TEST (2023) 32:1–33
https://doi.org/10.1007/s11749-022-00824-z

ORIGINAL PAPER

A general procedure for change-point detection in multivariate time series

Mamadou Lamine Diop[1] · **William Kengne**[1]

Received: 21 November 2021 / Accepted: 28 June 2022 / Published online: 21 July 2022

Abstract

We consider the change-point detection in a general class of time series models, including multivariate continuous and integer- valued time series. We propose a Wald-type statistic based on the estimator performed by a general contrast function, which can be constructed from the likelihood, a quasi-likelihood, a least squares method, etc. Sufficient conditions are provided to ensure that the test statistic convergences to a well-known distribution under the null hypothesis (of no change) and diverges to infinity under the alternative, which establishes the consistency of the procedure. Some examples of models are detailed to illustrate the scope of application of the proposed change-point detection tool. The procedure is applied to simulated and real data examples for numerical illustration.

Keywords Change-point · Multivariate time series · Minimum contrast estimation · Consistency · Causal processes · Integer-valued time series

Mathematics Subject Classification 62M10 · 62F05 · 62F12

1 Introduction

Since Page (1955), the change-point problem has been widely studied. Several approaches and procedures have been developed for univariate and multivariate processes with continuous or integer- valued variables.

M. L. DIOP: Supported by the MME-DII center of excellence (ANR-11-LABEX-0023-01) and the ANR BREAKRISK: ANR-17-CE26-0001-01.
M. L. DIOP and W. KENGNE: Developed within the CY Initiative of Excellence (grant "Investissements d'Avenir" ANR-16-IDEX-0008), Project "EcoDep" PSI-AAP2020-0000000013.

✉ William Kengne
william.kengne@u-cergy.fr

Mamadou Lamine Diop
mamadou-lamine.diop@u-cergy.fr

1 THEMA, CY Cergy Paris Université, 33 Boulevard du Port, 95011 Cergy-Pontoise Cedex, France

Consider observations $(Y_1, \ldots, Y_n)$, generated from a multivariate continuous or integer-valued process $Y = \{Y_t,\ t \in \mathbb{Z}\}$. These observations depend on a parameter $\theta^* \in \Theta \subset \mathbb{R}^d$ $(d \in \mathbb{N})$ which may change over time. More precisely, consider the following test hypotheses:

- H_0: $(Y_1, \ldots, Y_n)$ is a trajectory of the process $Y = \{Y_t,\ t \in \mathbb{Z}\}$ which depends on θ^*.
- H_1: There exists $((\theta_1^*, \theta_2^*), t^*) \in \Theta^2 \times \{2, 3, \ldots, n-1\}$ (with $\theta_1^* \neq \theta_2^*$) such that $(Y_1, \ldots, Y_{t^*})$ is a trajectory of a process $Y^{(1)} = \{Y_t^{(1)},\ t \in \mathbb{Z}\}$ that depends on θ_1^* and $(Y_{t^*+1}, \ldots, Y_n)$ is a trajectory of a process $Y^{(2)} = \{Y_t^{(2)},\ t \in \mathbb{Z}\}$ that depends on θ_2^*.

Note that under H_1, $(Y_1, \ldots, Y_n)$ is a trajectory of the process $\{(Y_t^{(1)})_{t \leq t^*}, (Y_t^{(2)})_{t > t^*}\}$ which depends on θ_1^* and θ_2^*. In the whole paper, it is assumed that Θ is a fixed compact subset of $\mathbb{R}^d$ $(d \in \mathbb{N})$.

This test for change-point detection is often addressed with a Wald-type statistic based on the likelihood, quasi-likelihood, conditional least-squares or density power divergence estimator. Likelihood estimate-based procedure has been proposed for continuous and integer-valued time series; see, for instance, Lee and Lee (2004), Kang and Lee (2014), Doukhan and Kengne (2015), Diop and Kengne (2017), Lee et al. (2018). Several authors have pointed out some restrictions of these procedures and proposed a Wald-type statistic based on a quasi-likelihood estimators; see, among others papers, Lee and Song (2008), Kengne (2012), Diop and Kengne (2021). Other procedures have been developed with the (conditional) least-squares estimator (see, for instance, Lee and Na 2005a; Kang and Lee 2009) or the density power divergence estimator (see, among others, Lee and Na 2005b; Kang and Song 2015). Lee et al. (2003) proposed a procedure for change-point detection in a large class of time series models, but this procedure does not take into account the change-point alternative and does not ensure the consistency in power. We refer also to the works of Qu and Perron (2007) and Kim and Lee (2020) and the references therein, for some procedures for change-point detection in multivariate regressions and systems, and to Franke et al. (2012), Hudecová (2013), Fokianos et al. (2014), Hudecová et al. (2017), for other procedures for change-point detection in time series of counts.

In this new contribution, we consider a multivariate continuous or integer-valued process and deal with a general contrast, where the likelihood, quasi-likelihood, conditional least-squares or density power divergence can be seen as a specific case.

Let $\widehat{C}\big((Y_t)_{t\in T}, \theta\big)$ be a contrast function defined for any segment $T \subset \{1, \ldots, n\}$ and $\theta \in \Theta$ by:

$$\widehat{C}\big((Y_t)_{t\in T}, \theta\big) = \sum_{t\in T} \widehat{\varphi}_t(\theta), \tag{1}$$

where $\widehat{\varphi}_t$ depends on $Y_1, \ldots, Y_t$, and is such that the minimum contrast estimator (MCE), computed on a segment $T \subset \{1, \ldots, n\}$ is given by

$$\widehat{\theta}(T) := \underset{\theta\in\Theta}{\operatorname{argmin}}\Big(\widehat{C}\big((Y_t)_{t\in T}, \theta\big)\Big). \tag{2}$$

See also Killick et al. (2012) for the use of a general cost/contrast function in a context of the multiple change-points detection. In the sequel, we use the notation $\widehat{C}(T, \theta) = \widehat{C}\big((Y_t)_{t\in T}, \theta\big)$ and address the following issues.

(i) We propose a Wald-type statistic based on the MCE for testing H_0 against H_1. The asymptotic studies under the null and the alternative hypotheses show that the test has correct size asymptotically and is consistent in power. This test unifies the treatment of a large class of models, including multivariate continuous and count processes, and many existing results in the literature can be seen as specific cases of the results obtained below.
(ii) Application to a large class of multivariate causal processes is carried out. We provide sufficient conditions under which the asymptotic results of the change-point detection hold.
(iii) A general class of multivariate integer-valued models is considered. In the case where the conditional distribution belongs to the m-parameter exponential family, we provide sufficient conditions that ensure the existence of a stationary and ergodic τ-weakly dependent solution. The inference is carried out, and the consistency and the asymptotic normality of the Poisson quasi maximum likelihood estimator (PQMLE) are established. This inference question has been addressed by Ahmad (2016) with the equation-by-equation PQMLE, Lee et al. (2018) for bivariate Poisson INGARCH model, Cui et al. (2020) for flexible bivariate Poisson integer-valued GARCH model, Fokianos et al. (2020) for linear and log-linear multivariate Poisson autoregressive models. The model considered in Sect. 4 appears to be more general, and the conditions imposed for asymptotic studies seem to be more straightforward. Also, we show that the asymptotic results of the change-point detection hold for this class of models.

The paper is structured as follows. Section 2 contains the general assumptions and the construction of the test statistic for change-point detection, as well as the main asymptotic results under H_0 and H_1. Section 3 is devoted to the application of the proposed change-point detection procedure to a general class of continuous-valued processes. Section 4 focuses on a general class of observation-driven integer-valued time series. In Sect. 5, we present some numerical results.

Section 6 contains the proofs of the main results.

2 General change-point detection procedure

2.1 Assumptions

Throughout the sequel, the following norms will be used:

- $\|x\| := \sum_{i=1}^{p} |x_i|$ *for any* $x \in \mathbb{R}^p$ (*with* $p \in \mathbb{N}$);

- $\|x\| := \max_{1\leq j\leq q} \sum_{i=1}^{p} |x_{i,j}|$*for any matrix* $x = (x_{i,j}) \in M_{p,q}(\mathbb{R})$;*where* $M_{p,q}(\mathbb{R})$*denotes the set of matrices of dimension* $p \times q$*with coefficients in* $\mathbb{R}$;
- $\|g\|_{\mathcal{K}} := \sup_{\theta\in\mathcal{K}} (\|g(\theta)\|)$*for any compact set* $\mathcal{K} \subseteq \Theta$*and function* $g : \mathcal{K} \longrightarrow M_{p,q}(\mathbb{R})$;
- $\|Y\|_r := \mathbb{E}\left(\|Y\|^r\right)^{1/r}$*for any random vector* Y*with finite* $r-$*order moments.*

Let $Y = \{Y_t,\ t \in \mathbb{Z}\}$ be a multivariate continuous or integer- valued process depending on a parameter $\theta^* \in \Theta$ and denote by $\mathcal{F}_{t-1} = \sigma\{Y_{t-1}, \ldots\}$ the σ-field generated by the whole past at time $t-1$. In the sequel, we assume that $(j_n)_{n\geq 1}$ and $(k_n)_{n\geq 1}$ are two integer-valued sequences such that $j_n \leq k_n$, $k_n \to \infty$ and $k_n - j_n \to \infty$ as $n \to \infty$, and use the notation $T_{\ell,\ell'} = \{\ell, \ell+1, \ldots, \ell'\}$ for any $(\ell, \ell') \in \mathbb{N}^2$ such as $\ell \leq \ell'$. We consider a segment T_{j_n,k_n} and set the following assumptions for (Y, θ^*) under H_0.

(**A1**): The process $Y = \{Y_t,\ t \in \mathbb{Z}\}$ is assumed to be stationary and ergodic.

(**A2**): Assume that the MCE $\widehat{\theta}(T_{j_n,k_n})$ (defined in (2)) converges $a.s.$ to θ^*.

(**A3**): For all $t \in T_{j_n,k_n}$, the function $\theta \mapsto \widehat{\varphi}_t(\theta)$ (see (1)) is assumed to be continuously differentiable on Θ, in addition, assume there exists a sequence of random function $(\varphi_t(\cdot))_{t\in\mathbb{Z}}$ such that the mapping $\theta \mapsto \varphi_t(\theta)$ is continuously differentiable on Θ and for all $\theta \in \Theta$, the sequence $(\partial\varphi_t(\theta)/\partial\theta)_{t\in\mathbb{Z}}$ is stationary and ergodic, satisfying:

$$\mathbb{E}\Big\|\frac{\partial}{\partial\theta}\varphi_t(\theta)\Big\|_{\Theta}^{2} < \infty;\ \ \frac{1}{\sqrt{k_n - j_n}} \sum_{t\in T_{j_n,k_n}} \Big\|\frac{\partial}{\partial\theta}\widehat{\varphi}_t(\theta) - \frac{\partial}{\partial\theta}\varphi_t(\theta)\Big\|_{\Theta} = o_P(1) \ \text{ and}$$
$$\frac{1}{k_n - j_n} \sum_{t\in T_{j_n,k_n}} \Big\|\frac{\partial}{\partial\theta}\widehat{\varphi}_t(\theta)\frac{\partial}{\partial\theta^T}\widehat{\varphi}_t(\theta) - \frac{\partial}{\partial\theta}\varphi_t(\theta)\frac{\partial}{\partial\theta^T}\varphi_t(\theta)\Big\|_{\Theta} = o(1)\ a.s.. \quad (3)$$

Furthermore, assume that $\left(\frac{\partial}{\partial\theta}\varphi_t(\theta^*), \mathcal{F}_t\right)_{t\in\mathbb{Z}}$ is a stationary ergodic, square integrable martingale difference sequence with covariance $G = \mathbb{E}\Big[\frac{\partial\varphi_0(\theta^*)}{\partial\theta}\frac{\partial\varphi_0(\theta^*)}{\partial\theta^T}\Big]$ assumed to be positive definite.

(**A4**): For all $t \in T_{j_n,k_n}$, the function $\theta \mapsto \widehat{\varphi}_t(\theta)$ is assumed to be 2 times continuously differentiable on Θ, moreover, under the assumption (**A3**), assume that the function $\theta \mapsto \frac{\partial\varphi_t(\theta)}{\partial\theta}$ is continuously differentiable on Θ, such that the sequence $(\partial^2\varphi_t(\theta)/\partial\theta\partial\theta^T)_{t\in\mathbb{Z}}$ is stationary and ergodic, satisfying:

$$\mathbb{E}\Big\|\frac{\partial^2\varphi_t(\theta)}{\partial\theta\partial\theta^T}\Big\|_{\Theta} < \infty,\ \ \frac{1}{k_n - j_n} \sum_{t\in T_{j_n,k_n}} \Big\|\frac{\partial^2\widehat{\varphi}_t(\theta)}{\partial\theta\partial\theta^T} - \frac{\partial^2\varphi_t(\theta)}{\partial\theta\partial\theta^T}\Big\|_{\Theta} = o(1)\ a.s., \quad (4)$$

and the matrix $F = \mathbb{E}\Big[\frac{\partial^2\varphi_0(\theta^*)}{\partial\theta\partial\theta^T}\Big]$ assumed to be invertible.

The conditions (**A1**) and (**A2**) assume the stationarity of the process under H_0 and ensure that the MCE computed on each segment converges to the parameter of the stationary solution of the segment. Assumptions (**A3**) and (**A4**) allow to unify the theory for any model that satisfies such conditions and ensures the asymptotic normality of the MCE. More precisely, under (**A1**)–(**A4**), standard arguments can be used to get,

$$\sqrt{k_n - j_n}\left(\widehat{\theta}(T_{j_n,k_n}) - \theta^*\right) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, \Omega^{-1}) \text{ with } \Omega := FG^{-1}F. \tag{5}$$

The examples detailed in Sects. 3 and 4 show that the assumptions (**A1**)–(**A4**) hold for many classical time series models.

Now, for any $\ell, \ell' \in \mathbb{N}$ with $\ell \leq \ell'$, define the matrices:

$$\widehat{G}(T_{\ell,\ell'}) = \frac{1}{\ell' - \ell + 1} \sum_{t \in T_{\ell,\ell'}} \frac{\partial}{\partial\theta}\widehat{\varphi}_t(\widehat{\theta}(T_{\ell,\ell'}))\frac{\partial}{\partial\theta^T}\widehat{\varphi}_t(\widehat{\theta}(T_{\ell,\ell'})) \quad \text{and}$$

$$\widehat{F}(T_{\ell,\ell'}) = \frac{1}{\ell' - \ell + 1} \sum_{t \in T_{\ell,\ell'}} \frac{\partial^2}{\partial\theta\partial\theta^T}\widehat{\varphi}_t(\widehat{\theta}(T_{\ell,\ell'})).$$

According to (**A1**)–(**A4**), $\widehat{F}(T_{j_n,k_n})$ and $\widehat{G}(T_{j_n,k_n})$ converges almost surely to F and G, respectively. Indeed, for example,

$$\begin{aligned}\|\widehat{F}(T_{j_n,k_n}) - F\| &\leq \frac{1}{k_n - j_n} \sum_{t \in T_{j_n,k_n}} \left\| \frac{\partial^2 \widehat{\varphi}_t(\theta)}{\partial\theta\partial\theta^T} - \frac{\partial^2 \varphi_t(\theta)}{\partial\theta\partial\theta^T} \right\|_\Theta \\ &\quad + \left\| \sum_{t \in T_{j_n,k_n}} \frac{1}{k_n - j_n} \frac{\partial^2 \varphi_t(\widehat{\theta}(T_{j_n,k_n}))}{\partial\theta\partial\theta^T} - \mathbb{E}\left[\frac{\partial^2 \varphi_0(\theta^*)}{\partial\theta\partial\theta^T}\right] \right\| \\ &= o(1) + o(1) = o(1) \; a.s..\end{aligned}$$

The first term of the right-hand side of the above inequality is $a.s.\ o(1)$ from (4) and the second term is also $a.s.\ o(1)$ since $\widehat{\theta}(T_{j_n,k_n}) \xrightarrow{a.s.} n \to \infty\theta^*$ and by applying the uniform law of large numbers to the sequence $(\partial^2\varphi_t(\theta)/\partial\theta\partial\theta^T)_{t\in\mathbb{Z}}$. Under the assumption (**A3**), similar arguments yield $\|\widehat{G}(T_{j_n,k_n}) - G\| = o(1)\ a.s.$.

Therefore, $\widehat{F}(T_{j_n,k_n})\widehat{G}(T_{j_n,k_n})^{-1}\widehat{F}(T_{j_n,k_n})$ is a consistent estimator of the covariance matrix Ω.

2.2 Change-point test and asymptotic results

We derive a retrospective test procedure based on the MCE of the parameter.

For all $n \geq 1$, define the matrix $\widehat{\Omega}(u_n)$ and the subset $\mathcal{T}_n$ by

$$\begin{aligned}\widehat{\Omega}(u_n) = \frac{1}{2}\big[&\widehat{F}(T_{1,u_n})\widehat{G}(T_{1,u_n})^{-1}\widehat{F}(T_{1,u_n}) \\ &+\widehat{F}(T_{u_n+1,n})\widehat{G}(T_{u_n+1,n})^{-1}\widehat{F}(T_{u_n+1,n})\big] \text{ and } \mathcal{T}_n = [v_n, n - v_n] \cap \mathbb{N},\end{aligned}$$

where $(u_n, v_n)_{n\geq 1}$ is a bivariate integer-valued sequence such that: $(u_n, v_n) = o(n)$ and $u_n, v_n \underset{n\to\infty}{\longrightarrow} +\infty$. Note that the asymptotic properties of $\widehat{\Omega}(u_n)$ are very important to prove consistency of the procedure. Indeed, under H_0, $\widehat{\Omega}(u_n)$ is a consistent estimator of the matrix Ω. Under the alternative and the classical Assumption **B** (see below), one can show that the first component of $\widehat{\Omega}(u_n)$ converges to the inverse of the covariance matrix of the stationary model of the first regime and second component (even if its consistency is not ensured) is positive semi-definite. This will play a key role in proving the consistency under the alternative.

For any $1 < k < n$, let us introduce

$$\widehat{Q}_{n,k} = \frac{(k(n-k))^2}{n^3}\big(\widehat{\theta}(T_{1,k}) - \widehat{\theta}(T_{k+1,n})\big)^T \widehat{\Omega}(u_n)\big(\widehat{\theta}(T_{1,k}) - \widehat{\theta}(T_{k+1,n})\big).$$

Therefore, consider the following test statistic:

$$\widehat{Q}_n = \max_{k\in\mathcal{T}_n}\big(\widehat{Q}_{n,k}\big). \tag{6}$$

The construction of this statistic follows the approach of Doukhan and Kengne (2015); that is, $\widehat{Q}_{n,k}$ evaluates a distance between $\widehat{\theta}(T_{1,k})$ and $\widehat{\theta}(T_{k+1,n})$ for all $k \in \mathcal{T}_n$. Let us stress that for n large enough, $\widehat{\theta}(T_{1,k})$ and $\widehat{\theta}(T_{k+1,n})$ are close to $\widehat{\theta}(T_{1,n})$ which converges to θ^* under H_0 (from the consistency of the MCE in Assumption (**A2**)). The null hypothesis will thus be rejected if there exists a time $k \in \mathcal{T}_n$ such that the distance between $\widehat{\theta}(T_{1,k})$ and $\widehat{\theta}(T_{k+1,n})$ is too large.

The following theorem gives the asymptotic behavior of the test statistic under H_0.

Theorem 1 *Under H_0 with $\theta^* \in \overset{\circ}{\Theta}$, assume that (**A1**)–(**A4**) hold for (Y, θ^*). Then,*

$$\widehat{Q}_n \underset{n\to\infty}{\overset{\mathcal{D}}{\longrightarrow}} \sup_{0\leq\tau\leq 1} \|W_d(\tau)\|^2, \tag{7}$$

where W_d is a d-dimensional Brownian bridge.

For a nominal level $\alpha \in (0, 1)$, the critical region of the test is then $(\widehat{Q}_n > c_{d,\alpha})$, where $c_{d,\alpha}$ is the $(1-\alpha)$-quantile of the distribution of $\sup_{0\leq\tau\leq 1} \|W_d(\tau)\|^2$. The critical values $c_{d,\alpha}$ can be easily obtained through a Monte Carlo simulation; see, for instance, Lee et al. (2003).

Under the alternative hypothesis, we consider the following additional condition for the break instant.

Assumption B: *There exists $\tau^* \in (0, 1)$ such that $t^* = [n\tau^*]$,where $[x]$denotes the integer part of x.*

We obtain the following main result under H_1.

Theorem 2 *Under H_1 with θ_1^* and θ_2^* belonging to $\overset{\circ}{\Theta}$, assume that (**A1**)–(**A4**) hold for $(Y^{(1)}, \theta_1^*)$ and $(Y^{(2)}, \theta_2^*)$. If Assumption **B** is satisfied, then*

$$\widehat{Q}_n \underset{n\to\infty}{\overset{\mathcal{P}}{\longrightarrow}} +\infty. \tag{8}$$

Note that under H_1, Theorem 2 needs, in particular, the stationarity of the processes $Y^{(1)}$ and $Y^{(2)}$, but the independence between these two processes is not needed.

In the next two sections, we will detail some examples of classes of multivariate time series with a quasi-likelihood contrast function. We also show that under some regularity conditions, the general assumptions required for Theorems 1 and 2 are satisfied for these classes.

Let us stress that the scope of the proposed procedure is quite extensive and is not only restricted to the examples below. This procedure can be applied for instance, for change-point detection in models with exogenous covariates (see Diop and Kengne 2022a; Aknouche and Francq 2021), for integer-valued time series with negative binomial quasi-likelihood contrast (see Aknouche et al. 2018) or with density power divergence contrast (see Kim and Lee 2020), for general time series model with the conditional least-squares contrast (see Klimko and Nelson 1978). In fact, one can easily see in these papers that the assumptions **(A1)**–**(A4)** hold.

3 Application to a class of multidimensional causal processes

Let $\{Y_t,\ t \in \mathbb{Z}\}$ be a multivariate time series of dimension $m \in \mathbb{N}$. For any $\mathcal{T} \subseteq \mathbb{Z}$ and $\theta \in \Theta$, consider the general class of causal processes defined by

Class $\mathcal{AC}_{\mathcal{T}}(M_\theta, f_\theta)$: A process $\{Y_t,\ t \in \mathcal{T}\}$ belongs to $\mathcal{AC}_{\mathcal{T}}(M_\theta, f_\theta)$ if it satisfies:

$$Y_t = M_\theta(Y_{t-1}, Y_{t-2}, \ldots) \cdot \xi_t + f_\theta(Y_{t-1}, Y_{t-2}, \ldots) \ \ \forall t \in \mathcal{T}, \tag{9}$$

where $M_\theta(Y_{t-1}, Y_{t-2}, \ldots)$ is a $m \times p$ random matrix having almost everywhere ($a.e$) full rank m, $f_\theta(Y_{t-1}, Y_{t-2}, \ldots)$ is a $\mathbb{R}^m$-random vector and $(\xi_t)_{t\in\mathbb{Z}}$ is a sequence of $\mathbb{R}^p$-random vector with zero-mean, independent, identically distributed ($i.i.d$) satisfying $\xi_t = (\xi_t^{(k)})_{1\leq k\leq p}$ with $\mathbb{E}\big[\xi_0^{(k)}\xi_0^{(k')}\big] = 0$ for $k \neq k'$ and $\mathbb{E}\big[{\xi_0^{(k)}}^2\big] = \mathrm{Var}(\xi_0^{(k)}) = 1$ for $1 \leq k \leq p$. $M_\theta(\cdot)$ and $f_\theta(\cdot)$ are assumed to be known up to the parameter θ. This class has been studied in Doukhan and Wintenberger (2008), Bardet and Wintenberger (2009).

We would like to carry out the change-point test presented in Sect. 1 for the class $\mathcal{AC}_{\mathcal{T}}(M_\theta, f_\theta)$. For this purpose, we assume that $(Y_1, \ldots, Y_n)$ is a trajectory generated from one or two processes satisfying (9).

For all $t \in \mathbb{Z}$, denote by $\mathcal{F}_t = \sigma(Y_s,\ s \leq t)$ the σ-field generated by the whole past at time t. For any segment $T \subset \{1, \ldots, n\}$ and $\theta \in \Theta$, we define the contrast function based on the conditional Gaussian quasi-log-likelihood given by (up to an additional constant)

$$\begin{aligned}
\widehat{C}(T,\theta) &= \frac{1}{2}\sum_{t\in T}\widehat{\varphi}_t(\theta) \quad \text{with} \\
\widehat{\varphi}_t(\theta) &= (Y_t - \widehat{f}_\theta^t)^T(\widehat{H}_\theta^t)^{-1}(Y_t - \widehat{f}_\theta^t) + \log(\det(\widehat{H}_\theta^t)),
\end{aligned} \tag{10}$$

where $\widehat{f}_\theta^t := f_\theta(Y_{t-1}, \ldots, Y_1, 0, \ldots)$, $\widehat{M}_\theta^t := M_\theta(Y_{t-1}, \ldots, Y_1, 0, \ldots)$, $\widehat{H}_\theta^t := \widehat{M}_\theta^t(\widehat{M}_\theta^t)^T$.

Thus, the MCE computed on T is defined by

$$\widehat{\theta}(T) = \underset{\theta\in\Theta}{\operatorname{argmin}}\big(\widehat{C}(T,\theta)\big). \tag{11}$$

Let Ψ_θ be a generic symbol for any of the functions f_θ, M_θ or $H_\theta = M_\theta M_\theta^T$ and $\mathcal{K} \subseteq \Theta$ be a compact subset. To study the stability properties of the class (9), Bardet and Wintenberger (2009) imposed the following classical Lipschitz-type conditions on the function Ψ_θ.

Assumption $\mathbf{A}_i(\Psi_\theta,\mathcal{K})$ $(i=0,1,2)$: For any $y\in(\mathbb{R}^m)^\infty$, the function $\theta\mapsto\Psi_\theta(y)$ is i times continuously differentiable on $\mathcal{K}$ with $\big\|\frac{\partial^i\Psi_\theta(0)}{\partial\theta^i}\big\|_{\mathcal{K}}<\infty$, and there exists a sequence of nonnegative real numbers $(\alpha_k^{(i)}(\Psi_\theta,\mathcal{K}))_{k\in\mathbb{N}}$ satisfying: $\sum_{k=1}^{\infty}\alpha_k^{(i)}(\Psi_\theta,\mathcal{K})<\infty$, for $i=0,1,2$; such that for any $x,y\in(\mathbb{R}^m)^\infty$,

$$\Big\|\frac{\partial^i}{\partial\theta^i}\Psi_\theta(x)-\frac{\partial^i}{\partial\theta^i}\Psi_\theta(y)\Big\|_{\mathcal{K}} \le \sum_{k=1}^{\infty}\alpha_k^{(i)}(\Psi_\theta,\mathcal{K})\|x_k-y_k\|,$$

where x, y, x_k, y_k are, respectively, replaced by xx^T, yy^T, $x_kx_k^T$, $y_ky_k^T$ if $\Psi_\theta=H_\theta$.

For $r\ge1$, define the set

$$\begin{aligned}\Theta(r) = \Big\{\theta\in\mathbb{R}^d \,\big/\, &\mathbf{A}_0(f_\theta,\{\theta\}) \text{ and } \mathbf{A}_0(M_\theta,\{\theta\}) \text{ hold with}\\ &\sum_{k=1}^{\infty}\Big\{\alpha_k^{(0)}(f_\theta,\{\theta\})+\|\xi_0\|_r\alpha_k^{(0)}(M_\theta,\{\theta\}\Big\}<1\Big\}\\ &\bigcup\Big\{\theta\in\mathbb{R}^d \,\big/\, f_\theta=0 \text{ and } \mathbf{A}_0(H_\theta,\{\theta\}) \text{ holds with } \|\xi_0\|_r^2\\ &\quad\sum_{k=1}^{\infty}\alpha_k^{(0)}(H_\theta,\{\theta\})<1\Big\}.\end{aligned}$$

The following regularity conditions are also considered in Bardet and Wintenberger (2009) to assure the consistency and the asymptotic normality of $\widehat{\theta}(T_{1,n})$ under H_0.

($\mathcal{AC}$.**A0**): For all $\theta\in\Theta$ and some $t\in\mathbb{Z}$, $\big(f_{\theta^*}^t=f_\theta^t$ and $H_{\theta^*}^t=H_\theta^t\ a.s.\big)\Rightarrow\theta=\theta^*$.

($\mathcal{AC}$.**A1**): $\exists\underline{H}>0$ such that $\inf_{\theta\in\Theta}\det(H_\theta(y))\ge\underline{H}$, for all $y\in(\mathbb{R}^m)^\infty$.

($\mathcal{AC}$.**A2**): $\alpha_k^{(i)}(f_\theta,\Theta)+\alpha_k^{(i)}(M_\theta,\Theta)+\alpha_k^{(i)}(H_\theta,\Theta)=O(k^{-\gamma})$ for $i=0,1,2$ and some $\gamma>3/2$.

($\mathcal{AC}$.**A3**): One of the families $\big(\frac{\partial f_{\theta^*}^0}{\partial\theta_i}\big)_{1\le i\le d}$ or $\big(\frac{\partial H_{\theta^*}^0}{\partial\theta_i}\big)_{1\le i\le d}$ is $a.e$ linearly independent.

Under $\mathbf{A}_0(\Psi_\theta,\Theta)$ (for $\Psi_\theta=f_\theta,M_\theta,H_\theta$) with $\theta^*\in\Theta\cap\Theta(1)$, Bardet and Wintenberger (2009) established the existence of a strictly stationary and ergodic solution to the class $\mathcal{AC}_{\mathbb{Z}}(M_{\theta^*},f_{\theta^*})$, which shows that the assumption (**A1**) holds. Under H_0, if $\mathbf{A}_0(f_\theta,\Theta)$, $\mathbf{A}_0(M_\theta,\Theta)$ (or $\mathbf{A}_0(H_\theta,\Theta)$) and ($\mathcal{AC}$.**A0**)–($\mathcal{AC}$.**A2**) hold with $\theta^*\in\Theta\cap\Theta(2)$, then $\widehat{\theta}(T_{j_n,k_n})\xrightarrow{a.s.}n\to\infty\theta^*$ (from Theorem 1 of Bardet and Wintenberger 2009). Therefore, (**A2**) is satisfied.

Let us define

$$\varphi_t(\theta) := (Y_t - f_\theta^t)^T (H_\theta^t)^{-1} (Y_t - f_\theta^t) + \log(\det(H_\theta^t)) \tag{12}$$

with $f_\theta^t := f_\theta(Y_{t-1}, \ldots)$, $M_\theta^t := M_\theta(Y_{t-1}, \ldots)$ and $H_\theta^t := M_\theta^t (M_\theta^t)^T$.

Now, consider the change-point test presented in Section, where the observations $(Y_1, \ldots, Y_n)$ depend on θ^* under H_0 and on (θ_1^*,θ_2^*) under H_1. Under H_0, $\mathbf{A}_i(f_\theta, \Theta)$, $\mathbf{A}_i(M_\theta, \Theta)$ (or $\mathbf{A}_i(M_\theta, \Theta)$) for $i = 0, 1, 2$ and $(\mathcal{AC}.\mathbf{A0})$–$(\mathcal{AC}.\mathbf{A3})$ with $\theta^* \in \overset{\circ}{\Theta} \cap \Theta(4)$, Bardet and Wintenberger (2009) have proved that $\widehat{\theta}(T_{j_n,k_n})$ is asymptotically normal. Then, using the sequence of functions $(\varphi_t(\cdot))_{t\in\mathbb{Z}}$ defined in (12), one can see that the assumptions (**A3**) and (**A4**) also hold. For the condition (3) and those imposed on the sequence $(\frac{\partial}{\partial\theta}\varphi_t(\theta^*), \mathcal{F}_t)_{t\in\mathbb{Z}}$ in (**A3**), see the proof of their Theorem 2 and the arguments in the proof of Lemma 2 (ii). Thus, under the null hypothesis, all the required assumptions (**A1**)–(**A4**) are verified for (Y, θ^*), which assures that Theorem 1 applies to this class of models. Note that, by the same arguments, one can also see that these assumptions hold for $(Y^{(1)}, \theta_1^*)$ and $(Y^{(2)}, \theta_2^*)$ under H_1. Therefore, Theorem 2 also applies to this class.

The models $VAR(1)$ considered in (20) and (22) are examples of processes belonging to the class $\mathcal{AC}_\mathcal{T}(M_\theta, f_\theta)$. Such examples have been studied; see, for instance, Dvořák and Prášková (2013) and Kirch et al. (2015). But, the models (20) and (22) below are quite general, since the matrix M_θ is part of the parameters of the model and a change might occur in this matrix.

4 Inference and application in general multivariate count process

4.1 Model formulation and inference

Consider a multivariate count time series $\{Y_t = (Y_{t,1}, \ldots, Y_{t,m})^T,\ t \in \mathbb{Z}\}$ with value in $\mathbb{N}_0^m$ (with $m \in \mathbb{N}$, $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$) and denote by $\mathcal{F}_{t-1} = \sigma\{Y_{t-1}, \ldots\}$ the σ-field generated by the whole past at time $t-1$. For any $\mathcal{T} \subseteq \mathbb{Z}$ and $\theta \in \Theta$, define the class of multivariate observation-driven integer-valued time series given by

Class $\mathcal{MOD}_\mathcal{T}(f_\theta)$: The multivariate count process $Y = \{Y_t,\ t \in \mathcal{T}\}$ belongs to $\mathcal{MOD}_\mathcal{T}(f_\theta)$ if it satisfies:

$$\mathbb{E}(Y_t | \mathcal{F}_{t-1}) = f_\theta(Y_{t-1}, Y_{t-2}, \ldots)\ \ \forall t \in \mathcal{T}, \tag{13}$$

where $f_\theta(\cdot)$ is a measurable multivariate function with nonnegative components, assumed to be known up to the parameter θ.

In this section, it is assumed that any $\{Y_t,\ t \in \mathbb{Z}\}$ belonging to $\mathcal{MOD}_\mathcal{T}(f_\theta)$ is a stationary and ergodic process (i.e., the condition (**A1**) imposed for the change-point detection holds) satisfying:

$$\exists C > 0, \epsilon > 0,\ \text{such that}\ \forall t \in \mathbb{Z},\ \ \|Y_t\|_{1+\epsilon} < C. \tag{14}$$

Proposition 1 provides sufficient conditions for the existence of a stationary and ergodic solution of (13) when the conditional distribution belongs to a m-parameter exponential family. The condition (14) is a classical assumption which ensures that the process $\{Y_t,\ t \in \mathcal{T}\}$ has moments of order slightly greater than 1 (see, for instance, Ahmad and Francq 2016).

Let $(Y_1, \ldots, Y_n)$ be observations generated from $\mathcal{MOD}_{\mathbb{Z}}(f_{\theta^*})$ with $\theta^* \in \Theta$. The conditional Poisson quasi-log-likelihood computed on $\{1, \ldots, n\}$ is given by (up to a constant)

$$L_n(\theta) := \sum_{t=1}^{n} \ell_t(\theta) \ \text{ with } \ \ell_t(\theta) = \sum_{i=1}^{m} \big(Y_{t,i} \log \lambda_{t,i}(\theta) - \lambda_{t,i}(\theta)\big),$$

where $\lambda_t(\theta) := \big(\lambda_{t,1}(\theta), \ldots, \lambda_{t,m}(\theta)\big) = f_\theta(Y_{t-1}, Y_{t-2}, \ldots)$. An approximated conditional quasi-log-likelihood is given by

$$\widehat{L}_n(\theta) := \sum_{t=1}^{n} \widehat{\ell}_t(\theta) \ \text{ with } \ \widehat{\ell}_t(\theta) = \sum_{i=1}^{m} \big(Y_{t,i} \log \widehat{\lambda}_{t,i}(\theta) - \widehat{\lambda}_{t,i}(\theta)\big),$$

where $\widehat{\lambda}_t(\theta) := \big(\widehat{\lambda}_{t,1}(\theta), \ldots, \widehat{\lambda}_{t,m}(\theta)\big)^T = f_\theta(Y_{t-1}, \ldots, Y_1, 0, \ldots)$. Therefore, the Poisson quasi-maximum likelihood estimator (QMLE) of θ^* is defined by

$$\widehat{\theta}_n := \underset{\theta \in \Theta}{\operatorname{argmax}} \big(\widehat{L}_n(\theta)\big).$$

Note that under the assumption of independence among components and conditionally Poisson distributed, this Poisson QMLE is equivalent to the maximum likelihood estimator. Let us highlight that we deal with an arbitrary dependence among components and arbitrary conditional distribution; that is, the distribution of the components could differ from each other.

For a process $\{Y_t,\ t \in \mathbb{Z}\}$ belonging to $\mathcal{MOD}_{\mathbb{Z}}(f_{\theta^*})$, we set the following assumptions in order to establish the consistency and the asymptotic normality of the Poisson QMLE.

Assumption $\mathbf{A}_i(\Theta)$ ($i = 0, 1, 2$): For any $y \in \big(\mathbb{N}_0^m\big)^\infty$, the function $\theta \mapsto f_\theta(y)$ is i times continuously differentiable on Θ with $\big\|\partial^i f_\theta(0)/\partial\theta^i\big\|_\Theta < \infty$, and there exists a sequence of nonnegative real numbers $(\alpha_k^{(i)})_{k\geq 1}$ satisfying $\sum_{k=1}^\infty \alpha_k^{(0)} < 1$ (or $\sum_{k=1}^\infty \alpha_k^{(i)} < \infty$ for $i = 1, 2$); such that for any $y, y' \in \big(\mathbb{N}_0^m\big)^\infty$,

$$\left\| \frac{\partial^i f_\theta(y)}{\partial\theta^i} - \frac{\partial^i f_\theta(y')}{\partial\theta^i} \right\|_\Theta \leq \sum_{k=1}^{\infty} \alpha_k^{(i)} \|y_k - y'_k\|.$$

($\mathcal{MOD}$.**A0**): For all $\theta \in \Theta$, $\big(f_{\theta^*}(Y_{t-1}, Y_{t-2}, \ldots) \overset{a.s.}{=} f_\theta(Y_{t-1}, Y_{t-2}, \ldots)$ for some $t \in \mathbb{Z}\big) \Rightarrow \theta^* = \theta$; moreover, $\exists \underline{c} > 0$ such that $f_\theta(y) \geq \underline{c}\mathbf{1}_m$ componentwise, for all $\theta \in \Theta$, $y \in \big(\mathbb{N}_0^m\big)^\infty$, where $\mathbf{1}_m^T = (1, \ldots, 1)$ is a vector of dimension m.

($\mathcal{MOD}$.**A1**): θ^* is an interior point of $\Theta \subset \mathbb{R}^d$.
($\mathcal{MOD}$.**A2**): The family $\left(\frac{\partial \lambda_t(\theta^*)}{\partial \theta_i}\right)_{1\leq i\leq d}$ is $a.e.$ linearly independent.

Proposition 1 establishes the existence of a stationary and ergodic solution of the model (13) for the m-parameter exponential family conditional distribution. Consider a m-dimensional process $\{Y_t,\ t\in\mathbb{Z}\}$ satisfying

$$Y_t|\mathcal{F}_{t-1} \sim p(y|\eta_t) \;\text{ with }\; \lambda_t(\theta) := \mathbb{E}(Y_t|\mathcal{F}_{t-1}) = f_\theta(Y_{t-1}, Y_{t-2}, \ldots) \tag{15}$$

where $p(\cdot|\cdot)$ is a multivariate discrete distribution belonging to the m-parameter exponential family; that is

$$p(y|\eta) = \exp\{\eta^T y - A(\eta)\}h(y),\;\; y\in\mathbb{N}_0^m$$

where η is the natural parameter (i.e., η_t is the natural parameter of the distribution of $Y_t|\mathcal{F}_{t-1}$) and $A(\eta), h(y)$ are known functions. It is assumed that the function $\eta\mapsto A(\eta)$ is twice continuously differentiable on the natural parameter space; therefore, the mean and variance of this distribution are $\partial A(\eta)/\partial\eta$ and $\partial^2 A(\eta)/\partial\eta^2$, respectively. See Khatri (1983) for more details on such class of distribution. For the model (15), it holds that

$$\mathbb{E}(Y_t|\mathcal{F}_{t-1}) = f_\theta(Y_{t-1}, Y_{t-2}, \ldots) = \frac{\partial A(\eta_t)}{\partial \eta}.$$

Proposition 1 *Assume that $\boldsymbol{A}_0(\Theta)$ holds. Then, there exists a $\tau-$weakly dependent, stationary and ergodic solution $\{Y_t,\ t\in\mathbb{Z}\}$ to (15), satisfying $\mathbb{E}\|Y_t\|<\infty$.*

Let $(S,\mathcal{A},\mathbb{P})$ be a probability space, $\mathcal{M}$ a σ-subalgebra of $\mathcal{A}$ and Z a random variable with values in a Banach space $(E,\|\cdot\|)$. Assume that $\|Z\|_1<\infty$ and define the coefficient τ as

$$\tau(\mathcal{M}, Z) = \left\| \sup_{h\in\Lambda_1(E)} \left\{ \left| \int h(x)\mathbb{P}_{Z|\mathcal{M}}(dx) - \int h(x)\mathbb{P}_Z(dx) \right| \right\} \right\|_1,$$

where $\Lambda_1(E)$ is the set of functions $h : E \to \mathbb{R}$ such that $\mathrm{Lip}(h) := \sup x, y\in E,\ x\neq y|h(x)-h(y)|/\|x-y\| \leq 1$. Consider an E-valued strictly stationary process $(Z_t)_{t\in\mathbb{Z}}$ and set for all $i\in\mathbb{Z}$, $\mathcal{M}_i = \sigma(Z_t,\ t\leq i)$. The dependence between the past of the process $(Z_t)_{t\in\mathbb{Z}}$ and its future k-tuples may be assessed as follows. Consider the norm $\|x-y\| = \|x_1-y_1\|+\cdots+\|x_k-y_k\|$ on E^k and define

$$\tau_k(s) = \max_{1\leq\ell\leq k} \frac{1}{\ell} \sup\left\{\tau(\mathcal{M}_i, (Z_{j_1},\ldots,Z_{j_\ell})) \text{ with } i+s\leq j_1<\cdots<j_\ell\right\}$$
$$\text{and}\quad \tau(s) = \sup_{k>0}\tau_k(s).$$

If $\tau(s)$ tends to 0 as $s\to\infty$, then the process $(Z_t)_{t\in\mathbb{Z}}$ is said to be τ-weakly dependent. The weak dependence concept has been introduced by Doukhan and Louhichi

(1999) for the purpose of taking into account some situations where the mixing conditions are not satisfied. We refer to the lecture notes Dedecker et al. (2007) for an overview on this dependence concept.

In the sequel, we deal with the more general class of model (13), where the distribution of $Y_t|\mathcal{F}_{t-1}$ may be outside the m-parameter exponential family. The following theorem shows that the Poisson QMLE for the class of models (13) is strongly consistent.

Theorem 3 *Assume that* $\boldsymbol{A}_0(\Theta)$, *(*$\mathcal{MOD}$*.**A0**) and* (14) *(with* $\epsilon \geq 1$*) hold with*

$$\alpha_k^{(0)} = \mathcal{O}(k^{-\gamma}) \quad \textit{for some} \quad \gamma > 3/2. \tag{16}$$

Then

$$\widehat{\theta}_n \xrightarrow{a.s.} n \to \infty \theta^*.$$

For any $t \in \mathbb{Z}$ and $\theta \in \Theta$, denote $\Gamma_t(\theta) := (Y_t - \lambda_t(\theta))(Y_t - \lambda_t(\theta))^T$ and $D_t(\theta)$ the $m \times m$ diagonal matrix with the ith diagonal element is equal to $\lambda_{t,i}(\theta)$ for any $i = 1, \ldots, m$. From the assumption ($\mathcal{MOD}$.**A0**), the matrix $D_t(\theta)$ is $a.s.$ positive definite. Combining all the regularity assumptions and notations given above, we obtain the asymptotic normality of the Poisson QMLE, as shown in the following theorem.

Theorem 4 *Assume that* $\boldsymbol{A}_i(\Theta)$ *(*$i = 0, 1, 2$*), (*$\mathcal{MOD}$*.**A0**)–(*$\mathcal{MOD}$*.**A2**) and* (14) *(with* $\epsilon \geq 3$*) hold with*

$$\alpha_k^{(0)} + \alpha_k^{(1)} + \alpha_k^{(2)} = \mathcal{O}(k^{-\gamma}) \quad \textit{for some} \quad \gamma > 3/2, \tag{17}$$

then

$$\sqrt{n}(\widehat{\theta}_n - \theta^*) \xrightarrow[n \to \infty]{\mathcal{D}} \mathcal{N}(0, \Sigma) \quad \textit{with} \quad \Sigma := J_{\theta^*}^{-1} I_{\theta^*} J_{\theta^*}^{-1},$$

where

$$J_{\theta^*} = \left[\frac{\partial \lambda_0^T(\theta^*)}{\partial \theta} D_0^{-1}(\theta^*) \frac{\partial \lambda_0(\theta^*)}{\partial \theta^T} \right] \quad \textit{and}$$
$$I_{\theta^*} = \mathbb{E}\left[\frac{\partial \lambda_0^T(\theta^*)}{\partial \theta} D_0^{-1}(\theta^*) \Gamma_0(\theta^*) D_0^{-1}(\theta^*) \frac{\partial \lambda_0(\theta^*)}{\partial \theta^T} \right].$$

In comparison with the results of Lee et al. (2018), Cui et al. (2020) and Fokianos et al. (2020), Theorems 3 and 4 are applied to the class $\mathcal{MOD}_{\mathbb{Z}}(f_{\theta^*})$ with a general structure of the conditional mean, whereas these authors deal with linear and log-linear models. Moreover, Lee et al. (2018) and Cui et al. (2020) consider a parametric framework, with the assumption that the conditional distribution given the whole past is known, which is quite restrictive in practice. Theorems 3 and 4 are established in a semi-parametric setting.

4.2 Change-point detection

Now, assume that the trajectory $(Y_1, \dots, Y_n)$ is generated from one or two processes satisfying the general model (13) and consider the change-point test of Sect. 1, where the observations depend on θ^* under H_0 and on (θ_1^*, θ_2^*) under H_1. Let us define the contrast function based on the conditional Poisson quasi-log-likelihood for any segment $T \subset \{1, \dots, n\}$ and $\theta \in \Theta$:

$$\widehat{C}(T, \theta) := \sum_{t \in T} \widehat{\varphi}_t(\theta) \quad \text{with} \quad \widehat{\varphi}_t(\theta) = -\widehat{\ell}_t(\theta) \quad \text{for all} \quad t \in \mathbb{Z}. \tag{18}$$

Thus, the MCE computed on T is given by

$$\widehat{\theta}(T) := \underset{\theta \in \Theta}{\operatorname{argmin}} \big(\widehat{C}(T, \theta)\big). \tag{19}$$

Under the null hypothesis, the assumption (**A2**) holds from Theorem 3. Letting $\varphi_t(\theta) := -\ell_t(\theta)$ for all $t \in \mathbb{Z}$ and $\theta \in \Theta$, one can see that (**A3**) and (**A4**) are also satisfied from Theorem 4. The relation (3) in (**A3**) holds from Lemma 2 (i) (see below), the proof of Lemma 3 (a) and the arguments in the proof of Lemma 2 (ii), whereas the relation (4) in (**A4**) holds from Lemme 2 (ii), Lemma 3 (c). See also Lemma 3(b) for the required properties about the sequence $(\frac{\partial}{\partial \theta} \varphi_t(\theta^*), \mathcal{F}_t)_{t \in \mathbb{Z}}$. Hence, in absence of change, all the conditions of Theorem 1 are verified for (Y, θ^*), which assures that the first result about the asymptotic behavior of the test statistic $\widehat{Q}_n$ applies to the class of models (13). Under the change point alternative H_1, one can go along similar lines to verify that (**A1**)–(**A4**) are satisfied for (Y_1, θ_1^*) and (Y_2, θ_2^*). This shows that Theorem 1 can also be applied to this class.

5 Numerical results

In this section, the statistic $\widehat{Q}_n$ will be computed with $u_n = [(\log(n))^2]$ and $v_n = [(\log(n))^{5/2}]$ for a sample size n. The procedure is implemented in the *R* software (developed by the CRAN project).

5.1 Simulation study

We investigate the performance (level and power) of the test statistic through two examples of two-dimensional processes, with sample size $n = 250, 500, 1000$ and the nominal level $\alpha = 0.05$. Let us consider the following models.

– *A bivariate AR (1) model.* Consider the two-dimensional AR(1) model (with zero-mean) expressed as

$$Y_t = A_0 Y_{t-1} + \gamma_0 \xi_t \quad \text{for all} \quad t \in \mathbb{Z}, \tag{20}$$

where $Y_t = (Y_{t,1}, Y_{t,2})^T$, $A_0 = (a_{i,j})_{i,j=1,2}$ is a 2×2 matrix with eigenvalues inside the complex unit circle, γ_0 is a nonzero real number and $(\xi_t = (\xi_{t,1}, \xi_{t,2})^T)_{t \in \mathbb{Z}}$ is a bivariate white noise satisfying the conditions of the class (9). This process belongs to the class $\mathcal{AC}_{\mathcal{T}}(M_\theta, f_\theta)$ with $f_\theta(Y_{t-1}, \ldots) = A_0 Y_{t-1}$ and $M_\theta(Y_{t-1}, \ldots) = \begin{pmatrix} \gamma_0 & 0 \\ 0 & \gamma_0 \end{pmatrix}$. The parameter of the model is denoted by $\theta_0 = (a_{1,1}, a_{1,2}, a_{2,1}, a_{2,2}, \gamma_0)$. At the nominal level $\alpha = 0.05$, the critical value of the test is therefore $c_{5,\alpha} \approx 3.899$ (see Lee et al. 2003). The performance will be evaluated in cases where the innovation $(\xi_t)_{t \in \mathbb{Z}}$ is obtained from the standardized Student distributions with 5 and 8 degrees of freedom for the first and the second component, respectively. In the scenarios of change, we assume that the parameter $\theta_0 \equiv (A_0, \gamma_0)$ changes to $\theta_1 \equiv (A_1, \gamma_1)$.

– *A bivariate INARCH(1) model.*
Assume that $\{Y_t = (Y_{t,1}, Y_{t,2})^T,\ t \in \mathbb{Z}\}$ is a count time series with value in $\mathbb{N}_0^2$, where $\{Y_{t,1},\ t \in \mathbb{Z}\}$ and $\{Y_{t,2},\ t \in \mathbb{Z}\}$ are two processes with conditional distribution following a Poisson distribution and a negative binomial distribution, respectively. More precisely,

$$\begin{cases} Y_{t,1} | \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_{t,1}) \\ Y_{t,2} | \mathcal{F}_{t-1} \sim \text{NB}(r, r/(r + \lambda_{t,2})) \end{cases} \quad \text{with} \quad \lambda_t := (\lambda_{t,1}, \lambda_{t,2})^T = d_0 + B_0 Y_{t-1}, \tag{21}$$

where $d_0 = (d^{(1)}, d^{(2)})^T \in (0, \infty)^2$, $B_0 = (b_{i,j})_{i,j=1,2}$ is a 2×2 matrix with nonnegative coefficients, and $NB(r, p)$ denotes the negative binomial distribution with parameter (r, p) and mean $r(1 - p)/p$. It is assumed that for all $t \in \mathbb{Z}$, $Y_{t,1}$, and $Y_{t,2}$ are conditionally independent given $\mathcal{F}_{t-1}$ and that the parameter r is known for each simulation; that is, the parameter of interest is $\theta_0 = (d^{(1)}, d^{(2)}, b_{1,1}, b_{1,2}, b_{2,1}, b_{2,2})$ and the critical value of the test is $c_{6,\alpha} \approx 4.375$ (see also Lee et al. 2003).
In situations of break, we also assume that the parameter changes from θ_0 (which is characterized here by (d_0, B_0)) to θ_1 that we will characterize by (d_1, B_1) for this model.

Figure 1 is an illustration of a typical realizations of the statistics $\widehat{Q}_{n,k}$ for two trajectories of length 1000 generated from bivariate AR(1) processes: a trajectory without change and a trajectory with a change at time $t^* = 500$. One can see that, for the trajectory without change, the statistics $\widehat{Q}_{n,k}$ are well below the critical value (see Fig. 1a). For the scenario with change, the maximum (which represents the value of $\widehat{Q}_n$) of the statistics $\widehat{Q}_{n,k}$ is higher than the limit of the critical region and that it is obtained at a point very close to the instant of break (see Fig. 1b). This empirically comforts the common use of the classical estimator $\widehat{t}_n = \operatorname{argmax} k \in \mathcal{T}_n \left(\widehat{Q}_{n,k}\right)$ to determine the break-point.

To evaluate the empirical level and power, we consider trajectories generated from the two models (20) and (21) in the following situations: (i) scenarios with a constant parameter θ_0 and (ii) scenarios with a parameter change $(\theta_0 \to \theta_1)$ at time $t^* = n/2$. The replication number in each simulation is 500. For different scenarios, Table 1

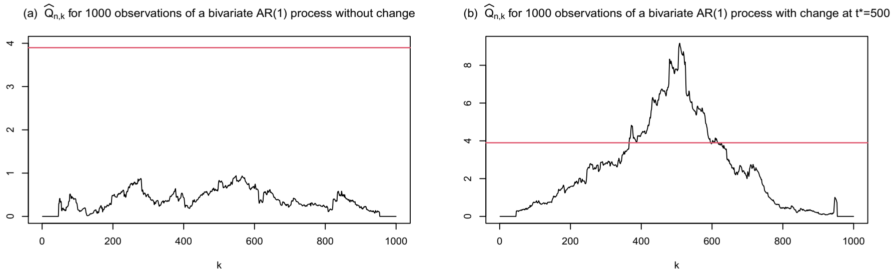

Fig. 1 Typical realizations of the statistics $\widehat{Q}_{n,k}$ for two trajectories generated from bivariate AR(1) processes defined in (20). **a** Is a realization for 1000 observations with a constant parameter $\theta_0 = (0.5, -0.2, 0.35, 0.1, 1)$. **b** Is a realization for 1000 observations in a scenario where the parameter changes from $\theta_0 = (0.5, -0.2, 0.35, 0.1, 1)$ to $\theta_1 = (0.5, -0.2, 0.1, 0.1, 1)$ at $t^* = 500$. The horizontal line represents the limit of the critical region of the test

indicates the proportion of the number of rejections of the null hypothesis computed under H_0 (for the levels) and H_1 (for the powers). As can be seen from this table, the empirical levels are close to the nominal level for each of the two models. One can see that the statistic is quite sensitive for detecting the change for both the cases considered under the alternative: the scenario with dependent components and independent components (i.e., the scenario where the matrix A_1 or B_1 is diagonal) after the breakpoint. For both the classes of models, the results of the test are quite accurate; the empirical level approaching the nominal one when n increases and the empirical power increases with n and is close to 1 when $n = 1000$. This is consistent with the asymptotic results of Theorem 1 and 2.

5.2 Real data example

We consider the bivariate time series whose variables represent the average daily concentrations of particulate matter with a diameter less than $10\mu m$ and carbon monoxide (PM_{10},CO), collected at some monitoring stations in the Vitória metropolitan area. We deal with the data from January 31, 2010 to December 30, 2010 (observations on 334 days); see Fig. 2a and b. This series is a part of a dataset obtained from the State Environment and Water Resources Institute (available at https://rss.onlinelibrary.wiley.com/pb-assets/hub-assets/rss/Datasets/RSSC%2067.2/C1239deSouza-1531120585220.zip), which were analyzed by de Souza et al. (2018).

To apply the proposed test procedure, we consider a two-dimensional AR(1) model (with nonzero mean) given for $t \in \mathbb{Z}$, by

$$Y_t = \omega_0 + A_0 Y_{t-1} + M\xi_t, \tag{22}$$

where $Y_t = (PM_{10,t}, CO_t)^T$ (the value of the corresponding vector at day t), ω_0 is a 2-dimensional vector, A_0, M are 2×2 matrices and (ξ_t) a bivariate white noise satisfying the conditions of the class (9). The parameter of the model is $\theta = (\omega_0, A_0, M) \in \mathbb{R}^{10}$.

The realizations of $\widehat{Q}_{n,k}$ (for all $k \in \mathcal{T}_n$) displayed in Fig. 2c show that the resulting test statistic $\widehat{Q}_n$ is higher than the critical value of the test, which indicates that a

Table 1 Empirical levels and powers at the nominal level 0.05 for the change-point test in the models (20) and (21)

		Scenarios	$n = 250$	$n = 500$	$n = 1000$
	Levels:				
Model (20)		$A_0 = \begin{pmatrix} 0.5 & -0.2 \\ 0.35 & 0.1 \end{pmatrix}$ and $\gamma_0 = 1$	0.058	0.054	0.052
		$A_0 = \begin{pmatrix} 0.6 & 0.3 \\ 0.0 & 0.4 \end{pmatrix}$ and $\gamma_0 = 2.5$	0.058	0.056	0.048
	Powers:				
		$A_0 = \begin{pmatrix} 0.5 & -0.2 \\ 0.35 & 0.1 \end{pmatrix}$ and $\gamma_0 = 1$; $A_1 = \begin{pmatrix} 0.5 & -0.2 \\ 0.1 & 0.1 \end{pmatrix}$ and $\gamma_1 = \gamma_0$	0.456	0.676	0.952
		$A_0 = \begin{pmatrix} 0.5 & -0.2 \\ 0.35 & 0.1 \end{pmatrix}$ and $\gamma_0 = 1$; $A_1 = A_0$ and $\gamma_1 = 0.5$	0.804	0.896	0.996
		$A_0 = \begin{pmatrix} 0.6 & 0.3 \\ 0.0 & 0.4 \end{pmatrix}$ and $\gamma_0 = 2.5$; $A_1 = \begin{pmatrix} 0.6 & 0.0 \\ 0.0 & 0.4 \end{pmatrix}$ and $\gamma_1 = 2.3$	0.602	0.926	0.990
	Levels:				
Model (21)		$d_0 = (1, 0.3)^T$ and $B_0 = \begin{pmatrix} 0.5 & 0.2 \\ 0.1 & 0.4 \end{pmatrix}$	0.076	0.064	0.056
		$d_0 = (0.5, 0.5)^T$ and $B_0 = \begin{pmatrix} 0.25 & 0.5 \\ 0.1 & 0.35 \end{pmatrix}$	0.078	0.062	0.052
	Powers:				
		$d_0 = (1, 0.3)^T$ and $B_0 = \begin{pmatrix} 0.5 & 0.2 \\ 0.1 & 0.4 \end{pmatrix}$; $d_1 = d_0$ and $B_1 = \begin{pmatrix} 0.5 & 0.0 \\ 0.0 & 0.4 \end{pmatrix}$	0.634	0.840	0.982
		$d_0 = (0.5, 0.5)^T$ and $B_0 = \begin{pmatrix} 0.25 & 0.5 \\ 0.1 & 0.35 \end{pmatrix}$; $d_1 = (0.5, 1)^T$ and $B_1 = B_0$	0.824	0.970	0.996

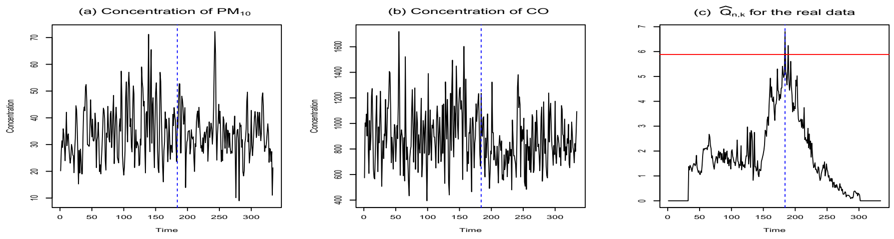

Fig. 2 Plot of the time series and the statistic $\widehat{Q}_{n,k}$ for the change-point detection applied to the bivariate real data (PM_{10},CO) with the VAR(1) process defined in (22). The horizontal line represents the limit of the critical region of the test. The vertical line represents the estimated breakpoint

change-point is detected in this series. The breakpoint is estimated as $\widehat{t_n} = 184$ (see Fig. 2), which corresponds to the date August 02, 2010. The estimated model with two regimes is given by:

$$\begin{cases} Y_t = \big(\underset{[2.854]}{16.254}, \underset{[41.240]}{664.266}\big)^T + \begin{pmatrix} \underset{[0.0668]}{0.520} & \underset{[0.002]}{0.003} \\ \underset{[1.056]}{1.462} & \underset{[0.048]}{0.234} \end{pmatrix} Y_{t-1} \\ \quad + \begin{pmatrix} \underset{[3.427]}{6.877} & \underset{[3.312]}{3.563} \\ \underset{[2.528]}{90.140} & \underset{[2.412]}{-84.360} \end{pmatrix} \eta_t \quad \text{for} \quad t \leq 184, \\ Y_t = \big(\underset{[5.729]}{19.362}, \underset{[49.210]}{667.930}\big)^T + \begin{pmatrix} \underset{[0.096]}{0.352} & \underset{[0.002]}{0.004} \\ \underset{[0.320]}{-0.457} & \underset{[0.046]}{0.209} \end{pmatrix} Y_{t-1} \\ \quad + \begin{pmatrix} \underset{[3.322]}{6.322} & \underset{[2.371]}{3.807} \\ \underset{[1.969]}{66.160} & \underset{[2.407]}{-60.031} \end{pmatrix} \eta_t \quad \text{for} \quad t > 184, \end{cases} \tag{23}$$

where in brackets are the standard errors of the estimators obtained from the sandwich matrix $\widehat{\Omega}_T^{-1}$, a consistent estimator of the covariance matrix Ω^{-1} defined in (5), computed on the segment $T \subset \{1, \ldots, n\}$. Simulations carried out with the parameters in (23) show that the procedure works well (in term of empirical level and power) in that case. Also, this result is in accordance with those obtained by Diop and Kengne (2022b) who have found a break on August 06, 2010 with an epidemic procedure in the carbon monoxide series. The first regime (from January 31, 2010 to August 08, 2010) includes the austral winter and a period where the winds are weaker. These meteorological factors are known to increase the concentration of some pollutants (such as the carbon monoxide), which are important determinants associated to the PM_{10} concentration (see, for instance, Ng and Awang 2018 and the references therein).

6 Proofs of the main results

Let $(\psi_n)_{n\in\mathbb{N}}$ and $(r_n)_{n\in\mathbb{N}}$ be sequences of random variables or vectors. Throughout this section, we use the notation $\psi_n = o_P(r_n)$ to mean: for all $\varepsilon > 0$, $\mathbb{P}(\|\psi_n\| \geq$

$\varepsilon\|r_n\|)_{n\to\infty}\!\!\!\longrightarrow 0$. Write $\psi_n = O_P(r_n)$ to mean: for all $\varepsilon > 0$, there exists $C > 0$ such that $\mathbb{P}(\|\psi_n\| \geq C\|r_n\|) \leq \varepsilon$ for n large enough. In the sequel, C denotes a positive constant whose the value may differ from one inequality to another.

6.1 Proof of the results of Section 2

6.1.1 Proof of Theorem 1

Define the statistic

$$Q_n = \max_{k\in\mathcal{T}_n}\left(Q_{n,k}\right) \quad \text{with}$$
$$Q_{n,k} = \frac{(k(n-k))^2}{n^3}\left(\widehat{\theta}(T_{1,k}) - \widehat{\theta}(T_{k+1,n})\right)^T \Omega\left(\widehat{\theta}(T_{1,k}) - \widehat{\theta}(T_{k+1,n})\right),$$

where Ω is the covariance matrix defined in the assumption (**A4**). For any segment $T \subset \{1,\ldots,n\}$ and $\theta \in \Theta$, we also define the function

$$C(T,\theta) = \sum_{t\in T}\varphi_t(\theta), \quad \text{where} \quad (\varphi_t(\cdot))_{t\in\mathbb{Z}} \text{ is given in } (\mathbf{A3}).$$

Let $1 \leq k \leq k' \leq n, \bar{\theta} \in \Theta$ and $i \in \{1,2,\ldots,d\}$. By the mean value theorem applied to the function $\theta \mapsto \frac{\partial}{\partial\theta_i}C(T_{k,k'},\theta)$, there exists $\theta_{n,i}$ between $\bar{\theta}$ and θ^* such that

$$\frac{\partial}{\partial\theta_i}C(T_{k,k'},\bar{\theta}) = \frac{\partial}{\partial\theta_i}C(T_{k,k'},\theta^*) + \frac{\partial^2}{\partial\theta\partial\theta_i}C(T_{k,k'},\theta_{n,i})(\bar{\theta}-\theta^*),$$

which implies

$$(k'-k+1)F_n(T_{k,k'},\bar{\theta})(\theta^*-\bar{\theta}) = \frac{\partial}{\partial\theta}C(T_{k,k'},\theta^*) - \frac{\partial}{\partial\theta}C(T_{k,k'},\bar{\theta}) \tag{24}$$

with

$$F_n(T_{k,k'},\bar{\theta}) = \frac{1}{(k'-k+1)}\frac{\partial^2}{\partial\theta\partial\theta_i}C(T_{k,k'},\theta_{n,i})_{1\leq i\leq d}. \tag{25}$$

The following lemma will be useful in the sequel.

Lemma 1 *Assume that the conditions of Theorem 1 hold.*

(i) $\max_{k\in\mathcal{T}_n}\left|\widehat{Q}_{n,k} - Q_{n,k}\right| = o_P(1)$.

(ii) *If $(j_n)_{n\geq 1}$ and $(k_n)_{n\geq 1}$ are two integer-valued sequences such that $j_n \leq k_n$, $k_n \underset{n\to\infty}{\longrightarrow} \infty$ and $k_n - j_n \underset{n\to\infty}{\longrightarrow} \infty$, then $F_n(T_{j_n,k_n},\widehat{\theta}(T_{j_n,k_n})) \overset{a.s.}{\longrightarrow} n \to \infty F$, where F is the matrix defined in (**A4**).*

Proof (i) Let $k \in \mathcal{T}_n$. As $n \to \infty$, from the asymptotic normality of the MCE and the consistency of $\widehat{\Omega}(u_n)$, we obtain:

$$\begin{aligned}\big\|\sqrt{k}\big(\widehat{\theta}(T_{1,k}) - \theta^*\big)\big\| = O_P(1),\ \big\|\sqrt{n-k}\big(\widehat{\theta}(T_{k+1,n}) - \theta^*\big)\big\| \\ = O_P(1) \quad \text{and} \quad \big\|\widehat{\Omega}(u_n) - \Omega\big\| = o(1).\end{aligned} \tag{26}$$

Then, it holds that

$$\begin{aligned}&\big|\widehat{Q}_{n,k} - Q_{n,k}\big| \\ &= \frac{(k(n-k))^2}{n^3}\Big|\big(\widehat{\theta}(T_{1,k}) - \widehat{\theta}(T_{k+1,n})\big)^T\big(\widehat{\Omega}(u_n) - \Omega\big)\big(\widehat{\theta}(T_{1,k}) - \widehat{\theta}(T_{k+1,n})\big)\Big| \\ &\leq C\frac{(k(n-k))^2}{n^3}\big\|\widehat{\Omega}(u_n) - \Omega\big\|\big\|\widehat{\theta}(T_{1,k}) - \widehat{\theta}(T_{k+1,n})\big\|^2 \\ &\leq C\big\|\widehat{\Omega}(u_n) - \Omega\big\|\Big[\frac{k(n-k)^2}{n^3}\big\|\sqrt{k}\big(\widehat{\theta}(T_{1,k}) - \theta^*\big)\big\|^2 \\ &\quad + \frac{k^2(n-k)}{n^3}\big\|\sqrt{n-k}\big(\widehat{\theta}(T_{k+1,n}) - \theta^*\big)\big\|^2\Big] \\ &\leq o(1)O_P(1);\end{aligned}$$

which allows to conclude.

(ii) Applying (25) with $\bar{\theta} = \widehat{\theta}(T_{j_n,k_n})$, we obtain

$$\begin{aligned}F_n(T_{j_n,k_n}, \widehat{\theta}(T_{j_n,k_n})) &= \left(\frac{1}{k_n - j_n + 1}\frac{\partial^2}{\partial\theta\partial\theta_i}C(T_{j_n,k_n}, \theta_{n,i})\right)_{1\leq i\leq d} \\ &= \frac{1}{k_n - j_n + 1}\left(\sum_{t\in T_{j_n,k_n}}\frac{\partial^2\varphi_t(\theta_{n,i})}{\partial\theta\partial\theta_i}\right)_{1\leq i\leq d},\end{aligned}$$

where $\theta_{n,i}$ belongs between $\widehat{\theta}(T_{j_n,k_n})$ and θ^*. Since $\widehat{\theta}(T_{j_n,k_n}) \xrightarrow{a.s.} n \to \infty\theta^*$, $\theta_{n,i} \xrightarrow{a.s.} n \to \infty\theta^*$ (for any $i = 1, \ldots, d$) and that $F = \mathbb{E}\big[\frac{\partial^2\varphi_0(\theta^*)}{\partial\theta\partial\theta^T}\big]$ exists (see the assumption (**A4**)), by the uniform strong law of large numbers, for any $i = 1, \ldots, d$, we get

$$\begin{aligned}&\left\|\frac{1}{k_n - j_n + 1}\sum_{t\in T_{j_n,k_n}}\frac{\partial^2\varphi_t(\theta_{n,i})}{\partial\theta\partial\theta_i} - \mathbb{E}\left[\frac{\partial^2\varphi_0(\theta^*)}{\partial\theta\partial\theta_i}\right]\right\| \\ &\leq \left\|\frac{1}{k_n - j_n + 1}\sum_{t\in T_{j_n,k_n}}\frac{\partial^2\varphi_t(\theta_{n,i})}{\partial\theta\partial\theta_i} - \mathbb{E}\Big[\frac{\partial^2\varphi_0(\theta_{n,i})}{\partial\theta\partial\theta_i}\Big]\right\| \\ &\quad + \left\|\mathbb{E}\left[\frac{\partial^2\varphi_0(\theta_{n,i})}{\partial\theta\partial\theta_i}\right] - \mathbb{E}\left[\frac{\partial^2\varphi_0(\theta^*)}{\partial\theta\partial\theta_i}\right]\right\|\end{aligned}$$

$$\leq \left\| \frac{1}{k_n - j_n + 1} \sum_{t \in T_{j_n,k_n}} \frac{\partial^2 \varphi_t(\theta)}{\partial\theta\partial\theta_i} - \mathbb{E}\left[\frac{\partial^2 \varphi_0(\theta)}{\partial\theta\partial\theta_i}\right] \right\|_\Theta$$
$$+ o(1) = o(1) + o(1) = o(1).$$

This completes the proof of the lemma.

□

Now, we use (26) and the part (ii) of Lemma 1 to show that

$$Q_n \underset{n\to\infty}{\overset{\mathcal{D}}{\longrightarrow}} \sup_{0\leq\tau\leq 1} \|W_d(\tau)\|^2. \tag{27}$$

Let $k \in \mathcal{T}_n$. Applying (24) with $\bar{\theta} = \widehat{\theta}(T_{1,k})$ and $T_{k,k'} = T_{1,k}$, we get

$$F_n(T_{1,k}, \widehat{\theta}(T_{1,k})) \cdot (\theta^* - \widehat{\theta}(T_{1,k})) = \frac{1}{k}\left(\frac{\partial}{\partial\theta} C(T_{1,k}, \theta_1^*) - \frac{\partial}{\partial\theta} C(T_{1,k}, \widehat{\theta}(T_{1,k}))\right). \tag{28}$$

With $\bar{\theta} = \widehat{\theta}(T_{k+1,n})$ and $T_{k,k'} = T_{k+1,n}$, (24) becomes

$$\begin{aligned} &F_n(T_{k+1,n}, \widehat{\theta}(T_{k+1,n})) \cdot (\theta^* - \widehat{\theta}(T_{k+1,n})) \\ &\quad = \frac{1}{n-k}\left(\frac{\partial}{\partial\theta} C(T_{k+1,n}, \theta_1^*) - \frac{\partial}{\partial\theta} C(T_{k+1,n}, \widehat{\theta}(T_{k+1,n}))\right). \end{aligned} \tag{29}$$

Moreover, as $n \to +\infty$, Lemma 1(ii) implies

$$\left\|F_n(T_{1,k}, \widehat{\theta}(T_{1,k})) - F\right\| = o(1) \quad \text{and} \quad \left\|F_n(T_{k+1,n}, \widehat{\theta}(T_{k+1,n})) - F\right\| = o(1).$$

Then, according to (26), for n large enough, (28) gives

$$\begin{aligned} \sqrt{k} F\left(\theta^* - \widehat{\theta}(T_{1,k})\right) &= \frac{1}{\sqrt{k}}\left(\frac{\partial}{\partial\theta} C(T_{1,k}, \theta^*) - \frac{\partial}{\partial\theta} C(T_{1,k}, \widehat{\theta}(T_{1,k}))\right) \\ &\quad - \sqrt{k}\left(\left(F_n(T_{1,k}, \widehat{\theta}(T_{1,k})) - J\right)\left(\widehat{\theta}(T_{1,k}) - \theta_0\right)\right) \\ &= \frac{1}{\sqrt{k}}\left(\frac{\partial}{\partial\theta} C(T_{1,k}, \theta^*) - \frac{\partial}{\partial\theta} C(T_{1,k}, \widehat{\theta}(T_{1,k}))\right) + o_P(1) \\ &= \frac{1}{\sqrt{k}}\left(\frac{\partial}{\partial\theta} C(T_{1,k}, \theta^*) - \frac{\partial}{\partial\theta} \widehat{C}(T_{1,k}, \widehat{\theta}(T_{1,k}))\right) + o_P(1) \\ &\quad + \frac{1}{\sqrt{k}}\left(\frac{\partial}{\partial\theta} \widehat{C}(T_{1,k}, \widehat{\theta}(T_{1,k})) - \frac{\partial}{\partial\theta} C(T_{1,k}, \widehat{\theta}(T_{1,k}))\right) \\ &= \frac{1}{\sqrt{k}}\left(\frac{\partial}{\partial\theta} C(T_{1,k}, \theta^*) - \frac{\partial}{\partial\theta} \widehat{C}(T_{1,k}, \widehat{\theta}(T_{1,k}))\right) \\ &\quad + o_P(1) \quad \text{(from the condition (3) in (\textbf{A3}))}. \end{aligned}$$

This is equivalent to

$$F\left(\theta^* - \widehat{\theta}(T_{1,k})\right) = \frac{1}{k}\Big(\frac{\partial}{\partial\theta}C(T_{1,k},\theta^*) - \frac{\partial}{\partial\theta}\widehat{C}(T_{1,k},\widehat{\theta}(T_{1,k}))\Big) + o_P\Big(\frac{1}{\sqrt{k}}\Big). \tag{30}$$

For n large enough, $\widehat{\theta}(T_{1,k})$ is an interior point of Θ and we have $\frac{\partial}{\partial\theta}\widehat{C}(T_{1,k},\widehat{\theta}(T_{1,k})) = 0$.

Hence, for n large enough, we get from (30)

$$F\left(\theta^* - \widehat{\theta}(T_{1,k})\right) = \frac{1}{k}\frac{\partial}{\partial\theta}C(T_{1,k},\theta^*) + o_P\Big(\frac{1}{\sqrt{k}}\Big). \tag{31}$$

Similarly, we can use (29) to obtain

$$F\left(\theta^* - \widehat{\theta}(T_{k+1,n})\right) = \frac{1}{n-k}\frac{\partial}{\partial\theta}C(T_{k+1,n},\theta^*) + o_P\Big(\frac{1}{\sqrt{n-k}}\Big). \tag{32}$$

The subtraction of (31) and (32) gives

$$\begin{aligned}
&-F\left(\widehat{\theta}(T_{1,k}) - \widehat{\theta}(T_{k+1,n})\right)\\
&= \frac{1}{k}\frac{\partial}{\partial\theta}C(T_{1,k},\theta^*) - \frac{1}{n-k}\frac{\partial}{\partial\theta}C(T_{k+1,n},\theta^*)\\
&\quad + o_P\left(\frac{1}{\sqrt{k}} + \frac{1}{\sqrt{n-k}}\right)\\
&= \frac{1}{k}\frac{\partial}{\partial\theta}C(T_{1,k},\theta^*) - \frac{1}{n-k}\\
&\quad \left(\frac{\partial}{\partial\theta}C(T_{1,n},\theta^*) - \frac{\partial}{\partial\theta}C(T_{1,k},\theta^*)\right) + o_P\left(\frac{1}{\sqrt{k}} + \frac{1}{\sqrt{n-k}}\right)\\
&= \frac{n}{k(n-k)}\left(\frac{\partial}{\partial\theta}C(T_{1,k},\theta^*) - \frac{k}{n}\cdot\frac{\partial}{\partial\theta}C(T_{1,n},\theta^*)\right) + o_P\left(\frac{1}{\sqrt{k}} + \frac{1}{\sqrt{n-k}}\right).
\end{aligned}$$

Since the matrix G is positive definite (see (**A3**)), the above equality is equivalent to

$$\begin{aligned}
&-\frac{k(n-k)}{n^{3/2}}G^{-1/2}F\left(\widehat{\theta}(T_{1,k}) - \widehat{\theta}(T_{k+1,n})\right)\\
&= \frac{G^{-1/2}}{\sqrt{n}}\left(\frac{\partial}{\partial\theta}C(T_{1,k},\theta^*) - \frac{k}{n}\cdot\frac{\partial}{\partial\theta}C(T_{1,n},\theta^*)\right)\\
&\quad + o_P\left(\frac{\sqrt{k(n-k)}}{n} + \frac{\sqrt{n-k}}{\sqrt{n}}\right)\\
&= \frac{G^{-1/2}}{\sqrt{n}}\left(\frac{\partial}{\partial\theta}C(T_{1,k},\theta^*) - \frac{k}{n}\cdot\frac{\partial}{\partial\theta}C(T_{1,n},\theta^*)\right) + o_P(1)
\end{aligned} \tag{33}$$

and that $Q_{n,k}$ can be rewritten as

$$Q_{n,k} = \left\| \frac{k(n-k)}{n^{3/2}} G^{-1/2} F \left(\widehat{\theta}(T_{1,k}) - \widehat{\theta}(T_{k+1,n}) \right) \right\|^2 \quad \text{for all} \quad k \in \mathcal{T}_n. \tag{34}$$

Moreover, applying the central limit theorem for the martingale difference sequence $\left(\frac{\partial}{\partial\theta}\varphi_t(\theta^*), \mathcal{F}_t\right)_{t\in\mathbb{Z}}$ (see Billingsley 1968), we have

$$\begin{aligned} &\frac{1}{\sqrt{n}} \left(\frac{\partial}{\partial\theta} C(T_{1,[n\tau]}, \theta^*) - \frac{[n\tau]}{n} \frac{\partial}{\partial\theta} C(T_{1,n}, \theta^*) \right) \\ &= \frac{1}{\sqrt{n}} \left(\sum_{t=1}^{[n\tau]} \frac{\partial}{\partial\theta} \varphi_t(\theta^*) - \frac{[n\tau]}{n} \sum_{t=1}^{n} \frac{\partial}{\partial\theta} \varphi_t(\theta^*) \right) \\ &\xrightarrow[n\to\infty]{\mathcal{D}} B_G(\tau) - \tau B_G(1), \end{aligned}$$

where $[x]$ denotes the integer part of x and B_G is a Gaussian process with covariance matrix $\min(s,t)G$.

Then,

$$\begin{aligned} &\frac{G^{-1/2}}{\sqrt{n}} \left(\frac{\partial}{\partial\theta} C(T_{1,[n\tau]}, \theta^*) - \frac{[n\tau]}{n} \frac{\partial}{\partial\theta} C(T_{1,n}, \theta^*) \right) \xrightarrow[n\to\infty]{\mathcal{D}} B_d(\tau) - \tau B_d(1) \\ &= W_d(\tau) \quad \text{in} \quad D([0,1]), \end{aligned}$$

where B_d is a d-dimensional standard motion and W_d is a d-dimensional Brownian bridge.

Therefore, using (33) and (34), we obtain

$$\begin{aligned} Q_{n,[n\tau]} &= \left\| \frac{[n\tau](n-[n\tau])}{n^{3/2}} G^{-1/2} F \left(\widehat{\theta}(T_{1,[n\tau]}) - \widehat{\theta}(T_{[n\tau]+1,n}) \right) \right\|^2 \\ &\xrightarrow[n\to\infty]{\mathcal{D}} \sup_{0\le\tau\le 1} \|W_d(\tau)\|^2 \quad \text{in} \quad D([0,1]). \end{aligned}$$

For n large enough, we deduce

$$Q_n = \max_{v_n \le k \le n-v_n} \left(Q_{n,k} \right) = \sup_{\frac{v_n}{n} \le \tau \le 1 - \frac{v_n}{n}} Q_{n,[n\tau]} \xrightarrow[n\to\infty]{\mathcal{D}} \sup_{0\le\tau\le 1} \|W_d(\tau)\|^2 \quad \text{in} \quad D([0,1]);$$

which shows that (27) holds. Hence, we can conclude the proof of the theorem from Lemma 1(i). □

6.1.2 Proof of Theorem 2

Under the alternative, we can write

$$Y_t = \begin{cases} Y_t^{(1)} & \text{for} \quad t \leq t^*, \\ Y_t^{(2)} & \text{for} \quad t > t^*, \end{cases}$$

where $t^* = [\tau^* n]$ (with $0 < \tau^* < 1$) and $\{Y_t^{(j)}, t \in \mathbb{Z}\}$ $(j = 1, 2)$ is a stationary and ergodic process depending on the parameter θ_j^* (with $\theta_1^* \neq \theta_2^*$) satisfying the assumptions (**A1**)–(**A4**).

Remark that $\widehat{Q}_n = \max_{k \in \mathcal{T}_n} \big(\widehat{Q}_{n,k}\big) \geq \widehat{Q}_{n,t^*}$. Then, to prove the theorem, we will show that $\widehat{Q}_{n,t^*} \xrightarrow[n\to\infty]{\mathcal{P}} +\infty$.

For any $n \in \mathbb{N}$, we have

$$\widehat{Q}_{n,t^*} = \frac{(t^*(n-t^*))^2}{n^3} \big(\widehat{\theta}(T_{1,t^*}) - \widehat{\theta}(T_{t^*+1,n})\big)^T \widehat{\Omega}(u_n) \big(\widehat{\theta}(T_{1,t^*}) - \widehat{\theta}(T_{t^*+1,n})\big)$$

with

$$\widehat{\Omega}(u_n) = \frac{1}{2} \Big[\widehat{F}(T_{1,u_n}) \widehat{G}(T_{1,u_n})^{-1} \widehat{F}(T_{1,u_n}) + \widehat{F}(T_{u_n+1,n}) \widehat{G}(T_{u_n+1,n})^{-1} \widehat{F}(T_{u_n+1,n}) \Big].$$

Moreover, the two matrices in the formula of $\widehat{\Omega}(u_n)$ are positive semi-definite. Then, we obtain

$$\begin{aligned} \widehat{Q}_{n,t^*} = & \frac{([\tau^* n](n - [\tau^* n]))^2}{n^3} \left(\widehat{\theta}(T_{1,t^*}) - \widehat{\theta}(T_{t^*+1,n})\right)^T \\ & \times \Big[\widehat{F}(T_{1,u_n}) \widehat{G}(T_{1,u_n})^{-1} \widehat{F}(T_{1,u_n}) + \widehat{F}(T_{u_n+1,n}) \widehat{G}(T_{u_n+1,n})^{-1} \widehat{F}(T_{u_n+1,n}) \Big] \\ & \times \left(\widehat{\theta}(T_{1,t^*}) - \widehat{\theta}(T_{t^*+1,n})\right) \\ \geq & \, n \left(\widehat{\theta}(T_{1,t^*}) - \widehat{\theta}(T_{t^*+1,n})\right)^T \Big[\widehat{F}(T_{1,u_n}) \widehat{G}(T_{1,u_n})^{-1} \widehat{F}(T_{1,u_n}) \Big] \\ & \times \left(\widehat{\theta}(T_{1,t^*}) - \widehat{\theta}(T_{t^*+1,n})\right). \end{aligned} \tag{35}$$

By the consistency and asymptotic normality of the MCE, we have: (i) $\widehat{\theta}(T_{1,t^*}) - \widehat{\theta}(T_{t^*+1,n}) \xrightarrow{a.s.} n \to \infty \theta_1^* - \theta_2^* \neq 0$ and (ii) $\widehat{F}(T_{1,u_n}) \widehat{G}(T_{1,u_n})^{-1} \widehat{F}(T_{1,u_n})$ converges to the covariance matrix of the stationary model of the first regime which is positive definite. Therefore, (35) implies $\widehat{Q}_{n,t^*} \xrightarrow{a.s.} n \to \infty + \infty$. This establishes the theorem. □

6.2 Proof of the results of Section 4

6.3 Proof of Proposition 1

Let $F_\lambda(y)$ be the cumulative distribution function of $p(y|\eta)$ with marginals $F_{\lambda_1,1},\ldots,F_{\lambda_m,m}$, where $\lambda=(\lambda_1,\ldots,\lambda_m)^T=\partial A(\eta)/\partial\eta$. From the Sklar's theorem (see Sklar 1959), one can find a copula $\mathcal{C}$ such that, for all $y=(y_1,\ldots,y_m)\in\mathbb{R}^m$

$$F_\lambda(y)=\mathcal{C}\big(F_{\lambda_1,1}(y_1),\ldots,F_{\lambda_m,m}(y_m)\big).$$

For $i=1,\ldots,m$, denote by $F^{-1}_{\lambda,i}(u):=\inf\{y_i\geq 0,\ F_{\lambda,i}(y_i)\geq u\}$ for all $u\in[0,1]$. Let $\{U_t=(U_{t,1},\ldots,U_{t,m})^T,\ t\in\mathbb{Z}\}$ be a sequence of independent random vectors with distribution $\mathcal{C}$. We will prove that there exists a τ-weakly dependent, stationary and ergodic solution (Y_t,λ_t) of (15) satisfying:

$$Y_t=\Big(F^{-1}_{\lambda_{t,1},1}(U_{t,1}),\ldots,F^{-1}_{\lambda_{t,m},m}(U_{t,m})\Big)^T \tag{36}$$

with $\lambda_t=\big(\lambda_{t,1},\ldots,\lambda_{t,m}\big)^T=f_\theta(Y_{t-1},\ldots)$. For a process $(Y_t)_{t\in\mathbb{Z}}$ that fulfills (15) and (36), we get,

$$Y_t=\Big(F^{-1}_{\lambda_{t,1},1}(U_{t,1}),\ldots,F^{-1}_{\lambda_{t,m},m}(U_{t,m})\Big)^T:=\Psi(Y_{t-1},\ldots;U_t), \tag{37}$$

where Ψ is a function defined in $(\mathbb{N}_0^m)^\infty\times[0,1]^m$. According to Doukhan and Wintenberger (2008), it suffices to show that: (i) $\mathbb{E}\|\Psi(\boldsymbol{y};U_t)\|<\infty$ for some $\boldsymbol{y}\in(\mathbb{N}_0^m)^\infty$ and (ii) there exists a sequence of nonnegative real numbers $(\alpha_k(\Psi))_{k\geq1}$ satisfying $\sum_{k\geq1}\alpha_k(\Psi)<1$ such that, for all $\boldsymbol{y},\boldsymbol{y}'\in(\mathbb{N}_0^m)^\infty$, $\mathbb{E}\|\Psi(\boldsymbol{y};U_t)-\Psi(\boldsymbol{y}';U_t)\|\leq\sum_{k\geq1}\alpha_k(\Psi)\|y_k-y'_k\|$.

Proof of (i). Set $f_\theta(0,\ldots)=\lambda=(\lambda_1,\ldots,\lambda_m)^T$. The random vector $\big(F^{-1}_{\lambda_1,1}(U_{t,1},\ldots,F^{-1}_{\lambda_m,m}(U_{t,m})\big)^T$ is F_λ distributed. Thus,

$$\begin{aligned}\mathbb{E}\|\Psi(0,\ldots;U_t)\|&=\mathbb{E}\left\|\Big(F^{-1}_{\lambda_1,1}(U_{t,1}),\ldots,F^{-1}_{\lambda_m,m}(U_{t,m})\Big)\right\|=\|\lambda\|\\&=\|f_\theta(0,\ldots)\|<\infty,\end{aligned}$$

where this inequality holds from the assumption $\mathbf{A}_0(\Theta)$. □

Proof of (ii). For all $\boldsymbol{y},\boldsymbol{y}'\in(\mathbb{N}_0^m)^\infty$, set $\lambda=f_\theta(\boldsymbol{y},\ldots)=(\lambda_1,\ldots,\lambda_m)^T$ and $\lambda'=f_\theta(\boldsymbol{y}',\ldots)=(\lambda'_1,\ldots,\lambda'_m)^T$. We have,

$$
\begin{aligned}
&\mathbb{E}\|\Psi(\mathbf{y}; U_t) - \Psi(\mathbf{y}'; U_t)\| \\
&\quad = \mathbb{E}\left\| \left(F^{-1}_{\lambda_1,1}(U_{t,1}), \ldots, F^{-1}_{\lambda_m,m}(U_{t,m})\right) - \left(F^{-1}_{\lambda'_1,1}(U_{t,1}), \ldots, F^{-1}_{\lambda'_m,m}(U_{t,m})\right) \right\| \\
&\quad \le \sum_{i=1}^{m} \mathbb{E}\left|F^{-1}_{\lambda_i,i}(U_{t,i}) - F^{-1}_{\lambda'_i,i}(U_{t,i})\right| = \sum_{i=1}^{m} |\lambda_i - \lambda'_i| \qquad (38) \\
&\quad = \|\lambda - \lambda'\| = \|f_\theta(\mathbf{y}, \ldots) - f_\theta(\mathbf{y}', \ldots)\| \le \sum_{k=1}^{\infty} \alpha_k^{(0)} \|y_k - y'_k\|, \qquad (39)
\end{aligned}
$$

where the equality in (38) holds from the Proposition A.2 of Davis and Liu (2016) and the inequality in (39) holds from the assumption $\mathbf{A}_0(\Theta)$. Thus, take $\alpha_k(\Psi) = \alpha_k^{(0)}$, which completes the proof of the proposition. □

6.3.1 Proof of Theorem 3

To simplify, we will use the following notations in the sequel:

$$
\begin{aligned}
\ell_{t,i}(\theta) &:= Y_{t,i} \log \lambda_{t,i}(\theta) - \lambda_{t,i}(\theta) = Y_{t,i} \log f_\theta^{t,i} - f_\theta^{t,i}, \\
\widehat{\ell}_{t,i}(\theta) &:= Y_{t,i} \log \widehat{\lambda}_{t,i}(\theta) - \widehat{\lambda}_{t,i}(\theta) = Y_{t,i} \log \widehat{f}_\theta^{t,i} - \widehat{f}_\theta^{t,i},
\end{aligned}
$$

where $f_\theta^{t,i}$ and $\widehat{f}_\theta^{t,i}$ (for $i = 1, \ldots, m$) represent the ith component of $f_\theta^t \equiv f_\theta(Y_{t-1}, Y_{t-2}, \ldots)$ and $\widehat{f}_\theta^t \equiv f_\theta(Y_{t-1}, Y_{t-2}, \ldots, Y_1)$, respectively.

(i) Firstly, we will show that

$$
\frac{1}{n}\left\|\widehat{L}_n(\theta) - L_n(\theta)\right\|_\Theta \xrightarrow{a.s.} n \to \infty 0. \qquad (40)
$$

Remark that

$$
\begin{aligned}
\frac{1}{n}\left\|\widehat{L}_n(\theta) - L_n(\theta)\right\|_\Theta &\le \frac{1}{n}\sum_{t=1}^{n} \|\widehat{\ell}_t(\theta) - \ell_t(\theta)\|_\Theta \\
&\le \frac{1}{n}\sum_{i=1}^{m}\sum_{t=1}^{n} \|\widehat{\ell}_{t,i}(\theta) - \ell_{t,i}(\theta)\|_\Theta. \qquad (41)
\end{aligned}
$$

Using $\mathbf{A}_0(\Theta)$ with the condition (16) and the existence of the moment of order 2 (i.e., (14) with $\epsilon \ge 1$), one can proceed as in the proof of Theorem 3.1 in Doukhan and Kengne (2015) to prove that

$$
\frac{1}{n}\sum_{t=1}^{n} \|\widehat{\ell}_{t,i}(\theta) - \ell_{t,i}(\theta)\|_\Theta \xrightarrow{a.s.} n \to \infty 0 \quad \text{for all} \quad i = 1, \ldots, m.
$$

Therefore, (40) is obtained by using (41).

(ii) Let us establish that: for all $t \in \mathbb{Z}$,

$$\mathbb{E}\big[\|\ell_t(\theta)\|_\Theta\big] < \infty. \tag{42}$$

We have

$$\mathbb{E}\big[\|\ell_t(\theta)\|_\Theta\big] \leq \sum_{i=1}^{m} \mathbb{E}\big[\sup_{\theta\in\Theta} |\ell_{t,i}(\theta)|\big],$$

From $\mathbf{A}_0(\Theta)$, ($\mathcal{MOD}$.**A0**), (14) (with $\epsilon \leq 1$) and by going along similar lines as in the proof of Theorem 3.1 in Doukhan and Kengne (2015), we get: $\mathbb{E}\big[\sup_{\theta\in\Theta} |\ell_{t,i}(\theta)|\big] < \infty$ for all $i = 1, \ldots, m$. Thus, (42) holds.
Since $\{Y_t,\ t \in \mathbb{Z}\}$ is stationary and ergodic, the process $\{\ell_t(\theta),\ t \in \mathbb{Z}\}$ is also a stationary and ergodic sequence. Then, by the uniform strong law of large numbers applied to $\{\ell_t(\theta),\ t \in \mathbb{Z}\}$, it holds that

$$\left\|\frac{1}{n}L_n(\theta) - \mathbb{E}[\ell_0(\theta)]\right\|_\Theta = \left\|\frac{1}{n}\sum_{t=1}^{n}\ell_t(\theta) - \mathbb{E}[\ell_0(\theta)]\right\|_\Theta \xrightarrow{a.s.} n \to \infty 0.$$

Thus, from (40), we obtain

$$\begin{aligned}\left\|\frac{1}{n}\widehat{L}_n(\theta) - \mathbb{E}[\ell_0(\theta)]\right\|_\Theta &\leq \frac{1}{n}\left\|\widehat{L}_n(\theta) - L_n(\theta)\right\|_\Theta \\ &\quad + \left\|\frac{1}{n}L_n(\theta) - \mathbb{E}[\ell_0(\theta)]\right\|_\Theta \xrightarrow{a.s.} n \to \infty 0.\end{aligned}$$

(iii) To complete the proof of the theorem, it suffices to show that the function $\theta \mapsto L(\theta) = \mathbb{E}[\ell_0(\theta)]$ has a unique maximum at θ^*. Let $\theta \in \Theta$, such that $\theta \neq \theta^*$. We have

$$\begin{aligned}L(\theta^*) - L(\theta) &= \sum_{i=1}^{m}\left(\mathbb{E}\ell_{0,i}(\theta^*) - \mathbb{E}\ell_{0,i}(\theta)\right) \\ &= \sum_{i=1}^{m}\left(\mathbb{E}\left[Y_{0,i}\log f_{\theta^*}^{0,i} - f_{\theta^*}^{0,i}\right] - \mathbb{E}\left[Y_{0,i}\log f_{\theta}^{0,i} - f_{\theta}^{0,i}\right]\right) \\ &= \sum_{i=1}^{m}\left(\mathbb{E}\left[f_{\theta^*}^{0,i}\log f_{\theta^*}^{0,i} - f_{\theta^*}^{0,i}\right] - \mathbb{E}\left[f_{\theta^*}^{0,i}\log f_{\theta}^{0,i} - f_{\theta}^{0,i}\right]\right) \\ &= \sum_{i=1}^{m}\left(\mathbb{E}\left[f_{\theta^*}^{0,i}\left(\log f_{\theta^*}^{0,i} - \log f_{\theta}^{0,i}\right)\right] - \mathbb{E}\left(f_{\theta^*}^{0,i} - f_{\theta}^{0,i}\right)\right).\end{aligned}$$

According to the identifiability assumption $\mathbf{A}_0(\Theta)$ and since $\theta \neq \theta^*$, there exists i_0 such that $f_\theta^{0,i_0} \neq f_{\theta^*}^{0,i_0}$. By going as in the proof of Theorem 3.1 in Doukhan and Kengne (2015), we get $\mathbb{E}\big[f_{\theta^*}^{0,i_0}\big(\log f_{\theta^*}^{0,i_0} - \log f_\theta^{0,i_0}\big)\big] - \mathbb{E}[f_{\theta^*}^{0,i_0} - f_\theta^{0,i_0}] > 0$

and $\mathbb{E}\big[f_{\theta^*}^{0,i}\big(\log f_{\theta^*}^{0,i} - \log f_{\theta}^{0,i}\big)\big] - \mathbb{E}[f_{\theta^*}^{0,i} - f_{\theta}^{0,i}] \geq 0$ for $i = 1, \ldots, m,\ i \neq i_0$. This establishes (iii), which consequently yields the theorem. □

6.3.2 Proof of Theorem 4

Applying the mean value theorem to the function $\theta \mapsto \frac{\partial}{\partial\theta_i} L_n(\theta)$ for all $i \in \{1, \ldots, d\}$, there exists $\bar{\theta}_{n,i}$ between $\widehat{\theta}_n$ and θ^* such that

$$\frac{\partial}{\partial\theta_i} L_n(\widehat{\theta}_n) = \frac{\partial}{\partial\theta_i} L_n(\theta^*) + \frac{\partial^2}{\partial\theta\partial\theta_i} L_n(\bar{\theta}_{n,i})(\widehat{\theta}_n - \theta^*),$$

which is equivalent to

$$\begin{aligned}\sqrt{n} J(\widehat{\theta}_n)(\widehat{\theta}_n - \theta^*) &= \frac{1}{\sqrt{n}}\left(\frac{\partial}{\partial\theta} L_n(\theta^*) - \frac{\partial}{\partial\theta}\widehat{L}_n(\widehat{\theta}_n)\right) \\ &\quad + \frac{1}{\sqrt{n}}\left(\frac{\partial}{\partial\theta}\widehat{L}_n(\widehat{\theta}_n) - \frac{\partial}{\partial\theta} L_n(\widehat{\theta}_n)\right)\end{aligned} \tag{43}$$

with

$$J(\widehat{\theta}_n) = \left(-\frac{1}{n}\frac{\partial^2}{\partial\theta\partial\theta_i} L_n(\bar{\theta}_{n,i})\right)_{1\leq i\leq d}. \tag{44}$$

The following lemma is needed.

Lemma 2 *Assume that the conditions of Theorem 4 hold. Then,*

(i) $\mathbb{E}\Big[\frac{1}{\sqrt{n}}\Big\|\frac{\partial}{\partial\theta}\widehat{L}_n(\theta) - \frac{\partial}{\partial\theta} L_n(\theta)\Big\|_\Theta\Big] \underset{n\to\infty}{\longrightarrow} 0.$

(ii) $\frac{1}{n}\Big\|\frac{\partial^2}{\partial\theta\partial\theta^T}\widehat{L}_n(\theta) - \frac{\partial^2}{\partial\theta\partial\theta^T} L_n(\theta)\Big\|_\Theta = o(1)\ a.s..$

Proof (i) We have

$$\begin{aligned}&\frac{1}{\sqrt{n}}\left\|\frac{\partial}{\partial\theta}\widehat{L}_n(\theta) - \frac{\partial}{\partial\theta} L_n(\theta)\right\|_\Theta \\ &\quad \leq \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\|\frac{\partial}{\partial\theta}\widehat{\ell}_t(\theta) - \frac{\partial}{\partial\theta}\ell_t(\theta)\right\|_\Theta \\ &\quad \leq \frac{1}{\sqrt{n}}\sum_{i=1}^{m}\sum_{t=1}^{n}\left\|\frac{\partial}{\partial\theta}\widehat{\ell}_{t,i}(\theta) - \frac{\partial}{\partial\theta}\ell_{t,i}(\theta)\right\|_\Theta.\end{aligned} \tag{45}$$

Moreover, by proceeding as in Lemma 7.1 of Doukhan and Kengne (2015), we can use $\mathbf{A}_i(\Theta)$ $(i = 0, 1)$, (14) and the condition (17) to establish that

$$\mathbb{E}\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\|\frac{\partial}{\partial\theta}\widehat{\ell}_{t,i}(\theta) - \frac{\partial}{\partial\theta}\ell_{t,i}(\theta)\right\|_\Theta\right] \underset{n\to\infty}{\longrightarrow} 0 \quad \text{for all} \quad i = 1, \ldots, m.$$

Thus, we can conclude the proof of (i) from (45).

(ii) It holds that

$$\begin{aligned}\frac{1}{n}\Big\|\frac{\partial^2}{\partial\theta\partial\theta^T}\widehat{L}_n(\theta)-\frac{\partial^2}{\partial\theta\partial\theta^T}L_n(\theta)\Big\|_\Theta &\leq \frac{1}{n}\sum_{t=1}^{n}\Big\|\frac{\partial^2}{\partial\theta\partial\theta^T}\widehat{\ell}_t(\theta)-\frac{\partial^2}{\partial\theta\partial\theta^T}\ell_t(\theta)\Big\|_\Theta\\ &\leq \frac{1}{n}\sum_{i=1}^{m}\sum_{t=1}^{n}\Big\|\frac{\partial^2}{\partial\theta\partial\theta^T}\widehat{\ell}_{t,i}(\theta)-\frac{\partial^2}{\partial\theta\partial\theta^T}\ell_{t,i}(\theta)\Big\|_\Theta.\end{aligned}$$

By going as in the proof of Lemma 7.1 of Doukhan and Kengne (2015), one easily get for $i=1,\ldots,m$, $\frac{1}{n}\sum_{t=1}^{n}\Big\|\frac{\partial^2}{\partial\theta\partial\theta^T}\widehat{\ell}_{t,i}(\theta)-\frac{\partial^2}{\partial\theta\partial\theta^T}\ell_{t,i}(\theta)\Big\|_\Theta=o(1)$, which shows that (ii) holds. $\square$

The following lemma will also be needed.

Lemma 3 *If the assumptions of Theorem 4 hold, then*

(a) *the matrices* $J_{\theta^*}=\mathbb{E}\big[\frac{\partial\lambda_0^T(\theta^*)}{\partial\theta}D_0^{-1}(\theta^*)\frac{\partial\lambda_0(\theta^*)}{\partial\theta^T}\big]$ *and* $I_{\theta^*}=\mathbb{E}\big[\frac{\partial\lambda_0^T(\theta^*)}{\partial\theta}D_0^{-1}(\theta^*)\,\Gamma_0(\theta^*)D_0^{-1}(\theta^*)\frac{\partial\lambda_0(\theta^*)}{\partial\theta^T}\big]$ *exist and are positive definite;*

(b) $\Big(\frac{\partial\ell_t(\theta^*)}{\partial\theta},\mathcal{F}_t\Big)_{t\in\mathbb{Z}}$ *is a stationary ergodic, square integrable martingale difference sequence with covariance matrix* I_{θ^*}*;*

(c) $\mathbb{E}\big[\|\frac{\partial^2\ell_0(\theta)}{\partial\theta\partial\theta^T}\|_\Theta\big]<\infty$ *and* $\mathbb{E}\big[\frac{\partial^2\ell_0(\theta^*)}{\partial\theta\partial\theta^T}\big]=-J_{\theta^*}$*;*

(d) $J(\widehat{\theta}_n)\xrightarrow{a.s.} n\to\infty J_{\theta^*}$ *and that the matrix* J_{θ^*} *is invertible.*

Proof (*a*) From the assumption ($\mathcal{MOD}$.**A0**), we can find a constant $C>0$ such that it holds a.s.

$$\begin{aligned}\mathbb{E}\Big\|\frac{\partial\lambda_0^T(\theta)}{\partial\theta}D_0^{-1}(\theta)\frac{\partial\lambda_0(\theta)}{\partial\theta^T}\Big\|_\Theta &\leq C\mathbb{E}\Big\|\frac{\partial\lambda_0(\theta)}{\partial\theta}\Big\|_\Theta^2\\ &\leq C\sum_{i=1}^{m}\mathbb{E}\Big\|\frac{\partial\lambda_{0,i}(\theta)}{\partial\theta}\Big\|_\Theta^2.\end{aligned}$$

One can show as in the proof of Lemma 7.1 of Doukhan and Kengne (2015) that for any $i=1,\ldots,m$, $\mathbb{E}\big\|\frac{\partial\lambda_{0,i}(\theta)}{\partial\theta}\big\|_\Theta^2<\infty$. Hence,

$$\mathbb{E}\Big\|\frac{\partial\lambda_0^T(\theta)}{\partial\theta}D_0^{-1}(\theta)\frac{\partial\lambda_0(\theta)}{\partial\theta^T}\Big\|_\Theta<\infty,$$

which establishes the existence of J_{θ^*}.

Using ($\mathcal{MOD}$.**A0**) and Hölder's inequality, we have

$$
\begin{aligned}
&\mathbb{E}\Big\| \frac{\partial \lambda_0^T(\theta)}{\partial \theta} D_0^{-1}(\theta)\Gamma_0(\theta)D_0^{-1}(\theta)\frac{\partial \lambda_0(\theta)}{\partial \theta^T}\Big\|_\Theta \leq \mathbb{E} \\
&\Big[\Big\| D_0^{-1}(\theta)(Y_0 - \lambda_0(\theta))\Big\|_\Theta^2 \Big\| \frac{\partial \lambda_0(\theta)}{\partial \theta}\Big\|_\Theta^2 \Big] \\
&\leq C\mathbb{E}\Big[\big(\|Y_0\|^2 + \|\lambda_0(\theta)\|_\Theta^2 \big) \Big\| \frac{\partial \lambda_0(\theta)}{\partial \theta}\Big\|_\Theta^2 \Big] \\
&\leq C\left(\mathbb{E}\|Y_0\|^4 + \mathbb{E}\|\lambda_0(\theta)\|_\Theta^4\right)^{1/2} \left(\mathbb{E}\Big\| \frac{\partial \lambda_0(\theta)}{\partial \theta}\Big\|_\Theta^4\right)^{1/2} \\
&\leq C\left(\mathbb{E}\|Y_0\|^4 + \sum_{i=1}^m \mathbb{E}\|\lambda_{0,i}(\theta)\|_\Theta^4\right)^{1/2} \left(\sum_{i=1}^m \mathbb{E}\Big\| \frac{\partial \lambda_{0,i}(\theta)}{\partial \theta}\Big\|_\Theta^4\right)^{1/2}.
\end{aligned}
$$

According to the existence of the moment of order 4 (from (14) with $\epsilon \geq 3$), $\mathbb{E}\|Y_0\|^4 < \infty$. Furthermore, proceeding as in the proof of Theorem 3.1 and Lemma 7.1 in Doukhan and Kengne (2015), one can also get $\mathbb{E}\|\lambda_{0,i}(\theta)\|_\Theta^4 < \infty$ and $\mathbb{E}\|\frac{\partial \lambda_{0,i}(\theta)}{\partial \theta}\|_\Theta^4 < \infty$ for any $i = 1, \ldots, m$. Therefore,

$$
\mathbb{E}\Big\| \frac{\partial \lambda_0^T(\theta)}{\partial \theta} D_0^{-1}(\theta)\Gamma_0(\theta)D_0^{-1}(\theta)\frac{\partial \lambda_0(\theta)}{\partial \theta^T}\Big\|_\Theta < \infty, \tag{46}
$$

which establishes that I_{θ^*} exists.

Now, let $U \in \mathbb{R}^d$ be a nonzero vector. We have $\frac{\partial \lambda_0(\theta^*)}{\partial \theta^T} \cdot U \neq 0$ a.s. from the assumption ($\mathcal{MOD}$.**A2**), which implies

$$
\begin{aligned}
U^T J_{\theta^*} U &= \mathbb{E}\left[U^T \frac{\partial \lambda_0^T(\theta^*)}{\partial \theta} D_0^{-1}(\theta^*)\frac{\partial \lambda_0(\theta^*)}{\partial \theta^T} U\right] \\
&= \mathbb{E}\left[\left(D_0^{-1/2}(\theta^*)\frac{\partial \lambda_0(\theta^*)}{\partial \theta^T} U\right)^T \cdot \left(D_0^{-1/2}(\theta^*)\frac{\partial \lambda_0(\theta^*)}{\partial \theta^T} U\right)\right] > 0
\end{aligned}
$$

and

$$
\begin{aligned}
U^T I_{\theta^*} U &= \mathbb{E}\left[U^T \frac{\partial \lambda_0^T(\theta^*)}{\partial \theta} D_0^{-1}(\theta^*)\Gamma_0(\theta^*)D_0^{-1}(\theta^*)\frac{\partial \lambda_0(\theta^*)}{\partial \theta^T} U\right] \\
&= \mathbb{E}\Bigg[\left(\big(Y_0 - \lambda_0(\theta^*)\big)^T D_0^{-1/2}(\theta^*)\frac{\partial \lambda_0(\theta^*)}{\partial \theta^T} U\right)^T \cdot \\
&\qquad \left(\big(Y_0 - \lambda_0(\theta^*)\big)^T D_0^{-1/2}(\theta^*)\frac{\partial \lambda_0(\theta^*)}{\partial \theta^T} U\right)\Bigg] > 0.
\end{aligned}
$$

Hence, J_{θ^*} and I_{θ^*} are positive definite.

(b) For any $\theta \in \Theta$, we have

$$\frac{\partial \ell_t(\theta)}{\partial \theta} = \sum_{i=1}^{m} \left(\frac{Y_{t,i}}{\lambda_{t,i}(\theta)} - 1 \right) \frac{\partial}{\partial \theta} \lambda_{t,i}(\theta) = \frac{\partial \lambda_t^T(\theta)}{\partial \theta} D_t^{-1}(\theta)(Y_t - \lambda_t(\theta)). \tag{47}$$

Then, according to the stability properties of $\{Y_t,\ t \in \mathbb{Z}\}$, the process $\left\{\frac{\partial \ell_t(\theta)}{\partial \theta},\ t \in \mathbb{Z}\right\}$ is also stationary and ergodic. Moreover, since $\lambda_t(\theta^*)$ and $\frac{\partial \lambda_t(\theta^*)}{\partial \theta}$ are $\mathcal{F}_{t-1}$-measurable, we have

$$\begin{aligned} \mathbb{E}\left[\frac{\partial \ell_t(\theta^*)}{\partial \theta}\right] &= \sum_{i=1}^{m} \mathbb{E}\left[\mathbb{E}\left[\frac{\partial \lambda_t^T(\theta^*)}{\partial \theta} D_t^{-1}(\theta^*)(Y_t - \lambda_t(\theta^*))|\mathcal{F}_{t-1}\right]\right] \\ &= \sum_{i=1}^{m} \mathbb{E}\left[\frac{\partial \lambda_t^T(\theta^*)}{\partial \theta} D_t^{-1}(\theta^*) \cdot \mathbb{E}\left[\left(Y_t - \lambda_t(\theta^*)\right)|\mathcal{F}_{t-1}\right]\right] = 0 \end{aligned}$$

In addition, $\mathbb{E}\left[\frac{\partial \ell_0(\theta^*)}{\partial \theta}\frac{\partial \ell_0(\theta^*)}{\partial \theta^T}\right] = I_{\theta^*}$. Hence, the part second part of the lemma holds.

(c) We have,

$$\mathbb{E}\left\|\frac{\partial^2 \ell_0(\theta)}{\partial \theta \partial \theta^T}\right\|_\Theta = \sum_{i=1}^{m} \mathbb{E}\left\|\frac{\partial^2 \ell_{0,i}(\theta)}{\partial \theta \partial \theta^T}\right\|_\Theta < \infty,$$

where the above inequality holds since $\mathbb{E}\left\|\frac{\partial^2 \ell_{0,i}(\theta)}{\partial \theta \partial \theta^T}\right\|_\Theta < \infty$ for $i = 1, \ldots, m$ by going as in the proof of Lemma 7.2 in Doukhan and Kengne (2015). Moreover, according to (47), for any $\theta \in \Theta$, we have

$$\frac{\partial^2 \ell_t(\theta)}{\partial \theta \partial \theta^T} = \sum_{i=1}^{m} \left(\frac{Y_{t,i}}{\lambda_{t,i}(\theta)} - 1 \right) \frac{\partial^2 \lambda_{t,i}(\theta)}{\partial \theta \partial \theta^T} - \sum_{i=1}^{m} \frac{Y_{t,i}}{\lambda_{t,i}^2(\theta)} \frac{\partial \lambda_{t,i}(\theta)}{\partial \theta} \frac{\partial \lambda_{t,i}(\theta)}{\partial \theta^T}.$$

Then, using conditional expectations, we obtain

$$\begin{aligned} \mathbb{E}\left[\frac{\partial^2 \ell_0(\theta^*)}{\partial \theta \partial \theta^T}\right] &= -\mathbb{E}\left[\sum_{i=1}^{m} \frac{1}{\lambda_{0,i}(\theta^*)} \frac{\partial \lambda_{0,i}(\theta^*)}{\partial \theta} \frac{\partial \lambda_{0,i}(\theta^*)}{\partial \theta^T}\right] \\ &= -\left[\frac{\partial \lambda_0^T(\theta^*)}{\partial \theta} D_0^{-1}(\theta^*) \frac{\partial \lambda_0(\theta^*)}{\partial \theta^T}\right] = -J_{\theta^*}. \end{aligned}$$

(d) We have

$$J(\widehat{\theta}_n) = \left(-\frac{1}{n} \frac{\partial^2}{\partial \theta \partial \theta_i} L_n(\bar{\theta}_{n,i})\right)_{1 \le i \le d} = \left(-\frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 \ell_t(\bar{\theta}_{n,i})}{\partial \theta \partial \theta_i}\right)_{1 \le i \le d}.$$

Since $\widehat{\theta}_n \xrightarrow{a.s.} n \to \infty\theta^*$, $\bar{\theta}_{n,i} \xrightarrow{a.s.} n \to \infty\theta^*$ (for any $i = 1, \ldots, d$) and that $\mathbb{E}[\frac{\partial^2 \ell_0(\theta^*)}{\partial\theta\partial\theta^T}] = -J_{\theta^*}$ exists, by the uniform strong law of large numbers, for any $i = 1, \ldots, d$, we get

$$\frac{1}{n}\sum_{t=1}^{n} \frac{\partial^2 \ell_t(\bar{\theta}_{n,i})}{\partial\theta\partial\theta_i} \overset{a.s.}{=} \frac{1}{n}\sum_{t=1}^{n} \frac{\partial^2 \ell_t(\theta^*)}{\partial\theta\partial\theta_i} \xrightarrow{a.s.} \mathbb{E}\left[\frac{\partial^2 \ell_0(\theta^*)}{\partial\theta\partial\theta_i}\right] \quad \text{as} \quad n \to \infty.$$

Therefore,

$$J(\widehat{\theta}_n) = \left(-\frac{1}{n}\sum_{t=1}^{n} \frac{\partial^2 \ell_t(\bar{\theta}_{n,i})}{\partial\theta\partial\theta_i}\right)_{1\le i\le d} \xrightarrow{a.s.} n \to \infty - \left(\mathbb{E}\left[\frac{\partial^2 \ell_0(\theta^*)}{\partial\theta\partial\theta_i}\right]\right)_{1\le i\le d}$$
$$= J_{\theta^*}.$$

This completes the proof of the lemma. □

Now, let us use the results of Lemmas 2 and 3 to complete the proof of Theorem 4.

Since $\widehat{\theta}_n$ is a local maximum of the function $\theta \mapsto \widehat{L}_n(\theta)$ for n large enough (from the assumption ($\mathcal{MOD}$.**A1**) and the consistency of $\widehat{\theta}_n$), $\frac{\partial}{\partial\theta}\widehat{L}_n(\widehat{\theta}_n) = 0$.

Thus, according to Lemma 2, the relation (43) becomes

$$\sqrt{n}J(\widehat{\theta}_n)(\widehat{\theta}_n - \theta^*) = \frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}L_n(\theta^*) + o_P(1). \tag{48}$$

Moreover, applying the central limit theorem to the sequence $\left(\frac{\partial\ell_t(\theta^*)}{\partial\theta}, \mathcal{F}_t\right)_{t\in\mathbb{Z}}$, it holds that

$$\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}L_n(\theta^*) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\ell_t(\theta^*) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_d(0, I_{\theta^*}).$$

Therefore, for n large enough, using Lemma 3(d) and the relation (48), we obtain

$$\sqrt{n}(\widehat{\theta}_n - \theta^*) = J_{\theta^*}^{-1}\left[\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}L_n(\theta^*)\right] + o_P(1) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_d\left(0, J_{\theta^*}^{-1}I_{\theta^*}J_{\theta^*}^{-1}\right).$$

This establishes the theorem. □

Acknowledgements The authors are grateful to the two anonymous referees for many relevant suggestions and comments which helped to improve the contents of this paper.

References

Ahmad A (2016) Contribution à l'économétrie des séries temporelles à valeurs entières. Ph.D. Thesis, Université de Lille

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

Aknouche A, Francq C (2021) Count and duration time series with equal conditional stochastic and mean orders. Economet Theor 37(2):248–280
Aknouche A, Bendjeddou S, Touche N (2018) Negative binomial quasi-likelihood inference for general integer-valued time series models. J Time Ser Anal 39(2):192–211
Bardet JM, Wintenberger O (2009) Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes. Ann Stat 37(5B):2730–2759
Billingsley P (1968) Convergence of probability measures. Wiley, New York
Cui Y, Li Q, Zhu F (2020) Flexible bivariate Poisson integer-valued Garch model. Ann Inst Stat Math 72(6):1449–1477
Davis RA, Liu H (2016) Theory and inference for a class of nonlinear models with application to time series of counts. Stat. Sinica 1673–1707
de Souza JB, Reisen VA, Franco GC, Ispány M, Bondon P, Santos JM (2018) Generalized additive models with principal component analysis: an application to time series of respiratory disease and air pollution data. J Roy Stat Soc: Ser C (Appl Stat) 67(2):453–480
Dedecker J, Doukhan P, Lang G, José Rafael LR, Louhichi S, Prieur C (2007) Weak dependence. In: Weak dependence: with examples and applications. Springer, pp 9–20
Diop ML, Kengne W (2021) Poisson QMLE for change-point detection in general integer-valued time series models. Metrika 1–31
Diop ML, Kengne W (2017) Testing parameter change in general integer-valued time series. J Time Ser Anal 38(6):880–894
Diop ML, Kengne W (2022) Inference and model selection in general causal time series with exogenous covariates. Electron J Stat 16(1):116–157
Diop ML, Kengne W (2022) Epidemic change-point detection in general causal time series. Stat Probab Lett 184:109416
Doukhan P, Kengne W (2015) Inference and testing for structural change in general Poisson autoregressive models. Electron J Stat 9:1267–1314
Doukhan P, Louhichi S (1999) A new weak dependence condition and applications to moment inequalities. Stoch Process Their Appl 84(2):313–342
Doukhan P, Wintenberger O (2008) Weakly dependent chains with infinite memory. Stoch Process Their Appl 118(11):1997–2013
Dvořák M, Prášková Z (2013) On testing changes in autoregressive parameters of a VAR model. Commun Stat Theory Methods 42(7):1208–1226
Fokianos K, Gombay E, Hussein A (2014) Retrospective change detection for binary time series models. J Stat Plan Inference 145:102–112
Fokianos K, Støve B, Tjøstheim D, Doukhan P (2020) Multivariate count autoregression. Bernoulli 26(1):471–499
Franke J, Kirch C, Kamgaing JT (2012) Changepoints in times series of counts. J Time Ser Anal 33(5):757–770
Hudecová Š (2013) Structural changes in autoregressive models for binary time series. J Stat Plan Inference 143(10):1744–1752
Hudecová Š, Hušková M, Meintanis SG (2017) Tests for structural changes in time series of counts. Scand J Stat 44(4):843–865
Kang J, Lee S (2009) Parameter change test for random coefficient integer-valued autoregressive processes with application to polio data analysis. J Time Ser Anal 30(2):239–258
Kang J, Lee S (2014) Parameter change test for Poisson autoregressive models. Scand J Stat 41(4):1136–1152
Kang J, Song J (2015) Robust parameter change test for Poisson autoregressive models. Stat Probab Lett 104:14–21
Kengne WC (2012) Testing for parameter constancy in general causal time-series models. J Time Ser Anal 33(3):503–518
Khatri CG (1983) Multivariate discrete exponential family of distributions and their properties. Commun Stat Theory Methods 12(8):877–893
Killick R, Fearnhead P, Eckley IA (2012) Optimal detection of changepoints with a linear computational cost. J Am Stat Assoc 107(500):1590–1598
Kim B, Lee S (2020) Robust estimation for general integer-valued time series models. Ann Inst Stat Math 72(6):1371–1396

Kirch C, Muhsal B, Ombao H (2015) Detection of changes in multivariate time series with application to EEG data. J Am Stat Assoc 110(511):1197–1216

Klimko LA, Nelson PI (1978) On conditional least squares estimation for stochastic processes. Ann Stat 629–642

Lee S, Lee T (2004) Cusum test for parameter change based on the maximum likelihood estimator. Seq Anal 23(2):239–256

Lee S, Na O (2005) Test for parameter change in stochastic processes based on conditional least-squares estimator. J Multivar Anal 93(2):375–393

Lee S, Na O (2005) Test for parameter change based on the estimator minimizing density-based divergence measures. Ann Inst Stat Math 57(3):553–573

Lee S, Song J (2008) Test for parameter change in ARMA models with GARCH innovations. Stat Probab Lett 78(13):1990–1998

Lee S, Ha J, Na O, Na S (2003) The cusum test for parameter change in time series models. Scand J Stat 30(4):781–796

Lee Y, Lee S, Tjøstheim D (2018) Asymptotic normality and parameter change test for bivariate Poisson INGARCH models. Test 27(1):52–69

Ng KY, Awang N (2018) Multiple linear regression and regression with time series error models in forecasting pm10 concentrations in peninsular Malaysia. Environ Monit Assess 190(2):1–11

Page E (1955) A test for a change in a parameter occurring at an unknown point. Biometrika 42(3/4):523–527

Qu Z, Perron P (2007) Estimating and testing structural changes in multivariate regressions. Econometrica 75(2):459–502

Sklar M (1959) Fonctions de repartition an dimensions et leurs marges. Publ Inst Statist Univ Paris 8:229–231