consistent with the test results based on separate families of hypotheses.

To probe further, we plot the binary sequence of datum's ID cards generated by the two models separately (i.e. R_t) in Fig. 4. Table 3 presents the contingency table of ID cards derived from the two models. Using the BIC (Katz, 1981), the ID sequences of both models exhibit first-order Markov chain properties. Furthermore, a likelihood ratio test (Anderson and Goodman, 1957; Billingsley, 1961) was conducted to examine the null hypothesis of the two binary sequences being derived from the same Markov

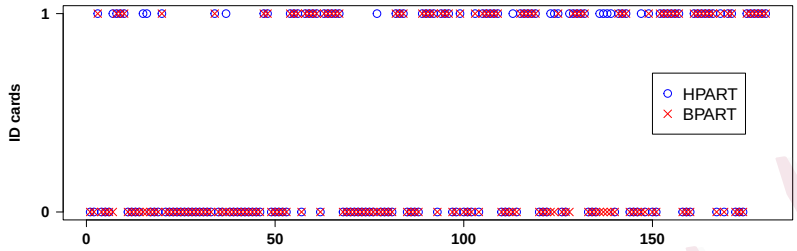

Figure 4: The sequence of datum's ID cards

chain. The test yielded a test statistic of 2.46 and a p-value of 0.29, leading to non-rejection of the null hypothesis. Therefore, within the confines of our data, the two models may be considered equivalent in terms of their buffer and hysteretic mechanisms.

Table 3: Contingency table of ID cards

	BPART:ID=1	BPART:ID=0	total
HPART:ID=1	72	14	86
HPART:ID=0	0	94	94
total	72	108	180

5.2 Hepatitis B Data

Consider the weekly number of hepatitis B cases reported in the state of Bremen from January 2023 to December 2024. We shall not repeat reasons for fitting the various models. Now, the data consist of 104 observations,

available at https://survstat.rki.de. The data are plotted in Fig. 5. The sample mean and variance are 6.99 and 12.24, respectively, suggesting mild over-dispersion in the data. Although the ACF in Fig. 6 does not show significance, the BDS test with smaller epsilon values reveals significant nonlinear dependence (p-values < 0.01) at different dimensions (e.g., 3- and 4-dim). This result suggests the presence of nonlinear structure in the data.

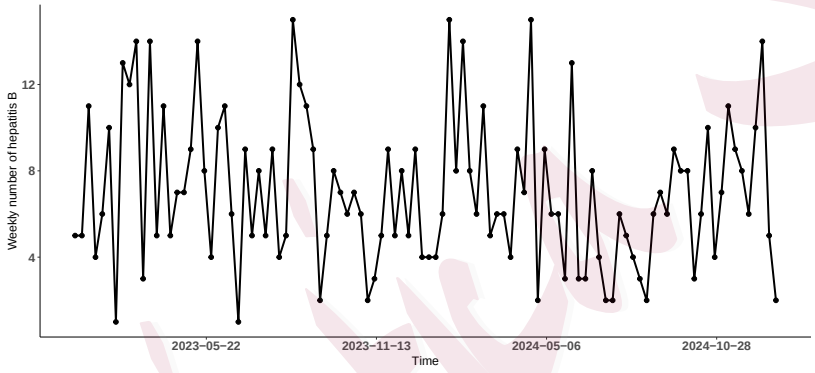

Figure 5: The weekly number of hepatitis B during 2023-2024.

We use the first 94 data to fit models and the last 10 data for rolling out-of-sample one-step-ahead predictions. The fitted PAR model is

$$\mathcal{L}(y_t \mid \mathcal{F}_{t-1}) = \text{Poisson}(\lambda_t),$$

$$\lambda_t = 0.02_{(0.38)} + 0.001_{(0.04)} y_{t-1} + 0.99_{(0.09)} \lambda_{t-1}.$$

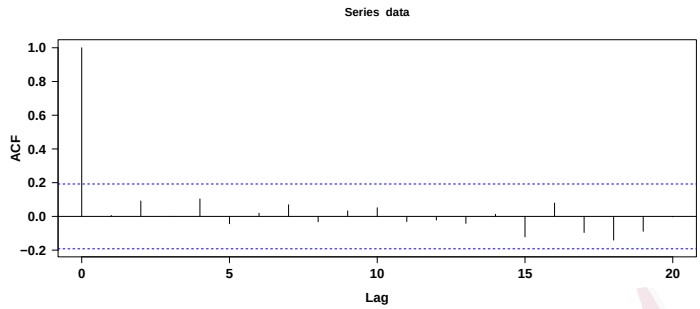

Figure 6: ACF of the Hepatitis B data.

The fitted SETPAR model is

$$\mathcal{L}(y_t \mid \mathcal{F}_{t-1}) = \text{Poisson}(\lambda_t),$$

$$\lambda_t = \begin{cases} \underset{(5.71)}{0.01} + \underset{(0.43)}{0.003} y_{t-1} + \underset{(0.84)}{0.92} \lambda_{t-1}, & \text{if } y_{t-1} \le 4, \\ \underset{(3.76)}{3.63} + \underset{(0.09)}{0.05} y_{t-1} + \underset{(0.54)}{0.44} \lambda_{t-1}, & \text{if } y_{t-1} > 4. \end{cases}$$

Now, consider tests between the BPART model and the HPART model.

$$H_0: \quad \mathcal{L}(y_t \mid \mathcal{F}_{t-1}) = \text{Poisson}(\lambda_t),$$

$$\lambda_t = \begin{cases} \underset{(1.51)}{0.0001} + \underset{(0.15)}{0.25} y_{t-1} + \underset{(0.23)}{0.88} \lambda_{t-1}, & \text{if } R_t = 1, \\ \underset{(1.66)}{0.001} + \underset{(0.08)}{0.08} y_{t-1} + \underset{(0.25)}{0.85} \lambda_{t-1}, & \text{if } R_t = 0. \end{cases} \qquad R_t = \begin{cases} 1, & \text{if } y_{t-1} \le 5, \\ 0, & \text{if } y_{t-1} > 7, \\ R_{t-1}, & \text{otherwise,} \end{cases} \tag{5.7}$$

$$\widetilde{H}_0: \quad \mathcal{L}(y_t \mid \mathcal{F}_{t-1}) = \text{Poisson}(\lambda_t),$$

$$\lambda_t = \begin{cases} \underset{(1.76)}{0.001} + \underset{(0.19)}{0.24} y_{t-1} + \underset{(0.25)}{0.92} \lambda_{t-1}, & \text{if } \{\Delta y_{t-1} \geq 2, y_{t-1} \leq 7\} \\ & \text{or } \{\Delta y_{t-1} < 2, y_{t-1} \leq 5\}, \\ \underset{(1.54)}{0.001} + \underset{(0.06)}{0.002} y_{t-1} + \underset{(0.23)}{0.95} \lambda_{t-1}, & \text{if } \{\Delta y_{t-1} \geq 2, y_{t-1} > 7\} \\ & \text{or } \{\Delta y_{t-1} < 2, y_{t-1} > 5\}. \end{cases} \tag{5.8}$$

The AIC values and Pearson residual analysis jointly confirm the adequacy of all fitted models (See Subsection S2.2 in the Supplementary Material).

Table 4 shows that we reach similar conclusions for the tests as in the case of the escape custody data.

Table 4: Test results

	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
Test (5.7)	Not rejected	Not rejected	Not rejected
Test (5.8)	Not rejected	Not rejected	Not rejected

Before moving forward, an initial inspection of the four models shows the following features: (i) except for the SETPAR model, the coefficient for λ_{t-1} is largely comparable across the other three models, implying that the differences in their λ_t are primarily attributable to the combined influence

of the intercept and y_{t-1}; (ii) the BPART model and the HPART model share the same $\{r, s\}$ parameters and a similar linear dynamics in the lower regime, but a different linear dynamics in the upper regime, a feature that mighy have an impact on its predication performance; (iii) the PAR model practically stands out on its own, and we might arguably consider $\lambda_t \approx \lambda_{t-1}$, in which case the PAR model might perhaps represent a steady state; (iv) the two linear dynamics of the SETPAR model bear no resemblance to those of either the BPART model or the HPART model; (v) in both the fitted BPART model and the fitted HPART model, the intercept parameters ω_1, ω_2 are very close to 0, and, for the upper regime of the HPART model, the coefficient of y_{t-1} is also close to 0 implying that $\lambda_t \approx \lambda_{t-1}$ possibly leading to a different prediction performance compared with the BPART model; (vi) similar back-of-envelope calculations give the following approximate combined effects: PAR model (0.03), SEPTAR model (0.03, 4.0), BPART model (1.8, 0.5) and HPART model (1.7, 0.01).

As for the performance of prediction, we can get the following results from Table 5. The PAR model performs the worst among the four models in the prediction exercise. This might be due to feature (iii) mentioned above. Despite feature (iv), the prediction performance of the SETPAR model is, somewhat to our surprise, identical to that of the BPART model.

Overall, the HPART model performs by far the best in the prediction exercise, substantially improving the PAR model by 20.7% and 14.8% in terms of MSE and MAE respectively. The fact that the BPART model and the HPART model are different seems to underline their difference in prediction performance.

Table 5: Summary of prediction error.

	PAR	SETPAR	BPART	HPART
MSE	13.64	11.89	11.89	10.81
MAE	2.97	2.87	2.66	2.53

The coefficient of y_{t-1} captures the short-term feedback effect of last week's realized cases: In the BPART model, this effect is stronger ($\alpha_1 = 0.25$) in the lower regime, suggesting that even small counts may propagate into the next week. In contrast, the effect is weaker ($\alpha_2 = 0.08$) in the upper regime, indicating that the process is less sensitive to additional cases once the epidemic is already elevated, possibly reflecting saturation or control measures. A high coefficient of λ_{t-1} indicates persistence of the potential outbreak intensity from the previous week. This implies that the number of hepatitis B cases has significant inertia, and last week's risk level impacts this week's risk level.

For the HPART model, $\alpha_1 = 0.24$ and $\alpha_2 = 0.002$ implying that once

the number of hepatitis B cases is high, the consitional mean of the process of next week is driven almost entirely by its previous intensity level. This means that the subsequent direct impact of the high number of hepatitis B cases is weakened, like cases on August 19, 2023 and January 17, 2024. Under high incidence condition, $\beta_2 = 0.95$ is larger, indicating that once a high incidence state is entered, the risk declines very slowly, showing a significant hysteresis effect, and it takes a long time to return to a low level. In addition, the relatively small magnitude of c and the closeness of the two thresholds, namely $r = 5$, $s = 7$, suggest that the model will switch to a high state more frequently and predict longer high-risk periods. In practical public health contexts, this setting is equivalent to a 'precautionary strategy': as long as there is a clear upward trend in cases, the alert level is raised, rather than waiting for a larger increase before switching.

Finally, among the first 94 data, 20 are in the buffer/hysteretic area and 14 of them have the same ID card 1/0 marking. This high proportion at 70% hints at the possibility that, for the hepatitis B data, the controlling factor of the HPART model could be similar to the mechanism of the BPART model governing the latter's buffer regime. This hint is consistent with the test results based on separate families of hypotheses. However, despite the possibly similar common controlling factor, unlike the previous case, for the

present case feature (v) noted above appears to have allowed the HPART model to edge over the BPART model in prediction performance.

The binary sequence of datum's ID cards generated separately by the HPART and BPART models are plotted in Fig. 7. Table 6 presents the

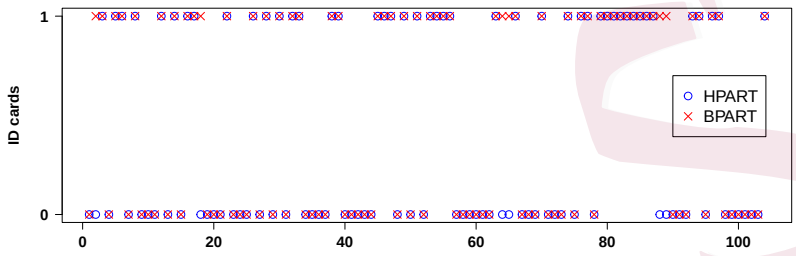

Figure 7: The sequence of datum's ID cards.

Table 6: Contingency Table of ID cards.

	BPART:ID=1	BPART:ID=0	total
HPART:ID=1	45	0	45
HPART:ID=0	6	53	59
total	51	53	104

two models' ID card contingency table. Using the BIC, both sequences are identified to have a Markov chain order of 0, which implies they can be treated as independent data. The p-value of the exact binomial test is 0.03. At the significance level of 0.05, we reject the null hypothesis that the two ID card sequences are identical. This suggests that within the confines of

our data, the two models differ with respect to their buffering and hysteretic mechanisms.

6. Conclusion and Discussion

In this article, we have proposed a new hysteretic Poisson autoregressive model with thresholds and further analysed the existing buffered Poisson autoregressive model with thresholds. We have studied the properties of the maximum likelihood estimators of parameters of both models in a unified manner. In the simulation studies presented in the Supplementary Material, we have confirmed the asymptotic results of the parameter estimates, and that the size and power of the test approach their desired values under different scenarios, as the length of observations increases. Most importantly, we have demonstrated the advantages of incorporating a hysteretic regime in the model by (i) providing a physically meaningful mechanism governing the regime-switching therein, (ii) a plausible interpretation and (iii) a better prediction performance.

Indeed, we have showcased two real-data studies to confirm the importance and effectiveness of a hysteretic regime. Further, these studies have also provided evidence in support of the view that the hysteretic mechanism of the HPART model could, in some cases, reveal the origin of the

switching mechanism operated by the BPART model in the buffer regime. Specifically, we have highlighted instances when the R_{t-1} of the BPART model is underlined by the controlling factor $I(\Delta y_{t-1} \geq c)$ of the HPART model.

The Fig. 8 gives a graphic representation of the hysteretic mechanism of the HPART model. The region 1 represents the lower regime while

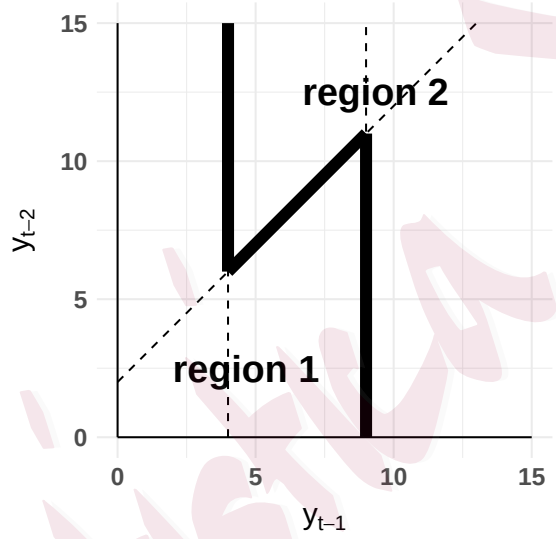

Figure 8: Schematic diagram of the hysteretic regime lying between the two vertical dotted lines.

the region 2 the upper regime. The two vertical lines represent the two thresholds r and s, while the diagonal lines represent the line of first-order difference equal to c, where c is taken as -2 for illustration. Similarly, the maximum range of y is set at 15 just for illustration. Note that the

hysteretic regime for a HPART model is neatly separated into a region 1 and a region 2. In general, for the buffer regime of a BPART model, two regions are expected to be more randomly distributed.

In a HPART model, when $r < y_{t-1} \leq s$, which linear model that y_t will follow is determined by the controlling factor $y_{t-1} - y_{t-2}$. If a BPART model and a HPART model fitted to the same time series share almost the same thresholds, then a substantial overlap of the ID cards of data in the hysteretic/buffer regime suggests that the two models might have a similar controlling factor.

Supplementary Material

The Supplementary Material contains simulation study, additional details for real data analysis results, and proofs of all theorems in the article with some useful technical lemmas.

Acknowledgements

We are grateful to the editor, associate editor, and two referees for their insightful and valuable comments and suggestions, which have helped us substantially improve the presentation and quality of our article. Li's work is supported by the Beijing Natural Science Foundation (No.F251002).

References

Ahmad, A. and C. Francq (2016). Poisson QMLE of count time series models. *J. Time Series Anal. 37*(3), 291–314.

Alzahrani, N., P. Neal, S. E. F. Spencer, T. J. McKinley, and P. Touloupou (2018). Model selection for time series of count data. *Comput. Statist. Data Anal. 122*, 33–44.

Anderson, T. W. and L. A. Goodman (1957). Statistical inference about Markov chains. *Ann. Math. Statist. 28*, 89–110.

Armillotta, M. and K. Fokianos (2023). Nonlinear network autoregression. *Ann. Statist. 51*(6), 2526–2552.

Armillotta, M. and K. Fokianos (2024). Count network autoregression. *J. Time Series Anal. 45*(4), 584–612.

Billingsley, P. (1961). Statistical methods in Markov chains. *Ann. Math. Statist. 32*, 12–40.

Brokate, M. and J. Sprekels (1996). *Hysteresis and Phase Transitions*. Springer, New York.

Chen, C. W. S., S. Lee, and K. Khamthong (2021). Bayesian inference of nonlinear hysteretic integer-valued GARCH models for disease counts. *Comput. Statist. 36*(1), 261–281.

Christou, V. and K. Fokianos (2015). Estimation and testing linearity for non-linear mixed Poisson autoregressions. *Electron. J. Stat. 9*(1), 1357–1377.

Cox, D. R. (1960). Tests of separate families of hypotheses. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I*, pp. 105–123. Univ. California Press, Berkeley-Los Angeles, Calif.

Cox, D. R. (1962). Further results on tests of separate families of hypotheses. *J. Roy. Statist. Soc. Ser. B 24*, 406–424.

Cox, D. R. (2013). A return to an old paper: 'Tests of separate families of hypotheses'. *J. R. Stat. Soc. Ser. B. Stat. Methodol. 75*(2), 207–215.

Davis, R. A., K. Fokianos, S. H. Holan, H. Joe, J. Livsey, R. Lund, V. Pipiras, and N. Ravishanker (2021). Count time series: A methodological review. *J. Amer. Statist. Assoc. 116*(535), 1533–1547.

Davis, R. A., S. H. Holan, R. Lund, and N. Ravishanker (Eds.) (2016). *Handbook of Discrete-Valued Time Series*. CRC Press, Boca Raton, FL.

Davis, R. A. and H. Liu (2016). Theory and inference for a class of nonlinear models with application to time series of counts. *Statist. Sinica 26*(4), 1673–1707.

Diop, M. L. and W. Kengne (2021). Piecewise autoregression for general integer-valued time series. *J. Statist. Plann. Inference 211*, 271–286.

Douc, R., K. Fokianos, and E. Moulines (2017). Asymptotic properties of quasi-maximum likelihood estimators in observation-driven time series models. *Electron. J. Stat. 11*(2), 2707–2740.

Doukhan, P., K. Fokianos, and J. Rynkiewicz (2021). Mixtures of nonlinear Poisson autoregressions. *J. Time Series Anal. 42*(1), 107–135.

Doukhan, P., A. Leucht, and M. H. Neumann (2022). Mixing properties of non-stationary

INGARCH(1,1) processes. *Bernoulli 28*(1), 663–688.

Ewing, J. A. (1885). Experimental researches in magnetism. *Philosophical Transactions of the Royal Society of London 176*, 523–640.

Ferland, R., A. Latour, and D. Oraichi (2006). Integer-valued GARCH process. *J. Time Ser. Anal. 27*(6), 923–942.

Fokianos, K., R. Fried, Y. Kharin, and V. Voloshko (2022). Statistical analysis of multivariate discrete-valued time series. *J. Multivariate Anal. 188*, Paper No. 104805, 15.

Fokianos, K., A. Rahbek, and D. Tjøstheim (2009). Poisson autoregression. *J. Amer. Statist. Assoc. 104*(488), 1430–1439.

Fokianos, K., B. r. Støve, D. Tjøstheim, and P. Doukhan (2020). Multivariate count autoregression. *Bernoulli 26*(1), 471–499.

Fokianos, K. and D. Tjøstheim (2011). Log-linear Poisson autoregression. *J. Multivariate Anal. 102*(3), 563–578.

Fokianos, K. and D. Tjøstheim (2012). Nonlinear Poisson autoregression. *Ann. Inst. Statist. Math. 64*(6), 1205–1225.

Heinen, A. (2003). Modelling time series count data: An autoregressive conditional Poisson model. *CORE Discussion Paper 2003/62, University of Louvain, Belgium*.

Huang, L. and M. Khabou (2023). Nonlinear Poisson autoregression and nonlinear Hawkes processes. *Stochastic Process. Appl. 161*, 201–241.

Jia, Y., S. Kechagias, J. Livsey, R. Lund, and V. Pipiras (2023). Latent Gaussian count time series. *J. Amer. Statist. Assoc. 118*(541), 596–606.

Karlis, D. and N. Mamode Khan (2023). Models for integer data. *Annu. Rev. Stat. Appl. 10*, 297–323.

Katz, R. W. (1981). On some criteria for estimating the order of a Markov chain. *Technometrics 23*(3), 243–249.

Kennedy, M. P. and L. O. Chua (1991). Hysteresis in electronic circuits: A circuit theorist's perspective. *International Journal of Circuit Theory and Applications 19*(5), 471–515.

Kong, J. and R. Lund (2023). Seasonal count time series. *J. Time Series Anal. 44*(1), 93–124.

Li, D., R. Zeng, L. Zhang, W. K. Li, and G. Li (2020). Conditional quantile estimation for hysteretic autoregressive models. *Statist. Sinica 30*(2), 809–827.

Li, G., B. Guan, W. K. Li, and P. L. H. Yu (2015). Hysteretic autoregressive time series models. *Biometrika 102*(3), 717–723.

Liu, M., Q. Li, and F. Zhu (2019). Threshold negative binomial autoregressive model. *Statistics 53*(1), 1–25.

Liu, M., Q. Li, and F. Zhu (2020). Self-excited hysteretic negative binomial autoregression. *AStA Adv. Stat. Anal. 104*(3), 385–415.

Liu, M., F. Zhu, J. Li, and C. Sun (2023). A systematic review of INGARCH models for integer-valued time series. *Entropy 25*(6), Paper No. 922, 27.

Lo, P. H., W. K. Li, P. L. H. Yu, and G. Li (2016). On buffered threshold GARCH models. *Statist. Sinica 26*(4), 1555–1567.

Morris, K. A. (2012). What is hysteresis? *Applied Mechanics Reviews 64*(5), 050801.

Neumann, M. H. (2011). Absolute regularity and ergodicity of Poisson count processes. *Bernoulli 17*(4), 1268–1284.

Rydberg, T. H. and N. Shephard (2000). Bin models for trade-by-trade data. modelling the number of trades in a fixed interval of time. *Econometric Society World Congress 2000, Contributed Papers No 0740, Econometric Society*.

Sellers, K. (2023). *The Conway-Maxwell-Poisson Distribution*. Cambridge University Press, Cambridge.

Smith, R. C. (2005). *Smart Material Systems*. SIAM, Philadelphia.

Tjøstheim, D. (2012). Some recent theory for autoregressive count time series. *TEST 21*(3), 413–438.

Tong, H. (1978). On a threshold model. *In: Chen, C.H. (Ed.), Pattern Recognition and Signal Processing, Sijthoff and Noordhoff, Amsterdam*, 575–586.

Tong, H. and K. S. Lim (1980). Threshold autoregression, limit cycles and cyclical data. *J. R. Stat. Soc. Ser. B. Stat. Methodol. 42*(3), 245–292.

Truong, B.-C., C. W. S. Chen, and S. Sriboonchitta (2017). Hysteretic Poisson INGARCH model for integer-valued time series. *Stat. Model. 17*(6), 401–422.

Wang, C., H. Liu, J.-f. Yao, R. A. Davis, and W. K. Li (2014). Self-excited threshold Poisson autoregression. *J. Amer. Statist. Assoc. 109*(506), 777–787.

Wang, D. and W. K. Li (2020). Unit root testing on buffered autoregressive model. *Statist. Sinica 30*(2), 977–1003.

Weiß, C. (2018). *An Introduction to Discrete-Valued Time Series*. Wiley, Hoboken, NJ.

Weiß, C. H. and F. Zhu (2024). Conditional-mean multiplicative operator models for count time series. *Comput. Statist. Data Anal. 191*, Paper No. 107885, 19.

Weiß, C. H., F. Zhu, and A. Hoshiyar (2022). Softplus INGARCH models. *Statist. Sinica 32*(2), 1099–1120.

Yang, K., X. Chen, H. Li, C. Xia, and X. Wang (2025). On bivariate self-exciting hysteretic integer-valued autoregressive processes. *J. Syst. Sci. Complex. 38*(5), 2204–2225.

Yang, K., X. Zhao, X. Dong, and C. H. Weiß(2024). Self-exciting hysteretic binomial autoregressive processes. *Statist. Papers 65*(3), 1197–1231.

Zeeman, E. C. (1977). *Catastrophe Theory*. Addison-Wesley Publishing Co., Reading, Mass.-London-Amsterdam. Selected papers, 1972–1977.

Zhang, R. and X. Dong (2026). An MCMC algorithm for bounded count time series hysteretic models with an application to disease infection. *J. Stat. Comput. Simul. 96*(2), 321–340.

Zhu, K., W. K. Li, and P. L. H. Yu (2017). Buffered autoregressive models with conditional heteroscedasticity: An application to exchange rates. *J. Bus. Econom. Statist. 35*(4),

528–542.

Zhu, K., P. L. H. Yu, and W. K. Li (2014). Testing for the buffered autoregressive processes. *Statist. Sinica 24* (2), 971–984.

Department of Statistics and Data Science, Tsinghua University, Beijing 100084, China.

E-mail: mxt22@mails.tsinghua.edu.cn

Department of Statistics and Data Science, Tsinghua University, Beijing 100084, China.

E-mail: malidong@tsinghua.edu.cn

Department of Statistics and Data Science, Tsinghua University, Beijing 100084, China

and

Paula and Gregory Chow Institute for Studies in Economics, Xiamen University, Xiamen, China

and

Department of Statistics, The London School of Economics and Political Science, London, U.K.

E-mail: howell.tong@gmail.com