Piecewise autoregression for general integer-valued time series

Mamadou Lamine Diop, William Kengne

Piecewise autoregression for general integer-valued time series

March 24, 2020

Mamadou Lamine DIOP[a, b 1] and William KENGNE[a 2]

[a] *THEMA, CY Cergy Paris Université, 33 Boulevard du Port, 95011 Cergy-Pontoise Cedex, France*
[b] *LERSTAD, Université Gaston Berger, BP 234 Saint-Louis, Sénégal*
E-mail: mamadou-lamine.diop@u-cergy.fr ; william.kengne@u-cergy.fr

Abstract : This paper proposes a piecewise autoregression for general integer-valued time series. The conditional mean of the process depends on a parameter which is piecewise constant over time. We derive an inference procedure based on a penalized contrast that is constructed from the Poisson quasi-maximum likelihood of the model. The consistency of the proposed estimator is established. From practical applications, we derive a data-driven procedure based on the slope heuristic to calibrate the penalty term of the contrast; and the implementation is carried out through the dynamic programming algorithm, which leads to a procedure of $\mathcal{O}(n^2)$ time complexity. Some simulation results are provided, as well as the applications to the US recession data and the number of trades in the stock of Technofirst.

Keywords: Multiple change-points, model selection, integer-valued time series, Poisson quasi-maximum likelihood, penalized quasi-likelihood, slope heuristic.

1 Introduction

We consider a $\mathbb{N}_0$-valued ($\mathbb{N}_0 = \mathbb{N} \cup \{0\}$) process $Y = \{Y_t,\, t \in \mathbb{Z}\}$ where the conditional mean

$$\lambda_t = \lambda_t(\theta_t^*) = \mathbb{E}(Y_t|\mathcal{F}_{t-1}) \tag{1.1}$$

is a function (see below) of the whole information $\mathcal{F}_{t-1}$ up to time $t-1$ and of an unknown parameter θ_t^* belongs to a compact subset $\Theta \subset \mathbb{R}^d$ ($d \in \mathbb{N}$). The inference in the cases where $\theta_t^* = \theta^*$ is constant or the distribution of $Y_t|\mathcal{F}_{t-1}$ is known have been studied by many authors in several directions; see for instance, Fokianos *et al.* (2009), Fokianos and Tjøstheim (2011, 2012), Davis and Liu (2016), Douc *et al.* (2017) among others, for some recent works. We consider here a more general setting where θ_t^* is piecewise constant (multiple change-point problem) and that the distribution of $Y_t|\mathcal{F}_{t-1}$ is unknown. We refer to Franke *et al.* (2012), Kang and Lee (2014), Doukhan and Kengne (2015), Leung *et al.* (2017) and the references therein for some tests for change-point detection in integer-valued time series.

[1]Supported by the Institute for advanced studies - IAS (CY Cergy Paris Université, France), the MME-DII center of excellence (ANR-11-LABEX-0023-01) and by the CEA-MITIC (Université Gaston Berger, Sénégal)
[2]Developed within the ANR BREAKRISK : ANR-17-CE26-0001-01

Let $(Y_1, \cdots, Y_n)$ be a trajectory generated as in model (1.1) and assume that the parameter θ_t^* is piecewise constant. Also, assume that $\exists K^* \in \mathbb{N}$, $\underline{\theta}^* = (\theta_1^*, \cdots, \theta_{K^*}^*) \in \Theta^{K^*}$ and $0 < t_1^* < \cdots < t_{K^*-1}^* < n$ such that, $\{Y_t,\ t_{j-1}^* < t \leq t_j^*\}$ is generated from the jth stationary regime ; i.e., it is a trajectory of the process $\{Y_{t,j},\ t \in \mathbb{Z}\}$ (which are not actually observed for $j = 1, \cdots, K^*$, see Section 2 for some details) satisfying:

$$\mathbb{E}(Y_{t,j}|\mathcal{F}_{t-1}) = f(Y_{t-1,j}, Y_{t-2,j}, \cdots ; \theta_j^*),\ \forall\ t_{j-1}^* < t \leq t_j^* \tag{1.2}$$

where $\mathcal{F}_t = \sigma(Y_{s,j}, s \leq t,\ j = 1, \cdots, K^* - 1)$ is the σ-field generated by the whole information up to time t and f is a measurable non-negative function assumed to be known up to the parameter θ_t^*. K^* is the number of segments (or regimes) of the model; the jth segment corresponds to $\{t_{j-1}^* + 1, t_{j-1}^* + 2, \cdots, t_j^*\}$ and depends on the parameter θ_j^*. $t_1^*, \cdots, t_{K^*-1}^*$ are the change-point locations; by convention, $t_0^* = -\infty$ and $t_{K^*}^* = \infty$. To ensure the identifiability of the change-point locations, it is reasonable to assume that $\theta_j^* \neq \theta_{j+1}^*$ for $j = 1, \cdots, K^* - 1$. The case $K^* = 1$ corresponds to the model without change. In the sequel, we assume that the random variables Y_t, $t \in \mathbb{Z}$ have the same (up to the parameter θ_t^*) distribution P and denote by $P(\cdot|\mathcal{F}_{t-1})$ the distribution of $Y_t|\mathcal{F}_{t-1}$. For instance, for an INGARCH(p^*, q^*) representation, we have

$$\lambda_t = \alpha_{0,j}^* + \sum_{i=1}^{q^*} \alpha_{i,j}^* Y_{t-i} + \sum_{i=1}^{p^*} \beta_{i,j}^* \lambda_{t-i}, \quad \text{for all } t_{j-1}^* < t \leq t_j^*,$$

where $\alpha_{0,j}^* > 0$, $\alpha_{1,j}^*, \cdots, \alpha_{q^*,j}^*, \beta_{1,j}^*, \cdots, \beta_{p^*,j}^* \geq 0$. The parameters vector of the jth regime is $\theta_j^* = (\alpha_{0,j}^*, \alpha_{1,j}^*, \cdots, \alpha_{q^*,j}^*, \beta_{1,j}^*, \cdots, \beta_{p^*,j}^*)$. Therefore, Θ is a compact subset of $(0, \infty) \times [0, \infty)^{p^*+q^*}$ such that for all $\theta = (\alpha_0, \alpha_1, \cdots, \alpha_{q^*}, \beta_1, \cdots, \beta_{p^*}) \in \Theta$, $\sum_{i=1}^{q^*} \alpha_i + \sum_{i=1}^{p^*} \beta_i < 1$. For all $j = 1, \cdots, K^*$, we assume that $\theta_j^* \in \Theta$; hence, there exists a sequence of non-negative real numbers $(\psi_k(\theta_j^*))_{k \geq 0}$ such that $\lambda_t = \psi_0(\theta_j^*) + \sum_{k \geq 1} \psi_k(\theta_j^*) Y_{t-k}$. Then, $f(y_1, y_2, \cdots; \theta_j^*) = \psi_0(\theta_j^*) + \sum_{k \geq 1} \psi_k(\theta_j^*) y_k$ for any $(y_1, y_2, \cdots) \in \mathbb{N}_0^{\infty}$. For instance, if the distribution P is Poisson, negative binomial or binary, then we get respectively a Poisson, negative binomial, binary INGARCH process; see some examples in Section 4.

Our main focus of interest is the estimation of the unknown parameters $\big(K^*, (t_j^*)_{1 \leq j \leq K^*-1}, (\theta_j^*)_{1 \leq j \leq K^*}\big)$ in the model (1.2). This can be viewed as a classical model selection problem. Assume that the observations $Y_1, \cdots, Y_n$ are generated from (1.2). Let $K_{\max}$ be the upper bound of the number of segments (note that $K_{\max} < n$). Denote by $\mathcal{M}_n$ the set of partitions of $[\![1, n]\!]$ into at most $K_{\max}$ contiguous segments. Set $m = \{T_1, \cdots, T_K\}$ a generic element of K segments in $\mathcal{M}_n$. Consider the collection $\{\mathcal{S}_m,\ m \in \mathcal{M}_n\}$ where, for a given $m \in \mathcal{M}_n$, $\mathcal{S}_m$ is the families of sequence (θ_t) which are piecewise constant on the partition m. Any $\vartheta = (\theta_t) \in \mathcal{S}_m$ depends on the parameter $\underline{\theta} = (\theta_1, \cdots, \theta_K)$ which is the piecewise values of θ_t on each segment. Set $\mathcal{S} = \cup_{m \in \mathcal{M}_n} \mathcal{S}_m$. Denote by ϑ a generic element of $\mathcal{S}$, with partition m and parameter $\underline{\theta}$. $|\underline{\theta}| = K$ denotes the number of the piecewise segments, also called the dimension of ϑ. The true model ϑ^* with dimension K^*, depends on a partition m^* and the parameter $\underline{\theta}^*$.

For any $\vartheta \in \mathcal{S}$, set $\lambda_t^{\vartheta} = \sum_{k=1}^{K} \lambda_t(\theta_k) \mathbb{1}_{t \in T_k}$ and denote by $P(\cdot|\mathcal{F}_{t-1}, \vartheta)$ the distribution of $Y_t|\mathcal{F}_{t-1}, \vartheta$; let $p(\cdot|\mathcal{F}_{t-1}, \vartheta) = p(\cdot; \lambda_t^{\vartheta})$ be the probability density function of this distribution. For $\vartheta \in \mathcal{S}$, let $P_{n,\vartheta}$ be the conditional distribution of $(Y_1, \cdots, Y_n)|\mathcal{F}_{n-1}, \vartheta$. We consider the log-likelihood contrast conditioned to $Y_0, Y_{-1}, \cdots$: $\forall \vartheta \in \mathcal{S}$,

$$\gamma_n(\vartheta) := \gamma_n(P_{n,\vartheta}) = -\log P_{n,\vartheta}(Y_1, \cdots, Y_n) = -\sum_{t=1}^{n} \log p(Y_t|\mathcal{F}_{t-1}, \vartheta) = -\sum_{t=1}^{n} \log p(Y_t; \lambda_t^{\vartheta}).$$

Thus, the minimal contrast estimator $\widehat{\vartheta}_m$ of ϑ^* on the collection $\mathcal{S}_m$ is obtained by minimizing the contrast $\gamma_n(\vartheta)$ over $\vartheta \in \mathcal{S}_m$; that is, $\widehat{\vartheta}_m = \underset{\vartheta \in \mathcal{S}_m}{\operatorname{argmin}}\ \gamma_n(\vartheta)$. The main approaches of the model selection procedures take into account the model complexity and select the estimator $\widehat{\vartheta}_{m_n}$ such that, m_n minimizes the penalized criterion

$$\mathrm{crit}_n(m) = \gamma_n(\widehat{\vartheta}_m) + \mathrm{pen}_n(m), \quad \text{for all } m \in \mathcal{M}_n \tag{1.3}$$

where $\mathrm{pen}_n : \mathcal{M}_n \to \mathbb{R}_+$ is a penalty function, possibly data-dependent. We now address the following issues.

(i) **Semi-parametric setting**. Kashikar *et al.* (2013) have carried out structural breaks in Poisson INAR process from the MCMC and Gibbs sampling approach. Cleynen and Lebarbier (2014 and 2017) have recently considered the change-point type problem (1.2) with i.i.d. observations; in their works, the distribution P is assumed to be known and could be Poisson, Negative binomial or belongs to the exponential family distribution. From the practical viewpoint, we consider the case where P is unknown and deal with the Poisson quasi-likelihood (see for instance, Ahmad and Francq (2016)). So in the sequel, γ_n is the Poisson quasi-likelihood contrast and $\widehat{\vartheta}_m$ is the Poisson quasi-maximum likelihood estimator (PQMLE).

(ii) **Multiple change-point problem from a non-asymptotic point of view**. This question is tacked by model selection approach. Numerous works have been devoted to this issue; see among others, Lebarbier (2005), Arlot and Massart (2009), Cleynen and Lebarbier (2014 and 2017), Arlot and Celisse (2016).
In this (quasi)log-likelihood framework, it is more usual to consider the Kullback-Leibler risk. For any $\vartheta \in \mathcal{S}$, the Kullback-Leibler divergence between P_{n,ϑ^*} and $P_{n,\vartheta}$ is

$$\begin{aligned} KL(\vartheta^*, \vartheta) := KL(P_{n,\vartheta^*}, P_{n,\vartheta}) &= \mathbb{E}\Big[\log \frac{P_{n,\vartheta^*}(Y_1, \cdots, Y_n)}{P_{n,\vartheta}(Y_1, \cdots, Y_n)}\Big] = \sum_{t=1}^{n} \mathbb{E}\Big[\log \frac{p(Y_t|\mathcal{F}_{t-1}, \vartheta^*)}{p(Y_t|\mathcal{F}_{t-1}, \vartheta)}\Big] \\ &= \sum_{t=1}^{n} \mathbb{E}\big[\log p(Y_t; \lambda_t^{\vartheta^*})\big] - \sum_{t=1}^{n} \mathbb{E}\big[\log p(Y_t; \lambda_t^{\vartheta})\big], \end{aligned}$$

where $\mathbb{E}$ denotes the expectation with respect to the true distribution of the observations. In the case where γ_n is the likelihood contrast, we get $KL(\vartheta^*, \vartheta) = \mathbb{E}[\gamma_n(\vartheta) - \gamma_n(\vartheta^*)]$. The "ideal" partition $m(\vartheta^*)$ (the one whose estimator is closest to ϑ^* according to the Kullback-Leibler risk) satisfying:

$$m(\vartheta^*) = \underset{m \in \mathcal{M}_n}{\operatorname{argmin}}\ \mathbb{E}[KL(\vartheta^*, \widehat{\vartheta}_m)].$$

The corresponding estimator $\widehat{\vartheta}_{m(\vartheta^*)}$, called the *oracle*, depends on the true sample distribution, and cannot be computed in practice. The goal is to calibrate the penalty term, such that the segmentation $\widehat{m}$ provides an estimator $\widehat{\vartheta}_{\widehat{m}}$ where the risk of $\widehat{\vartheta}_{\widehat{m}}$ is close as possible to the risk of the *oracle*, namely such that

$$\mathbb{E}[KL(\vartheta^*, \widehat{\vartheta}_{\widehat{m}})] \leq C\ \mathbb{E}[KL(\vartheta^*, \widehat{\vartheta}_{m(\vartheta^*)})] \tag{1.4}$$

for a non-negative constant C, expected close to 1. This issue is addressed in the above mentioned papers, and the results obtained are heavily relied on the independence of the observations. In our setting here, it seems to be a more difficult task. But, we believe that the coupling method can be used as in Lerasle (2011) to overcome this difficulty. We leave this question as the topic of a different research project.

(iii) **Multiple change-point problem from an asymptotic point of view**. The aim here is to consistently estimate the parameters of the change-point model. This issue has been addressed by several authors using the classical contrast/criteria optimization or binary/sequential segmentation/estimation; see for instance, Bai and Perron (1998), Davis *et al.* (2008), Harchaoui and Lévy-Leduc (2010), Bardet *et al.* (2012), Davis and Yau (2013), Davis *et al.* (2016), Ma and Yau (2016), Yau and Zhao (2016), Inclan and Tiao (1994), Bai (1997), Fryzlewicz and Subba Rao (2014), Fryzlewicz (2014), among others, for some advanced towards this issue. These works and many other papers in the literature on the asymptotic study of multiple change-point problem are often focused on continuous valued time series; moreover, the case of a large class of semi-parametric model for discrete-valued time series (such as those discussed earlier) have not yet addressed.

We consider (1.2) and derive a penalized contrast of type (1.3). We assume that there exists a partition $\underline{\tau}^*$ of $[0,1]$ such that $[\underline{\tau}^* n] = m^*$, where $[\underline{\tau}^* n]$ is the corresponding partition of $[\![1,n]\!]$ obtained from $\underline{\tau}^*$. We provide sufficient conditions on the penalty pen_n, for which the estimators $\widehat{m}$ and $\widehat{\vartheta}_{\widehat{m}}$ are consistent; that is:

$$\Big(|\widehat{m}|, \frac{\widehat{m}}{n}, \widehat{\vartheta}_{\widehat{m}}\Big) \underset{n\to\infty}{\overset{\mathcal{P}}{\longrightarrow}} \big(K^*, \underline{\tau}^*, \vartheta^*\big)$$

where $\frac{\widehat{m}}{n}$ is the corresponding partition of $[0,1]$ obtained from $\widehat{m}$.

The paper is organized as follows. In Section 2, we set some notations, assumptions and define the Poisson QMLE. In Section 3, we derive the estimation procedure and provide the main results. Some simulations results are displayed in Section 4 whereas Section 5 focus on applications on the US recession data and the daily number of trades in the stock of Technofirst. Section 6 is devoted to a summary and conclusion. The Supporting Information provides the proofs of the main results.

2 Notations and Poisson QMLE

We set the following classical Lipschitz-type condition on the function f.

Assumption $\mathbf{A}_i(\Theta)$ $(i = 0,1,2)$: For any $y \in \mathbb{N}_0^{\mathbb{N}}$, the function $\theta \mapsto f(y;\theta)$ is i times continuously differentiable on Θ and there exists a sequence of non-negative real numbers $(\alpha_k^{(i)})_{k\geq 1}$ satisfying $\sum\limits_{k=1}^{\infty} \alpha_k^{(0)} < 1$ (or $\sum\limits_{k=1}^{\infty} \alpha_k^{(i)} < \infty$ for $i = 1,2$); such that for any $y, y' \in \mathbb{N}_0^{\mathbb{N}}$,

$$\sup_{\theta\in\Theta} \Big\| \frac{\partial^i f(y;\theta)}{\partial \theta^i} - \frac{\partial^i f(y';\theta)}{\partial \theta^i} \Big\| \leq \sum_{k=1}^{\infty} \alpha_k^{(i)} |y_k - y'_k|;$$

where $\|\cdot\|$ denotes any vector, matrix norm.

In the whole paper, it is assumed that for $j = 1, \cdots, K^*$, there exists a stationary and ergodic process $\{Y_{t,j},\, t \in \mathbb{Z}\}$ satisfying

$$\mathbb{E}(Y_{t,j}|\mathcal{F}_{t-1,j}) = f(Y_{t-1,j}, Y_{t-2,j}, \cdots; \theta_j^*),\ \forall\, t \in \mathbb{Z} \tag{2.1}$$

where $\mathcal{F}_{t,j} = \sigma(Y_{s,j}, s \leq t)$ is the σ-field generated by $\{Y_{s,j}, s \leq t\}$; and

$$\exists C > 0, \epsilon > 1,\ \text{such that}\ \forall t \in \mathbb{Z},\quad \mathbb{E}Y_{t,j}^{1+\epsilon} < C. \tag{2.2}$$

$\{Y_{t,j},\, t \in \mathbb{Z}\}$ is a stationary solution of the jth regime. The focus process $Y = \{Y_t,\, t \in \mathbb{Z}\}$ is modelled by these stationary regimes; that is, for any $j = 1, \cdots, K^*$, $\{Y_t, t_{j-1}^* < t \leq t_j^*\}$ is a trajectory of the process $\{Y_{t,j}, t \in \mathbb{Z}\}$. Note that, the main assumption here is that, the observations on each segment j are stationary and depend on θ_j^*. Therefore, the process Y is then constructed such that for any $j = 1, \cdots, K^*$, $Y_t = Y_{t,j}$

for $t^*_{j-1} < t \leq t^*_j$. This stationary assumption on each segment has been discussed in the literature. In the classical autoregressive models (including ARMA and GARCH process), Bardet *et al.* (2012) have avoided this assumption by an approximation study between the stationary and the nonstationary regime. As pointed out by Doukhan and Kengne (2015) (see Remark 4.1), such study can be carried out in the Poisson autoregressive model. But, it seems to be not easy (or not possible) in the general class of model considered here, since the conditional distribution is assumed to be unknown.

Ahmad and Francq (2016) (Section 3) have discussed about the stationarity and the ergodicity issues. In many classical integer-valued time series, the assumption $\mathbf{A}_0(\Theta)$ allows the existence of a stationary and ergodic process satisfying (2.1) (see for instance, Doukhan *et al.* (2012, 2013), Davis and Liu (2016)).

2.1 Notations

Assume that a trajectory $(Y_1, \cdots, Y_n)$ of Y is observed; with $0 < t^*_1 < \cdots < t^*_{K^*-1} < n$. By convention $t^*_0 = -\infty$ and $t^*_{K^*} = \infty$. We will use the following notations.

- For any finite set A, $|A|$ denotes the cardinality of A.

- For $a, b \in \mathbb{R}$ (with $a \leq b$), $[\![a, b]\!] = \mathbb{N} \cap [a, b]$ is the set of integers between a and b.

- For any $(p, q) \in \mathbb{N}^2$, $M_{p,q}(\mathbb{R})$ denotes the set of matrices of dimension $p \times q$ with coefficients in $\mathbb{R}$.

- For $K \in \mathbb{N}$, $\mathcal{M}_n(K) = \big\{\underline{t} = (t_1, \ldots, t_{K-1})\ ;\ 0 < t_1 < \ldots < t_{K-1} < n,\ t_{j+1} - t_j > u_n \text{ for } j = 1, \cdots, K-1 \text{ and } n - t_{K-1} > u_n\big\}$ for some sequence (u_n) with values in $\mathbb{N}$ (see Subsection 3.1); in particular, $\underline{t}^* = \big(t^*_1, \ldots, t^*_{K^*-1}\big) \in \mathcal{M}_n(K^*)$ is the true vector of the locations of breaks. When $K = 1$, $\mathcal{M}_n(1)$ corresponds to the model with no break.
  In the sequel, any configuration $\underline{t} = (t_1, \ldots, t_{K-1}) \in \mathcal{M}_n(K)$ is also used as a partition $\{T_1, T_2, \cdots, T_K\}$ of $[\![1, n]\!]$ into K contiguous segments, where $T_1 = \{1, \cdots, t_1\}$, $T_j = \{t_{j-1}+1, \cdots, t_j\}$ for $j = 2, \cdots, K-1$, $T_K = \{t_{K-1} + 1, \cdots, n\}$. In particular, $T^*_1 = \{1, \cdots, t^*_1\}$, $T^*_j = \{t^*_{j-1} + 1, \cdots, t^*_j\}$ for $j = 2, \cdots, K^* - 1$ and $T_{K^*} = \{t_{K^*-1} + 1, \cdots, n\}$. $\mathcal{M}_n(K)$ corresponds to the set of partitions of $[\![1, n]\!]$ into K contiguous segments.

- For $K \in \mathbb{N}^*$ and $\underline{t} \in \mathcal{M}_n(K)$ fixed, we set $n_k = |T_k|$ for $1 \leq k \leq K$. In particular, $n^*_j = |T^*_j|$ for $1 \leq j \leq K^*$. For $1 \leq k \leq K$ and $1 \leq j \leq K^*$, let $n_{k,j} = |T^*_j \cap T_k|$.

- Let $\underline{\theta}^* = (\theta^*_1, \cdots, \theta^*_{K^*}) \in \Theta^{K^*}$ be the vector of the true parameters of the model (1.2).

Throughout the sequel, the following norms will be used:

- $\|f\|_\Theta := \sup_{\theta \in \Theta} (\|f(\theta)\|)$ *for any function* $f : \Theta \longrightarrow M_{p,q}(\mathbb{R})$;

- $\|x\|_m = \max\limits_{1 \leq i \leq K} |x_i|$ *for* $x = (x_1, \cdots, x_K) \in \mathbb{R}^K$;

- *if* Y *is a random vector with finite* $r-$*order moments, we set* $\|Y_t\|_r = \mathbb{E}\left(\|Y\|^r\right)^{1/r}$.

2.2 Poisson QMLE

Let $(Y_1, \dots, Y_n)$ be a trajectory generated from the model (1.2). Since the conditional distribution is assumed to be unknown, the likelihood of the model is unknown. The estimation procedure of the parameters θ_j^* is based on the Poisson quasi-maximum likelihood introduced by Ahmad and Francq (2016). The conditional Poisson (quasi)log-likelihood of the model (1.2) computed on a segment $T \subset \{1, \dots, n\}$ is given (up to a constant) by

$$\widehat{L}_n(T,\theta) := \sum_{t\in T}(Y_t \log \widehat{\lambda}_t(\theta) - \widehat{\lambda}_t(\theta)) = \sum_{t\in T} \widehat{\ell}_t(\theta) \quad \text{with } \widehat{\ell}_t(\theta) = Y_t \log \widehat{\lambda}_t(\theta) - \widehat{\lambda}_t(\theta) \tag{2.3}$$

where $\widehat{\lambda}_t(\theta) = \widehat{f}_t^\theta = f(Y_{t-1}, \cdots Y_1, 0, \cdots, 0; \theta)$.
According to (2.3), the Poisson quasi-likelihood estimator (PQMLE) of θ_j^* computed on T is defined by

$$\widehat{\theta}_n(T) := \underset{\theta\in\Theta}{\operatorname{argmax}}(\widehat{L}_n(T,\theta)). \tag{2.4}$$

Now, for $j = 1, \cdots, K^*$, define the Poisson (quasi)log-likelihood of the jth regime by

$$L_{n,j}(T_j^*,\theta) := \sum_{t\in T_j^*}(Y_{t,j} \log \lambda_{t,j}(\theta) - \lambda_{t,j}(\theta)) = \sum_{t\in T_j^*} \ell_{t,j}(\theta) \quad \text{with } \ell_{t,j}(\theta) = Y_{t,j} \log \lambda_{t,j}(\theta) - \lambda_{t,j}(\theta)$$

where $\lambda_{t,j}(\theta) = f_{t,j}^\theta = f(Y_{t-1,j}, Y_{t-2,j}, \cdots; \theta)$. It can be approximated by

$$\widehat{L}_{n,j}(T_j^*,\theta) := \sum_{t\in T_j^*}(Y_{t,j} \log \widehat{\lambda}_{t,j}(\theta) - \widehat{\lambda}_{t,j}(\theta)) = \sum_{t\in T_j^*} \widehat{\ell}_{t,j}(\theta) \quad \text{with } \widehat{\ell}_{t,j}(\theta) = Y_{t,j} \log \widehat{\lambda}_{t,j}(\theta) - \widehat{\lambda}_{t,j}(\theta) \tag{2.5}$$

where $\widehat{\lambda}_{t,j}(\theta) = \widehat{f}_{t,j}^\theta = f(Y_{t-1,j}, \cdots, Y_{t_{j-1}^*+1,j}, 0 \cdots 0; \theta)$.
According to (2.5), the PQMLE of θ_j^* computed on T_j^* is defined by

$$\widetilde{\theta}_n(T_j^*) := \underset{\theta\in\Theta}{\operatorname{argmax}}(\widehat{L}_{n,j}(T_j^*,\theta)). \tag{2.6}$$

Let us stress that $\widehat{\theta}_n(T)$ is the PQMLE of the parameter computed on the segment T from the model with multiple change-point ; i.e., $\widehat{\lambda}_t(\theta)$ (used in $\widehat{L}_n(T,\theta)$) depends on the observations of all the previous regimes; whereas $\widetilde{\theta}_n(T_j^*)$ represents the PQMLE of the parameter computed from the jth stationary regime ; i.e., $\widehat{\lambda}_{t,j}(\theta)$ (used in $\widehat{L}_{n,j}(T_j^*,\theta)$) depends on only the observations of the jth stationary regime.

To avoid the problems of parameter identifiability and to study the asymptotic normality of the PQMLE, we shall assume:
(**A0**): for all $(\theta,\theta') \in \Theta^2$, $\Big(f(Y_{t-1}, Y_{t-2}, \cdots; \theta) = f(Y_{t-1}, Y_{t-2}, \cdots; \theta')$ a.s. for some $t \in \mathbb{N}\Big) \Rightarrow \theta = \theta'$; moreover, $\exists \underline{c} > 0$ such that $\inf_{\theta\in\Theta} f(y_1, y_2, \cdots; \theta) \geq \underline{c}$, for all $y \in \mathbb{N}_0^{\mathbb{N}}$.

In order to ensure the consistency and the asymptotic normality of the PQMLE, we set the following assumptions for each segment $j = 1, \cdots, K^*$ (see also Ahmad and Francq (2016)):

(**A1**): θ_j^* is an interior point of $\Theta \subset \mathbb{R}^d$;

(**A2**): $a_{t,j} \xrightarrow{a.s} 0$ and $Y_{t,j} a_{t,j} \xrightarrow{a.s} 0$ as $t \to \infty$, where $a_{t,j} = \sup_{\theta\in\Theta} \left|\widehat{\lambda}_{t,j}(\theta) - \lambda_{t,j}(\theta)\right|$;

(**A3**): $J_j = \mathbb{E}\Big[\frac{1}{\lambda_{t,j}(\theta_j^*)} \frac{\partial \lambda_{t,j}(\theta_j^*)}{\partial\theta} \frac{\partial \lambda_{t,j}(\theta_j^*)}{\partial\theta'}\Big] < \infty$ and $I_j = \mathbb{E}\Big[\frac{\mathrm{Var}(Y_{t,j}|\mathcal{F}_{t-1})}{\lambda_{t,j}^2(\theta_j^*)} \frac{\partial \lambda_{t,j}(\theta_j^*)}{\partial\theta} \frac{\partial \lambda_{t,j}(\theta_j^*)}{\partial\theta'}\Big] < \infty$;

(**A4**): for all $c' \in \mathbb{R}$, $c' \frac{\partial \lambda_{t,j}(\theta_j^*)}{\partial \theta} = 0$ a.s $\Rightarrow$ $c' = 0$;

(**A5**): there exists a neighborhood $V(\theta_j^*)$ of θ_j^* such that: for all $i, k \in \{1, \cdots, d\}$,

$$\mathbb{E}\left[\sup_{\theta \in V(\theta_j^*)} \left| \frac{\partial^2}{\partial \theta_i \partial \theta_k} \ell_{t,j}(\theta) \right| \right] < \infty;$$

(**A6**): $b_{t,j}$, $b_{t,j} Y_{t,j}$ and $a_{t,j} d_{t,j} Y_{t,j}$ are of order $O(t^{-h})$ for some $h > 1/2$, where

$$b_{t,j} = \sup_{\theta \in \Theta} \left\{ \mathbb{E}\left[\left\| \frac{\partial \widehat{\lambda}_{t,j}(\theta)}{\partial \theta} - \frac{\partial \lambda_{t,j}(\theta)}{\partial \theta} \right\| \right] \right\} \text{ and } d_{t,j} = \sup_{\theta \in \Theta} \max \left\{ \mathbb{E}\left[\left\| \frac{1}{\widehat{\lambda}_{t,j}(\theta)} \frac{\partial \widehat{\lambda}_{t,j}(\theta)}{\partial \theta} \right\| \right], \mathbb{E}\left[\left\| \frac{1}{\lambda_{t,j}(\theta)} \frac{\partial \lambda_{t,j}(\theta)}{\partial \theta} \right\| \right] \right\}.$$

These aforementioned assumptions hold for many classical models, see Ahmad and Francq (2016). These authors have established that the estimator $\widetilde{\theta}_n(T_j^*)$ is strongly consistent, for each regime $j \in \{1, \cdots, K^*\}$; that is,

$$\widetilde{\theta}_n(T_j^*) \xrightarrow[n \to \infty]{a.s.} \theta_j^*.$$

They have also proved the asymptotic normality of $\widetilde{\theta}_n(T_j^*)$; that is,

$$\sqrt{n_j^*}(\widetilde{\theta}_n(T_j^*) - \theta_j^*) \xrightarrow[n \to \infty]{\mathcal{D}} \mathcal{N}(0, \Sigma_j), \ \forall j = 1, \cdots, K^*,$$

where $\Sigma_j := J_j^{-1} I_j J_j^{-1}$. Under the above assumptions, for any $j = 1, \cdots, K^*$, the matrix Σ_j can be consistently estimated by

$$\begin{aligned}
\widehat{\Sigma}_j &= \widehat{J}_j^{-1} \widehat{I}_j \widehat{J}_j^{-1}, \text{ where} \\
\widehat{J}_j &= \frac{1}{n} \sum_{t=1}^{n} \frac{1}{\widehat{\lambda}_{t,j}(\widetilde{\theta}_n(T_j^*))} \frac{\partial \widehat{\lambda}_{t,j}(\widetilde{\theta}_n(T_j^*))}{\partial \theta} \frac{\partial \widehat{\lambda}_{t,j}(\widetilde{\theta}_n(T_j^*))}{\partial \theta'}, \\
\widehat{I}_j &= \frac{1}{n} \sum_{t=1}^{n} \Big(\frac{Y_t}{\widehat{\lambda}_{t,j}(\widetilde{\theta}_n(T_j^*))} - 1 \Big)^2 \frac{\partial \widehat{\lambda}_{t,j}(\widetilde{\theta}_n(T_j^*))}{\partial \theta} \frac{\partial \widehat{\lambda}_{t,j}(\widetilde{\theta}_n(T_j^*))}{\partial \theta'}.
\end{aligned} \tag{2.7}$$

If we consider the process $\{Y_t, \, t \in \mathbb{Z}\}$, these properties are also verified on the segment T_1^* since it is easy to see that $\{(Y_t, \lambda_t), \, t \in T_1^*\}$ is a stationary process while $\{(Y_t, \lambda_t), \, t > t_1^*\}$ is not.

The following proposition establishes the consistency of the estimator $\widehat{\theta}_n(T_j^*)$, for any $j \in \{1, \cdots K^*\}$.

Proposition 2.1 *Assume that (**A0**)-(**A2**) and ($\boldsymbol{A}_0(\Theta)$) hold. Then*

$$\widehat{\theta}_n(T_j^*) \xrightarrow[n \to \infty]{a.s.} \theta_j^*, \ \forall j = 1, \cdots, K^*.$$

The results of this proposition have been obtained by Ahmad and Francq (2016) when (Y_t, λ_t) is strictly stationary.

3 Estimation procedure and main results

In this section, we carry out the estimation of the number of breaks $K^* - 1$ and the instants of breaks $\underline{t}^*$ by using a penalized contrast. Some asymptotic studies are also reported.

3.1 Penalized Poisson quasi-likelihood estimator

For any configuration of regimes $K \geq 1$, $\underline{t} \in \mathcal{M}_n(K)$ and $\underline{\theta} = (\theta_1, \cdots, \theta_K) \in \Theta^K$, we define the contrast

$$(QLIK) \qquad \widehat{J}_n(K, \underline{t}, \underline{\theta}) := -2 \sum_{k=1}^{K} \widehat{L}_n(T_k, \theta_k). \tag{3.1}$$

According to the proprieties of the PQMLE, when K^* is known, a natural estimator of $(\underline{t}^*, \underline{\theta}^*) = ((t_j^*)_{1 \leq j \leq K^*-1}, (\theta_j^*)_{1 \leq j \leq K^*})$ for the model (1.2) is therefore the PQMLE on every interval $[t_j + 1, \cdots, t_{j+1}]$ and every parameter θ_j, for $1 \leq j \leq K^*$. But, since K^* is assumed to be unknown, we cannot directly use such method. To take into account the estimation of K^*, the most classical solution is to penalize the contrast by an additional term $\kappa_n K$, where κ_n represents a regularization parameter.
Now, define the penalized contrast $QLIK$, called $penQLIK$, by

$$(penQLIK) \qquad \widetilde{J}_n(K, \underline{t}, \underline{\theta}) := \widehat{J}_n(K, \underline{t}, \underline{\theta}) + \kappa_n K, \tag{3.2}$$

with $\kappa_n \leq n$ and $\kappa_n \underset{n \to \infty}{\longrightarrow} +\infty$.
The estimator of $(K^*, \underline{t}^*, \underline{\theta}^*)$ is defined as one of the minimizers of the penalized contrast:

$$\left(\widehat{K}_n, \widehat{\underline{t}}_n, \widehat{\underline{\theta}}_n\right) \in \underset{1 \leq K \leq K_{\max}}{\operatorname{argmin}} \ \underset{(\underline{t}, \underline{\theta}) \in \mathcal{M}_n(K) \times \Theta^K}{\operatorname{argmin}} \left(\widetilde{J}_n(K, \underline{t}, \underline{\theta})\right) \ \text{and} \ \widehat{\underline{\tau}}_n = \frac{\widehat{\underline{t}}_n}{n}. \tag{3.3}$$

We will also carry out a data-driven method, based on the slope heuristic procedure (see Baudry *et al.* (2010)) to calibrate the penalty term. In this procedure, it is expected that the criteria $QLIK$ is a linear transformation of the penalty (here the number of regimes K) for the most complex models (with K close to $K_{\max}$). Two times the slope of the linear part of $-QLIK(K)$ with $K \leq K_{\max}$ should be close to the "optimal" penalty term (which the corresponding estimator fulfills an oracle inequality as (1.4)). Note that, in practice, a numerical algorithm can be used to compute the estimator on each segment; therefore, a minimum size is needed for the numerical computation of the criteria. Thus, we consider only the segments of length larger than some u_n and we can a priori fix $K_{\max}$ smaller than $[n/u_n]$, where (u_n) is a sequence with values in $\mathbb{N}$, satisfying $u_n \underset{n \to \infty}{\longrightarrow} \infty$. The complete procedure can be summarized as follows:

1. For each $1 \leq K \leq K_{\max}$, draw $\left(K, -\min_{\underline{t}, \underline{\theta}} QLIK(K)\right)$. Then compute the slope of the linear part: this slope is $\widehat{\kappa}_n/2$.
2. Using $\kappa_n = \widehat{\kappa}_n$, draw $\left(K, -\min_{\underline{t}, \underline{\theta}} penQLIK(K)\right)_{1 \leq K \leq K_{\max}}$. This curve has a global minimum at $\widehat{K}_n$.

Let us point out that the theoretical validity of this slope heuristic procedure has been established in some model selection problem; see for instance, Arlot and Massart (2009), Lerasle (2011). The theoretical validity in a large class of semi-parametric model for integer-valued time series considered here has not yet been established. But this heuristic has been successfully applied for change-point detection in some settings (see for example, Baudry *et al.* (2010), Bardet *et al.* (2012)).

3.2 Asymptotic behavior

Under some assumptions, we will establish the asymptotic behavior of the estimator $(\widehat{K}_n, \widehat{\underline{t}}_n, \widehat{\underline{\theta}}_n)$. Throughout this article, we set the following classical assumption in the problem of break detection:

Assumption B. $\min_{1\leq j\leq K^*-1}\left\|\theta^*_{j+1}-\theta^*_j\right\|>0$. Also, there exists a vector $\underline{\tau}^*=(\tau^*_1,\cdots,\tau^*_{K-1})$ with $0<\tau^*_1<\cdots<\tau^*_{K-1}<1$, called the vector of breaks such that $t^*_j=\left[n\tau^*_j\right]$, for $j=1,\cdots,K$ (where $[\cdot]$ is the integer part).

The following theorem gives the consistency of $(\widehat{K}_n,\widehat{\underline{t}}_n,\widehat{\underline{\theta}}_n)$.

Theorem 3.1 *Assume that $K_{\max}>K^*$ and **(A0)**-**(A2)**, **B**. If $\boldsymbol{A}_i(\Theta)$ $(i=0,1,2)$, (2.2) (with $\epsilon>1$) hold and (κ_ℓ) satisfies*

$$\sum_{\ell\geq 1}\frac{1}{\kappa_\ell}\sum_{k\geq\ell}\alpha^{(0)}_k<\infty, \tag{3.4}$$

then

$$\left(\widehat{K}_n,\widehat{\underline{\tau}}_n,\widehat{\underline{\theta}}_n\right)\ \xrightarrow[n\to\infty]{\mathcal{P}}\ (K^*,\underline{\tau}^*,\underline{\theta}^*)\,.$$

By convention, throughout the sequel, if the vectors $\widehat{\underline{t}}_n$ and $\underline{t}^*$ do not have the same length, complete the shorter of the two vectors with 0 before computing the norm $\left\|\widehat{\underline{t}}_n-\underline{t}^*\right\|_m$. The following theorem establishes the rates of convergence of the estimator $\widehat{\underline{\tau}}_n$.

Theorem 3.2 *Assume that $K_{\max}>K^*$ and **(A0)**-**(A2)**, **B**. If $\boldsymbol{A}_i(\Theta)$ $(i=0,1,2)$, (2.2) (with $\epsilon>1$), (3.4) hold and*

$$\sum_{\ell\geq 1}\frac{1}{\sqrt{\ell}}\sum_{k\geq\ell}\alpha^{(i)}_k<\infty,\ \ for\ i=1,2, \tag{3.5}$$

then the sequence $\left(\left\|\widehat{\underline{t}}_n-\underline{t}^*\right\|_m\right)_{n>1}$ is uniformly tight in probability; that is,

$$\lim_{\delta\to\infty}\ \lim_{n\to\infty}\mathbb{P}\left(\left\|\widehat{\underline{t}}_n-\underline{t}^*\right\|_m>\delta\right)=0.$$

This result implies that for any sequence of non-negative real numbers $(w_n)_n$ such that $w_n\to\infty$, $\|\widehat{\underline{t}}_n-\underline{t}^*\|_m=o_P(w_n)$; that is, a convergence rate close to $O_P(1)$. As pointed out by Bardet *et al.* (2012), this is the rate obtained when $(Y_t)_t$ is a sequence of independent random variables.

Now, we give the convergence in distribution of the estimator $\widehat{\underline{\theta}}_n$. By convention, if $\widehat{K}_n<K^*$, set $\widehat{T}_j=\widehat{T}_{\widehat{K}_n}$, for $j\in\left\{\widehat{K}_n,\cdots,K^*\right\}$. The following theorem establishes the asymptotic normality of $\widehat{\theta}_n(\widehat{T}_j)$.

Theorem 3.3 *Assume that $K_{\max}>K^*$ and **(A0)**-**(A6)** and **B**. If $\boldsymbol{A}_i(\Theta)$ $(i=0,1,2)$, (2.2) (with $\epsilon>2$), (3.4) and (3.5) hold, then*

$$\sqrt{n^*_j}\left(\widehat{\theta}_n(\widehat{T}_j)-\theta^*_j\right)\ \xrightarrow[n\to\infty]{\mathcal{D}}\ \mathcal{N}_d\left(0,\Sigma_j\right),\quad\forall\, j=1,\cdots,K^*,$$

where $\Sigma_j:=J^{-1}_j(\theta^*_j)I_j(\theta^*_j)J^{-1}_j(\theta^*_j)$ with

$$J_j(\theta^*_j)=\mathbb{E}\Big[\frac{1}{\lambda_{t,j}(\theta^*_j)}\frac{\partial\lambda_{t,j}(\theta^*_j)}{\partial\theta}\frac{\partial\lambda_{t,j}(\theta^*_j)}{\partial\theta'}\Big]\quad and\quad I_j(\theta^*_j)=\mathbb{E}\Big[\frac{\mathrm{Var}(Y_{t,j}|\mathcal{F}_{t-1})}{\lambda^2_{t,j}(\theta^*_j)}\frac{\partial\lambda_{t,j}(\theta^*_j)}{\partial\theta}\frac{\partial\lambda_{t,j}(\theta^*_j)}{\partial\theta'}\Big].$$

Remark 3.4 *The conditions on the regularization parameter $(\kappa_n)_{n\in\mathbb{N}}$ can be obtained if the Lipschitzian coefficients of $f(\cdot\,;\theta)$ and its derivatives are bounded by a geometric or Riemanian sequence:*

1. *the geometric case: if* $\alpha_k^{(i)} = \mathcal{O}(a^k)$ *(*$i = 0, 1, 2$*) with* $0 \leq a < 1$*, then any choice of* $(\kappa_n)_{n\in\mathbb{N}}$ *such that* $\kappa_n \leq n$ *and* $\kappa_n \to \infty$ *satisfies (3.4) (for instance,* κ_n *of order* $\log n$ *as in the BIC approach). Also, (3.5) holds.*

2. *the Riemanian case: if* $\alpha_k^{(i)} = \mathcal{O}(k^{-\gamma})$ *(*$i = 0, 1, 2$*) with* $\gamma > 3/2$*, then, (3.5) holds. Moreover,*

   - *if* $\gamma > 2$*, then the conditions (3.4) holds for any choice of* $(\kappa_n)_{n\in\mathbb{N}}$ *such that* $\kappa_n \leq n$ *and* $\kappa_n \to \infty$*.*
   - *if* $3/2 < \gamma \leq 2$*, then any choice such that* $\kappa_n = \mathcal{O}(n^{\delta})$ *with* $\delta > 2 - \gamma$ *or* $\kappa_n = \mathcal{O}(n^{2-\gamma}(\log n)^{\delta})$ *with* $\delta > 1$ *satisfies (3.4).*

4 Some simulations results

In this section, we implement the procedure on the R software (developed by the CRAN project). We will restrict our attention to the estimation of the vector $(K^*, \underline{t}^*)$; i.e, the number of segments K^* and the instants of breaks $\underline{t}^*$. For the performances of the estimator of the parameter $\underline{\theta}^*$, we refer to the works of Ahmad and Francq (2016). For each process, we generate 100 replications following the scenarios considered. The estimated number of segments is computed by using the $QLIK$ criteria penalized with $\kappa_n = \widehat{\kappa}_n$, $\kappa_n = \log n$ and $\kappa_n = n^{1/3}$. The value of the estimator $\widehat{\kappa}_n$ is calibrated by applying the slope estimation procedure (see Baudry *et al.* (2010)) as described above. Once the regularization parameter κ_n is obtained, the dynamic programming algorithm is used to minimize the criteria. With this algorithm, the complexity of the procedure declines from $\mathcal{O}(n^{K_{\max}})$ to $\mathcal{O}(n^2)$.

4.1 Implementation procedure

We give the steps of the dynamic programming algorithm for computing the number of segments $\widehat{K}_n$ and the optimal configuration of the break-points vector $\widehat{\underline{t}}$. This algorithm is such that if $(t_1, \cdots, t_{K-1}, t)$ represents the optimal configuration of $Y_1, \cdots, Y_t$ into K segments, then $(t_1, \cdots, t_{K-1})$ is the optimal configuration of $Y_1, \cdots, Y_{t_{K-1}}$ into $K - 1$ segments. Assume that the regularization parameter κ_n is known and let ML be the upper triangular matrix of dimension $n \times n$ with $ML_{i,l} = \widehat{L}(T_{i,l}, \widehat{\theta}_n(T_{i,l}))$, where $T_{i,l} = \{i, i+1, \cdots, l\}$, for $1 \leq i \leq l \leq n$. We summarize the implementation of the procedure as follows:

- *The number of segments* $\widehat{K}_n$: Let C be an upper triangular matrix of dimension $K_{\max} \times n$. For $1 \leq K \leq K_{\max}$ and $K \leq t \leq n$, $C_{K,t}$ will be the minimum penalized criteria of $Y_1, \cdots, Y_t$ into K segments. For $t = 1, \cdots, n$, $C_{1,t} = -2ML_{1,t} + \kappa_n$ and the relation $C_{K+1,t} = \min_{K \leq l \leq t-1} (C_{K,l} - 2ML_{l+1,t} + \kappa_n)$ is satisfied. Hence, $\widehat{K}_n = \operatorname{argmin}_{1 \leq K \leq K_{\max}} (C_{K,n})$.

- *The change-point locations* $\widehat{\underline{t}}_n$: Let Z be an upper triangular matrix of dimension $(K_{\max} - 1) \times n$. For $1 \leq K \leq (K_{\max} - 1)$ and $K + 1 \leq t \leq n$, $Z_{K,t}$ will be the Kth potential break-point of $Y_1, \cdots, Y_t$. Therefore, the relation $Z_{K,t} = \min_{K \leq l \leq t-1} (C_{K,l} - 2ML_{l+1,t} + \kappa_n)$ is satisfied for $K = 1, \cdots, (K_{\max} - 1)$. The break-points are obtained as follows: set $\widehat{t}_{\widehat{K}_n} = n$ and for $K = \widehat{K}_n - 1, \cdots, 1$, $\widehat{t}_K = Z_{K, \widehat{t}_{K+1}}$.

4.2 Results of simulations

4.2.1 Poisson-INARCH models

We consider the problem (1.2) for a Poisson-INARCH(1); i.e., $(Y_1, \cdots, Y_n)$ is a trajectory of the process $Y = \{Y_t,\, t \in \mathbb{Z}\}$ satisfying:

$$Y_t|\mathcal{F}_{t-1} \sim \mathcal{P}(\lambda_t) \ \ ; \ \ \lambda_t = f(Y_{t-1}, Y_{t-2}, \cdots; \theta^*_j) = \alpha_0^{(j)} + \alpha^{(j)} Y_{t-1}, \quad \forall\, t \in T^*_j, \ \ \forall\, j \in \{1, \cdots, K^*\}\,. \tag{4.1}$$

The parameters vector is $\theta^*_j = (\alpha_0^{(j)}, \alpha^{(j)})$, for all $j \in \{1, \cdots, K^*\}$.

For $n = 500$ and $n = 1000$, we generate a sample $(Y_1, \cdots, Y_n)$ in the following situations:

- **scenario IA**$_0$: $\theta^*_1 = (0.5, 0.6)$ is constant $(K^* = 1)$;
- **scenario IA**$_1$: $\theta^*_1 = (0.5, 0.6)$ changes to $\theta^*_2 = (1.0, 0.6)$ at $t^* = 0.5n$ $(K^* = 2)$;
- **scenario IA**$_2$: $\theta^*_1 = (0.5, 0.6)$ changes to $\theta^*_2 = (1.0, 0.6)$ at $t^*_1 = 0.3n$ which changes to $\theta^*_3 = (1.0, 0.25)$ at $t^*_2 = 0.7n$ $(K^* = 3)$.

For the scenario **IA**$_2$, we generate a series with $n = 1000$. Figure 1 shows the slope of the linear part of the $-QLIK$ criteria minimized in $(\underline{t}, \underline{\theta})$. We obtain $\widehat{\kappa}_n \approx 5.9$; this value of $\widehat{\kappa}_n$ is data-dependent ; i.e., it is the value that has been used for this particular realization. By using this above value for κ_n, we minimize the $penQLIK$ in $(K, \underline{t}, \underline{\theta})$, with $1 \leq K \leq K_{\max}$. Figure 1 also displays the points $(K, \min_{\underline{t},\underline{\theta}} penQLIK(K))$ for $1 \leq K \leq K_{\max} = 15$. One can see that the estimated number of segments is $\widehat{K}_n = 3$. Also, the estimated instants of breaks vector is $\widehat{\underline{t}}_n = (291, 702)$ $(\underline{t}^* = (300, 700))$ (see Figure 2).

Now, we are going to generate 100 replications of a Poisson-INGARCH(1,1) process following the scenarios **IA**$_0$-**IA**$_2$. Table 1 indicates the frequencies of the number of replications where $\widehat{K}_n = K^*$, $\widehat{K}_n < K^*$ and $\widehat{K}_n > K^*$, for the regularization parameter $\kappa_n = \widehat{\kappa}_n,\ \log n,\ n^{1/3}$. For the scenarios **IA**$_1$ and **IA**$_2$, we also consider the replications where the true number of breaks is achieved (i.e., $\widehat{K}_n = K^*$) and we present some elementary statistics of the estimated instants of breaks (see Table 1).

The results in Table 1 show that for the penalties considered, the performances increase with n in all scenarios. In accordance with Theorem 3.1, the consistency of the penalties $\log n$ and $n^{1/3}$ is numerically convincing. Moreover, the $n^{1/3}-$penalty outperforms the other procedures when $n = 1000$.

4.2.2 Poisson-INGARCH models

We consider the problem (1.2) for a Poisson-INGARCH(1,1); i.e., $(Y_1, \cdots, Y_n)$ is a trajectory of the process $Y = \{Y_t,\, t \in \mathbb{Z}\}$ satisfying:

$$Y_t|\mathcal{F}_{t-1} \sim \mathcal{P}(\lambda_t) \ ; \ \lambda_t = \alpha_0^{(j)} + \alpha^{(j)} Y_{t-1} + \beta^{(j)} \lambda_{t-1}, \quad \forall\, t \in T^*_j, \ \ \forall\, j \in \{1, \cdots, K^*\}\,. \tag{4.2}$$

The parameters vector is $\theta^*_j = (\alpha_0^{(j)}, \alpha^{(j)}, \beta^{(j)})$, for all $j \in \{1, \cdots, K^*\}$.

For $n = 500$ and $n = 1000$, we generate 100 replications of the model (4.2) in the following situations:

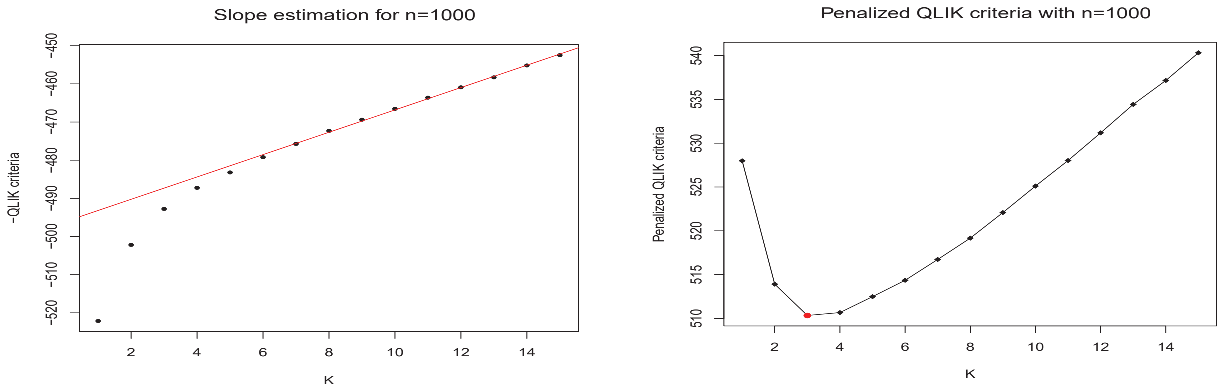

Figure 1: *A typical example for a Poisson-INARCH(1) process in the scenario* $\mathbf{IA}_2$ *with* $n = 1000$*. The left-hand graph is the curve of* $-\min_{\underline{t},\underline{\theta}} QLIK(K)$*, for* $1 \leq K \leq K_{\max}$*; the solid line represents the linear part of this curve with the slope* $\widehat{\kappa}_n/2 = 2.928$*. This value of* $\widehat{\kappa}_n$ *depends on this particular realization. The right-hand side is the graph* $(K, \min_{\underline{t},\underline{\theta}} penQLIK(K))$*, for* $1 \leq K \leq K_{\max}$*.*

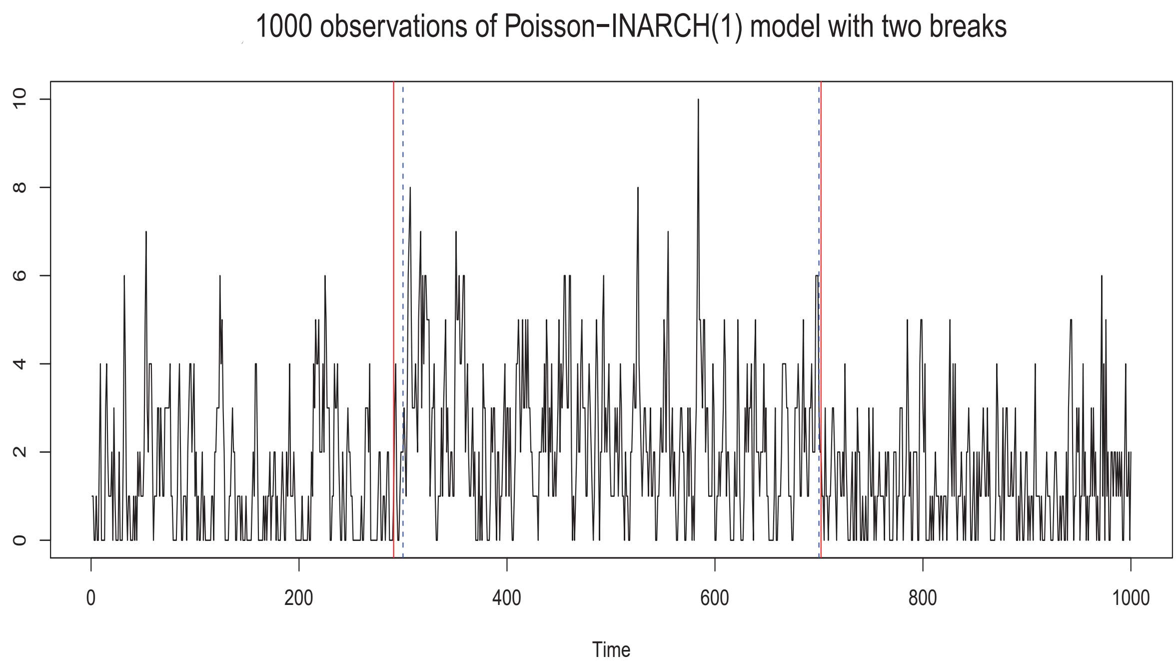

Figure 2: *The estimation of the break-points for a trajectory of a Poisson-INARCH(1) process in the scenario* $\mathbf{IA}_2$*. The solid lines represent the estimated instants of breaks and the dotted lines represent the true ones.*

- **scenario** $\mathbf{IG}_0$: $\theta_1^* = (1.0, 0.2, 0.15)$ is constant $(K^* = 1)$;
- **scenario** $\mathbf{IG}_1$: $\theta_1^* = (1.0, 0.2, 0.15)$ changes to $\theta_2^* = (1.0, 0.45, 0.15)$ at $t^* = 0.5n$ $(K^* = 2)$;
- **scenario** $\mathbf{IG}_2$: $\theta_1^* = (0.1, 0.3, 0.6)$ changes to $\theta_2^* = (0.5, 0.3, 0.6)$ at $t_1^* = 0.3n$ which changes to $\theta_3^* = (0.5, 0.3, 0.2)$ at $t_2^* = 0.7n$ $(K^* = 3)$.

Table 1: *Breaks estimated after 100 replications for a Poisson-INARCH(1) process following the scenarios* $\mathbf{IA}_0$-$\mathbf{IA}_2$*. The first three columns show the frequencies of the estimation of the true, low and high number of breaks. The last three columns give some elementary statistics of the change-point locations when the true number of breaks is achieved.*

			Frequencies			Mean ± s.d.		Mean
Scenarios			$\widehat{K}_n = K^*$	$\widehat{K}_n < K^*$	$\widehat{K}_n > K^*$	$\widehat{\tau}_1$	$\widehat{\tau}_2$	$\lVert \underline{\widehat{\tau}}_n - \underline{\tau}^* \rVert$
$\mathbf{IA}_0$ ($K^*=1$)	$n=500$	$\kappa_n = \widehat{\kappa}_n$	0.72	0.00	0.28			
		$\kappa_n = \log n$	0.76	0.00	0.24			
		$\kappa_n = n^{1/3}$	0.94	0.00	0.06			
	$n=1000$	$\kappa_n = \widehat{\kappa}_n$	0.94	0.00	0.06			
		$\kappa_n = \log n$	0.90	0.00	0.10			
		$\kappa_n = n^{1/3}$	1.00	0.00	0.00			
$\mathbf{IA}_1$ ($K^*=2$)	$n=500$	$\kappa_n = \widehat{\kappa}_n$	0.76	0.03	0.21	0.497 ± 0.064		0.038
		$\kappa_n = \log n$	0.83	0.03	0.14	0.495 ± 0.066		0.040
		$\kappa_n = n^{1/3}$	0.87	0.09	0.04	0.495 ± 0.064		0.038
	$n=1000$	$\kappa_n = \widehat{\kappa}_n$	0.89	0.00	0.11	0.507 ± 0.033		0.019
		$\kappa_n = \log n$	0.87	0.00	0.13	0.507 ± 0.034		0.020
		$\kappa_n = n^{1/3}$	0.98	0.00	0.02	0.506 ± 0.032		0.019
$\mathbf{IA}_2$ ($K^*=3$)	$n=500$	$\kappa_n = \widehat{\kappa}_n$	0.62	0.13	0.25	0.311 ± 0.071	0.689 ± 0.060	0.061
		$\kappa_n = \log n$	0.73	0.12	0.15	0.317 ± 0.073	0.690 ± 0.072	0.067
		$\kappa_n = n^{1/3}$	0.64	0.33	0.03	0.310 ± 0.058	0.685 ± 0.070	0.061
	$n=1000$	$\kappa_n = \widehat{\kappa}_n$	0.87	0.00	0.13	0.300 ± 0.034	0.693 ± 0.030	0.034
		$\kappa_n = \log n$	0.84	0.00	0.16	0.302 ± 0.043	0.692 ± 0.030	0.038
		$\kappa_n = n^{1/3}$	0.93	0.05	0.02	0.300 ± 0.051	0.694 ± 0.028	0.037

Table 2: *Breaks estimated after 100 replications for a Poisson-INGARCH(1,1) process following the scenarios* $\mathbf{IG}_0$-$\mathbf{IG}_2$*. The first three columns show the frequencies of the estimation of the true, low and high number of breaks. The last three columns give some elementary statistics of the change-point locations when the true number of breaks is achieved.*

			Frequencies			Mean ± s.d.		Mean
Scenarios			$\widehat{K}_n = K^*$	$\widehat{K}_n < K^*$	$\widehat{K}_n > K^*$	$\widehat{\tau}_1$	$\widehat{\tau}_2$	$\lVert \underline{\widehat{\tau}}_n - \underline{\tau}^* \rVert$
$\mathbf{IG}_0$ ($K^*=1$)	$n=500$	$\kappa_n = \widehat{\kappa}_n$	0.86	0.00	0.14			
		$\kappa_n = \log n$	0.96	0.00	0.04			
		$\kappa_n = n^{1/3}$	1.00	0.00	0.00			
	$n=1000$	$\kappa_n = \widehat{\kappa}_n$	0.92	0.00	0.08			
		$\kappa_n = \log n$	0.96	0.00	0.04			
		$\kappa_n = n^{1/3}$	1.00	0.00	0.00			
$\mathbf{IG}_1$ ($K^*=2$)	$n=500$	$\kappa_n = \widehat{\kappa}_n$	0.75	0.05	0.20	0.515 ± 0.066		0.038
		$\kappa_n = \log n$	0.70	0.03	0.27	0.514 ± 0.073		0.040
		$\kappa_n = n^{1/3}$	0.78	0.06	0.16	0.512 ± 0.066		0.038
	$n=1000$	$\kappa_n = \widehat{\kappa}_n$	0.75	0.05	0.20	0.507 ± 0.031		0.019
		$\kappa_n = \log n$	0.58	0.00	0.42	0.508 ± 0.034		0.021
		$\kappa_n = n^{1/3}$	0.83	0.03	0.13	0.501 ± 0.048		0.022
$\mathbf{IG}_2$ ($K^*=3$)	$n=500$	$\kappa_n = \widehat{\kappa}_n$	0.53	0.41	0.06	0.299 ± 0.078	0.691 ± 0.073	0.053
		$\kappa_n = \log n$	0.58	0.23	0.19	0.299 ± 0.074	0.693 ± 0.070	0.049
		$\kappa_n = n^{1/3}$	0.37	0.49	0.14	0.300 ± 0.076	0.697 ± 0.015	0.047
	$n=1000$	$\kappa_n = \widehat{\kappa}_n$	0.62	0.26	0.12	0.293 ± 0.050	0.702 ± 0.010	0.025
		$\kappa_n = \log n$	0.60	0.06	0.34	0.293 ± 0.051	0.702 ± 0.010	0.026
		$\kappa_n = n^{1/3}$	0.56	0.29	0.15	0.301 ± 0.029	0.699 ± 0.011	0.016

Table 2 indicates the frequencies of the true number of breaks estimated and some elementary statistics of the estimators of the change-point locations. It appears that the results of the $n^{1/3}$-penalty and the slope procedure are quite satisfactory except for the case of two breaks. In this later case, the $n^{1/3}$-penalty and the slope procedure over-penalize the number of breaks, while the $\log n$-penalty under-penalizes. But, overall, the performances of the proposed procedures increase with n and the estimation of the break-points locations is well achieved.

4.2.3 Negative binomial INGARCH models

We consider the problem (1.2) for a negative binomial INGARCH(1,1) (NB-INGARCH(1,1)); i.e., $(Y_1, \cdots, Y_n)$ is a trajectory of the process $Y = \{Y_t,\, t \in \mathbb{Z}\}$ satisfying:

$$Y_t|\mathcal{F}_{t-1} \sim NB(r, p_t) \;;\; r\frac{(1-p_t)}{p_t} = \lambda_t = \alpha_0^{(j)} + \alpha^{(j)} Y_{t-1} + \beta^{(j)} \lambda_{t-1}, \quad \forall\, t \in T_j^*, \quad \forall\, j \in \{1, \cdots, K^*\}\,; \tag{4.3}$$

where the parameters vector is $\theta_j^* = (\alpha_0^{(j)}, \alpha^{(j)}, \beta^{(j)})$, for all $j \in \{1, \cdots, K^*\}$ and $NB(r,p)$ denotes the negative binomial distribution with parameters r and p. Here, we use the probability mass function of $NB(r,p)$ given by

$$\mathbb{P}(Y = y) = \binom{y+r-1}{r-1} p^r (1-p)^y, \text{ for all } y = 0, 1, \ldots$$

For $r = 14$ (used for transaction data, see Diop and Kengne (2017)), $n = 500$ and $n = 1000$, we generate a sample $(Y_1, \cdots, Y_n)$ in the following situations:

- **scenario NB-IG**$_0$: $\theta_1^* = (1.0, 0.2, 0.15)$ is constant $(K^* = 1)$;
- **scenario NB-IG**$_1$: $\theta_1^* = (1, 0.2, 0.15)$ changes to $\theta_2^* = (1, 0.45, 0.15)$ at $t^* = 0.5n$ $(K^* = 2)$;
- **scenario NB-IG**$_2$: $\theta_1^* = (0.1, 0.3, 0.6)$ changes to $\theta_2^* = (0.5, 0.3, 0.6)$ at $t_1^* = 0.3n$ which changes to $\theta_3^* = (0.5, 0.3, 0.2)$ at $t_2^* = 0.7n$ $(K^* = 3)$.

Table 3: *Breaks estimated after 100 replications for a NB-INGARCH(1,1) process following the scenarios* $\mathbf{NB-IG}_0$-$\mathbf{NB-IG}_2$. *The first three columns show the frequencies of the estimation of the true, low and high number of breaks. The last three columns give some elementary statistics of the change-point locations when the true number of breaks is achieved.*

Scenarios			Frequencies $\widehat{K}_n = K^*$	$\widehat{K}_n < K^*$	$\widehat{K}_n > K^*$	Mean ± s.d. $\widehat{\tau}_1$	$\widehat{\tau}_2$	Mean $\lVert\widehat{\underline{\tau}}_n - \underline{\tau}^*\rVert$
NB-IG$_0$ ($K^*=1$)	$n=500$	$\kappa_n = \widehat{\kappa}_n$	0.90	0.00	0.10			
		$\kappa_n = \log n$	0.95	0.00	0.05			
		$\kappa_n = n^{1/3}$	0.98	0.00	0.02			
	$n=1000$	$\kappa_n = \widehat{\kappa}_n$	0.92	0.00	0.08			
		$\kappa_n = \log n$	0.94	0.00	0.06			
		$\kappa_n = n^{1/3}$	0.98	0.00	0.02			
NB-IG$_1$ ($K^*=2$)	$n=500$	$\kappa_n = \widehat{\kappa}_n$	0.60	0.09	0.31	0.512 ± 0.122		0.072
		$\kappa_n = \log n$	0.55	0.04	0.41	0.514 ± 0.110		0.063
		$\kappa_n = n^{1/3}$	0.65	0.12	0.23	0.519 ± 0.106		0.060
	$n=1000$	$\kappa_n = \widehat{\kappa}_n$	0.69	0.02	0.29	0.507 ± 0.065		0.037
		$\kappa_n = \log n$	0.56	0.00	0.44	0.500 ± 0.054		0.030
		$\kappa_n = n^{1/3}$	0.83	0.02	0.15	0.505 ± 0.061		0.037
NB-IG$_2$ ($K^*=3$)	$n=500$	$\kappa_n = \widehat{\kappa}_n$	0.45	0.51	0.04	0.330 ± 0.084	0.696 ± 0.040	0.057
		$\kappa_n = \log n$	0.41	0.13	0.46	0.319 ± 0.080	0.700 ± 0.026	0.057
		$\kappa_n = n^{1/3}$	0.43	0.28	0.29	0.328 ± 0.061	0.685 ± 0.066	0.055
	$n=1000$	$\kappa_n = \widehat{\kappa}_n$	0.70	0.25	0.05	0.304 ± 0.060	0.699 ± 0.020	0.033
		$\kappa_n = \log n$	0.54	0.02	0.44	0.299 ± 0.066	0.699 ± 0.014	0.033
		$\kappa_n = n^{1/3}$	0.68	0.19	0.13	0.325 ± 0.090	0.697 ± 0.018	0.043

Once again, it appears in Table 3 that the performances of the proposed procedures increase with n and the estimation of the break-points locations remain satisfactory even in this case where the Poisson quasi-likelihood used is quite different from the true distribution of the observations.

4.2.4 Binary Time Series

Consider the problem (1.2) for a binary INARCH(1) (BIN-INARCH(1)) time series model; i.e., $(Y_1, \cdots, Y_n)$ is a trajectory of the process $Y = \{Y_t, t \in \mathbb{Z}\}$ satisfying:

$$Y_t|\mathcal{F}_{t-1} \sim \mathcal{B}(p_t) \ ; \ p_t = \lambda_t = \alpha_0^{(j)} + \alpha^{(j)} Y_{t-1}, \ \ \forall \, t \in T_j^*, \ \ \forall \, j \in \{1, \cdots, K^*\} \, ; \tag{4.4}$$

where the parameters vector is $\theta_j^* = (\alpha_0^{(j)}, \alpha^{(j)})$ (with $0 < \alpha_0^{(j)} + \alpha^{(j)} < 1$), for all $j \in \{1, \cdots, K^*\}$ and $\mathcal{B}(p)$ denotes the Bernoulli distribution with parameter p.
For $n = 500$ and $n = 1000$, we generate a sample $(Y_1, \cdots, Y_n)$ in the following situations:

- **scenario BIN-IA$_0$**: $\theta_1^* = (0.15, 0.75)$ is constant ($K^* = 1$) ;
- **scenario BIN-IA$_1$**: $\theta_1^* = (0.15, 0.75)$ changes to $\theta_2^* = (0.04, 0.60)$ at $t^* = 0.5n$ ($K^* = 2$) ;
- **scenario BIN-IA$_2$**: $\theta_1^* = (0.15, 0.75)$ changes to $\theta_2^* = (0.04, 0.60)$ at $t_1^* = 0.3n$ which changes to $\theta_3^* = (0.25, 0.35)$ at $t_2^* = 0.7n$ ($K^* = 3$).

The scenario **BIN-IA$_1$** is related and close to the real data example (see below).

Table 4: *Breaks estimated after 100 replications for a BIN-INARCH(1) process following the scenarios* $\mathbf{BIN-IA}_0$*-*$\mathbf{BIN-IA}_2$*. The first three columns show the frequencies of the estimation of the true, low and high number of breaks. The last three columns give some elementary statistics of the change-point locations when the true number of breaks is achieved.*

Scenarios			Frequencies $\widehat{K}_n = K^*$	$\widehat{K}_n < K^*$	$\widehat{K}_n > K^*$	Mean ± s.d. $\widehat{\tau}_1$	$\widehat{\tau}_2$	Mean $\lVert\widehat{\underline{\tau}}_n - \underline{\tau}^*\rVert$
BIN-IA$_0$ ($K^* = 1$)	$n = 500$	$\kappa_n = \widehat{\kappa}_n$	0.84	0.00	0.16			
		$\kappa_n = \log n$	0.98	0.00	0.02			
		$\kappa_n = n^{1/3}$	0.98	0.00	0.02			
	$n = 1000$	$\kappa_n = \widehat{\kappa}_n$	0.86	0.00	0.14			
		$\kappa_n = \log n$	1.00	0.00	0.00			
		$\kappa_n = n^{1/3}$	1.00	0.00	0.00			
BIN-IA$_1$ ($K^* = 2$)	$n = 500$	$\kappa_n = \widehat{\kappa}_n$	0.69	0.10	0.21	0.499 ± 0.091		0.055
		$\kappa_n = \log n$	0.72	0.27	0.01	0.491 ± 0.087		0.051
		$\kappa_n = n^{1/3}$	0.52	0.48	0.00	0.484 ± 0.091		0.054
	$n = 1000$	$\kappa_n = \widehat{\kappa}_n$	0.89	0.00	0.11	0.499 ± 0.036		0.020
		$\kappa_n = \log n$	0.96	0.01	0.03	0.500 ± 0.035		0.019
		$\kappa_n = n^{1/3}$	0.85	0.15	0.00	0.484 ± 0.091		0.054
BIN-IA$_2$ ($K^* = 3$)	$n = 500$	$\kappa_n = \widehat{\kappa}_n$	0.75	0.18	0.07	0.324 ± 0.094	0.695 ± 0.044	0.060
		$\kappa_n = \log n$	0.43	0.57	0.00	0.312 ± 0.044	0.694 ± 0.030	0.035
		$\kappa_n = n^{1/3}$	0.23	0.77	0.00	0.306 ± 0.033	0.702 ± 0.017	0.026
	$n = 1000$	$\kappa_n = \widehat{\kappa}_n$	0.95	0.03	0.02	0.303 ± 0.046	0.696 ± 0.019	0.028
		$\kappa_n = \log n$	0.90	0.10	0.00	0.299 ± 0.037	0.697 ± 0.017	0.025
		$\kappa_n = n^{1/3}$	0.52	0.48	0.00	0.296 ± 0.021	0.697 ± 0.019	0.020

Table 4 shows that the procedures provide satisfactory results with BIN-INARCH(1) model, except that the $n^{1/3}$-penalty in the case of two breaks. But, the performances of these procedures increase with n and the break-points locations are overall well estimated.

4.2.5 INARCH(∞) models

Now, consider a Poisson-INARCH(∞); i.e., $(Y_1, \cdots, Y_n)$ is a trajectory of the process $Y = \{Y_t, t \in \mathbb{Z}\}$ satisfying:

$$Y_t|\mathcal{F}_{t-1} \sim \mathcal{P}(\lambda_t) \ \ ; \ \ \lambda_t = \alpha_0^{(j)} + \sum_{k=1}^{\infty} \alpha_k Y_{t-k}, \quad \forall \, t \in T_j^*, \ \ \forall \, j \in \{1, \cdots, K^*\} \, , \tag{4.5}$$

where $\alpha_0^{(j)} > 0$ (for all $j = 1, \cdots, K^*$), $\alpha_k \geq 0$ (for all $k \geq 1$) and $\sum_{k\geq 1} \alpha_k < 1$; that is, we focus on the change in the parameter $\alpha_0^{(j)}$. This process corresponds to a particular case of the problem (1.2) with $f(y_1, y_2, \cdots, \alpha_0^{(j)}) = \alpha_0^{(j)} + \sum_{k=1}^{\infty} \alpha_k y_k$ for each regime $j \in \{1, \cdots, K^*\}$. We deal with a scenario where the consistency of the BIC procedure is not ensured. Therefore, we consider the Riemanian case with $\alpha_k = \mathcal{O}(k^{-1.7})$ (in the scenario detailed below). More precisely, we consider the model (4.5) with

$$\lambda_t = \alpha_0^{(j)} + \frac{1}{2.2} \sum_{k=1}^{\infty} \frac{1}{k^{1.7}} Y_{t-k}.$$

The number $1/2.2$ is obtained from the values of the Riemann zeta function, and allows the condition $\frac{1}{2.2} \sum_{k=1}^{\infty} \frac{1}{k^{1.7}} < 1$. According to Theorem 3.1 and Remark 3.4, if the regularization parameter verifies $\kappa_n = \mathcal{O}(n^{\delta})$ with $\delta > 0.3$, then the consistency holds. Thus, the consistency of the BIC penalty is not ensured.

For $n = 500$ and $n = 100$, we generate a trajectory $(Y_1, \cdots, Y_n)$ of the model (4.5) in the following scenarios:

- **scenario IA-INF**$_0$: $\alpha_0^{(1)} = 0.5$ is constant ($K^* = 1$) ;
- **scenario IA-INF**$_1$: $\alpha_0^{(1)} = 0.5$ changes to $\alpha_0^{(2)} = 0.1$ at $t^* = 0.5n$ ($K^* = 2$).

Table 5: *Breaks estimated after 100 replications for an INARCH(∞) process following the scenarios* $\mathbf{IA} - \mathbf{INF}_0$ *and* $\mathbf{IA} - \mathbf{INF}_1$. *The first three columns show the frequencies of the estimation of the true, low and high number of breaks. The last two columns give some elementary statistics of the change-point locations when the true number of breaks is achieved.*

Scenarios			Frequencies $\widehat{K}_n = K^*$	$\widehat{K}_n < K^*$	$\widehat{K}_n > K^*$	Mean ± s.d. $\widehat{\tau}_1$	Mean $\Vert \widehat{\underline{\tau}}_n - \underline{\tau}^* \Vert$
IA-INF$_0$ ($K^* = 1$)	$n = 500$	$\kappa_n = \widehat{\kappa}_n$	0.15	0.00	0.85		
		$\kappa_n = \log n$	0.56	0.00	0.44		
		$\kappa_n = n^{1/3}$	0.84	0.00	0.16		
	$n = 1000$	$\kappa_n = \widehat{\kappa}_n$	0.17	0.00	0.83		
		$\kappa_n = \log n$	0.57	0.00	0.43		
		$\kappa_n = n^{1/3}$	0.95	0.00	0.05		
IA-INF$_1$ ($K^* = 2$)	$n = 500$	$\kappa_n = \widehat{\kappa}_n$	0.66	0.04	0.34	0.498 ± 0.0057	0.003
		$\kappa_n = \log n$	0.49	0.02	0.49	0.498 ± 0.0013	0.002
		$\kappa_n = n^{1/3}$	0.82	0.03	0.15	0.497 ± 0.0062	0.003
	$n = 1000$	$\kappa_n = \widehat{\kappa}_n$	0.79	0.00	0.21	0.499 ± 0.0004	0.001
		$\kappa_n = \log n$	0.43	0.00	0.57	0.499 ± 0.0006	0.001
		$\kappa_n = n^{1/3}$	0.93	0.00	0.07	0.499 ± 0.0005	0.001

In Table 5, one can see that the $n^{1/3}$-penalty uniformly outperforms the other two procedures. Moreover, the performances of the proposed procedures increase with n, except the $\log n$-penalty whose the performances decrease with n. Hence, the consistency of the BIC procedure is quite questionable in this case.

5 Real data application

We apply our change-point procedure to two examples of real data series. To compute the estimator $\widehat{K}_n$, the $\widehat{\kappa}_n$-penalty is used with $u_n = \left[(\log(n))^{\delta}\right]$ (where $3/2 \leq \delta \leq 2$) and $K_{\max} = 15$.

5.1 The US recession data

Firstly, we consider the series of the quarterly recession data from the USA for the period 1855-2013 (see Figure 4). This series (Y_t) represents a binary variable that is equal to 1 if there is a recession in at least 1 month in the quarter and 0 otherwise. There are 636 quarterly observations obtained from The National Bureau of Economic Research. These data have already been analyzed by several authors. Hudecová (2013) has applied a change-point procedure based on a normalized cumulative sums of residuals and has found a break in the first quarter of 1933. Recently, Diop and Kengne (2017) have applied a change-point test based on the maximum likelihood estimator of the model's parameter and have detected a break in the last quarter of 1932.

We consider the INARCH(1) representation and apply the *penQLIK* contrast procedure. This choice is motivated by the fact that the estimation of the last component of θ (i.e., the parameter β) is not significant in the INGARCH(1,1) representation (see Diop and Kengne (2017)). The test of nullity of one coefficient (TNOC) proposed by Ahmad and Francq (2016), applied a posteriori (after change-point detection) also confirms these results. As noted in the implementation of the dynamic programming algorithm, we begin by the calibration of the regularization parameter κ_n. The slope estimation procedure applied with $u_n = \left[(\log n)^2\right]$ returns the value $\widehat{\kappa}_n \approx 3.21$ and the estimation of the number of segments is $\widehat{K}_n = 2$; i.e., one break is detected (see Figure 3). The location of the break-point estimated is $\widehat{t} = 313$. The change detected at $\widehat{t} = 313$ corresponds to the first quarter of 1933 (see Figure 4). These results are in concordance with those obtained by Diop and Kengne (2017) and Hudecová (2013). The estimated model with one break-point is

$$\mathbb{E}(Y_t|\mathcal{F}_{t-1}) = \lambda_t = \begin{cases} \underset{(0.028)}{0.120} + \underset{(0.215)}{0.749} Y_{t-1}, & \text{for } t \leq 313 \\ \\ \underset{(0.013)}{0.047} + \underset{(0.230)}{0.681} Y_{t-1}, & \text{for } t > 313, \end{cases}$$

where in parentheses are the standard errors of the estimators obtained from the robust sandwich matrix $\widehat{H}_j^{-1}\widehat{\Sigma}_j^{-1}\widehat{H}_j^{-1}$ computed on each regime j, with $\widehat{\Sigma}_j$ is given by (2.7) and $\widehat{H}_j = \frac{1}{n}\sum_{t=1}^{n} \frac{\partial^2 \widehat{\ell}_{t,j}(\widehat{\theta}_n(\widehat{T}_j))}{\partial\theta\partial\theta'}$. The estimation of the parameters displays a distortion in term of standard errors; it can be explained by the fact that the true distribution of the observations (which is binary), is quite different from the Poisson quasi-likelihood used.

5.2 Number of trades in the stock of Technofirst

Secondly, we apply our change-point detection procedure to a financial time series data. We consider the daily number of trades in the stock of Technofirst listed in the NYSE Euronext group. It is a series of 1000 observations from 04 January 2010 to 20 April 2016 (see Figure 6). The data are available online at the website "https://www.euronext.com/en/products/equities/FR0011651819-ALXP". These data have been analyzed by Ahmad and Francq (2016) with the PQMLE, and have concluded that the INGARCH(1,3) is more appropriate. We carry out an INGARCH(1,1) representation with the possibility of change in the observations.

The slope estimation procedure obtained with $u_n = \left[(\log(n))^2\right]$ returns $\widehat{\kappa}_n \approx 23.04$ and the estimation of the number of segments is $\widehat{K}_n = 3$; i.e., two changes are detected (see Figure 5). The locations of the break-points estimated are $\widehat{t}_1 = 230$ (06 April 2011) and $\widehat{t}_2 = 311$ (06 September 2011), see also Figure 6.

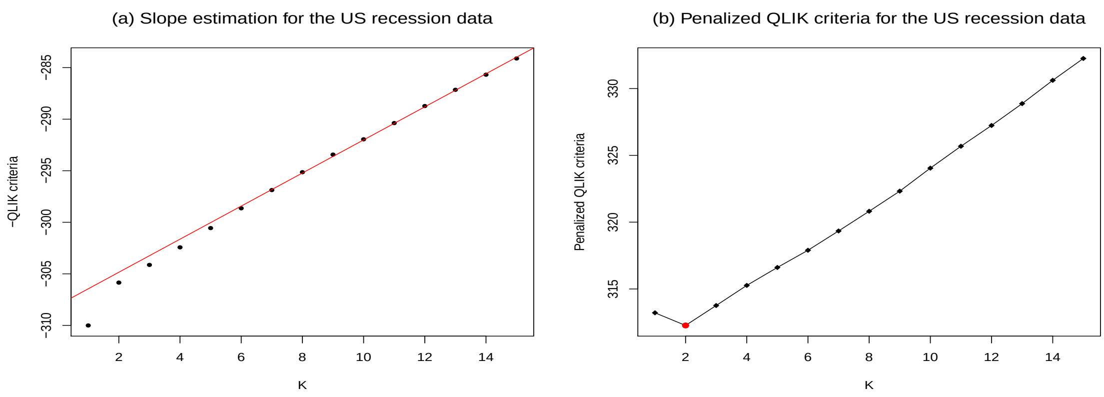

Figure 3: *The curve of* $-\min_{\underline{t},\underline{\theta}} QLIK(K)$ *and the graph* $(K, \min_{\underline{t},\underline{\theta}} penQLIK(K))$ *for the US recession data with an INARCH(1) representation. The solid line represents the linear part of this curve with slope* $\widehat{\kappa}_n/2 = 1.605$.

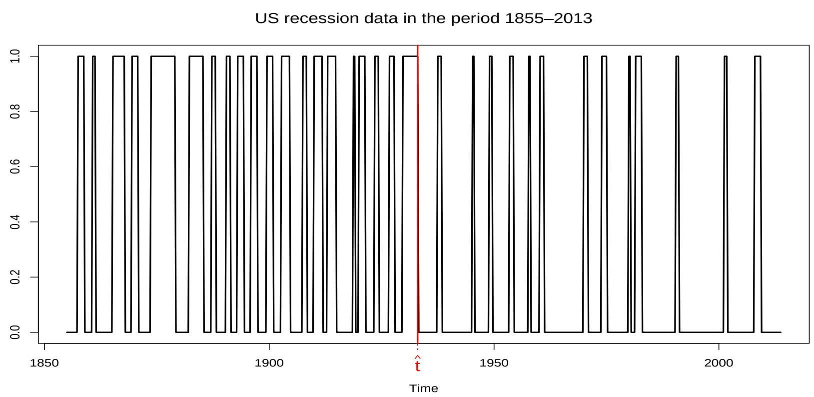

Figure 4: *The US recession data with the estimated location of the break-point* $\widehat{t}$.

The estimated model with the change-points is

$$\mathbb{E}(Y_t|\mathcal{F}_{t-1}) = \lambda_t = \begin{cases} \begin{array}{l} 2.436 + 0.368 Y_{t-1}, \text{ for } t \leq 230 \\ (0.126) \quad (0.032) \\ \\ 4.643 + 0.607 Y_{t-1} + 0.032 \lambda_{t-1}, \text{ for } 230 < t \leq 311 \\ (0.649) \quad (0.022) \qquad (0.033) \\ \\ 1.113 + 0.166 Y_{t-1} + 0.531 \lambda_{t-1}, \text{ for } t > 311, \\ (0.226) \quad (0.016) \qquad (0.071) \end{array} \end{cases}$$

where in parentheses are the robust standard errors of the estimators. Let us note that, we have applied the TNOC, which fund that the INARCH(1) representation is the most appropriate for the first regime.

6 Summary and conclusion

This paper focuses on the multiple change-point problem in a general class of integer-valued time series. A penalized contrast estimator based on the Poisson quasi-maximum likelihood of the model is proposed. The

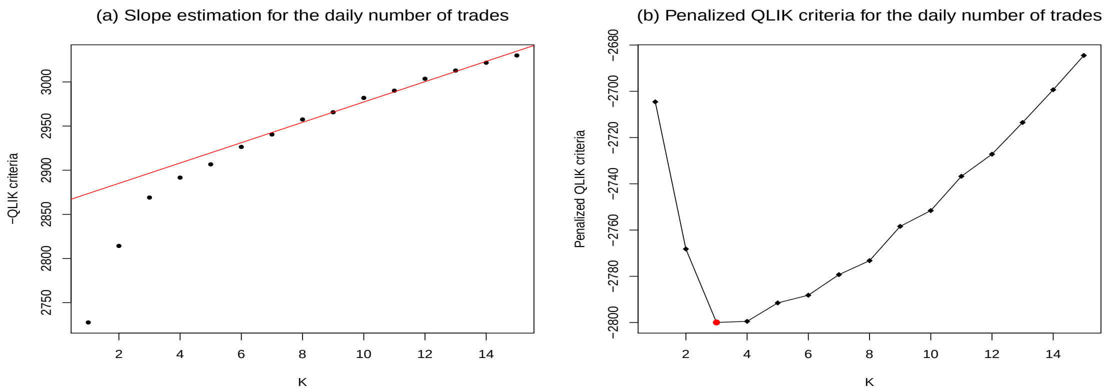

Figure 5: *The curve of* $-\min_{\underline{t},\underline{\theta}} QLIK(K)$ *and the graph* $(K, \min_{\underline{t},\underline{\theta}} penQLIK(K))$ *for the daily number of trades in the stock of Technofirst. The solid line represents the linear part of this curve with slope* $\widehat{\kappa}_n/2 = 11.52$.

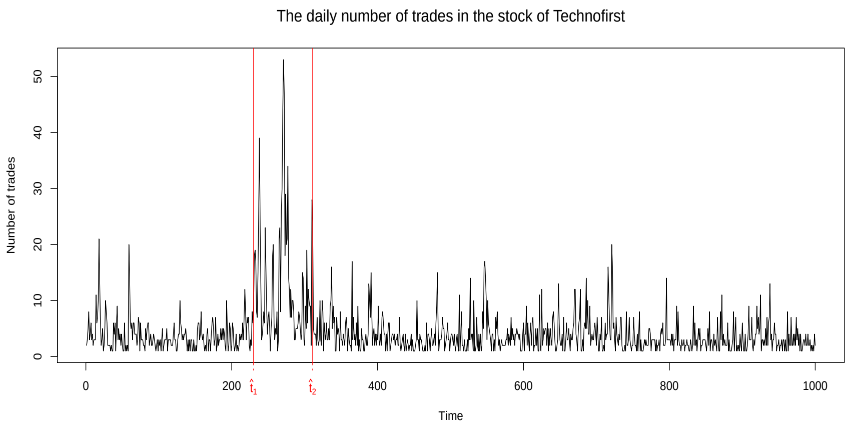

Figure 6: *The daily number of trades in the stock of the Technofirst with the estimated locations of the breakpoints* $\widehat{t}_1$ *and* $\widehat{t}_2$.

theoretical study establishes the consistency of the proposed estimator. A data-driven procedure based on the slope heuristic is also proposed to calibrate the penalty term of the contrast. The simulation study based on three penalty procedures (BIC, $n^{1/3}$ and slope heuristic) displays satisfactory results in the cases of Poisson, Negative binomial and binary INARCH (except the $n^{1/3}$-penalty in the case of two breaks for BIN-INARCH); these models are the most used in practice. The performances of these procedures increase with n and the estimated probability of detecting the true number of breaks is overall approaching 1 when $n = 1000$; this is consistent with the results of Theorem 3.1. In the case of the Poisson-INARCH(∞), the $n^{1/3}$-penalty outperforms the other two procedures. This is not too surprising, since according to Theorem 3.1, the consistence of the BIC procedure is not ensured in this case. Applications to real data examples show that these procedures work in practice. Also, for all the scenarios with break in Section 4 as well as for the real data examples, the slope heuristic procedure provides satisfactory results. Such data-driven procedure is today one of the most successful methods for practical calibration in the model selection problems. A possible extension of this work is the study of the theoretical validity of a data-driven calibration procedure in such models.

Acknowledgements

The authors are grateful to the Executive Editors, Co-Editors and the two anonymous Referees for many relevant suggestions and comments which helped to improve the contents of this article.

Appendix. Supplementary material: Proofs of the main results

This Supplementary material may be found in the online version of this article.

References

[1] Ahmad, A. and Francq, C. Poisson QMLE of count time series models. *Journal of Time Series Analysis 37*, (2016), 291-314.

[2] Arlot, S., Celisse, A. and Celisse, Z. A kernel multiple change-point algorithm via model selection. *arXiv:1202.3878v2 : 2016*.

[3] Arlot, S. and Massart, P. Data-driven Calibration of Penalties for Least-Squares Regression. *Journal of Machine Learning Research 10*, (2009), 245-279.

[4] Bai, J. Estimating multiple breaks one at a time. *Econometric theory 13 (3)*, (1997), 315-352.

[5] Bai, J. and Perron, P. Estimating and testing linear models with multiple structural changes. *Econometrica 66*, (1998), 47-78.

[6] Bardet, J. M., Kengne, K. and Wintenberger, O. Multiple breaks detection in general causal time series using penalized quasi-likelihood. *Electronic Journal of Statistics 6*, (2012), 435-477.

[7] Baudry, J. P., and Maugis, C. and Michel, B. Slope Heuristics: overview and implementation. *RR-INRIA n^o 7223*, 2010.

[8] Cleynen, A. and Lebarbier, E. Segmentation of the Poisson and negative binomial rate models: a penalized estimator. *ESAIM: Probability and Statistics 18*, (2014), 750-769.

[9] Cleynen, A. and Lebarbier, E. Model selection for the segmentation of multiparameter exponential family distributions. *Electronic Journal of Statistics 11*, (2017), 800-842.

[10] Davis, R.A. and Liu, H. Theory and Inference for a Class of Observation-Driven Models with Application to Time Series of Counts. *Statistica Sinica 26*, (2016) 1673-1707.

[11] Davis, R. A., Hancock, S. A. and Yao, Y.-C. On consistency of minimum description length model selection for piecewise autoregressions. *Journal of Econometrics 194*, (2016) 360-368.

[12] Davis, R. A., Lee, T. C. M. and Rodriguez-Yam, G. A. Break detection for a class of nonlinear time series models. *Journal of Time Series Analysis 29 (5)*, (2008), 834-867.

[13] Davis, R. A. and Yau, C. Y. Consistency of minimum description length model selection for piecewise stationary time series models. *Electronic Journal of Statistics 7*, (2013) 381-411.

[14] Diop, M.L. and Kengne, W. Testing parameter change in general integer-valued time series. *J. Time Ser. Anal. 38*, (2017) 880-894.

[15] Douc, R., Fokianos, K., and Moulines, E. Asymptotic properties of quasi-maximum likelihood estimators in observation-driven time series models. *Electronic Journal of Statistics 11*, (2017), 2707-2740.

[16] Doukhan, P., Fokianos, K., and Tjøstheim, D. On Weak Dependence Conditions for Poisson autoregressions. *Statist. and Probab. Letters 82*, (2012), 942-948.

[17] Doukhan, P., Fokianos, K., and Tjøstheim, D. Correction to "On weak dependence conditions for Poisson autoregressions" [Statist. Probab. Lett. 82 (2012) 942-948]. *Statist. and Probab. Letters 83*, (2013), 1926-1927.

[18] Doukhan, P. and Kengne, W. Inference and testing for structural change in general poisson autoregressive models. *Electronic Journal of Statistics 9*, (2015), 1267-1314.

[19] Fokianos, K., Rahbek, A. and Tjøstheim, D. Poisson autoregression. *Journal of the American Statistical Association 104*, (2009), 1430-1439.

[20] Fokianos, K. and Tjøstheim, D. Log-linear Poisson autoregression. *Journal of multivariate analysis 102*, (2011) 563-578.

[21] Fokianos, K., and Tjøstheim, D. Nonlinear Poisson autoregression. *Ann. Inst. Stat. Math. 64*, (2012) 1205-1225.

[22] Franke, J., Kirch, C. and Tadjuidje Kamgaing, J. Changepoints in times series of counts. *J. Time Ser. Anal. 33*, (2012) 757-770.

[23] Fryzlewicz, P. Wild binary segmentation for multiple change-point detection. *The Annals of Statistics 42 (6)*, (2014) 2243-228.

[24] Fryzlewicz, P., and Subba Rao, S. Multiple-change-point detection for auto-regressive conditional heteroscedastic processes. *Journal of the Royal Statistical Society: series B 76 (5)*, (2014) 903-924.

[25] Harchaoui, Z. and Lévy-Leduc, C. Multiple Change-Point Estimation With a Total Variation Penalty. *Journal of the American Statistical Association 105*, (2010) 1480-1493.

[26] Hudecová, Š. Structural changes in autoregressive models for binary time series. *Journal of Statistical Planning and Inference 143*, (2013) 1744-1752.

[27] Inclán, C. and Tiao, G. C. Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association 89 (427)*, (1994), 913-923.

[28] Kang, J. and Lee, S. Parameter change test for Poisson autoregressive models. *Scandinavian Journal of Statistics, 41(4)* , (2014), 1136-1152.

[29] Kashikar, A. S., Rohan, N., and Ramanathan, T. V. Integer autoregressive models with structural breaks. *Journal of Applied Statistics, 40(12)* , (2013), 2653-2669.

[30] Lavielle, M. and Ludena, C. The multiple change-points problem for the spectral distribution. *Bernoulli 6*, (2000), 845-869.

[31] Lebarbier E. Detecting multiple change-points in the mean of Gaussian process by model selection. *Signal Processing 85*, (2005), 717-736.

[32] Lerasle, M. Optimal model selection for density estimation of stationary data under various mixing conditions. *The Annals of Statistics 39 (4)*, (2011), 1852-1877.

[33] Leung, S. H., Ng, W. L., and Yau, C. Y. Sequential change-point detection in time series models based on pairwise likelihood. *Statistica Sinica 27(2)*, (2017), 575-605.

[34] Ma, T. F. and Yau, C. Y. A pairwise likelihood-based approach for changepoint detection in multivariate time series models. *Biometrika, 103(2)*, (2016), 409-421.

[35] Yau, C.Y. and Zhao, Z. Inference for multiple change points in time series via likelihood ratio scan statistics. *Journal of the Royal Statistical Society: Series B 78*, (2016), 895-916.