

# Inference and testing for structural change in general Poisson autoregressive models

Paul Doukhan\*

*AGM, Université de Cergy-Pontoise, 2 avenue Adolphe Chauvin. 95302 Cergy-Pontoise, France IUF, University Institute of France  
e-mail: [paul.doukhan@u-cergy.fr](mailto:paul.doukhan@u-cergy.fr)*

and

William Kengne\*

*THEMA, Université de Cergy-Pontoise, 33 Boulevard du Port, 95011 Cergy-Pontoise Cedex, France  
e-mail: [william.kengne@u-cergy.fr](mailto:william.kengne@u-cergy.fr)*

**Abstract:** We consider here together the inference questions and the change-point problem in a large class of Poisson autoregressive models (see Tjøstheim, 2012 [34]). The conditional mean (or intensity) of the process is involved as a non-linear function of its past values and the past observations. Under Lipschitz-type conditions, it can be written as a function of lagged observations. For the latter model, assume that the link function depends on an unknown parameter  $\theta_0$ . The consistency and the asymptotic normality of the maximum likelihood estimator of the parameter are proved. These results are used to study change-point problem in the parameter  $\theta_0$ . From the likelihood of the observations, two tests are proposed. Under the null hypothesis (i.e. no change), each of these tests statistics converges to an explicit distribution. Consistencies under alternatives are proved for both tests. Simulation results show how those procedures work in practice, and applications to real data are also processed.

**MSC 2010 subject classifications:** Primary 60G10, 62M10, 62F12; secondary 62F03, 62F05, 62F10.

**Keywords and phrases:** Time series of counts, Poisson autoregression, likelihood estimation, change-point, semi-parametric test.

Received February 2015.

## Contents

1	Introduction . . . . .	1268
2	Assumptions and examples . . . . .	1271
2.1	Assumptions . . . . .	1271
2.2	Examples . . . . .	1272

---

\*This work has been developed within the MME-DII center of excellence (ANR-11-LABEX-0023-01) <http://labex-mme-dii.u-cergy.fr/>.

2.2.1	Linear Poisson autoregression . . . . .	1272
2.2.2	Threshold Poisson autoregression . . . . .	1273
3	Likelihood inference . . . . .	1274
4	Testing for parameter changes . . . . .	1275
4.1	Asymptotic behavior under the null hypothesis . . . . .	1278
4.2	Asymptotic under the alternative . . . . .	1279
5	Some numerical results for inference in INTGARCH model . . . . .	1280
5.1	Estimation and identification . . . . .	1280
5.2	Some simulations results . . . . .	1282
5.3	Application to real data . . . . .	1283
6	Some numerical results for parameter change in INGARCH model . . . . .	1286
6.1	Testing for parameter change in INGARCH model . . . . .	1286
6.2	Real data application . . . . .	1287
7	Proofs of the main results . . . . .	1290
	Acknowledgements . . . . .	1312
	References . . . . .	1312

## 1. Introduction

Time series of counts appear as natural for modeling count events. Some examples can be found in epidemiology (number of new infections), in finance (number of transactions per minute), in industrial quality control (number of defects), just to name a few. We refer the reader to Held *et al.* (2005) [23], Brännäs and Quoreshi (2010) [5], Lambert (1992) [30] among others, for more details.

Real advances have been made in count time series modeling during the last two decades. Let  $Y = (Y_t)_{t \in \mathbb{Z}}$  be an integer-valued time series; denote  $\mathcal{F}_t = \sigma(Y_s, s \leq t)$  the  $\sigma$ -field generated by the whole past at time  $t$  and  $\mathcal{L}(Y_t/\mathcal{F}_{t-1})$  the conditional distribution of  $Y_t$  given the past. A model is characterized by the type of marginal distribution  $\mathcal{L}(Y_t/\mathcal{F}_{t-1})$ , and the dependence structure between  $\mathcal{L}(Y_t/\mathcal{F}_{t-1})$  and the past. Models with various marginal distributions and dependence structures have been studied; see for instance Kedem and Fokianos (2002) [26], Davis *et al.* (2005) [8], Ferland *et al.* (2006) [15], Davis and Wu (2009) [10], Weiß (2009) [35].

Fokianos *et al.* (2009) [19] considered the Poisson autoregression such that,  $\mathcal{L}(Y_t/\mathcal{F}_{t-1})$  is Poisson distributed with an intensity  $\lambda_t$  which is a function of  $\lambda_{t-1}$  and  $Y_{t-1}$ . Under linear autoregression, they proved both the consistency and the asymptotic normality of the maximum likelihood estimator of the regression parameter, by using a perturbation approach, which allows to use the standard Markov ergodic setting. Fokianos and Tjøstheim (2012) [20] extended the method to nonlinear Poisson autoregression with;  $\lambda_t = f(\lambda_{t-1}) + b(Y_{t-1})$  for nonlinear measurable functions  $f$  and  $b$ . In the same vein, Neumann (2011) [31] studied a much more general model where  $\lambda_t = f(\lambda_{t-1}, Y_{t-1})$ . He focused on the absolute regularity and the specification test for the intensity function,

while the recent works of Fokianos and Neumann (2013) [18] studied goodness-of-fit tests which are able to detect local alternatives. Doukhan *et al.* (2012) [13], considered a more general model with infinitely many lags. Stationarity and the existence of moments are proved by using a weak dependent approach and contraction properties.

Later, Davis and Liu (2012) [9] studied the model where, the distribution  $\mathcal{L}(Y_t/\mathcal{F}_{t-1})$  belongs to a class of one-parameter exponential family with finite order dependence. This class contains Poisson and negative binomial (with fixed number of failures) distribution. From the theory of iterated random functions, they established the stationarity and the absolute regularity properties of the process. They also proved the consistency and asymptotic normality of the maximum likelihood estimator of the parameter of the model.

Douc *et al.* (2013) [11] considered a class of observation-driven time series which covers linear, log-linear, and threshold Poisson autoregressions. Their approach is based on a recent theory for Markov chains based upon Hairer and Mattingly (2006) [22] recent work; this allows existence and uniqueness of the invariant distribution for Markov chains without irreducibility. Further, they proved the consistency of the conditional likelihood estimator of the model (even for mis-specified models); the asymptotic normality is not yet considered in this setting.

Asymptotic theory for inference on time series models usually needs the stationarity properties of the process. But in practice, real data often suffer from non-stationarity which may be due to structural changes occur during the data generating period. Several ways to consider such structural changes are possible, as this was demonstrated during the thematic cycle *Nonstationarity and Risk Management* held in Cergy-Pontoise during year 2012<sup>1</sup>. In the context of count models, Kang and Lee (2009) [25] proposed a CUSUM procedure for testing of parameter changes in a first-order random coefficient integer-valued autoregressive model defined through thinning operator. Fokianos and Fried (2010, 2012) [16, 17] studied mean shift in linear and log-linear Poisson autoregression. Dependence between the level shift and time allows their model to detect several types of interventions effects such as outliers. Franke *et al.* (2012) [21], considered parameter change in Poisson autoregression of order one. Their tests are based on the cumulative sum of residuals using conditional least-squares estimator.

Here, we shall first consider a time series of counts  $Y = (Y_t)_{t \in \mathbb{Z}}$  satisfying:

$$Y_t/\mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t) \text{ with } \lambda_t = F((\lambda_{t-1}, Y_{t-1}), (\lambda_{t-2}, Y_{t-2}) \dots) \quad (1)$$

where  $\mathcal{F}_t = \sigma(Y_s, s \leq t)$  and  $F$  a measurable non-negative function. The properties of the general class of Poisson autoregressive model (1) have been investigated in Doukhan *et al.* [13]. Such infinite order processes provided a large way to take into account dependence on the past observations. Proceeding as

<sup>1</sup><http://www.u-cergy.fr/en/advanced-studies-institute/thematic-cycles/thematic-cycle-2012/finance-cycle.html>

in Doukhan and Wintenberger (2008) [12], we show that under some Lipschitz-type conditions on  $F$ , the conditional mean  $\lambda_t$  can be written as a function of the past observations.

This yields to consider the model

$$Y_t/\mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t) \quad \text{with} \quad \lambda_t = f(Y_{t-1}, Y_{t-2} \dots) \quad (2)$$

where  $f$  is a measurable non-negative function. We assume that  $f$  is known up to a parameter  $\theta_0$  belonging to some compact set  $\Theta \subset \mathbb{R}^d$  with  $d \in \mathbb{N} - \{0\}$ . That is

$$Y_t/\mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t) \quad \text{with} \quad \lambda_t = f_{\theta_0}(Y_{t-1}, Y_{t-2} \dots) \quad \text{and} \quad \theta_0 \in \Theta. \quad (3)$$

Many classical integer-valued time series satisfying (3) (see examples below).

Remark that, model (3) (as well as models (1) and (2)) can be represented in terms of Poisson processes. Let  $\{N_t(\cdot) ; t = 1, 2, \dots\}$  be a sequence of independent Poisson processes of unit intensity.  $Y_t$  can be seen as the number (say  $N_t(\lambda_t)$ ) of events of  $N_t(\cdot)$  in the time interval  $[0, \lambda_t]$ . So, we have the representation

$$Y_t = N_t(\lambda_t) \quad \text{with} \quad \lambda_t = f_{\theta_0}(Y_{t-1}, Y_{t-2} \dots). \quad (4)$$

The Poisson autoregressive models are known to capture the overdispersion phenomenon in counts data, meaning if the process  $(Y_t)_{t \in \mathbb{Z}}$  is stationary it always occurs that  $\text{Var}(Y_t) \geq E(Y_t)$ .

The paper first works out the asymptotic properties of the maximum likelihood estimator of the model (3). Under some Lipschitz-type assumption on the function  $f_\theta$ , we investigate sufficient conditions for the consistency and the asymptotic normality of the maximum likelihood estimator of  $\theta_0$ . Note that, Doukhan *et al.* [13] have used weak dependence approach to prove the existence of stationary and ergodic solution of (1). By using their results, some assumptions (such as increasing condition on  $f_\theta(\cdot)$  or four times differentiability on the function  $\theta \mapsto f_\theta$ ) that often needed for perturbation technique (see for instance Fokianos *et al.* [19] or Fokianos *et al.* [20]) are relaxed. Although the models studied by Davis and Liu [9] and Douc *et al.* [11] allow large classes of marginal distributions, the infinitely many lags of model (1) (or model (3)) enables a higher order dependence structure.

The second contribution of this work is the two tests for change detection in model (3). We propose a new idea to take into account the change-point alternative. This implies that, the procedures proposed will be numerically easy applied than those proposed by Kengne (2012) [27]. The consistency under the alternative is proved. Contrary to Franke *et al.* [21], the multiple change alternative has been considered and independence between the observations before and after the change-point is not assumed. Note that, the intervention problem studied by Fokianos and Fried [16, 17] is intended to sudden shift in the conditional mean of the process. Such outlier could in some case be seen as a particular case of structural change problem that we develop here for a large class of models. However, if the intervention affects only a few data points, in

the classical change-point setting (where the length of each regime tends to infinity with the same rate as the sample size), such effects will be asymptotically negligible.

The forthcoming Section 2 provides some assumptions on model with examples. Section 3 is devoted to the definition of the maximum likelihood estimator with its asymptotic properties. In Section 4, we propose the tests for detecting change in parameter of model (3). Some simulation results and real data applications for inference and change-point detection are presented in Sections 5 and 6, lastly the proofs of the main results are provided in Section 7.

## 2. Assumptions and examples

### 2.1. Assumptions

We will use the following classical notations:

1.  $\|y\| := \sum_{j=1}^p |y_j|$  for any  $y \in \mathbb{R}^p$ ;
2. for any compact set  $\mathcal{K} \subseteq \mathbb{R}^d$  and for any function  $g : \mathcal{K} \rightarrow \mathbb{R}^{d'}$ ,  $\|g\|_{\mathcal{K}} = \sup_{\theta \in \mathcal{K}} (\|g(\theta)\|)$ ;
3. if  $Y$  is a random variable with finite moment of order  $r > 0$ , then  $\|Y\|_r = (E|Y|^r)^{1/r}$ ;
4. for any set  $\mathcal{K} \subseteq \mathbb{R}^d$ ,  $\overset{\circ}{\mathcal{K}}$  denotes the interior of  $\mathcal{K}$ .

A classical Lipschitz-type conditions is assumed on the model (1).

**Assumption  $A_F$ .** There exists a sequence of non-negative real numbers  $(\alpha_j)_{j \geq 1}$  satisfying

$$\sum_{j=1}^{\infty} \alpha_j < 1/(1 + \epsilon)$$

for some  $\epsilon > 0$  and such that for any  $x, x' \in ((0, \infty) \times \mathbb{N})^{\mathbb{N}}$ ,

$$|F(x) - F(x')| \leq \sum_{j=1}^{\infty} \alpha_j \|x_j - x'_j\|.$$

Under assumption  $(A_F)$ , Doukhan *et al.* [13, 14] prove that the above equation has a strictly stationary solution  $(Y_t, \lambda_t)_{t \in \mathbb{Z}}$ ; moreover this solution is  $\tau$ -weakly dependent with finite moment of any order. The following proposition show that the conditional mean  $\lambda_t$  of model (1) can be expressed as a function of only the past observations of the process.

**Proposition 2.1.** *Under  $(A_F)$ , the conditional mean of the strictly stationary and ergodic solution of (1) can be written as*

$$\lambda_t = f(Y_{t-1}, Y_{t-2} \dots)$$

where  $f : \mathbb{N} \rightarrow \mathbb{R}^+$  is a measurable function.

From Proposition 2.1, it appears that the information on the unobservable process  $(\lambda_{t-j})$  can be captured by the observable process  $(Y_{t-j})$ . Hence, we will

focus on the model (3); with the advantage to easily compute the derivative  $\partial\lambda_t/\partial\theta$  (very useful to derive the asymptotic covariance matrix in the inference study). Note that, if one carries inference on the model (1) by assuming that  $\lambda_t = F_{\theta_0}((\lambda_{t-1}, Y_{t-1}), (\lambda_{t-2}, Y_{t-2}), \dots)$ , then it will not be easy (or not possible) to compute  $\partial\lambda_t/\partial\theta$  or to express it as a function of  $\partial F_{\theta}/\partial\theta$  in the general case.

We focus on the model (3) with the following assumptions. For  $i = 0, 1, 2$  and for any compact set  $\mathcal{K} \subset \Theta$ , we introduce

**Assumption  $A_i(\mathcal{K})$ .**  $\|\partial^i f_{\theta}(0)/\partial\theta^i\|_{\Theta} < \infty$  and there exists a sequence of non-negative real numbers  $(\alpha_k^{(i)}(\mathcal{K}))_{k \geq 1}$  satisfying  $\sum_{j=1}^{\infty} \alpha_k^{(0)}(\mathcal{K}) < 1$  or  $\sum_{j=1}^{\infty} \alpha_k^{(i)}(\mathcal{K}) < \infty$  (for  $i = 1, 2$ ) such that

$$\left\| \frac{\partial^i f_{\theta}(y)}{\partial\theta^i} - \frac{\partial^i f_{\theta}(y')}{\partial\theta^i} \right\|_{\mathcal{K}} \leq \sum_{k=1}^{\infty} \alpha_k^{(i)}(\mathcal{K}) |y_k - y'_k| \quad \text{for all } y, y' \in (\mathbb{R}^+)^{\mathbb{N}}.$$

The Lipschitz-type condition  $A_0(\Theta)$  is the parametric version of the assumption  $A_F$ . It is classical when studying the existence of solutions of such model (see for instance [12, 1] or [13]). In particular,  $A_0(\Theta)$  implies that for all  $\theta \in \Theta$  and  $y \in (\mathbb{R}^+)^{\mathbb{N}}$ ,

$$f_{\theta}(y) \leq f_{\theta}(0) + \sum_{k=1}^{\infty} \alpha_k^{(i)}(\Theta) y_k.$$

The latter relation is a useful tool for proving that the stationary solution of (3) admits finite moments (see the proof of Theorem 2.1 of [13]). The assumptions  $A_1(\mathcal{K})$  and  $A_2(\mathcal{K})$  as well as the following assumptions  $D(\Theta)$ ,  $\text{Id}(\Theta)$  and  $\text{Var}(\Theta)$  are needed to define and to study the asymptotic properties of the maximum likelihood estimator of the model (3).

**Assumption  $D(\Theta)$ .**  $\exists c > 0$  such that  $\inf_{\theta \in \Theta} f_{\theta}(y) \geq c$  for all  $y \in (\mathbb{R}^+)^{\mathbb{N}}$ .

**Assumption  $\text{Id}(\Theta)$ .** For all  $(\theta, \theta') \in \Theta^2$ ,

$$\left( f_{\theta}(Y_{t-1}, \dots) = f_{\theta'}(Y_{t-1}, \dots) \text{ a.s. for some } t \in \mathbb{Z} \right) \Rightarrow \theta = \theta'.$$

**Assumption  $\text{Var}(\Theta)$ .** For all  $\theta \in \Theta$  and  $t \in \mathbb{Z}$ , the components of the vector  $\frac{\partial f_{\theta}}{\partial\theta^i}(Y_{t-1}, \dots)$  are a.s. linearly independent.

## 2.2. Examples

### 2.2.1. Linear Poisson autoregression

We consider an integer-valued time series  $(Y_t)_{t \in \mathbb{Z}}$  satisfying for any  $t \in \mathbb{Z}$

$$Y_t / \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t) \quad \text{with} \quad \lambda_t = \phi_0(\theta_0) + \sum_{k \geq 1} \phi_k(\theta_0) Y_{t-k} \quad (5)$$

where  $\theta_0 \in \Theta \subset \mathbb{R}^d$ , the functions  $\theta \mapsto \phi_k(\theta)$  are positive and satisfying  $\sum_{k \geq 1} \|\phi_k(\theta)\|_{\Theta} < 1$ . This model is also called an INARCH( $\infty$ ), due to its

similarity with the classical ARCH( $\infty$ ) model. Assumptions  $A_0(\Theta)$  holds automatically. If the function  $\phi_k$  are twice continuously differentiable such that  $\sum_{k \geq 1} \|\phi'_k(\theta)\|_{\Theta} < \infty$  and  $\sum_{k \geq 1} \|\phi''_k(\theta)\|_{\Theta} < \infty$  then  $A_1(\Theta)$  and  $A_2(\Theta)$  hold. If  $\inf_{\theta \in \Theta} \phi_0(\theta) > 0$  then  $D(\Theta)$  holds. Moreover, if there exists a finite subset  $I \subset \mathbb{N} - \{0\}$  such that the function  $\theta \mapsto (\phi_k(\theta))_{k \in I}$  is injective, then assumption  $\text{Id}(\Theta)$  holds i.e. model (5) is identifiable. Finally, assumption  $\text{Var}(\Theta)$  holds if for any  $\theta \in \Theta$  there exists  $d$  functions  $\phi_{k_1}, \dots, \phi_{k_d}$  such that the matrix  $(\frac{\partial \phi_{k_i}}{\partial \theta^j})_{1 \leq i, j \leq d}$  (computed at  $\theta$ ) has a full rank.

**Particular case of INGARCH( $p, q$ ) processes** Assume that

$$Y_t / \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t) \quad \text{with} \quad \lambda_t = \alpha_0^* + \sum_{k=1}^p \alpha_k^* \lambda_{t-k} + \sum_{k=1}^q \beta_k^* Y_{t-k} \quad (6)$$

where  $\theta_0 = (\alpha_0^*, \alpha_1^*, \dots, \alpha_p^*, \beta_1^*, \dots, \beta_q^*) \in \Theta$  with

$$\Theta = \left\{ \theta = (\alpha_0, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q) \in ]0, \infty[ \times ]0, \infty[^{p+q}, \sum_{k=1}^p \alpha_k + \sum_{k=1}^q \beta_k < 1 \right\}.$$

Hence, the Lipschitz-type condition ( $A_F$ ) is satisfied. In this case, we can find for any  $\theta \in \Theta$ , a sequence of non-negative real numbers  $(\psi_k(\theta))_{k \geq 0}$  satisfying  $\sum_{k \geq 1} \|\psi_k(\theta)\|_{\Theta} < 1$  such that

$$\lambda_t = \psi_0(\theta_0) + \sum_{k \geq 1} \psi_k(\theta_0) Y_{t-k}.$$

Therefore, assumptions  $A_0(\Theta)$  holds. Moreover, the functions  $\psi_k(\theta)$  are twice continuously differentiable with respect to  $\theta$  and its derivatives decay exponentially, hence  $A_1(\Theta)$  and  $A_2(\Theta)$  hold. If  $\inf_{\theta \in \Theta} (\alpha_0) > 0$  then  $D(\Theta)$  holds. For this particular case,  $\text{Id}(\Theta)$  holds automatically. See [15] and [35] for more details on this model.

The adequacy of this linear model to the number of transactions per minute for the stock Ericsson B during July 2, 2002 has been proved by Fokianos and Neumann [18].

### 2.2.2. Threshold Poisson autoregression

We consider a threshold Poisson autoregression model defined by:

$$Y_t / \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t) \quad \text{with} \quad \lambda_t = \phi_0(\theta_0) + \sum_{k \geq 1} \left( \phi_k^+(\theta_0) \max(Y_{t-k} - \ell, 0) + \phi_k^-(\theta_0) \min(Y_{t-k}, \ell) \right) \quad (7)$$

where  $\phi_0(\theta_0) > 0$ ,  $\phi_k^+(\theta_0), \phi_k^-(\theta_0) \geq 0$  and  $\ell \in \mathbb{N}$ . We can also write

$$\lambda_t = \phi_0(\theta_0) + \sum_{k \geq 1} \left( \phi_k^-(\theta_0) Y_{t-k} + (\phi_k^+(\theta_0) - \phi_k^-(\theta_0)) \max(Y_{t-k} - \ell, 0) \right).$$

This example of nonlinear model is also called an integer-valued threshold ARCH (or INTARCH) due to its definition like the threshold ARCH model proposed by Zakoïan (1994) [36]; see also [21] for INTARCH(1) model. In the INTARCH( $\infty$ ) model, the regression coefficient at lag  $t - k$  is  $\phi_k^-(\theta_0)$  if  $Y_{t-k} \leq \ell$  and  $\phi_k^+(\theta_0)$  if  $Y_{t-k} > \ell$ ; such model can then be used to capture a piecewise phenomenon.  $\ell$  is the threshold parameter of the model. If  $\sum_{k \geq 1} \max(\|\phi_k^+(\theta)\|_{\Theta}, \|\phi_k^-(\theta)\|_{\Theta}) < 1$  then  $A_0(\Theta)$  holds. Furthermore, if the functions  $\theta \mapsto \phi_k^+(\theta)$  and  $\theta \mapsto \phi_k^-(\theta)$  are twice continuously differentiable such that  $\sum_{k \geq 1} \max(\|\frac{\partial}{\partial \theta} \phi_k^+(\theta)\|_{\Theta}, \|\frac{\partial}{\partial \theta} \phi_k^-(\theta)\|_{\Theta}) < \infty$  and  $\sum_{k \geq 1} \max(\|\frac{\partial^2}{\partial \theta^2} \phi_k^+(\theta)\|_{\Theta}, \|\frac{\partial^2}{\partial \theta^2} \phi_k^-(\theta)\|_{\Theta}) < \infty$ , then  $A_1(\Theta)$  and  $A_2(\Theta)$  hold. Conditions on  $D(\Theta)$ ,  $\text{Id}(\Theta)$  and  $\text{Var}(\Theta)$  are obtained as above. A special case is the INTGARCH( $p, q$ ) model, see Section 5.

**Remark 2.1.** An association argument can be used to show that the INGARCH and INTGARCH models capture only positive dependence in time series count data. It is a frequent phenomenon in transactions data (see for instance the subsection 5.3 for the number of transactions for the stock Ericsson B).

### 3. Likelihood inference

Assume that the trajectory  $(Y_1, \dots, Y_n)$  is observed. The conditional (log)-likelihood of model (3) computed on  $T \subset \{1, \dots, n\}$ , is given (up to a constant) by

$$L_n(T, \theta) = \sum_{t \in T} (Y_t \log \lambda_t(\theta) - \lambda_t(\theta)) = \sum_{t \in T} \ell_t(\theta) \quad \text{with} \quad \ell_t(\theta) = Y_t \log \lambda_t(\theta) - \lambda_t(\theta)$$

where  $\lambda_t(\theta) = f_{\theta}(Y_{t-1}, \dots)$ . In the sequel, we use the notation  $f_{\theta}^t := f_{\theta}(Y_{t-1}, \dots)$ . Since only  $Y_1, \dots, Y_n$  are observed, the (log)-likelihood is approximated by

$$\widehat{L}_n(T, \theta) = \sum_{t \in T} (Y_t \log \widehat{\lambda}_t(\theta) - \widehat{\lambda}_t(\theta)) = \sum_{t \in T} \widehat{\ell}_t(\theta) \quad \text{with} \quad \widehat{\ell}_t(\theta) = Y_t \log \widehat{\lambda}_t(\theta) - \widehat{\lambda}_t(\theta) \quad (8)$$

where  $\widehat{\lambda}_t(\theta) := \widehat{f}_{\theta}^t := f_{\theta}(Y_{t-1}, \dots, Y_1, 0, \dots)$  and  $\widehat{\lambda}_1(\theta) = f_{\theta}(0, \dots)$ . The maximum likelihood estimator (MLE) of  $\theta_0$  computed on  $T$  is defined by

$$\widehat{\theta}_n(T) = \operatorname{argmax}_{\theta \in \Theta} (\widehat{L}_n(T, \theta)). \quad (9)$$

For any  $k, k' \in \mathbb{Z}$  such as  $k \leq k'$ , denote

$$T_{k, k'} = \{k, k+1, \dots, k'\}.$$

**Theorem 3.1.** Let  $(j_n)_{n \geq 1}$  and  $(k_n)_{n \geq 1}$  be two integer valued sequences such that  $j_n \leq k_n$ ,  $k_n \rightarrow +\infty$  and  $k_n - j_n \rightarrow +\infty$  as  $n \rightarrow +\infty$ . Assume  $\theta_0 \in \overset{\circ}{\Theta}$  and  $D(\Theta)$ ,  $\text{Id}(\Theta)$  and  $A_0(\Theta)$  hold with

$$\sum_{j \geq 1} \sqrt{j} \times \alpha_j^{(0)}(\Theta) < \infty. \quad (10)$$

It holds that

$$\widehat{\theta}_n(T_{j_n, k_n}) \xrightarrow[n \rightarrow +\infty]{a.s.} \theta_0.$$

The following theorem regarding the asymptotic normality of the MLE of model (3) holds.

**Theorem 3.2.** *Let  $(j_n)_{n \geq 1}$  and  $(k_n)_{n \geq 1}$  be two integer valued sequences such that  $j_n \leq k_n$ ,  $k_n \rightarrow +\infty$  and  $k_n - j_n \rightarrow +\infty$  as  $n \rightarrow +\infty$ . Under the assumptions of Theorem 3.1 and  $\text{Var}(\Theta)$ , if for  $i = 1, 2$   $A_i(\Theta)$  hold with*

$$\sum_{j \geq 1} \sqrt{j} \times \alpha_j^{(i)}(\Theta) < \infty \tag{11}$$

then

$$\sqrt{k_n - j_n}(\widehat{\theta}_n(T_{j_n, k_n}) - \theta_0) \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N}(0, \Sigma^{-1})$$

where  $\Sigma = E\left(\frac{1}{f_{\theta_0}^0} \left(\frac{\partial}{\partial \theta} f_{\theta_0}^0\right) \left(\frac{\partial}{\partial \theta} f_{\theta_0}^0\right)'\right)$ .

According to the Lemma 7.2 and the proof of Theorem 3.2, the matrix

$$\left(\frac{1}{k_n - j_n} \sum_{t=j_n}^{k_n} \frac{1}{\widehat{f}_{\theta}^t} \left(\frac{\partial}{\partial \theta} \widehat{f}_{\theta}^t\right) \left(\frac{\partial}{\partial \theta} \widehat{f}_{\theta}^t\right)'\right) \Big|_{\theta = \widehat{\theta}_n(T_{j_n, k_n})} \text{ and}$$

$$\left(-\frac{1}{k_n - j_n} \frac{\partial^2}{\partial \theta \partial \theta'} \widehat{L}_n(T_{j_n, k_n}, \theta)\right) \Big|_{\theta = \widehat{\theta}_n(T_{j_n, k_n})}$$

are both consistent estimators of  $\Sigma$ .

**Remark 3.1.** 1. Theorem 3.1 still holds if we replace (10) by the much weaker condition

$$\sum_{j \geq 1} \log j \times \alpha_j^{(0)}(\Theta) < \infty$$

(see the proof of the theorem).

2. In Theorems 3.1 and 3.2, the typical sequences  $j_n = 1$  and  $k_n = n$ ,  $\forall n \geq 1$  can be chosen. This choice is the case where the estimator is computed with all the observations. But in the change-point study and depending on the procedure used, one might need to compute the estimator on each regime. Results are written this way to cover this situation.

3. If the Lipschitz coefficients  $(\alpha_j^{(i)}(\Theta))_{j \geq 1}$  satisfy  $\alpha_j^{(i)}(\Theta) = \mathcal{O}(j^{-\gamma})$  with  $\gamma > 3/2$ , then conditions (10) and (11) hold.

#### 4. Testing for parameter changes

We consider the observations  $Y_1, \dots, Y_n$  generated as in model (3) and assume that the parameter  $\theta_0$  may change over time. More precisely, we assume that  $\exists K \geq 1, \theta_1^*, \dots, \theta_K^* \in \Theta, 0 = t_0^* < t_1^* < \dots < t_{K-1}^* < t_K^* = n$  such that  $Y_t = Y_t^{(j)}$  for  $t_{j-1}^* < t \leq t_j^*$ , where the process  $(Y_j^{(j)})_{t \in \mathbb{Z}}$  is a stationary solution of (3) depending on  $\theta_j^*$ . The case where the parameter does not change corresponds to  $K = 1$ . This problem leads to the following test hypotheses:

- H<sub>0</sub>: Observations  $(Y_1, \dots, Y_n)$  are a trajectory of a process  $(Y_t)_{t \in \mathbb{Z}}$  solution of (3), depending on  $\theta_0 \in \Theta$ .
- H<sub>1</sub>: There exists  $K \geq 2, \theta_1^*, \theta_2^*, \dots, \theta_K^*$  with  $\theta_1^* \neq \theta_2^* \neq \dots \neq \theta_K^*, 0 = t_0^* < t_1^* < \dots < t_{K-1}^* < t_K^* = n$  such that the observations  $(Y_t)_{t_{j-1}^* < t \leq t_j^*}$  are a trajectory of the process  $(Y_t^{(j)})_{t \in \mathbb{Z}}$  solution of (3), depending on  $\theta_j^*$ .

Let us note that, contrary to Franke *et al.* [21], the independence between the observations before and after the change-point is not assumed. Moreover, their assumption (A9) impose a change in the mean of the marginal distribution of the observed process. In the case of linear Poisson autoregression, this condition leads to a change in the unconditional mean. The procedures developed here include the situation where parameter can change but not the mean of the marginal distribution (see an example of empirical results in the Section 6). Therefore, the present change-point problem is more general.

**Remark 4.1.** Let  $\lambda_t^{(j)}$  be the conditional mean on the segment  $T_j^* = \{t_{j-1}^*, t_{j-1}^* + 1, \dots, t_j^*\}$ . By using the representation (4), we can write

$$Y_t^{(j)} = N_t(\lambda_t^{(j)}) \quad \text{with} \quad \lambda_t^{(j)} = f_{\theta_j^*}(Y_{t-1}^{(j)}, \dots).$$

We can also write

$$Y_t^{(j)} = \tilde{F}(Y_{t-1}^{(j)}, \dots; N_t) \quad \text{where}$$

$\tilde{F}(y_1, y_2, \dots; N_t) = N_t(f_{\theta_j^*}(y_1, y_2, \dots))$  for any  $y_k \in \mathbb{N}, k \geq 1$ . Hence for any  $y = (y_k)_{k \geq 1}$  and  $y' = (y'_k)_{k \geq 1}$ :

$$\begin{aligned} E|\tilde{F}(y; N_t) - \tilde{F}(y'; N_t)| &= E|N_t(f_{\theta_j^*}(y)) - N_t(f_{\theta_j^*}(y'))| = |f_{\theta_j^*}(y) - f_{\theta_j^*}(y')| \\ &\leq \sum_{k \geq 1} \alpha_k^{(0)} |y_k - y'_k|, \end{aligned} \quad (12)$$

where the second equality follows by seeing  $|N_t(f_{\theta_j^*}(y)) - N_t(f_{\theta_j^*}(y'))|$  as a number of events  $N_t$  that occur in the time interval  $[0, |f_{\theta_j^*}(y) - f_{\theta_j^*}(y')|]$ .

Let  $(\tilde{Y}_t, \tilde{\lambda}_t)_{t_{j-1}^* < t \leq t_j^*}$  be the nonstationary approximation of the process  $(Y_t^{(j)}, \lambda_t^{(j)})_{t \in \mathbb{Z}}$  on the segment  $T_j^*$ ; i.e.

$$\tilde{Y}_t = N_t(\tilde{\lambda}_t) \quad \text{with} \quad \tilde{\lambda}_t = f_{\theta_j^*}(Y_{t-1}^{(j)}, \dots, Y_{t-t_{j-1}^*}^{(j)}, Y_{t-t_{j-1}^*-1}^{(j-1)}, \dots, Y_1^{(1)}, 0, \dots).$$

By using assumption  $A_0(\Theta)$  and relation (12), one can show that the approximated process  $(\tilde{Y}_t, \tilde{\lambda}_t)_{t_{j-1}^* < t \leq t_j^*}$  converges (in  $L^r$  for any  $r \geq 1$ ) to the stationary regime (see for instance Bardet *et al.* [2] where similar approximation has been made for a class of causal time series models). So, the results of Section 4.2 can be extended (after an approximation study) by relaxing the stationarity assumption after change.

Recall that under  $H_0$ , the likelihood function of the model computed on  $T \subset \{1, \dots, n\}$  is given by

$$\hat{L}_n(T, \theta) = \sum_{t \in T} (Y_t \log \hat{f}_\theta^t - \hat{f}_\theta^t)$$

where  $\widehat{f}_\theta^t = f_\theta(Y_{t,1}, \dots)$  and the maximum likelihood estimator is given by  $\widehat{\theta}_n(T) = \operatorname{argmax}_{\theta \in \Theta} \widehat{L}_n(T, \theta)$ . It holds from Theorem 3.2 that, under  $H_0$ , the asymptotic covariance matrix of  $\widehat{\theta}_n(T_{1,n})$  is  $\widehat{\Sigma}_n^{-1}$  where

$$\widehat{\Sigma}_n = \left( \frac{1}{n} \sum_{t=1}^n \frac{1}{\widehat{f}_\theta^t} \left( \frac{\partial}{\partial \theta} \widehat{f}_\theta^t \right) \left( \frac{\partial}{\partial \theta} \widehat{f}_\theta^t \right)' \right) \Big|_{\theta = \widehat{\theta}_n(T_{1,n})}.$$

$\widehat{\Sigma}_n$  is a consistent estimator of

$$\Sigma = E \left( \frac{1}{f_{\theta_0}^0} \left( \frac{\partial}{\partial \theta} f_{\theta_0}^0 \right) \left( \frac{\partial}{\partial \theta} f_{\theta_0}^0 \right)' \right)$$

under  $H_0$  (see the proof of Theorem 3.2). The consistency of  $\widehat{\Sigma}_n$  under  $H_1$  is not ensured.  $\widehat{\Sigma}_n$  does not take into account the change-point alternative. So, the consistency under  $H_1$  of any test based on  $\widehat{\Sigma}_n$  will not be easy to prove.

Let  $(u_n)_{n \geq 1}$  and  $(v_n)_{n \geq 1}$  be two integer valued sequences satisfying  $u_n, v_n \rightarrow +\infty$ ,  $\frac{u_n}{n}, \frac{v_n}{n} \rightarrow 0$  as  $n \rightarrow +\infty$ . Our test statistic is based on the following matrix

$$\begin{aligned} \widehat{\Sigma}_n(u_n) = & \frac{1}{2} \left[ \left( \frac{1}{u_n} \sum_{t=1}^{u_n} \frac{1}{\widehat{f}_\theta^t} \left( \frac{\partial}{\partial \theta} \widehat{f}_\theta^t \right) \left( \frac{\partial}{\partial \theta} \widehat{f}_\theta^t \right)' \right) \Big|_{\theta = \widehat{\theta}_n(T_{1,u_n})} \right. \\ & \left. + \left( \frac{1}{n - u_n} \sum_{t=u_n+1}^n \frac{1}{\widehat{f}_\theta^t} \left( \frac{\partial}{\partial \theta} \widehat{f}_\theta^t \right) \left( \frac{\partial}{\partial \theta} \widehat{f}_\theta^t \right)' \right) \Big|_{\theta = \widehat{\theta}_n(T_{u_n+1,n})} \right]. \end{aligned}$$

Theorem 3.1 and Lemma 7.2 show that  $\widehat{\Sigma}_n(u_n)$  is consistent under  $H_0$ . Under  $H_1$ , we will use the classical assumption that the breakpoint gown at the rate  $n$ . This will allow us to show that the first component of  $\widehat{\Sigma}_n(u_n)$  converges to the covariance matrix of the stationary model of the first regime. It will be a key to prove the consistency under  $H_1$ .

Another way to deal is to consider the matrix

$$\begin{aligned} \widetilde{\Sigma}_n(u_n) = & \frac{1}{2} \left[ \left( \frac{1}{n - u_n} \sum_{t=1}^{n-u_n} \frac{1}{\widehat{f}_\theta^t} \left( \frac{\partial}{\partial \theta} \widehat{f}_\theta^t \right) \left( \frac{\partial}{\partial \theta} \widehat{f}_\theta^t \right)' \right) \Big|_{\theta = \widehat{\theta}_n(T_{1,n-u_n})} \right. \\ & \left. + \left( \frac{1}{u_n} \sum_{t=n-u_n+1}^n \frac{1}{\widehat{f}_\theta^t} \left( \frac{\partial}{\partial \theta} \widehat{f}_\theta^t \right) \left( \frac{\partial}{\partial \theta} \widehat{f}_\theta^t \right)' \right) \Big|_{\theta = \widehat{\theta}_n(T_{n-u_n+1,n})} \right]. \end{aligned}$$

Asymptotically, both the matrices  $\widehat{\Sigma}_n(u_n)$  and  $\widetilde{\Sigma}_n(u_n)$  have the same behavior under  $H_0$ . In the case of non stationarity after change, the procedure based on  $\widetilde{\Sigma}_n(u_n)$  can provide more distortion; because, according to the dependence on the past, the second component of  $\widetilde{\Sigma}_n(u_n)$  will converge very slowly than the first component of  $\widehat{\Sigma}_n(u_n)$ .

In the sequel, we will use a weight function  $q : (0, 1) \rightarrow (0, \infty)$  assuming to non-decreasing in a neighborhood of zero, non-increasing in a neighborhood of one and satisfying  $\inf_{\eta < \tau < 1-\eta} q(\tau) > 0$  for all  $0 < \eta < 1/2$ .

Let us define now the test statistics:

- $\widehat{C}_n = \max_{v_n \leq k \leq n-v_n} \widehat{C}_{n,k}$  where

$$\widehat{C}_{n,k} = \frac{1}{q^2(\frac{k}{n})} \frac{k^2(n-k)^2}{n^3} (\widehat{\theta}_n(T_{1,k}) - \widehat{\theta}_n(T_{k+1,n}))' \widehat{\Sigma}_n(u_n) (\widehat{\theta}_n(T_{1,k}) - \widehat{\theta}_n(T_{k+1,n}))$$

where  $q$  is a weight function define on  $(0, 1)$ , see bellow;

- $\widehat{Q}_n^{(1)} = \max_{v_n \leq k \leq n-v_n} \widehat{Q}_{n,k}^{(1)}$  where

$$\widehat{Q}_{n,k}^{(1)} = \frac{k^2}{n} (\widehat{\theta}_n(T_{1,k}) - \widehat{\theta}_n(T_{1,n}))' \widehat{\Sigma}_n(u_n) (\widehat{\theta}_n(T_{1,k}) - \widehat{\theta}_n(T_{1,n}));$$

- $\widehat{Q}_n^{(2)} = \max_{v_n \leq k \leq n-v_n} \widehat{Q}_{n,k}^{(2)}$  where

$$\widehat{Q}_{n,k}^{(2)} = \frac{(n-k)^2}{n} (\widehat{\theta}_n(T_{k+1,n}) - \widehat{\theta}_n(T_{1,n}))' \widehat{\Sigma}_n(u_n) (\widehat{\theta}_n(T_{k+1,n}) - \widehat{\theta}_n(T_{1,n})).$$

The first procedure is based on the statistic  $\widehat{C}_n$  and the other one is based on  $\widehat{Q}_n$  defined by

$$\widehat{Q}_n := \max(\widehat{Q}_n^{(1)}, \widehat{Q}_n^{(2)}).$$

The weight function  $q$  is used to increase the power of the test based on the statistic  $\widehat{C}_n$ . Its behavior can be controlled at the neighborhood of zero and one by the integral

$$I_{0,1}(q, c) = \int_0^1 \frac{1}{t(1-t)} \exp\left(-\frac{cq^2(t)}{t(1-t)}\right) dt, \quad c > 0$$

see Csörgo *et al.* [6] or Csörgo and Horváth [7]. The natural choice is  $q(t) = (t(1-t))^\gamma$  with  $0 \leq \gamma < 1/2$ .

Furthermore, in practice the sequences  $(u_n)_{n \geq 1}$  and  $(v_n)_{n \geq 1}$  are chosen to ensure the convergence of the numerical algorithm used to compute  $\widehat{\theta}_n(T_{1,u_n})$  and  $\widehat{\theta}_n(T_{1,v_n})$ . Hence, these procedures might not be accurate for smaller sample size (i.e. when  $n < 200$ ). For Poisson INGARCH model,  $u_n = v_n = [(\log n)^{\delta_0}]$  (with  $5/2 \leq \delta_0 \leq 3$ ) can be chosen (see also Remark 1 of [27]).

#### 4.1. Asymptotic behavior under the null hypothesis

The asymptotic distributions of these statistics under  $H_0$  are given in the next theorem.

**Theorem 4.1.** Assume  $D(\Theta)$ ,  $\text{Id}(\Theta)$ ,  $\text{Var}(\Theta)$  and  $A_i(\Theta)$   $i = 0, 1, 2$  hold with

$$\sum_{j \geq 1} \sqrt{j} \times \alpha_j^{(i)}(\Theta) < \infty.$$

Under  $H_0$  with  $\theta_0 \in \overset{\circ}{\Theta}$ ,

1. if  $I(q, c) < \infty$  for some  $c > 0$ , then

$$\widehat{C}_n \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \sup_{0 < \tau < 1} \frac{\|W_d(\tau)\|^2}{q^2(\tau)};$$

2. for  $j = 1, 2$ ,

$$\widehat{Q}_n^{(j)} \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \sup_{0 < \tau < 1} \|W_d(\tau)\|^2,$$

where  $W_d$  is a  $d$ -dimensional Brownian bridge.

The distribution of  $\sup_{0 \leq \tau \leq 1} \|W_d(\tau)\|^2$  is explicitly known. In the general case, the quantile values of the limit distribution of the first procedure (based on  $\widehat{C}_n$ ) can be computed through Monte-Carlo simulations. In the sequel, we will take  $q \equiv 1$ . The Theorem 4.2 below implies that the statistics  $\widehat{C}_n$  and  $\widehat{Q}_n$  are too large under the alternative. For any  $\alpha \in (0, 1)$ , denote  $c_\alpha$  the  $(1 - \alpha)$ -quantile of the distribution of  $\sup_{0 \leq \tau \leq 1} \|W_d(\tau)\|^2$ . Then at a nominal level  $\alpha \in (0, 1)$ , take  $(\widehat{C}_n > c_\alpha)$  as the critical region of the test procedure based on  $\widehat{C}_n$ . This test has correct size asymptotically. On the other hand, it holds that

$$\limsup_{n \rightarrow \infty} P(\widehat{Q}_n > c_{\alpha/2}) \leq \alpha.$$

So we can use  $c_{\alpha/2}$  as the critical value of the test based on  $\widehat{Q}_n$  i.e.  $(\widehat{Q}_n > c_{\alpha/2})$  as the critical region. This leads to an asymptotically conservative procedure. To get correct asymptotic size in the procedure based on  $\widehat{Q}_n$ , we have to study the asymptotic distribution of  $(\widehat{Q}_n^{(1)}, \widehat{Q}_n^{(2)})$ . This seems to be a very difficult problem in view of the dependence structure of the model and the general structure of the parameter. In the problem of discriminating between long-range dependence and changes in mean, Berkes *et al.* [3] have studied the limit distribution of such statistic (i.e. the maximum of the maximum between the statistic based on the estimator computed with the observations until the time  $k$  ( $X_1, \dots, X_k$ ) and the one computed with the observations after  $k$  ( $X_{k+1}, \dots, X_n$ )). This problem is the topic of a different research project.

#### 4.2. Asymptotic under the alternative

Under  $H_1$ , we assume

**Assumption B.** There exist  $\tau_1^*, \dots, \tau_{K-1}^*$  with  $0 < \tau_1^* < \dots < \tau_{K-1}^* < 1$  such that for  $j = 1, \dots, K$ ,  $t_j^* = [n\tau_j^*]$  (where  $[x]$  is the integer part of  $x$ ).

The asymptotic behaviors of these test statistics are given by the following theorem.

**Theorem 4.2.** Assume  $B$ ,  $D(\Theta)$ ,  $\text{Id}(\Theta)$ ,  $\text{Var}(\Theta)$  and  $A_i(\Theta)$   $i = 0, 1, 2$  hold with

$$\sum_{j \geq 1} \sqrt{j} \times \alpha_j^{(i)}(\Theta) < \infty.$$

1. Under  $H_1$  with  $K = 2$ , if  $\theta_1^* \neq \theta_2^*$  and  $\theta_1^*, \theta_2^* \in \overset{\circ}{\Theta}$  then

$$\widehat{C}_n \xrightarrow[n \rightarrow +\infty]{P} +\infty.$$

2. Under  $H_1$ , if  $\theta_1^* \neq \theta_K^*$  and  $\theta_1^*, \theta_2^*, \dots, \theta_K^* \in \overset{\circ}{\Theta}$  then

$$\widehat{Q}_n \xrightarrow[n \rightarrow +\infty]{P} +\infty.$$

It follows that the procedure based on  $\widehat{C}_n$  is consistent under a single change alternative while the statistic  $\widehat{Q}_n$  diverges to infinity even under multiple change-points alternative. So, combined with an iterated cumulative sums of squares type algorithm (see [24]), the latter procedure can be used to estimate the number and the break points in the multiple change-points problem.

The Figure 1 is an illustration of these tests for the linear Poisson autoregressive model of order 1

$$Y_t/\mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t) \quad \text{with} \quad \lambda_t = \alpha_0 + \beta_1 Y_{t-1}. \quad (13)$$

One can see that, under  $H_0$ , the statistics  $\widehat{C}_{n,k}$ ,  $\widehat{Q}_{n,k}^{(1)}$  and  $\widehat{Q}_{n,k}^{(2)}$  are all below the horizontal line (see c-), e-), g-)) which represents the limit of the critical region. These statistics are greater than the critical value in the neighborhood of the breakpoint under  $H_1$  (see d-), f-), h-)). In several situations, only one of the statistics  $\widehat{Q}_n^{(1)}$  and  $\widehat{Q}_n^{(2)}$  is greater than the critical value under the alternative; so the use of  $\widehat{Q}_n := \max(\widehat{Q}_n^{(1)}, \widehat{Q}_n^{(2)})$  is needed to get more powerful procedure (see for instance the real data application in Subsection 6.2).

## 5. Some numerical results for inference in INTGARCH model

We consider the integer-valued threshold GARCH( $p, q$ ) (or INTGARCH( $p, q$ )) process  $Y = (Y_t)_{t \in \mathbb{Z}}$  which satisfying

$Y_t/\mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t)$  with

$$\lambda_t = \alpha_0 + \sum_{k=1}^p \alpha_k \lambda_{t-k} + \sum_{k=1}^q \left( \beta_k^+ \max(Y_{t-k} - \ell, 0) + \beta_k^- \min(Y_{t-k}, \ell) \right)$$

where  $\alpha_0 > 0$ ,  $\alpha_k, \beta_k^+, \beta_k^- \geq 0$  and  $\sum_{k=1}^p \alpha_k + \sum_{k=1}^q \max(\beta_k^+, \beta_k^-) < 1$ .

Denote  $\theta_0 = (\alpha_0, \alpha_1, \dots, \alpha_p, \beta_1^+, \beta_1^-, \dots, \beta_q^+, \beta_q^-)$  the true parameter of the model. This model is a special case of the INTARCH( $\infty$ ) see (7).

### 5.1. Estimation and identification

Assume that a trajectory  $(Y_1, \dots, Y_n)$  of  $Y$  is observed. If the orders ( $p, q$ ) and the threshold  $\ell$  are known, then the parameter  $\theta_0$  can be estimated by maximizing the conditional log-likelihood defined in (8) and Theorem 3.1 and Theorem 3.2 are applied.

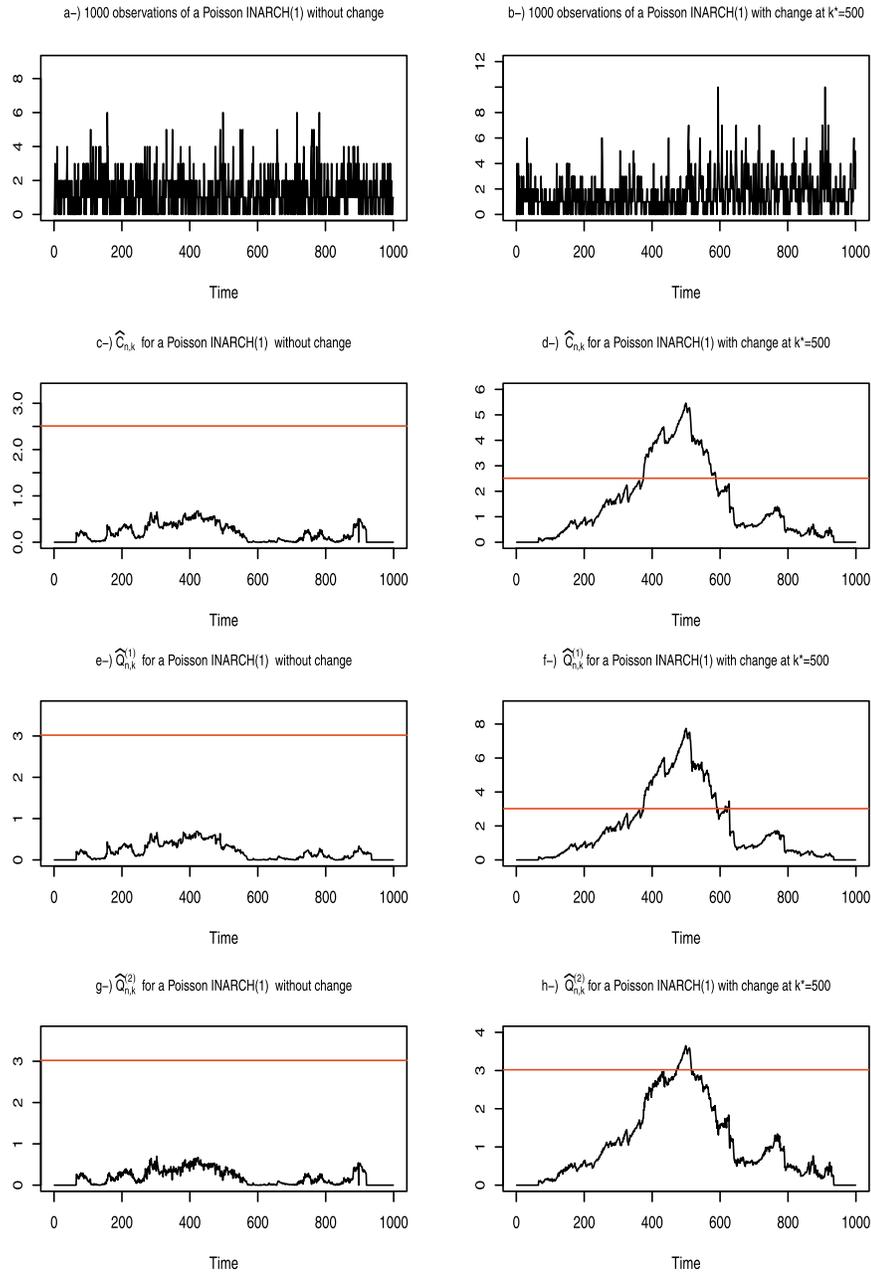


FIG 1. Typical realization of 1000 observations of two Poisson INARCH(1) processes and the corresponding statistics  $\widehat{C}_{n,k}^{(1)}$ ,  $\widehat{Q}_{n,k}^{(1)}$  and  $\widehat{Q}_{n,k}^{(2)}$ . a-) is a Poisson INARCH(1) process without change, where the parameter  $\theta_0 = (1, 0.2)$  is constant. b-) is a Poisson INARCH(1) process where the parameter  $\theta_0 = (1, 0.2)$  changes to  $(1, 0.45)$  at  $k^* = 500$ . c-), d-), e-), f-), g-) and h-) are their corresponding statistics  $\widehat{C}_{n,k}$ ,  $\widehat{Q}_{n,k}^{(1)}$  and  $\widehat{Q}_{n,k}^{(2)}$ . The horizontal line represents the limit of the critical region of the test.

The order  $(p, q)$  can be estimated by using an information criterion such as AIC or BIC; and the threshold  $\ell$  is estimated by maximizing the log-likelihood over a set of integer values  $\{0, 1, \dots, \ell_{\max}\}$  where  $\ell_{\max}$  is an upper bound of the true threshold  $\ell$ . In practice, we can choose an adapted  $\ell_{\max} = \max(Y_1, \dots, Y_n)$ .

Let  $p_{\max}$  and  $q_{\max}$  be the upper bound of the orders  $p$  and  $q$  respectively. Hence, the estimation of  $\theta_0$ ,  $(p, q)$  and  $\ell$  can be done in the following three steps:

- Step 1: For each  $(p, q, \ell) \in \{0, \dots, p_{\max}\} \times \{0, \dots, q_{\max}\} \times \{0, \dots, \ell_{\max}\}$  fixed, compute the estimation  $\hat{\theta}_{p, q, \ell}$  as in (9).
- Step 2: for each  $\ell \in \{0, \dots, \ell_{\max}\}$  fixed, select the “best” order  $(\hat{p}_\ell, \hat{q}_\ell)$  by minimizing the AIC or the BIC criteria.
- Step 3: Estimate the threshold by maximizing the log-likelihood as follows

$$\hat{\ell}_n = \operatorname{argmax}_{\ell \in \{0, \dots, \ell_{\max}\}} L(T_{1, n}, \hat{\theta}_{\hat{p}_\ell, \hat{q}_\ell, \ell}).$$

Therefore, the final estimated parameters of the model are  $\hat{\ell}_n$ ,  $(\hat{p}_{\hat{\ell}_n}, \hat{q}_{\hat{\ell}_n})$  and  $\hat{\theta}_{\hat{p}_{\hat{\ell}_n}, \hat{q}_{\hat{\ell}_n}, \hat{\ell}_n}$ . We have implemented this procedure on the R-software (developed by the CRAN project).

## 5.2. Some simulations results

We consider an INTGARCH(1, 1) process  $Y = (Y_t)_{t \in \mathbb{Z}}$  defined by

$$Y_t / \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t) \text{ with} \\ \lambda_t = 1 + 0.7\lambda_{t-1} + 0.2 \max(Y_{t-1} - 6, 0) + 0.1 \min(Y_{t-1}, 6). \quad (14)$$

This scenario is close to the real data example (see below). Note that, the INTGARCH(1, 1) model can be seen as a nonlinear Poisson autoregressive model with one knot. The MLE is reasonably good, as discussed in Davis and Liu [9]. We will focus on the estimation of the threshold  $\ell$  and the order  $(p, q)$  of an INTGARCH( $(p, q)$ ) model.

Let  $(Y_1, \dots, Y_n)$  be a trajectory generated according to (14). We fit an INTGARCH( $p, q$ ) model from these observations;  $p$ ,  $q$  and  $\ell$  are assumed to be unknown and are estimated as described above. For the problem of selecting the order  $(p, q)$ , Table 1 indicates the proportions of the true order (1, 1), low and high order models selected (using AIC and BIC) based on 200 replications

TABLE 1  
Frequencies of the true order, low and high order models selected (using AIC and BIC) based on 200 replications with  $n = 500$  and  $n = 1000$

Criteria	$(\hat{p}_n, \hat{q}_n) = (1, 1)$ (true order)	$\hat{p}_n + \hat{q}_n < 2$ (low order)	$(\hat{p}_n, \hat{q}_n) \neq (1, 1) \ \& \ \hat{p}_n + \hat{q}_n > 2$
			(high order)
$n = 500$	AIC	0.705	0.040
	BIC	0.670	0.250
$n = 1000$	AIC	0.870	0
	BIC	0.915	0.02

TABLE 2  
Some elementary statistics of the estimator  $\hat{\ell}_n$  for the model selected using AIC and BIC

$\hat{\ell}_n$		Mean	SD	Min	$Q_1$	Med	$Q_3$	Max
$n = 500$	AIC	5.645	1.905	0	4	6	8	11
	BIC	4.245	2.759	0	2	5	6	10
$n = 1000$	AIC	5.920	1.625	1	5	6	7	9
	BIC	5.770	1.873	0	5	6	7	9

with  $n = 500$  and  $n = 1000$ . We have used  $p_{\max} = q_{\max} = 5$ . Some empirical statistics of the estimator of (the threshold)  $\ell_n$  are reported in Table 2.

These results show that, the (empirical) probability of selecting the true order increases as  $n$  increases for both AIC and BIC. Even when  $n = 500$  (this length is close to the real data example, see below) this proportion is reasonably acceptable. It also appears that the BIC leads to select low orders models and less selects high orders models than the AIC; this is not surprising, since the penalty term of the BIC is greater than that of the AIC. Note that, the consistency of the BIC and the efficiency of the AIC have been proved in many situations in model selection theory; even if such results have not yet been proved in our model here, we can nevertheless see that the (empirical) probability to select the true model with the BIC increases (with  $n$ ) than that of the AIC.

The empirical statistics of  $\hat{\ell}_n$  displayed in Table 2 show that the estimation of the threshold is reasonably good (for both the AIC and the BIC) in terms of mean and quantiles. For example, when  $n = 1000$  for more than 60% of the replications, the estimation of the threshold belongs to the set  $\{5, 6, 7\}$  while the true value of the threshold is 6.

### 5.3. Application to real data

We consider the number of transactions per minute for the stock Ericsson B during July 2, 2002. There are 460 available observations which represent the transaction of approximately 8 hours (from 09:35 to 17:14), see Figure 2. The empirical mean and variance of the series are 9.909 and 32.836 respectively. That is the data are overdispersed and the positive dependence (see ACF in Figure 2) suggest that the models studied here are candidates for fitting these data.

Fokianos *et al.* [19] have fitted these data using linear Poisson autoregression and exponential autoregressive model with one lag autoregression. These two models describe the data reasonably well and the Pearson residuals provided appear to be white. The linear Poisson autoregression pass the goodness-of-fit test proposed by Fokianos and Neumann [18]. The nonlinear Poisson autoregression with one knot applied by Davis and Liu [9] seems also to describe well the data according to the Pearson residuals analysis.

Nevertheless, the autocorrelation function of the observations displays strong dependence between transactions. Therefore, we apply an INTGARCH( $p, q$ ) (with  $p$  and  $q$  unknown) and select the “best” model as described above. AIC leads to INTGARCH(2, 1) model and BIC to INTGARCH(1, 1). The conditional



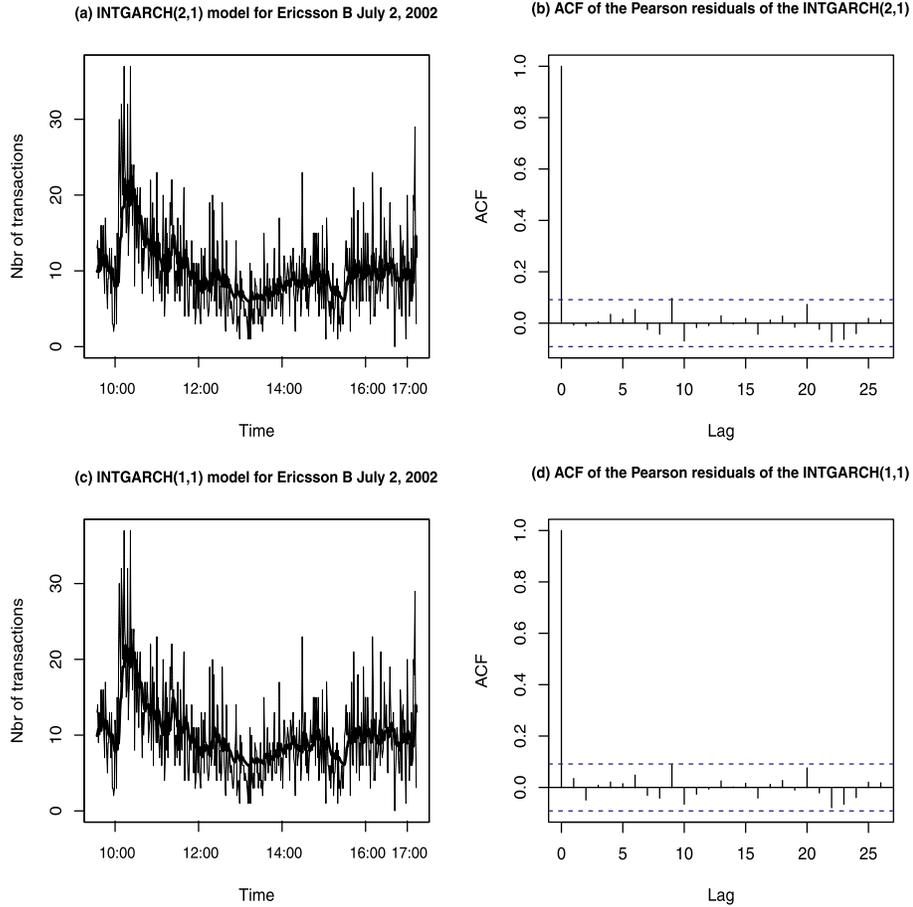


FIG 3. (a) Observed and fitted (thick line) values of the number of transactions of the stock Ericsson B on July 2, 2002 using INTGARCH(2,1) model; (b) is the corresponding autocorrelation function of the Pearson residuals. (c) Observed and fitted (thick line) values of the number of transactions of the stock Ericsson B on July 2, 2002 using INTGARCH(1,1) model; (d) is the corresponding autocorrelation function of the Pearson residuals.

if the Poisson INGARCH(1,1) pass the test proposed by Fokianos and Neumann, some doubt have been pointed about the linearity assumption of these data (see the test based on  $H_n$  in [18]) and the Poisson INTGARCH(2,1) and INTGARCH(1,1) models seem to well describe the data. Let us make some discussion according to the AIC, BIC and the Pearson residuals.

- First not that, the INGARCH(1,1) model can be seen as a particular case of the INTGARCH( $p, q$ ) with the threshold  $\ell = 0$ . Thus the INTGARCH(2,1) and INTGARCH(1,1) models outperform the INGARCH(1,1) in terms of the AIC and BIC respectively; see also Table 3 (the likelihood have been computed up to some constants).

TABLE 3  
 Model selection for the stock Ericsson B on July 2, 2002

	INGARCH(1, 1)	INTGARCH(1, 1)	INTGARCH(2, 1)
AIC	-12170.80	-12175.09	-12176.63
BIC	-12158.41	-12158.56	-12155.97
MSE (Pearson residuals)	2.36	2.31	2.33

- The Pearson residuals of these three models seem to be white (see [19] for the case of the INGARCH(1, 1)). The mean square error of the Pearson residuals (defined by  $\sum_{t=1}^n \hat{\xi}_t^2 / (n - d)$  where  $d$  is the number of estimated parameters) displayed in Table 3 show a slight gain with the INTGARCH(1, 1) model.

In conclusion, it appears that the Poisson INGARCH(1, 1), INTGARCH(1, 1) and INTGARCH(2, 1) fit the data adequately. None of these models outperforms the others, according to the three criteria (AIC, BIC and the MSE of the Pearson residuals). INTGARCH(1, 1) is more accurate in terms of the MSE of the Pearson residuals, INTGARCH(2, 1) outperforms in terms of AIC whereas INTGARCH(1, 1) and INGARCH(1, 1) models are preferable in terms of BIC.

## 6. Some numerical results for parameter change in INGARCH model

### 6.1. Testing for parameter change in INGARCH model

We provide some simulations results to show the empirical performance of the tests procedures described above. We consider the Poisson INGARCH(1,1) model:

$$Y_t / \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t) \text{ with } \lambda_t = \alpha_0 + \alpha_1 \lambda_{t-1} + \beta_1 Y_{t-1}. \quad (15)$$

For sample sizes  $n = 500, 1000$ , the statistics  $\hat{C}_n$  and  $\hat{Q}_n$  are computed with  $u_n = v_n = [(\log n)^{5/2}]$ . The empirical levels and powers reported in the followings table are obtained after 200 replications at the nominal level  $\alpha = 0.05$ .

#### 1. Poisson INGARCH(1,1) with one change-point alternative.

Denote by  $\theta = (\alpha_0, \alpha_1, \beta_1)$  the parameter of the model (15). Table 4 indicates the empirical levels computed when the parameter is  $\theta_0$  and the empirical powers computed when  $\theta_0$  changes to  $\theta_1$  at  $n/2$ .

#### 2. Poisson INARCH(1) with two change-points alternative.

We assume in (15) that  $\alpha_1 = 0$  and denote by  $\theta = (\alpha_0, \beta_1)$  the parameter of the model. Table 5 indicates the empirical levels computed when the parameter is  $\theta_0$  and the empirical powers computed when  $\theta_0$  changes to  $\theta_1$  at  $0.3n$  which changes to  $\theta_2$  at  $0.7n$ . The second alternative scenario is a case where the change in the parameters does not induce a change in the mean of the marginal distribution.

TABLE 4  
Empirical levels and powers at the nominal level 0.05 of test for parameter change in Poisson INGARCH(1,1) model with one change-point alternative

	Procedure	$n = 500$	$n = 1000$
Empirical levels			
$\theta_0 = (1, 0.1, 0.2)$	$\hat{C}_n$ statistic	0.020	0.040
	$\hat{Q}_n$ statistic	0.035	0.045
$\theta_0 = (0.3, 0.5, 0.1)$	$\hat{C}_n$ statistic	0.065	0.060
	$\hat{Q}_n$ statistic	0.080	0.055
Empirical powers			
$\theta_0 = (1, 0.1, 0.2); \theta_1 = (0.7, 0.1, 0.2)$	$\hat{C}_n$ statistic	0.415	0.840
	$\hat{Q}_n$ statistic	0.595	0.910
$\theta_0 = (0.3, 0.5, 0.1); \theta_1 = (0.3, 0.3, 0.4)$	$\hat{C}_n$ statistic	0.695	0.960
	$\hat{Q}_n$ statistic	0.865	0.995

TABLE 5  
Empirical levels and powers at the nominal level 0.05 of test for parameter change in Poisson INARCH(1) model with two change-points alternative

	Procedure	$n = 500$	$n = 1000$
Empirical levels			
$\theta_0 = (1, 0.2)$	$\hat{C}_n$ statistic	0.065	0.050
	$\hat{Q}_n$ statistic	0.060	0.040
$\theta_0 = (0.2, 0.5)$	$\hat{C}_n$ statistic	0.080	0.055
	$\hat{Q}_n$ statistic	0.095	0.060
Empirical powers			
$\theta_0 = (1, 0.2); \theta_1 = (1, 0.45); \theta_2 = (1, 0.15)$	$\hat{C}_n$ statistic	0.680	0.975
	$\hat{Q}_n$ statistic	0.565	0.985
$\theta_0 = (0.2, 0.5); \theta_1 = (0.3, 0.25); \theta_2 = (0.1, 0.75)$	$\hat{C}_n$ statistic	0.705	0.990
	$\hat{Q}_n$ statistic	0.740	0.995

It appears in Tables 4, 5 that these two procedures display a size distortion when  $n = 500$ ; but the empirical levels are close to the nominal one when  $n = 1000$ . One can also see that the empirical powers of these procedures increase with  $n$  and are more accurate even for the case that the break does not induce a change in the mean of the marginal distribution. Although the procedure based on  $\hat{Q}_n$  is little more powerful, the test based on  $\hat{C}_n$  provides satisfactory empirical powers even in the case of two change-points alternative. This could be a starting point for investigation of the consistency of this procedure under multiple change-points alternative.

## 6.2. Real data application

We consider the number of transactions per minute for the stock Ericsson B during July 16, 2002. There are 460 observations which represent trading from 09:35 to 17:14. Figure 4 displays the data and its autocorrelation function.

Several works (see for instance Fokianos *et al.* [19], Davis and Liu [9]) on the series of July 2, 2002 led to use an INGARCH(1,1) model (on the series of July 2, 2002). This model provides  $\hat{\alpha}_1 + \hat{\beta}_1$  close to unity. It can be seen in the slow

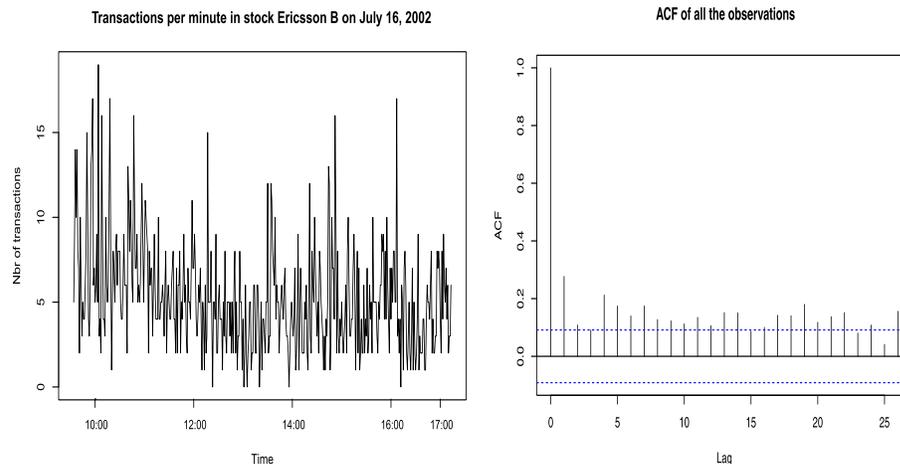


FIG 4. Number of transactions per minute for the stock Ericsson B during July 16, 2002 and their autocorrelation function.

decay of the autocorrelation function (see Figure 2). The series in the period 2–22 July 2002 have been studied by Brännäs and Quoreshi [5]. They have pointed out the presence of long memory in these data and applied INARMA model to both level and first difference forms.

For the transaction during July 16, 2002, we test the adequacy of the Poisson INGARCH(1,1) model by applying the goodness-of-fit test proposed by Fokianos and Neumann [18]. Let  $\hat{\theta}_n = (\hat{\alpha}_{0,n}, \hat{\alpha}_{1,n}, \hat{\beta}_{1,n})$  be the maximum likelihood estimator computed on the observations. Denote  $\hat{I}_t = (\hat{\lambda}_t, Y_t)$  where  $\hat{\lambda}_t = \hat{\alpha}_{0,n} + \hat{\alpha}_{1,n}\hat{\lambda}_{t-1} + \hat{\beta}_{1,n}Y_{t-1}$ . Recall that the estimated Pearson residuals is given by  $\hat{\xi}_t = (Y_t - \hat{\lambda}_t)/\sqrt{\hat{\lambda}_t}$ . The goodness-of-fit test is based on the statistic

$$\hat{T}_n = \sup_{x \in \Pi} |\hat{G}_n(x)| \quad \text{with} \quad \hat{G}_n(x) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \hat{\xi}_t w(x - \hat{I}_{t-1})$$

where  $\Pi = [0, \infty)^2$  and  $w(x) = w(x_1, x_2) = K(x_1)K(x_2)$  where  $K(\cdot)$  is a univariate kernel. A parametric bootstrap procedure can be used to compute p-values of this test. See [18] for more detail on this test procedure.

We have applied this test with  $B = 300$  bootstrap replications and the p-values 0.032 and 0.05 have been obtained respectively for uniform and Epanechnikov kernel. So, the linear Poisson INGARCH(1,1) model is rejected.

The previous test for change detection have been applied to the series. A change has been detected around the midday at  $t^* = 12:05$ . See the statistics  $\hat{C}_{n,k}$ ,  $\hat{Q}_{n,k}^{(1)}$  and  $\hat{Q}_{n,k}^{(2)}$  as well as the breakpoint and the autocorrelation function of each regime on Figure 5.

To assess the adequacy of the linear Poisson INGARCH(1,1) on each regime, we apply the goodness-of-fit test of Fokianos and Neumann [18]. The p-values

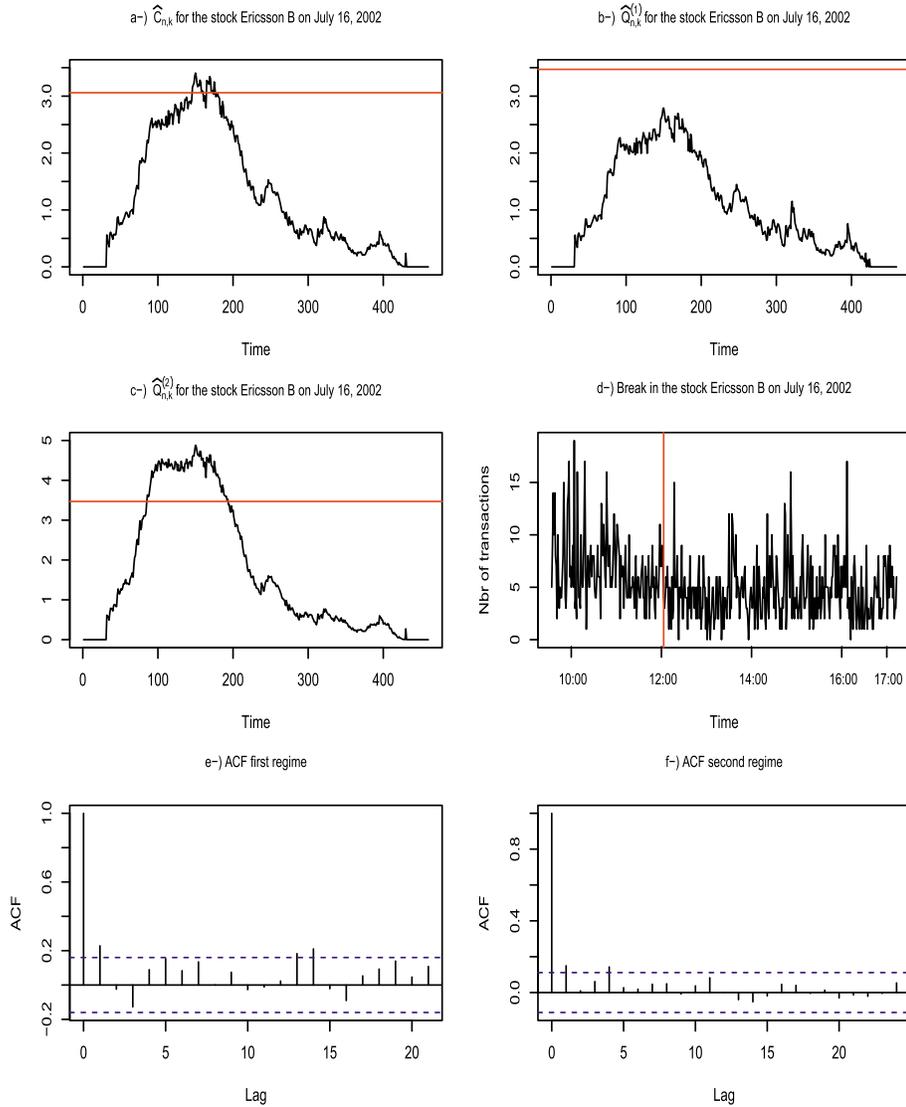


FIG 5. Break detection in the transactions per minute of the stock Ericsson B on July 16, 2002. a-), b-), c-) the corresponding statistics  $\hat{C}_{n,k}$ ,  $\hat{Q}_{n,k}^{(1)}$  and  $\hat{Q}_{n,k}^{(2)}$ . d-) Breakpoint in the data; e-), f-) the autocorrelation functions of each regime.

obtained are displayed in Table 6. These results point to the adequacy of the linear Poisson INGARCH(1,1) on of the first regime and raises some doubt about the linearity on the second regime. This shows that, the model structure of the transactions in the morning may be different to the structure of the transactions in the afternoon. Moreover, Figure 5 shows that, the autocorrelation function

TABLE 6  
*p*-values when testing adequacy of the Poisson INGARCH(1,1) on each regime after break detection in the transactions per minute of the stock Ericsson B on July 16, 2002. Results are based on  $B=300$  bootstrap replications

	Uniform kernel	Epanechnikov kernel
First regime (from 09:35 to 12:05)	0.250	0.213
Second regime (from 12:06 to 17:14)	0.006	0.006

of each regime decreases fast; this rules out the idea of the long memory in the series.

## 7. Proofs of the main results

**Proof of the Proposition 2.1.** We will use the same techniques as in [12]. Let  $p, q$  two fixed non-negative integers. Define the sequence  $(\lambda_t^{p,q})_{t \in \mathbb{Z}}$  by

$$\lambda_t^{p,q} = \begin{cases} 0 & \text{if } t \leq -q \\ F(\lambda_{t-1}^{p,q}, \dots, \lambda_{t-p}^{p,q}, 0, \dots; Y_{t-1}, \dots) & \text{otherwise.} \end{cases} \quad (16)$$

The existence of moment of any order of the process  $(Y_t, \lambda_t)_{t \in \mathbb{Z}}$  (see [13]) and assumption  $(A_F)$  imply the existence of moment of any order of  $(\lambda_t^{p,q})_{t \in \mathbb{Z}}$ . Let us show that  $(\lambda_0^{p,q})_{q \geq 0}$  is a Cauchy sequence in  $L^1$ . By using  $(A_F)$ , we have

$$E|\lambda_0^{p,q+1} - \lambda_0^{p,q}| \leq \sum_{j=1}^p \alpha_j E|\lambda_{-j}^{p,q+1} - \lambda_{-j}^{p,q}|.$$

By definition and the strictly stationarity of  $(Y_t)_{t \in \mathbb{Z}}$ , we can easily see that for  $j = 1, \dots, p$ , the couples  $(\lambda_{-j}^{p,q+1}, \lambda_{-j}^{p,q+1})$  and  $(\lambda_0^{p,q+1-j}, \lambda_0^{p,q-j})$  have the same distribution. Hence,

$$E|\lambda_0^{p,q+1} - \lambda_0^{p,q}| \leq \sum_{j=1}^p \alpha_j E|\lambda_0^{p,q+1-j} - \lambda_0^{p,q-j}|.$$

For any fixed  $p$ , denote  $v_q = E|\lambda_0^{p,q+1} - \lambda_0^{p,q}|$  for all  $q > p$ . It holds that

$$v_q \leq \sum_{j=1}^p \alpha_j v_{q-j}.$$

By applying the Lemma 5.4 of [12], we obtain

$$v_q \leq \alpha^{q/p} v_0 \text{ where } \alpha = \sum_{j=1}^{\infty} \alpha_j.$$

Hence,  $v_q \rightarrow 0$  as  $q \rightarrow \infty$ . Thus, for any  $p > 0$ , the sequence  $(\lambda_0^{p,q})$  is a Cauchy sequence in  $L^1$ . Therefore, it converges to some limit denoted  $\lambda_0^p$ . Moreover, since the sequence  $(\lambda_0^{p,q})_{q \geq 1}$  is measurable w.r.t to  $\sigma(Y_t, t < 0)$ , it is

the case of the limit  $\lambda_0^p$ . So, there exists a measurable function  $f^{(p)}$  such that  $\lambda_0^p = f^{(p)}(Y_{-1}, \dots)$ . By going along similar lines, it holds that for any  $t \in \mathbb{Z}$ , the sequence  $(\lambda_t^{p,q})_{q \geq 1}$  converges in  $L^1$  to some  $\lambda_t^p = f^{(p)}(Y_{t-1}, \dots)$  and since  $(Y_t)_{t \in \mathbb{Z}}$ , is stationary and ergodic, the process  $(\lambda_t^p)_{t \in \mathbb{Z}}$  is too stationary and ergodic.

Let  $p$  and  $t$  fixed. For  $q$  large enough, we have

$$\lambda_t^{p,q} = F(\lambda_{t-1}^{p,q}, \dots, \lambda_{t-p}^{p,q}, 0, \dots; Y_{t-1}, \dots)$$

(see (16)). By using the continuity (which comes from Lipschitz-type conditions) of  $(Y_1, \dots, Y_p) \mapsto F(Y_1, \dots, Y_p, 0, \dots; y)$  for any fixed  $y = (y_i)_{i \geq 1}$  and by carrying  $q$  to infinity, it holds that

$$\lambda_t^p = F(\lambda_{t-1}^p, \dots, \lambda_{t-p}^p, 0, \dots; Y_{t-1}, \dots). \tag{17}$$

Denote  $\mu_p = E\lambda_t^p$ ,  $\mu = \sup_{p \geq 1} \mu_p$ ,  $\Delta_{p,t} = E|\lambda_t^{p+1} - \lambda_t^p|$  and  $\Delta_p = \sup_{t \in \mathbb{Z}} \Delta_{p,t}$ . By going the same lines as in [12], we obtain  $\Delta_p \leq C\alpha_{p+1}$ . Therefore,  $\Delta_p \rightarrow 0$  as  $p \rightarrow \infty$ . This shows that for any fixed  $t \in \mathbb{Z}$ ,  $(\lambda_t^p)_{p \geq 1}$  is a Cauchy sequence in  $L^1$ . Thus it converges to some random  $\tilde{\lambda}_t \in L^1$ . Moreover,  $\tilde{\lambda}_t$  is measurable w.r.t  $\sigma(Y_j, j < t)$  (because it is the case of  $(\lambda_t^p)_{p \geq 1}$ ). Thus, there exists a measurable function  $f$  such that  $\tilde{\lambda}_t = f(Y_{t-1}, \dots)$  for any  $t \in \mathbb{Z}$ . This implies that  $(\tilde{\lambda}_t)_{t \in \mathbb{Z}}$  is strictly stationary and ergodic. Finally, by using equation (17) and continuity of  $F$ , it comes that

$$\tilde{\lambda}_t = F(\tilde{\lambda}_{t-1}, \dots; Y_{t-1}, \dots), \quad \text{for any } t \in \mathbb{Z}. \tag{18}$$

Hence, the process  $(Y_t, \tilde{\lambda}_t)_{t \in \mathbb{Z}}$  is strictly stationary ergodic and satisfying (1). By the uniqueness of the solution, it holds that  $\tilde{\lambda}_t = \lambda_t$  a.s.

Thus  $\lambda_t = f(Y_{t-1}, \dots)$  for any  $t \in \mathbb{Z}$ . ■

**Proof of the Theorem 3.1.** Without loss of generality, for simplifying notation, we will make the proof with  $T_{j_n, k_n} = T_{1, n}$ . The proof is divided into two parts. We will first show that  $\|\frac{1}{n} \sum_{t \in T_{1, n}} \hat{\ell}_t(\theta) - L(\theta)\|_{\Theta} \xrightarrow[n \rightarrow +\infty]{a.s.} 0$  where  $L(\Theta) := E(\ell_0(\theta))$ ; secondly, we will show that the function  $\theta \mapsto L(\Theta)$  has a unique maximum in  $\theta_0$ .

- (i) Let  $\theta \in \Theta$ , recall that  $\ell_t(\theta) = Y_t \log \lambda_t(\theta) - \lambda_t(\theta) = Y_t \log f_{\theta}^t - f_{\theta}^t$ . Since  $(Y_t)_{t \in \mathbb{Z}}$  is stationary and ergodic, for any  $\theta \in \Theta$ ,  $(\ell_t(\theta))_{t \in \mathbb{Z}}$  is also a stationary and ergodic sequence. Moreover, we have for any  $\theta \in \Theta$ ,

$$\begin{aligned} |\ell_t(\theta)| &\leq |Y_t| |\log f_{\theta}^t| + |f_{\theta}^t| \\ &\leq |Y_t| \left| \log \left( \frac{f_{\theta}^t}{\underline{c}} \times \underline{c} \right) \right| + |f_{\theta}^t| \\ &\leq |Y_t| \left( \left| \frac{f_{\theta}^t}{\underline{c}} - 1 \right| + |\log \underline{c}| \right) + |f_{\theta}^t| \quad (\text{for } x > 1, |\log x| \leq |x - 1|) \\ &\leq |Y_t| \left( \frac{1}{\underline{c}} |f_{\theta}^t| + 1 + |\log \underline{c}| \right) + |f_{\theta}^t|. \end{aligned}$$

Hence,

$$\sup_{\theta \in \Theta} |\ell_t(\theta)| \leq |Y_t| \left( \frac{1}{\underline{c}} \|f_\theta^t\|_\Theta + 1 + |\log \underline{c}| \right) + \|f_\theta^t\|_\Theta.$$

We will show that, for any  $r > 0$ ,  $E(\|f_\theta^t\|_\Theta^r) < \infty$ . Since  $A_0(\Theta)$  holds, we have

$$\|f_\theta^t\|_\Theta \leq \|f_\theta^t - f_\theta(0)\|_\Theta + \|f_\theta(0)\|_\Theta \leq \sum_{j \geq 1} \alpha_j^{(0)}(\Theta) |Y_{t-j}| + \|f_\theta(0)\|_\Theta.$$

Thus, by using the stationarity of the process  $(Y_t)_{t \in \mathbb{Z}}$ , it follows that

$$\left( E(\|f_\theta^t\|_\Theta^r) \right)^{1/r} \leq \|Y_0\|^r \sum_{j \geq 1} \alpha_j^{(0)}(\Theta) + \|f_\theta(0)\|_\Theta < \infty.$$

Therefore, we have

$$E \left( \sup_{\theta \in \Theta} |\ell_t(\theta)| \right) \leq \frac{1}{\underline{c}} (E|Y_t|^2)^{1/2} \cdot \left( E \|f_\theta^t\|_\Theta^2 \right)^{1/2} + (1 + |\log \underline{c}|) E|Y_t| + E \|f_\theta^t\|_\Theta < \infty.$$

By the uniform strong law of large number applied on  $(\ell_t(\theta))_{t \geq 1}$  (see Straumann and Mikosch (2006) [33]), it holds that

$$\left\| \frac{1}{n} \sum_{t \in T_{1,n}} \ell_t(\theta) - E\ell_0(\theta) \right\|_\Theta \xrightarrow[n \rightarrow +\infty]{a.s.} 0. \quad (19)$$

Now let us show that

$$\frac{1}{n} \left\| \sum_{t \in T_{1,n}} \ell_t(\theta) - \sum_{t \in T_{1,n}} \widehat{\ell}_t(\theta) \right\|_\Theta \xrightarrow[n \rightarrow +\infty]{a.s.} 0.$$

We have

$$\frac{1}{n} \left\| \sum_{t \in T_{1,n}} \ell_t(\theta) - \sum_{t \in T_{1,n}} \widehat{\ell}_t(\theta) \right\|_\Theta \leq \frac{1}{n} \sum_{t \in T_{1,n}} \left\| \ell_t(\theta) - \widehat{\ell}_t(\theta) \right\|_\Theta.$$

We will apply the Corollary 1 of Kounias and Weng (1969) [29]. So, it suffices to show that

$$\frac{1}{n} \sum_{t \geq 1} \frac{1}{t} E \left( \left\| \ell_t(\theta) - \widehat{\ell}_t(\theta) \right\|_\Theta \right) < \infty.$$

For  $t \in T_{1,n}$  and  $\theta \in \Theta$ , we have

$$\ell_t(\theta) - \widehat{\ell}_t(\theta) = Y_t \log f_\theta^t - f_\theta^t - Y_t \log \widehat{f}_\theta^t + \widehat{f}_\theta^t = Y_t (\log f_\theta^t - \log \widehat{f}_\theta^t) - (f_\theta^t - \widehat{f}_\theta^t).$$

By using the relation  $|\log f_\theta^t - \log \widehat{f}_\theta^t| \leq \frac{1}{\underline{c}} |f_\theta^t - \widehat{f}_\theta^t|$ , it comes that

$$\left\| \ell_t(\theta) - \widehat{\ell}_t(\theta) \right\|_\Theta \leq \frac{1}{\underline{c}} |Y_t| \left\| f_\theta^t - \widehat{f}_\theta^t \right\|_\Theta + \left\| f_\theta^t - \widehat{f}_\theta^t \right\|_\Theta$$

$$\leq \left(\frac{1}{\underline{c}} |Y_t| + 1\right) \left\| f_\theta^t - \widehat{f}_\theta^t \right\|_\Theta.$$

By Cauchy-Schwartz inequality,

$$\begin{aligned} E\left(\left\| \ell_t(\theta) - \widehat{\ell}_t(\theta) \right\|_\Theta\right) &\leq E\left[\left(\frac{1}{\underline{c}} |Y_t| + 1\right) \left\| f_\theta^t - \widehat{f}_\theta^t \right\|_\Theta\right] \\ &\leq \left(E\left(\frac{1}{\underline{c}} |Y_t| + 1\right)^2\right)^{1/2} \times \left(E\left\| f_\theta^t - \widehat{f}_\theta^t \right\|_\Theta^2\right)^{1/2}. \end{aligned}$$

We have (by Minkowski inequality),

$$\left(E\left(\frac{1}{\underline{c}} |Y_t| + 1\right)^2\right)^{1/2} \leq \frac{1}{\underline{c}} (E|Y_t|^2)^{1/2} + 1 < \infty.$$

Thus, it comes that

$$E\left(\left\| \ell_t(\theta) - \widehat{\ell}_t(\theta) \right\|_\Theta\right) \leq C \left(E\left\| f_\theta^t - \widehat{f}_\theta^t \right\|_\Theta^2\right)^{1/2}.$$

But, we have  $\|f_\theta^t - \widehat{f}_\theta^t\|_\Theta \leq \sum_{j \geq t} \alpha_j^{(0)}(\Theta) |Y_{t-j}|$ . By using Minkowski inequality, it comes that

$$\left(E\left\| f_\theta^t - \widehat{f}_\theta^t \right\|_\Theta^2\right)^{1/2} \leq (E|Y_0|^2)^{1/2} \sum_{j \geq t} \alpha_j^{(0)}(\Theta).$$

Hence

$$\left(E\left\| f_\theta^t - \widehat{f}_\theta^t \right\|_\Theta^2\right)^{1/2} \leq C \sum_{j \geq 0} \alpha_j^{(0)}(\Theta).$$

Thus

$$E\left(\left\| \ell_t(\theta) - \widehat{\ell}_t(\theta) \right\|_\Theta\right) \leq C \sum_{j \geq t} \alpha_j^{(0)}(\Theta).$$

Therefore

$$\begin{aligned} \sum_{t \geq 1} \frac{1}{t} E\left\| \ell_t(\theta) - \widehat{\ell}_t(\theta) \right\|_\Theta &\leq C \sum_{t \geq 1} \frac{1}{t} \sum_{j \geq t} \alpha_j^{(0)}(\Theta) = C \sum_{t \geq 1} \sum_{j \geq t} \frac{1}{t} \alpha_j^{(0)}(\Theta) \\ &\leq \sum_{j \geq 1} \sum_{t=1}^j \frac{1}{t} \alpha_j^{(0)}(\Theta) = C \sum_{j \geq 1} \alpha_j^{(0)}(\Theta) \sum_{t=1}^j \frac{1}{t} \\ &\leq C \sum_{j \geq 1} \alpha_j^{(0)}(\Theta) \cdot (1 + \log j) \\ &\leq C \sum_{j \geq 1} \alpha_j^{(0)}(\Theta) + C \sum_{j \geq 1} \alpha_j^{(0)}(\Theta) \sqrt{j} \\ &< \infty \text{ (according to } A_0(\Theta) \text{ and (10)).} \end{aligned}$$

Hence, it follows that

$$\frac{1}{n} \left\| \sum_{t \in T_{1,n}} \ell_t(\theta) - \sum_{t \in T_{1,n}} \widehat{\ell}_t(\theta) \right\|_{\Theta} \xrightarrow[n \rightarrow +\infty]{a.s.} 0. \quad (20)$$

From (19) and (20), we deduce that

$$\left\| \frac{1}{n} \sum_{t \in T_{1,n}} \widehat{\ell}_t(\theta) - E\ell_0(\theta) \right\|_{\Theta} \xrightarrow[n \rightarrow \infty]{a.s.} 0.$$

- (ii) We will now show that the function  $\theta \mapsto L(\theta) = E\ell_0(\theta)$  has a unique maximum at  $\theta_0$ . We will proceed as in [9]. Let  $\theta \in \Theta$ , with  $\theta \neq \theta_0$ . We have

$$\begin{aligned} L(\theta_0) - L(\theta) &= E\ell_0(\theta_0) - E\ell_0(\theta) \\ &= E(Y_0 \log f_{\theta_0}^0 - f_{\theta_0}^0) - E(Y_0 \log f_{\theta}^0 - f_{\theta}^0) \\ &= E[f_{\theta_0}^0 (\log f_{\theta_0}^0 - \log f_{\theta}^0)] - E(f_{\theta_0}^0 - f_{\theta}^0). \end{aligned}$$

By applying the mean value theorem at the function  $x \mapsto \log x$  defined on  $[\underline{c}, +\infty[$ , there exists  $\xi$  between  $f_{\theta_0}^0$  and  $f_{\theta}^0$  such that  $\log f_{\theta_0}^0 - \log f_{\theta}^0 = (f_{\theta_0}^0 - f_{\theta}^0) \frac{1}{\xi}$ . Hence, it comes that

$$\begin{aligned} L(\theta_0) - L(\theta) &= E\left(\frac{1}{\xi} f_{\theta_0}^0 (f_{\theta_0}^0 - f_{\theta}^0)\right) - E(f_{\theta_0}^0 - f_{\theta}^0) \\ &= E\left(\left(\frac{f_{\theta_0}^0}{\xi} - 1\right)(f_{\theta_0}^0 - f_{\theta}^0)\right) \\ &= E\left(\frac{1}{\xi} (f_{\theta_0}^0 - \xi)(f_{\theta_0}^0 - f_{\theta}^0)\right). \end{aligned}$$

Since  $\theta \neq \theta_0$ , it follows from assumption  $\text{Id}(\Theta)$  that  $\frac{1}{\xi} (f_{\theta_0}^0 - \xi)(f_{\theta_0}^0 - f_{\theta}^0) \neq 0$  a.s. Moreover

- if  $f_{\theta_0}^0 < f_{\theta}^0$ , then  $f_{\theta_0}^0 < \xi < f_{\theta}^0$  and hence  $\frac{1}{\xi} (f_{\theta_0}^0 - \xi)(f_{\theta_0}^0 - f_{\theta}^0) > 0$ ;
- if  $f_{\theta_0}^0 > f_{\theta}^0$ , then  $f_{\theta}^0 < \xi < f_{\theta_0}^0$  and hence  $\frac{1}{\xi} (f_{\theta_0}^0 - \xi)(f_{\theta_0}^0 - f_{\theta}^0) > 0$ .

We deduce that  $\frac{1}{\xi} (f_{\theta_0}^0 - \xi)(f_{\theta_0}^0 - f_{\theta}^0) > 0$  a.s.. Hence  $L(\theta_0) - L(\theta) = E\left(\frac{1}{\xi} (f_{\theta_0}^0 - \xi)(f_{\theta_0}^0 - f_{\theta}^0)\right) > 0$ . Thus, the function  $\theta \mapsto L(\theta)$  has a unique maximum at  $\theta_0$ .

- (i), (ii) and standard arguments lead to the consistency of  $\widehat{\theta}_n(T_{1,n})$ . ■

The following lemma are needed to prove the Theorem 3.2.

**Lemma 7.1.** *Let  $(j_n)_{n \geq 1}$  and  $(k_n)_{n \geq 1}$  two integer valued sequences such that  $(j_n)_{n \geq 1}$  is increasing,  $j_n \rightarrow \infty$  and  $k_n - j_n \rightarrow \infty$  as  $n \rightarrow \infty$ . Let  $n \geq 1$ , for any segment  $T = T_{j_n, k_n} \subset \{1, \dots, n\}$ , it holds under assumptions of Theorem 3.2 that*

$$E\left(\frac{1}{\sqrt{k_n - j_n}} \left\| \frac{\partial}{\partial \theta} L_n(T, \theta) - \frac{\partial}{\partial \theta} \widehat{L}_n(T, \theta) \right\|_{\Theta}\right) \xrightarrow[n \rightarrow \infty]{} 0.$$

**Proof.** Let  $i \in \{1, \dots, n\}$ . We have

$$\begin{aligned} \frac{\partial}{\partial \theta_i} L_n(T, \theta) &= \sum_{t \in T} \frac{\partial}{\partial \theta_i} (Y_t \log f_\theta^t - f_\theta^t) = \sum_{t \in T} \left( Y_t \frac{1}{f_\theta^t} \frac{\partial}{\partial \theta_i} f_\theta^t - \frac{\partial}{\partial \theta_i} f_\theta^t \right) \\ &= \sum_{t \in T} \frac{\partial}{\partial \theta_i} \ell_t(\theta) \end{aligned}$$

and

$$\frac{\partial}{\partial \theta_i} \widehat{L}_n(T, \theta) = \sum_{t \in T} \left( Y_t \frac{1}{\widehat{f}_\theta^t} \frac{\partial}{\partial \theta_i} \widehat{f}_\theta^t - \frac{\partial}{\partial \theta_i} \widehat{f}_\theta^t \right) = \sum_{t \in T} \frac{\partial}{\partial \theta_i} \widehat{\ell}_t(\theta).$$

Hence

$$\begin{aligned} \left| \frac{\partial}{\partial \theta_i} \ell_t(\theta) - \frac{\partial}{\partial \theta_i} \widehat{\ell}_t(\theta) \right| &= \left| Y_t \frac{1}{f_\theta^t} \frac{\partial}{\partial \theta_i} f_\theta^t - \frac{\partial}{\partial \theta_i} f_\theta^t - Y_t \frac{1}{\widehat{f}_\theta^t} \frac{\partial}{\partial \theta_i} \widehat{f}_\theta^t + \frac{\partial}{\partial \theta_i} \widehat{f}_\theta^t \right| \\ &\leq |Y_t| \left| \frac{1}{f_\theta^t} \frac{\partial}{\partial \theta_i} f_\theta^t - \frac{1}{\widehat{f}_\theta^t} \frac{\partial}{\partial \theta_i} \widehat{f}_\theta^t \right| + \left| \frac{\partial}{\partial \theta_i} f_\theta^t - \frac{\partial}{\partial \theta_i} \widehat{f}_\theta^t \right|. \quad (21) \end{aligned}$$

Using the relation  $|a_1 b_1 - a_2 b_2| \leq |a_1 - a_2| |b_2| + |b_1 - b_2| |a_1|$ , we have

$$\begin{aligned} \left| \frac{1}{f_\theta^t} \frac{\partial}{\partial \theta_i} f_\theta^t - \frac{1}{\widehat{f}_\theta^t} \frac{\partial}{\partial \theta_i} \widehat{f}_\theta^t \right| &\leq \left| \frac{1}{f_\theta^t} - \frac{1}{\widehat{f}_\theta^t} \right| \left| \frac{\partial}{\partial \theta_i} \widehat{f}_\theta^t \right| + \left| \frac{\partial}{\partial \theta_i} f_\theta^t - \frac{\partial}{\partial \theta_i} \widehat{f}_\theta^t \right| \left| \frac{1}{f_\theta^t} \right| \\ &\leq \frac{1}{\underline{c}^2} |f_\theta^t - \widehat{f}_\theta^t| \left| \frac{\partial}{\partial \theta_i} \widehat{f}_\theta^t \right| + \frac{1}{\underline{c}} \left| \frac{\partial}{\partial \theta_i} f_\theta^t - \frac{\partial}{\partial \theta_i} \widehat{f}_\theta^t \right|. \end{aligned}$$

Hence, (21) implies

$$\begin{aligned} \left\| \frac{\partial}{\partial \theta_i} \ell_t(\theta) - \frac{\partial}{\partial \theta_i} \widehat{\ell}_t(\theta) \right\| &\leq |Y_t| \left( \frac{1}{\underline{c}^2} \|f_\theta^t - \widehat{f}_\theta^t\|_\Theta \left\| \frac{\partial}{\partial \theta_i} \widehat{f}_\theta^t \right\|_\Theta \right. \\ &\quad \left. + \frac{1}{\underline{c}} \left\| \frac{\partial}{\partial \theta_i} f_\theta^t - \frac{\partial}{\partial \theta_i} \widehat{f}_\theta^t \right\|_\Theta \right) + \left\| \frac{\partial}{\partial \theta_i} f_\theta^t - \frac{\partial}{\partial \theta_i} \widehat{f}_\theta^t \right\|_\Theta \\ &\leq C |Y_t| \left\| \frac{\partial}{\partial \theta_i} \widehat{f}_\theta^t \right\|_\Theta \|f_\theta^t - \widehat{f}_\theta^t\|_\Theta + C(1 + |Y_t|) \left\| \frac{\partial}{\partial \theta_i} f_\theta^t - \frac{\partial}{\partial \theta_i} \widehat{f}_\theta^t \right\|_\Theta. \end{aligned}$$

Let  $r > 0$ . Using the Minkowski and Hölder's inequalities, it holds that

$$\begin{aligned} \left( E \left[ \left\| \frac{\partial}{\partial \theta_i} \ell_t(\theta) - \frac{\partial}{\partial \theta_i} \widehat{\ell}_t(\theta) \right\|_\Theta^r \right] \right)^{\frac{1}{r}} &\leq C \left( E \left[ |Y_t|^r \left\| \frac{\partial}{\partial \theta_i} \widehat{f}_\theta^t \right\|_\Theta^r \|f_\theta^t - \widehat{f}_\theta^t\|_\Theta^r \right] \right)^{1/r} \\ &\quad + C \left( E \left[ (1 + |Y_t|)^r \left\| \frac{\partial}{\partial \theta_i} f_\theta^t - \frac{\partial}{\partial \theta_i} \widehat{f}_\theta^t \right\|_\Theta^r \right] \right)^{1/r} \\ &\leq C \left( (E |Y_t|^{3r})^{1/3} \left( E \left\| \frac{\partial}{\partial \theta_i} \widehat{f}_\theta^t \right\|_\Theta^{3r} \right)^{1/3} \left( E \|f_\theta^t - \widehat{f}_\theta^t\|_\Theta^{3r} \right)^{1/3} \right)^{1/r} \end{aligned}$$

$$\begin{aligned}
& + C \left( (E(1 + |Y_t|)^{2r})^{1/r} \left( E \left\| \frac{\partial}{\partial \theta_i} f_\theta^t - \frac{\partial}{\partial \theta_i} \hat{f}_\theta^t \right\|_\Theta^{2r} \right)^{1/2} \right)^{1/r} \\
& \leq C \|Y_t\|_{3r} \left\| \left\| \frac{\partial}{\partial \theta_i} \hat{f}_\theta^t \right\|_\Theta \right\|_{3r} \cdot \left( E \left\| f_\theta^t - \hat{f}_\theta^t \right\|_\Theta^{3r} \right)^{1/3r} \\
& \quad + C \|1 + |Y_t|\|_{2r} \cdot \left( E \left\| \frac{\partial}{\partial \theta_i} f_\theta^t - \frac{\partial}{\partial \theta_i} \hat{f}_\theta^t \right\|_\Theta^{2r} \right)^{1/2r}.
\end{aligned}$$

But we have  $\|Y_t\|_{3r} = C < \infty$  and  $\|1 + |Y_t|\|_{1r} < \infty$ . Hence

$$\left\| \frac{\partial}{\partial \theta_i} \hat{f}_\theta^t \right\|_\Theta \leq \left\| \frac{\partial}{\partial \theta_i} f_\theta(0) \right\|_\Theta + \left\| \frac{\partial}{\partial \theta_i} \hat{f}_\theta^t - \frac{\partial}{\partial \theta_i} f_\theta(0) \right\|_\Theta \leq C + \sum_{j \geq 1} \alpha_j^{(1)}(\Theta) |Y_{t-j}|.$$

Thus,

$$\left\| \left\| \frac{\partial}{\partial \theta_i} \hat{f}_\theta^t \right\|_\Theta \right\|_{3r} \leq C + \|Y_0\|_{3r} \sum_{j \geq 1} \alpha_j^{(1)}(\Theta) \leq C(1 + \sum_{j \geq 1} \alpha_j^{(1)}(\Theta)) < \infty.$$

We also have  $\|f_\theta^t - \hat{f}_\theta^t\|_\Theta \leq \sum_{j \geq t} \alpha_j^{(0)}(\Theta) |Y_{t-j}|$ . Hence

$$\left( E \left\| f_\theta^t - \hat{f}_\theta^t \right\|_\Theta^{3r} \right)^{1/3r} = \left\| \left\| f_\theta^t - \hat{f}_\theta^t \right\|_\Theta \right\|_{3r} \leq C \sum_{j \geq t} \alpha_j^{(0)}(\Theta).$$

The same argument gives

$$\left( E \left\| \frac{\partial}{\partial \theta_i} f_\theta^t - \frac{\partial}{\partial \theta_i} \hat{f}_\theta^t \right\|_\Theta^{2r} \right)^{1/2r} = \left\| \left\| \frac{\partial}{\partial \theta_i} f_\theta^t - \frac{\partial}{\partial \theta_i} \hat{f}_\theta^t \right\|_\Theta \right\|_{2r} \leq C \sum_{j \geq t} \alpha_j^{(1)}(\Theta).$$

Hence,

$$\left( E \left[ \left\| \frac{\partial}{\partial \theta_i} \ell_t(\theta) - \frac{\partial}{\partial \theta_i} \hat{\ell}_t(\theta) \right\|_\Theta^r \right] \right)^{1/r} \leq C \sum_{j \geq t} \left( \alpha_j^{(0)}(\Theta) + \alpha_j^{(1)}(\Theta) \right).$$

Therefore, we have (with  $r = 1$ )

$$\begin{aligned}
& E \left( \frac{1}{\sqrt{k_n - j_n}} \left\| \frac{\partial}{\partial \theta_i} L_n(T, \theta) - \frac{\partial}{\partial \theta_i} \hat{L}_n(T, \theta) \right\|_\Theta \right) \\
& \leq \frac{1}{\sqrt{k_n - j_n}} \sum_{t \in T} E \left\| \ell_t(\theta) - \hat{\ell}_t(\theta) \right\|_\Theta \\
& \leq C \frac{1}{\sqrt{k_n - j_n}} \sum_{t \in T} \left( \sum_{j \geq t} \left( \alpha_j^{(0)}(\Theta) + \alpha_j^{(1)}(\Theta) \right) \right)
\end{aligned}$$

$$\begin{aligned}
 &\leq C \frac{1}{\sqrt{k_n - j_n}} \sum_{t \in T} \left[ \sum_{j=t}^{k_n} \left( \alpha_j^{(0)}(\Theta) + \alpha_j^{(1)}(\Theta) \right) + \sum_{j \geq k_n} \left( \alpha_j^{(0)}(\Theta) + \alpha_j^{(1)}(\Theta) \right) \right] \\
 &\leq C \frac{1}{\sqrt{k_n - j_n}} \left[ \sum_{t=j_n}^{k_n} \sum_{j=t}^{k_n} \left( \alpha_j^{(0)}(\Theta) + \alpha_j^{(1)}(\Theta) \right) + \sum_{t=j_n}^{k_n} \sum_{j \geq k_n} \left( \alpha_j^{(0)}(\Theta) + \alpha_j^{(1)}(\Theta) \right) \right] \\
 &\leq C \frac{1}{\sqrt{k_n - j_n}} \left[ \sum_{j=j_n}^{k_n} \sum_{t=j_n}^j \left( \alpha_j^{(0)}(\Theta) + \alpha_j^{(1)}(\Theta) \right) \right. \\
 &\qquad \qquad \qquad \left. + (k_n - j_n) \sum_{j \geq k_n} \left( \alpha_j^{(0)}(\Theta) + \alpha_j^{(1)}(\Theta) \right) \right] \\
 &\leq C \frac{1}{\sqrt{k_n - j_n}} \left[ \sum_{j=j_n}^{k_n} (j - j_n) \left( \alpha_j^{(0)}(\Theta) + \alpha_j^{(1)}(\Theta) \right) \right. \\
 &\qquad \qquad \qquad \left. + (k_n - j_n) \sum_{j \geq k_n} \left( \alpha_j^{(0)}(\Theta) + \alpha_j^{(1)}(\Theta) \right) \right] \\
 &\leq C \frac{1}{\sqrt{k_n - j_n}} \sum_{j=j_n}^{k_n} (j - j_n) \left( \alpha_j^{(0)}(\Theta) + \alpha_j^{(1)}(\Theta) \right) \\
 &\qquad \qquad \qquad + C \sqrt{k_n - j_n} \sum_{j \geq k_n} \left( \alpha_j^{(0)}(\Theta) + \alpha_j^{(1)}(\Theta) \right) \\
 &\leq \frac{C}{\sqrt{k_n - j_n}} \sum_{j=j_n}^{j_n + \log(k_n - j_n)} (j - j_n) \left( \alpha_j^{(0)}(\Theta) + \alpha_j^{(1)}(\Theta) \right) \\
 &\qquad \qquad \qquad + \frac{C}{\sqrt{k_n - j_n}} \sum_{j=j_n + \log(k_n - j_n)}^{k_n} (j - j_n) \left( \alpha_j^{(0)}(\Theta) + \alpha_j^{(1)}(\Theta) \right) \\
 &\qquad \qquad \qquad + C \sum_{j \geq k_n} \sqrt{j} \left( \alpha_j^{(0)}(\Theta) + \alpha_j^{(1)}(\Theta) \right) \\
 &\leq \frac{C \log(k_n - j_n)}{\sqrt{k_n - j_n}} \sum_{j \geq 1} \left( \alpha_j^{(0)}(\Theta) + \alpha_j^{(1)}(\Theta) \right) \\
 &\qquad \qquad \qquad + C \sum_{j \geq j_n + \log(k_n - j_n)} \sqrt{j - j_n} \left( \alpha_j^{(0)}(\Theta) + \alpha_j^{(1)}(\Theta) \right) \\
 &\qquad \qquad \qquad + C \sum_{j \geq k_n} \sqrt{j} \left( \alpha_j^{(0)}(\Theta) + \alpha_j^{(1)}(\Theta) \right) \\
 &\qquad \qquad \qquad \xrightarrow{n \rightarrow +\infty} 0.
 \end{aligned}$$

This holds for any coordinate  $i = 1 \dots, d$ ; and completes the proof of the lemma.



**Lemma 7.2.** Let  $(j_n)_{n \geq 1}$  and  $(k_n)_{n \geq 1}$  two integer valued sequences such that  $(j_n)_{n \geq 1}$  is increasing,  $j_n \rightarrow \infty$  and  $k_n - j_n \rightarrow \infty$  as  $n \rightarrow \infty$ . Let  $n \geq 1$ , for any segment  $T = T_{j_n, k_n} \subset \{1, \dots, n\}$ , it holds under assumptions of Theorem 3.2,

$$(i) \left\| \frac{1}{k_n - j_n} \frac{\partial^2}{\partial \theta \partial \theta'} L_n(T, \theta) - E \left( \frac{\partial^2 \ell_0(\theta)}{\partial \theta \partial \theta'} \right) \right\|_{\Theta} \xrightarrow[n \rightarrow \infty]{a.s.} 0;$$

$$(ii) \left\| \frac{1}{k_n - j_n} \sum_{t \in T} \frac{1}{f_\theta^t} \left( \frac{\partial}{\partial \theta} \hat{f}_\theta^t \right) \left( \frac{\partial}{\partial \theta} \hat{f}_\theta^t \right)' - E \left( \frac{1}{f_\theta^0} \left( \frac{\partial}{\partial \theta} f_\theta^0 \right) \left( \frac{\partial}{\partial \theta} f_\theta^0 \right)' \right) \right\|_{\Theta} \xrightarrow[n \rightarrow \infty]{a.s.} 0.$$

**Proof.**

(i) For  $i, j \in \{1, \dots, d\}$ , we have

$$\begin{aligned} \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ell_t(\theta) &= \frac{\partial}{\partial \theta_j} \left( Y_t \frac{1}{f_\theta^t} \frac{\partial}{\partial \theta_i} f_\theta^t - \frac{\partial}{\partial \theta_i} f_\theta^t \right) \\ &= Y_t \left[ \frac{\partial}{\partial \theta_j} \left( \frac{1}{f_\theta^t} \right) \times \frac{\partial}{\partial \theta_i} f_\theta^t + \frac{1}{f_\theta^t} \times \frac{\partial^2}{\partial \theta_i \partial \theta_j} f_\theta^t \right] - \frac{\partial^2}{\partial \theta_i \partial \theta_j} f_\theta^t \\ &= Y_t \left[ -\frac{1}{(f_\theta^t)^2} \left( \frac{\partial f_\theta^t}{\partial \theta_j} \right) \times \frac{\partial f_\theta^t}{\partial \theta_i} + \frac{1}{f_\theta^t} \times \frac{\partial^2}{\partial \theta_i \partial \theta_j} f_\theta^t \right] - \frac{\partial^2 f_\theta^t}{\partial \theta_i \partial \theta_j} \\ &= -\frac{Y_t}{(f_\theta^t)^2} \left( \frac{\partial}{\partial \theta_i} f_\theta^t \right) \left( \frac{\partial}{\partial \theta_j} f_\theta^t \right) + \left( \frac{Y_t}{f_\theta^t} - 1 \right) \frac{\partial^2}{\partial \theta_i \partial \theta_j} f_\theta^t. \quad (22) \end{aligned}$$

We will show that  $E[\|\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ell_t(\theta)\|] < +\infty$ . From the Hölder's inequality, we have

$$\begin{aligned} E \left[ \left\| \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ell_t(\theta) \right\|_{\Theta} \right] &\leq \frac{1}{c^2} \|Y_t\|_3 \left\| \left\| \frac{\partial}{\partial \theta_i} f_\theta^t \right\|_{\Theta} \right\|_3 \left\| \left\| \frac{\partial}{\partial \theta_j} f_\theta^t \right\|_{\Theta} \right\|_3 \\ &\quad + C(\|Y_t\|_2 + 1) \left\| \left\| \frac{\partial^2}{\partial \theta_i \partial \theta_j} f_\theta^t \right\|_{\Theta} \right\|_2. \end{aligned}$$

But, we have  $\|Y_t\|_3 = \|Y_0\|_3 < \infty$ , and  $\|Y_t\|_2 < \infty$ .

$$\begin{aligned} \left\| \left\| \frac{\partial}{\partial \theta_i} f_\theta^t \right\|_{\Theta} \right\|_3 &\leq \left\| \left\| \frac{\partial}{\partial \theta_i} f_\theta(0) \right\|_{\Theta} \right\|_3 + \left\| \left\| \frac{\partial}{\partial \theta_i} f_\theta^t - \frac{\partial}{\partial \theta_i} f_\theta(0) \right\|_{\Theta} \right\|_3 \\ &\leq \left\| \left\| \frac{\partial}{\partial \theta_i} f_\theta(0) \right\|_{\Theta} \right\|_3 + \left\| \left\| \sum_{j \geq 1} \alpha_j^{(1)}(\Theta) |Y_{t-j}| \right\|_{\Theta} \right\|_3 \\ &\leq \left\| \left\| \frac{\partial}{\partial \theta_i} f_\theta(0) \right\|_{\Theta} \right\|_3 + \|Y_0\|_3 \sum_{j \geq 1} \alpha_j^{(1)}(\Theta) < +\infty. \end{aligned}$$

Similarly, we have  $\left\| \left\| \frac{\partial}{\partial \theta_j} f_\theta^t \right\|_{\Theta} \right\|_3 < +\infty$ . Using the same argument, we obtain

$$\left\| \left\| \frac{\partial^2}{\partial \theta_i \partial \theta_j} f_\theta^t \right\|_{\Theta} \right\|_2 \leq \|Y_0\|_2 \sum_{j \geq 1} \alpha_j^{(2)}(\Theta) < +\infty.$$

Hence,  $E[\|\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\theta)\|_{\Theta}] < +\infty$ . Thus, for the stationary ergodicity of the sequence  $(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\theta))_{t\in\mathbb{Z}}$  and the uniform strong law of large numbers, it holds that

$$\left\| \frac{1}{k_n - j_n} \frac{\partial^2}{\partial\theta_i\partial\theta_j} \ell_t(\theta) - E \frac{\partial^2}{\partial\theta_i\partial\theta_j} \ell_0(\theta) \right\|_{\Theta} \xrightarrow[n \rightarrow +\infty]{a.s.} 0.$$

This completes the proof of (i).

(ii) Goes the same lines as in (i) and as in Lemma 7.1. ■

**Proof of Theorem 3.2.** Here again, without loss of generality, we will make the proof with  $T_{j_n, k_n} = T_{1, n}$ . Recall that  $\Theta \subset \mathbb{R}^d$ . Let  $T \subset \{1, \dots, n\}$ ; for any  $\bar{\theta} \in \Theta$  and  $i = 1, \dots, n$ , by applying the Taylor expansion to the function  $\theta \mapsto \frac{\partial}{\partial\theta} L_n(T, \theta)$ , there exists  $\theta_{n,i}$  between  $\bar{\theta}$  and  $\theta_0$  such that

$$\frac{\partial}{\partial\theta_i} L_n(T, \bar{\theta}) = \frac{\partial}{\partial\theta_i} L_n(T, \theta_0) + \frac{\partial^2}{\partial\theta\partial\theta_i} L_n(T, \theta_{n,i}) \cdot (\bar{\theta} - \theta_0).$$

Denote

$$G_n(T, \bar{\theta}) = -\frac{1}{Card(T)} \left( \frac{\partial^2}{\partial\theta\partial\theta_i} L_n(T, \theta_{n,i}) \right)_{1 \leq i \leq d}.$$

It comes that

$$Card(T)G_n(T, \bar{\theta}) \cdot (\bar{\theta} - \theta_0) = \frac{\partial}{\partial\theta} L_n(T, \theta_0) - \frac{\partial}{\partial\theta} L_n(T, \bar{\theta}). \tag{23}$$

By applying (23) with  $\bar{\theta} = \hat{\theta}_n(T)$  we obtain

$$Card(T)G_n(T, \hat{\theta}_n(T)) \cdot (\hat{\theta}_n(T) - \theta_0) = \frac{\partial}{\partial\theta} L_n(T, \theta_0) - \frac{\partial}{\partial\theta} L_n(T, \hat{\theta}_n(T)). \tag{24}$$

(24) holds for any  $T \subset \{1, \dots, n\}$ , thus

$$\begin{aligned} \sqrt{n}G_n(T_{1,n}, \hat{\theta}_n(T_{1,n})) \cdot (\hat{\theta}_n(T_{1,n}) - \theta_0) &= \frac{1}{\sqrt{n}} \left( \frac{\partial}{\partial\theta} L_n(T_{1,n}, \theta_0) \right. \\ &\quad \left. - \frac{\partial}{\partial\theta} L_n(T_{1,n}, \hat{\theta}_n(T_{1,n})) \right). \end{aligned} \tag{25}$$

We can rewrite (25) as

$$\begin{aligned} \sqrt{n}G_n(T_{1,n}, \hat{\theta}_n(T_{1,n})) (\hat{\theta}_n(T_{1,n}) - \theta_0) &= \frac{1}{\sqrt{n}} \frac{\partial}{\partial\theta} L_n(T_{1,n}, \theta_0) \\ &- \frac{1}{\sqrt{n}} \frac{\partial \hat{L}_n(T_{1,n}, \hat{\theta}_n(T_{1,n}))}{\partial\theta} + \frac{1}{\sqrt{n}} \left( \frac{\partial \hat{L}_n(T_{1,n}, \hat{\theta}_n(T_{1,n}))}{\partial\theta} - \frac{\partial L_n(T_{1,n}, \hat{\theta}_n(T_{1,n}))}{\partial\theta} \right). \end{aligned}$$

For  $n$  large enough,  $\frac{\partial}{\partial\theta} \hat{L}_n(T_{1,n}, \hat{\theta}_n(T_{1,n})) = 0$ , because  $\hat{\theta}_n(T_{1,n})$  is a local maximum of  $\theta \mapsto \hat{L}(T_{1,n}, \theta)$ .

Moreover, according to Lemma 7.1, it holds that

$$E \left( \frac{1}{\sqrt{n}} \left\| \frac{\partial}{\partial \theta} L_n(T_{1,n}, \hat{\theta}_n(T_{1,n})) - \frac{\partial}{\partial \theta} \hat{L}_n(T_{1,n}, \hat{\theta}_n(T_{1,n})) \right\| \right) \leq E \left( \frac{1}{\sqrt{n}} \left\| \frac{\partial}{\partial \theta} L_n(T_{1,n}, \theta) - \frac{\partial}{\partial \theta} \hat{L}_n(T_{1,n}, \theta) \right\|_{\Theta} \right) \xrightarrow{n \rightarrow +\infty} 0.$$

So, for  $n$  large enough, we have

$$\sqrt{n} G_n(T_{1,n}, \hat{\theta}_n(T_{1,n})) (\hat{\theta}_n(T_{1,n}) - \theta_0) = \frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} L_n(T_{1,n}, \theta_0) + o_P(1). \tag{26}$$

To complete the proof of Theorem 3.2, we have to show that

- (a)  $(\frac{\partial}{\partial \theta} \ell_t(\theta_0), \mathcal{F}_t)_{t \in \mathbb{Z}}$  is a stationary ergodic martingale difference sequence and  $E(\|\frac{\partial}{\partial \theta} \ell_t(\theta_0)\|^2) < \infty$ ;
- (b)  $\Sigma = -E(\frac{\partial^2}{\partial \theta \partial \theta'} \ell_0(\theta_0))$  and  $G_n(T_{1,n}, \hat{\theta}_n(T_{1,n})) \xrightarrow[n \rightarrow +\infty]{a.s.} \Sigma$ ;
- (c)  $\Sigma = E[\frac{1}{f_{\theta_0}^0} (\frac{\partial}{\partial \theta} f_{\theta_0}^0) (\frac{\partial}{\partial \theta} f_{\theta_0}^0)']$  is invertible.

- (a) Recall that  $\frac{\partial}{\partial \theta} \ell_t(\theta_0) = (\frac{Y_t}{f_{\theta_0}^t} - 1) \frac{\partial}{\partial \theta} f_{\theta_0}^t$  and  $\mathcal{F}_t = \sigma(Y_s, s \leq t)$ . Since the functions  $f_{\theta_0}^t$  and  $\frac{\partial}{\partial \theta} f_{\theta_0}^t$  are  $\mathcal{F}_{t-1}$ -measurable, we have

$$E \left( \frac{\partial}{\partial \theta} \ell_t(\theta_0) | \mathcal{F}_{t-1} \right) = \left( \frac{1}{f_{\theta_0}^t} E(Y_t | \mathcal{F}_{t-1}) - 1 \right) \frac{\partial}{\partial \theta} f_{\theta_0}^t = 0.$$

Moreover, since  $|Y_t|$  and  $\|\frac{\partial}{\partial \theta} f_{\theta_0}^t\|$  have moment of any order, we have

$$E \left( \left| \frac{\partial}{\partial \theta} \ell_t(\theta_0) \right|^2 \right) \leq E \left( \left( \frac{|Y_t|}{\underline{c}} + 1 \right)^2 \left\| \frac{\partial}{\partial \theta} f_{\theta_0}^t \right\|^2 \right) < \infty.$$

- (b) According to (22), we have

$$\frac{\partial^2}{\partial \theta \partial \theta'} \ell_t(\theta) = -\frac{Y_t}{(f_{\theta_0}^t)^2} \left( \frac{\partial}{\partial \theta} f_{\theta_0}^t \right) \left( \frac{\partial}{\partial \theta} f_{\theta_0}^t \right)' + \left( \frac{Y_t}{f_{\theta_0}^t} - 1 \right) \frac{\partial^2}{\partial \theta \partial \theta'} f_{\theta_0}^t. \tag{27}$$

But by using the same argument as in (a), we obtain

$$E \left( \left( \frac{Y_t}{f_{\theta_0}^0} - 1 \right) \frac{\partial^2}{\partial \theta \partial \theta'} f_{\theta_0}^0 | \mathcal{F}_{t-1} \right) = 0.$$

Hence, (27) implies

$$\begin{aligned} E \left( \frac{\partial^2}{\partial \theta \partial \theta'} \ell_t(\theta_0) \right) &= -E \left( \frac{Y_t}{(f_{\theta_0}^0)^2} \left( \frac{\partial}{\partial \theta} f_{\theta_0}^0 \right) \left( \frac{\partial}{\partial \theta} f_{\theta_0}^0 \right)' \right) \\ &= -E \left( \frac{1}{f_{\theta_0}^0} \left( \frac{\partial}{\partial \theta} f_{\theta_0}^0 \right) \left( \frac{\partial}{\partial \theta} f_{\theta_0}^0 \right)' \right) = -\Sigma. \end{aligned}$$

Moreover, recall that

$$\begin{aligned} G_n(T_{1,n}, \hat{\theta}_n(T_{1,n})) &= -\frac{1}{n} \left( \frac{\partial^2}{\partial\theta\partial\theta_i} L(T_{1,n}, \theta_{n,i}) \right)_{1 \leq i \leq d} \\ &= -\frac{1}{n} \left( \sum_{t=1}^n \frac{\partial^2}{\partial\theta\partial\theta_i} \ell_t(\theta_{n,i}) \right)_{1 \leq i \leq d}. \end{aligned}$$

For any  $j = 1, \dots, d$ , we have

$$\begin{aligned} & \left| \frac{1}{n} \sum_{t=1}^n \frac{\partial^2}{\partial\theta_j\partial\theta_i} \ell_t(\theta_{n,i}) - E \left( \frac{\partial^2}{\partial\theta_j\partial\theta_i} \ell_0(\theta_0) \right) \right| \\ & \leq \left| \frac{1}{n} \sum_{t=1}^n \frac{\partial^2}{\partial\theta_j\partial\theta_i} \ell_t(\theta_{n,i}) - E \left( \frac{\partial^2}{\partial\theta_j\partial\theta_i} \ell_0(\theta_{n,i}) \right) \right| \\ & \quad + \left| E \left( \frac{\partial^2}{\partial\theta_j\partial\theta_i} \ell_0(\theta_{n,i}) \right) - E \left( \frac{\partial^2}{\partial\theta_j\partial\theta_i} \ell_0(\theta_0) \right) \right| \\ & \leq \left| E \left( \frac{\partial^2}{\partial\theta_j\partial\theta_i} \ell_0(\theta_{n,i}) \right) - E \left( \frac{\partial^2}{\partial\theta_j\partial\theta_i} \ell_0(\theta_0) \right) \right| \\ & \quad + \left\| \frac{1}{n} \sum_{t=1}^n \frac{\partial^2}{\partial\theta_j\partial\theta_i} \ell_t(\theta) - E \left( \frac{\partial^2}{\partial\theta_j\partial\theta_i} \ell_0(\theta) \right) \right\|_{\Theta} \xrightarrow[n \rightarrow +\infty]{a.s.} 0. \end{aligned}$$

This holds for any  $1 \leq i, j \leq d$ . Thus,

$$\begin{aligned} G_n(T_{1,n}, \hat{\theta}_n(T_{1,n})) &= -\frac{1}{n} \left( \frac{\partial^2}{\partial\theta\partial\theta_i} \ell_t(\theta_{n,i}) \right)_{1 \leq i \leq d} \\ &\xrightarrow[n \rightarrow +\infty]{a.s.} -E \left( \frac{\partial^2}{\partial\theta\partial\theta'} \ell_0(\theta_0) \right) = E \left[ \frac{1}{f_{\theta_0}^0} \left( \frac{\partial}{\partial\theta} f_{\theta_0}^0 \right) \left( \frac{\partial}{\partial\theta} f_{\theta_0}^0 \right)' \right] = \Sigma. \end{aligned}$$

- (c) If  $U$  is a non-zero vector of  $\mathbb{R}^d$ , according to assumption Var, it holds that  $U \frac{\partial}{\partial\theta} f_{\theta_0}^0 \neq 0$  a.s. Hence

$$U \Sigma U' = E \left( \frac{1}{f_{\theta_0}^0} U \left( \frac{\partial}{\partial\theta} f_{\theta_0}^0 \right) \left( \frac{\partial}{\partial\theta} f_{\theta_0}^0 \right)' U' \right) > 0.$$

Thus  $\Sigma$  is positive definite.

From (a), apply the central limit theorem for stationary ergodic martingale difference sequence, it follows that

$$\begin{aligned} \frac{1}{\sqrt{n}} \frac{\partial}{\partial\theta} L_n(T_{1,n}, \theta_0) &= \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial}{\partial\theta} \ell_t(\theta_0) \\ &\xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N} \left( 0, E \left[ \left( \frac{\partial}{\partial\theta} \ell_0(\theta_0) \right) \left( \frac{\partial}{\partial\theta} \ell_0(\theta_0) \right)' \right] \right). \quad (28) \end{aligned}$$

Recall that for  $i = 1, \dots, d$ ,  $\frac{\partial}{\partial \theta_i} \ell_t(\theta) = \left(\frac{Y_t}{f_{\theta_0}^t} - 1\right) \frac{\partial}{\partial \theta_i} f_{\theta_0}^t$ . For  $1 \leq i, j \leq d$ , we have

$$\begin{aligned} E\left(\frac{\partial}{\partial \theta_i} \ell_t(\theta_0) \times \frac{\partial}{\partial \theta_j} \ell_t(\theta_0)\right) &= E\left[E\left(\left(\frac{Y_t}{f_{\theta_0}^t} - 1\right)^2 \frac{\partial}{\partial \theta_i} f_{\theta_0}^t \times \frac{\partial}{\partial \theta_j} f_{\theta_0}^t \middle| \mathcal{F}_{t-1}\right)\right] \\ &= E\left[E\left(\left(\frac{Y_t}{f_{\theta_0}^t} - 1\right)^2 \middle| \mathcal{F}_{t-1}\right) \times \frac{\partial}{\partial \theta_i} f_{\theta_0}^t \times \frac{\partial}{\partial \theta_j} f_{\theta_0}^t\right]. \end{aligned}$$

We have,

$$\begin{aligned} E\left[\left(\frac{Y_t}{f_{\theta_0}^t} - 1\right)^2 \middle| \mathcal{F}_{t-1}\right] &= \frac{1}{(f_{\theta_0}^t)^2} E(Y_t^2 | \mathcal{F}_{t-1}) - \frac{2}{f_{\theta_0}^t} \times f_{\theta_0}^t + 1 \\ &= \frac{1}{(f_{\theta_0}^t)^2} E(Y_t^2 | \mathcal{F}_{t-1}) - 1 = \frac{1}{(f_{\theta_0}^t)^2} \left(\text{Var}(Y_t | \mathcal{F}_{t-1}) + (E(Y_t | \mathcal{F}_{t-1}))^2\right) - 1 \\ &= \frac{1}{(f_{\theta_0}^t)^2} (f_{\theta_0}^t + (f_{\theta_0}^t)^2) - 1 = \frac{1}{f_{\theta_0}^t}. \end{aligned}$$

Thus,

$$E\left[\left(\frac{\partial}{\partial \theta_i} \ell_t(\theta_0)\right) \times \left(\frac{\partial}{\partial \theta_j} \ell_t(\theta_0)\right)'\right] = E\left[\frac{1}{f_{\theta_0}^t} \left(\frac{\partial}{\partial \theta_i} f_{\theta_0}^t\right) \times \left(\frac{\partial}{\partial \theta_j} f_{\theta_0}^t\right)'\right].$$

Hence,

$$E\left[\left(\frac{\partial}{\partial \theta} \ell_t(\theta_0)\right) \times \left(\frac{\partial}{\partial \theta} \ell_t(\theta_0)\right)'\right] = E\left[\frac{1}{f_{\theta_0}^t} \left(\frac{\partial}{\partial \theta} f_{\theta_0}^t\right) \times \left(\frac{\partial}{\partial \theta} f_{\theta_0}^t\right)'\right] = \Sigma.$$

Thus, (28) becomes

$$\frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} L_n(T_{1,n}, \theta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial}{\partial \theta} \ell_t(\theta_0) \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N}(0, \Sigma). \quad (29)$$

(b) and (c) implies that the matrix  $G_n(T_{1,n}, \widehat{\theta}_n(T_{1,n}))$  converges a.s. to  $\Sigma$  and  $G_n(T_{1,n}, \widehat{\theta}_n(T_{1,n}))$  is invertible for  $n$  large enough. Hence, from (26) and (29), we have

$$\begin{aligned} \sqrt{n}(\widehat{\theta}_n(T_{1,n}) - \theta_0) &= \frac{1}{\sqrt{n}} \left(G_n(T_{1,n}, \widehat{\theta}_n(T_{1,n}))\right)^{-1} \frac{\partial}{\partial \theta} L_n(T_{1,n}, \theta_0) + o_P(1) \\ &= \frac{1}{\sqrt{n}} \Sigma^{-1} \frac{\partial}{\partial \theta} L_n(T_{1,n}, \theta_0) + o_P(1) \\ &\xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N}(0, \Sigma^{-1}). \end{aligned}$$

■

Before proving the Theorem 4.1, let us prove some preliminary lemma. Under  $H_0$ , recall

$$\Sigma = E\left[\frac{1}{f_{\theta_0}^0} \left(\frac{\partial}{\partial \theta} f_{\theta_0}^0\right) \left(\frac{\partial}{\partial \theta} f_{\theta_0}^0\right)'\right] = E\left[\left(\frac{\partial}{\partial \theta} \ell_0(\theta_0)\right) \left(\frac{\partial}{\partial \theta} \ell_0(\theta_0)\right)'\right].$$

Define the statistics

- $C_n = \max_{v_n \leq k \leq n - v_n} C_{n,k}$  where

$$C_{n,k} = \frac{1}{q^2 \left(\frac{k}{n}\right)} \frac{k^2(n-k)^2}{n^3} \left( \hat{\theta}_n(T_{1,k}) - \hat{\theta}_n(T_{k+1,n}) \right)' \Sigma \left( \hat{\theta}_n(T_{1,k}) - \hat{\theta}_n(T_{k+1,n}) \right)$$

- $Q_n^{(1)} = \max_{v_n \leq k \leq n-v_n} Q_{n,k}^{(1)}$  where

$$Q_{n,k}^{(1)} = \frac{k^2}{n} \left( \hat{\theta}_n(T_{1,k}) - \hat{\theta}_n(T_{1,n}) \right)' \Sigma \left( \hat{\theta}_n(T_{1,k}) - \hat{\theta}_n(T_{1,n}) \right);$$

- $Q_n^{(2)} = \max_{v_n \leq k \leq n-v_n} Q_{n,k}^{(2)}$  where

$$Q_{n,k}^{(2)} = \frac{(n-k)^2}{n} \left( \hat{\theta}_n(T_{k+1,n}) - \hat{\theta}_n(T_{1,n}) \right)' \Sigma \left( \hat{\theta}_n(T_{k+1,n}) - \hat{\theta}_n(T_{1,n}) \right).$$

**Lemma 7.3.** Under assumptions of Theorem 4.1, as  $n \rightarrow +\infty$ ,

- (i)  $\max_{v_n \leq k \leq n-v_n} |\hat{C}_{n,k} - C_{n,k}| = o_P(1)$ ;
- (ii) for  $j = 1, 2$ ,  $\max_{v_n \leq k \leq n-v_n} |\hat{Q}_{n,k}^{(j)} - Q_{n,k}^{(j)}| = o_P(1)$ .

**Proof.**

- (i) For any  $v_n \leq k \leq n - v_n$ , we have as  $n \rightarrow \infty$

$$\begin{aligned} \left| \hat{C}_{n,k} - C_{n,k} \right| &= \frac{1}{q^2 \left(\frac{k}{n}\right)} \frac{k^2(n-k)^2}{n^3} \\ &\times \left| \left( \hat{\theta}_n(T_{1,k}) - \hat{\theta}_n(T_{k+1,n}) \right)' \left( \hat{\Sigma}_n(u_n) - \Sigma \right) \left( \hat{\theta}_n(T_{1,k}) - \hat{\theta}_n(T_{k+1,n}) \right) \right| \\ &\leq \frac{1}{q^2 \left(\frac{k}{n}\right)} \frac{k^2(n-k)^2}{n^3} \left\| \hat{\Sigma}_n(u_n) - \Sigma \right\| \left\| \hat{\theta}_n(T_{1,k}) - \hat{\theta}_n(T_{k+1,n}) \right\|^2 \\ &\leq C \frac{1}{q^2 \left(\frac{k}{n}\right)} \frac{k(n-k)}{n^2} \left\| \hat{\Sigma}_n(u_n) - \Sigma \right\| \left( \left\| \sqrt{k}(\hat{\theta}_n(T_{1,k}) - \theta_0) \right\|^2 \right. \\ &\quad \left. + \left\| \sqrt{n-k}(\hat{\theta}_n(T_{k+1,n}) - \theta_0) \right\|^2 \right) \\ &\leq C \frac{1}{q^2 \left(\frac{k}{n}\right)} \frac{k(n-k)}{n^2} o(1) O_P(1). \end{aligned}$$

Thus, as  $n \rightarrow \infty$ , it holds that

$$\begin{aligned} \max_{v_n \leq k \leq n-v_n} \left| \hat{C}_{n,k} - C_{n,k} \right| &\leq o_P(1) \max_{v_n \leq k \leq n-v_n} \frac{1}{q^2 \left(\frac{k}{n}\right)} \frac{k(n-k)}{n^2} \\ &\leq o_P(1) \max_{\frac{v_n}{n} \leq \frac{k}{n} \leq 1 - \frac{v_n}{n}} \frac{1}{q^2 \left(\frac{k}{n}\right)} \frac{k}{n} \left(1 - \frac{k}{n}\right) \\ &\leq o_P(1) \sup_{0 < \tau < 1} \left( \frac{\sqrt{\tau(1-\tau)}}{q(\tau)} \right)^2 = o_P(1). \end{aligned}$$

The last equality above holds because  $\sup_{0 < \tau < 1} \frac{\sqrt{\tau(1-\tau)}}{q(\tau)} < \infty$ ; it is a consequence of the properties of the function  $q$  when  $I_{0,1}(q, c)$  is finite for some  $c > 0$ .

(ii) Goes the same lines as in (i). ■

### Proof of Theorem 4.1.

1. According to Lemma 7.3, it suffices to show that

$$C_n \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \sup_{0 < \tau < 1} \frac{\|W_d(\tau)\|^2}{q^2(\tau)}.$$

Let  $v_n \leq k \leq n - v_n$ . By applying (24) with  $T = T_{1,k}$  and  $T_{k+1,n}$ , we have

$$\begin{aligned} G_n(T_{1,k}, \hat{\theta}_n(T_{1,k}))(\hat{\theta}_n(T_{1,k}) - \theta_0) \\ = \frac{1}{k} \left( \frac{\partial}{\partial \theta} L_n(T_{1,k}, \theta_0) - \frac{\partial}{\partial \theta} L_n(T_{1,k}, \hat{\theta}_n(T_{1,k})) \right) \end{aligned}$$

and

$$\begin{aligned} G_n(T_{k+1,n}, \hat{\theta}_n(T_{k+1,n}))(\hat{\theta}_n(T_{k+1,n}) - \theta_0) \\ = \frac{1}{n-k} \left( \frac{\partial}{\partial \theta} L_n(T_{k+1,n}, \theta_0) - \frac{\partial}{\partial \theta} L_n(T_{k+1,n}, \hat{\theta}_n(T_{k+1,n})) \right). \end{aligned}$$

As  $n \rightarrow +\infty$ , we have

$$\begin{aligned} \left\| G_n(T_{1,k}, \hat{\theta}_n(T_{1,k})) - \Sigma \right\| &= \left\| G_n(T_{k+1,n}, \hat{\theta}_n(T_{k+1,n})) - \Sigma \right\| = o(1) \\ \sqrt{k}(\hat{\theta}_n(T_{1,k}) - \theta_0) &= O_P(1) \quad \text{and} \quad \sqrt{n-k}(\hat{\theta}_n(T_{k+1,n}) - \theta_0) = O_P(1). \end{aligned}$$

Thus, we have

$$\begin{aligned} \sqrt{k} \Sigma(\hat{\theta}_n(T_{1,k}) - \theta_0) &= \frac{1}{\sqrt{k}} \left( \frac{\partial}{\partial \theta} L_n(T_{1,k}, \theta_0) - \frac{\partial}{\partial \theta} L_n(T_{1,k}, \hat{\theta}_n(T_{1,k})) \right) \\ &\quad - \sqrt{k}(G_n(T_{1,k}, \hat{\theta}_n(T_{1,k})) - \Sigma)(\hat{\theta}_n(T_{1,k}) - \theta_0) \\ &= \frac{1}{\sqrt{k}} \left( \frac{\partial}{\partial \theta} L_n(T_{1,k}, \theta_0) - \frac{\partial}{\partial \theta} L_n(T_{1,k}, \hat{\theta}_n(T_{1,k})) \right) + o_P(1) \end{aligned}$$

i.e.

$$\begin{aligned} \Sigma(\hat{\theta}_n(T_{1,k}) - \theta_0) &= \frac{1}{k} \left( \frac{\partial}{\partial \theta} L_n(T_{1,k}, \theta_0) - \frac{\partial}{\partial \theta} L_n(T_{1,k}, \hat{\theta}_n(T_{1,k})) \right) \\ &\quad + o_P\left(\frac{1}{\sqrt{k}}\right). \end{aligned}$$

Moreover we have (as  $n \rightarrow +\infty$ )

$$\begin{aligned} & \frac{1}{\sqrt{k}} \left\| \frac{\partial}{\partial \theta} L_n(T_{1,k}, \hat{\theta}_n(T_{1,k})) \right\| = \frac{1}{\sqrt{k}} \left\| \frac{\partial}{\partial \theta} L_n(T_{1,k}, \hat{\theta}_n(T_{1,k})) - 0 \right\| \\ & \leq \frac{1}{\sqrt{k}} \left\| \frac{\partial}{\partial \theta} L_n(T_{1,k}, \hat{\theta}_n(T_{1,k})) - \frac{\partial}{\partial \theta} \hat{L}_n(T_{1,k}, \hat{\theta}_n(T_{1,k})) \right\| \\ & \leq \frac{1}{\sqrt{k}} \left\| \frac{\partial}{\partial \theta} L_n(T_{1,k}, \theta) - \frac{\partial}{\partial \theta} \hat{L}_n(T_{1,k}, \theta) \right\|_{\Theta} = o_P(1). \end{aligned}$$

Hence

$$\begin{aligned} \Sigma(\hat{\theta}_n(T_{1,k}) - \theta_0) &= \frac{1}{\sqrt{k}} \left( \frac{1}{k} \frac{\partial}{\partial \theta} L_n(T_{1,k}, \theta_0) - \frac{1}{\sqrt{k}} \frac{\partial}{\partial \theta} L_n(T_{1,k}, \hat{\theta}_n(T_{1,k})) \right) \\ & \qquad \qquad \qquad + o_P \left( \frac{1}{\sqrt{k}} \right) \\ &= \frac{1}{k} \frac{\partial}{\partial \theta} L_n(T_{1,k}, \theta_0) + \frac{1}{\sqrt{k}} o_P(1) + o_p \left( \frac{1}{\sqrt{k}} \right) \\ &= \frac{1}{k} \frac{\partial}{\partial \theta} L_n(T_{1,k}, \theta_0) + o_P \left( \frac{1}{\sqrt{k}} \right). \end{aligned}$$

Thus, we have

$$\Sigma(\hat{\theta}_n(T_{1,k}) - \theta_0) = \frac{1}{k} \frac{\partial}{\partial \theta} L_n(T_{1,k}, \theta_0) + o_P \left( \frac{1}{\sqrt{k}} \right).$$

Using the same ideas we also get:

$$\Sigma(\hat{\theta}_n(T_{k+1,n}) - \theta_0) = \frac{1}{n-k} \frac{\partial}{\partial \theta} L_n(T_{k+1,n}, \theta_0) + o_P \left( \frac{1}{\sqrt{n-k}} \right).$$

By subtracting the two above equalities, it follows that

$$\begin{aligned} & \Sigma(\hat{\theta}_n(T_{1,k}) - \hat{\theta}_n(T_{k+1,n})) \\ &= \frac{1}{k} \frac{\partial}{\partial \theta} L_n(T_{1,k}, \theta_0) - \frac{1}{n-k} \frac{\partial}{\partial \theta} L_n(T_{k+1,n}, \theta_0) + o_P \left( \frac{1}{\sqrt{k}} + \frac{1}{\sqrt{n-k}} \right) \\ &= \frac{1}{k} \frac{\partial}{\partial \theta} L_n(T_{1,k}, \theta_0) - \frac{1}{n-k} \left( \frac{\partial}{\partial \theta} L_n(T_{1,n}, \theta_0) - \frac{\partial}{\partial \theta} L_n(T_{1,k}, \theta_0) \right) \\ & \qquad \qquad \qquad + o_P \left( \frac{1}{\sqrt{k}} + \frac{1}{\sqrt{n-k}} \right) \\ &= \frac{n}{k(n-k)} \left( \frac{\partial}{\partial \theta} L_n(T_{1,k}, \theta_0) - \frac{n}{k} \frac{\partial}{\partial \theta} L_n(T_{1,n}, \theta_0) \right) + o_P \left( \frac{1}{\sqrt{k}} + \frac{1}{\sqrt{n-k}} \right) \end{aligned}$$

i.e.

$$\frac{k(n-k)}{n^{\frac{3}{2}}} \Sigma(\hat{\theta}_n(T_{1,k}) - \hat{\theta}_n(T_{k+1,n}))$$

$$\begin{aligned}
&= \frac{1}{\sqrt{n}} \left( \frac{\partial}{\partial \theta} L_n(T_{1,k}, \theta_0) - \frac{k}{n} \frac{\partial}{\partial \theta} L_n(T_{1,n}, \theta_0) \right) \\
&\quad + o_P \left( \frac{\sqrt{k(n-k)} \sqrt{k} + \sqrt{n-k}}{\sqrt{n}} \right) \\
&= \frac{1}{\sqrt{n}} \left( \frac{\partial}{\partial \theta} L_n(T_{1,k}, \theta_0) - \frac{k}{n} \frac{\partial}{\partial \theta} L_n(T_{1,n}, \theta_0) \right) + o_P(1).
\end{aligned}$$

Thus

$$\begin{aligned}
&\frac{k(n-k)}{n^{\frac{3}{2}}} \Sigma^{-1/2} \Sigma (\hat{\theta}_n(T_{1,k}) - \hat{\theta}_n(T_{k+1,n})) \\
&= \frac{\Sigma^{-1/2}}{\sqrt{n}} \left( \frac{\partial}{\partial \theta} L_n(T_{1,k}, \theta_0) - \frac{k}{n} \frac{\partial}{\partial \theta} L_n(T_{1,n}, \theta_0) \right) + o_P(1). \quad (30)
\end{aligned}$$

For  $0 < \tau < 1$ , we have

$$\frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} L_n(T_{1,[n\tau]}, \theta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[n\tau]} \frac{\partial}{\partial \theta} \ell_t(\theta_0).$$

We have shown (see the proof of Theorem 3.2) that  $(\frac{\partial}{\partial \theta} \ell_t(\theta_0), \mathcal{F}_t)_{t \in \mathbb{Z}}$  is a stationary ergodic square integrable martingale difference process with covariance matrix  $\Sigma$ . By the Central limit theorem for the martingale difference sequence (see Billingsley, 1968 [4]), it holds that

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \left( \frac{\partial}{\partial \theta} L_n(T_{1,[n\tau]}, \theta_0) - \frac{[n\tau]}{n} \frac{\partial}{\partial \theta} L_n(T_{1,n}, \theta_0) \right) \\
&= \frac{1}{\sqrt{n}} \left( \sum_{t=1}^{[n\tau]} \frac{\partial}{\partial \theta} \ell_t(\theta_0) - \frac{[n\tau]}{n} \sum_{t=1}^n \frac{\partial}{\partial \theta} \ell_t(\theta_0) \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} B_\Sigma(\tau) - \tau B_\Sigma(1)
\end{aligned}$$

where  $B_\Sigma$  is a Gaussian process with covariance matrix  $\min(s, \tau)\Sigma$ . Thus it follows that

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \Sigma^{-1/2} \left( \frac{\partial}{\partial \theta} L_n(T_{1,[n\tau]}, \theta_0) - \frac{[n\tau]}{n} \frac{\partial}{\partial \theta} L_n(T_{1,n}, \theta_0) \right) \\
&\quad \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} B_d(\tau) - \tau B_d(1) = W_d(\tau)
\end{aligned}$$

in  $D([0, 1])$ , where  $B_d$  is a  $d$ -dimensional standard Brownian motion, and  $W_d$  is a  $d$ -dimensional Brownian bridge.

So, we have (see (30))

$$\begin{aligned}
C_{n,[n\tau]} &= \frac{[n\tau]^2 (n - [n\tau])^2}{n^3} \\
&\quad \times (\hat{\theta}_n(T_{1,[n\tau]}) - \hat{\theta}_n(T_{[n\tau]+1,n}))' \Sigma (\hat{\theta}_n(T_{1,[n\tau]}) - \hat{\theta}_n(T_{[n\tau]+1,n}))
\end{aligned}$$

$$\begin{aligned}
 &= \left\| \frac{[n\tau](n - [n\tau])}{n^{3/2}} \Sigma^{-1/2} \Sigma (\hat{\theta}_n(T_{1,[n\tau]}) - \hat{\theta}_n(T_{[n\tau]+1,n})) \right\|^2 \\
 &= \left\| \frac{1}{\sqrt{n}} \Sigma^{-1/2} \left( \sum_{t=1}^{[n\tau]} \frac{\partial}{\partial \theta} \ell_t(\theta_0) - \frac{[n\tau]}{n} \sum_{t=1}^n \frac{\partial}{\partial \theta} \ell_t(\theta_0) \right) \right\|^2 + o_P(1) \\
 &\xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \|W_d(\tau)\|^2 \text{ in } \mathcal{D}([0, 1]).
 \end{aligned}$$

Hence, according to the properties of  $q$ , we have for any  $0 < \epsilon < 1/2$

$$\begin{aligned}
 &\max_{[n\epsilon] \leq k \leq n - [n\epsilon]} C_{n,k} \\
 &= \max_{[n\epsilon] \leq k \leq n - [n\epsilon]} \left( \frac{1}{q^2\left(\frac{k}{n}\right)} \frac{k^2(n-k)^2}{n^3} \right. \\
 &\quad \left. \times (\hat{\theta}_n(T_{1,k}) - \hat{\theta}_n(T_{k+1,n}))' \Sigma (\hat{\theta}_n(T_{1,k}) - \hat{\theta}_n(T_{k+1,n})) \right) \\
 &= \sup_{\epsilon \leq \tau \leq 1 - \epsilon} \left( \frac{1}{q^2\left(\frac{[n\tau]}{n}\right)} \frac{[n\tau]^2(n - [n\tau])^2}{n^3} \right. \\
 &\quad \left. \times (\hat{\theta}_n(T_{1,[n\tau]}) - \hat{\theta}_n(T_{[n\tau]+1,n}))' \Sigma (\hat{\theta}_n(T_{1,[n\tau]}) - \hat{\theta}_n(T_{[n\tau]+1,n})) \right) \\
 &= \sup_{\epsilon \leq \tau \leq 1 - \epsilon} \left\| \frac{1}{q\left(\frac{[n\tau]}{n}\right)} \frac{[n\tau](n - [n\tau])}{n^{3/2}} \Sigma^{1/2} (\hat{\theta}_n(T_{1,[n\tau]}) - \hat{\theta}_n(T_{[n\tau]+1,n})) \right\|^2 \\
 &= \sup_{\epsilon \leq \tau \leq 1 - \epsilon} \left\| \frac{1}{q\left(\frac{[n\tau]}{n}\right)} \frac{1}{\sqrt{n}} \left( \sum_{t=1}^{[n\tau]} \frac{\partial}{\partial \theta} \ell_t(\theta_0) - \frac{[n\tau]}{n} \sum_{t=1}^n \frac{\partial}{\partial \theta} \ell_t(\theta_0) \right) \right\|^2 + o_P(1) \\
 &\xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \sup_{\epsilon \leq \tau \leq 1 - \epsilon} \frac{\|W_d(\tau)\|^2}{q^2(\tau)}.
 \end{aligned}$$

Therefore, we have shown that

$$C_{n,[n\tau]} \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \frac{\|W_d(\tau)\|^2}{q^2(\tau)} \text{ in } \mathcal{D}([0, 1])$$

and that for all  $\epsilon \in (0, 1/2)$ ,

$$\max_{[n\epsilon] \leq k \leq n - [n\epsilon]} C_{n,k} = \sup_{\epsilon \leq \tau \leq 1 - \epsilon} C_{n,[n\tau]} \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \sup_{\epsilon \leq \tau \leq 1 - \epsilon} \frac{\|W_d(\tau)\|^2}{q^2(\tau)}.$$

Moreover, since  $I_{0,1}(q, c) < +\infty$  for some  $c > 0$ , one can show that  $\lim_{\tau \rightarrow 0} \frac{\|W_d(\tau)\|}{q(\tau)} < \infty$  and  $\lim_{\tau \rightarrow 1} \frac{\|W_d(\tau)\|}{q(\tau)} < \infty$  almost surely (see for instance [6]). Hence, for  $n$  large enough we have

$$C_n = \max_{v_n \leq k \leq n - v_n} C_{n,k} = \sup_{\frac{v_n}{n} \leq \tau \leq 1 - \frac{v_n}{n}} C_{n,[n\tau]} \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \sup_{0 \leq \tau \leq 1} \frac{\|W_d(\tau)\|^2}{q^2(\tau)}.$$

2. Apply Lemma 7.3 and goes along similar lines as in the proof of Theorem 1 of [27]. ■

**Proof of Theorem 4.2.**

1. Assume the alternative with one change at  $t_1^* = [\tau_1^* n]$  with  $0 < \tau_1^* < 1$ . The observations satisfy

$$Y_t = \begin{cases} Y_t^{(1)} & \text{for } t \leq t_1^*, \\ Y_t^{(2)} & \text{for } t > t_1^* \end{cases}$$

where  $(Y_t^{(1)})$  and  $(Y_t^{(2)})$  satisfy the main model (3) i.e.

$$Y_t^{(i)} / \mathcal{F}_{t-1}^{(i)} \sim \mathcal{P}(\lambda_t^{(i)})$$

with  $\lambda_t^{(i)} = f_{\theta_1^*}(Y_{t-1}, Y_{t-2}, \dots), \theta_1^* \neq \theta_2^*$ , and  $\mathcal{F}_t^{(i)} = \sigma(Y_s^{(i)}, s \leq t)$ . Recall that  $\widehat{C}_n = \max_{v_n \leq k \leq n-v_n} \widehat{C}_{n,k} \geq \widehat{C}_{n,t_1^*}$ . It suffices to show that  $\widehat{C}_{n,t_1^*} \xrightarrow[n \rightarrow +\infty]{P} +\infty$ . We have

$$\begin{aligned} \widehat{C}_{n,t_1^*} &= \frac{1}{q\left(\frac{t_1^*}{n}\right)} \frac{t_1^{*2}(n-t_1^*)^2}{n^3} \\ &\quad \times \left( \widehat{\theta}_n(T_{1,t_1^*}) - \widehat{\theta}_n(T_{t_1^*+1,n}) \right)' \widehat{\Sigma}(u_n) \left( \widehat{\theta}_n(T_{1,t_1^*}) - \widehat{\theta}_n(T_{t_1^*+1,n}) \right). \end{aligned}$$

Recall that the likelihood function computed on any subset  $T \subset \{1, \dots, n\}$  is defined by

$$\widehat{L}_n(T, \theta) = \sum_{t \in T} \widehat{\ell}_t(\theta)$$

where  $\widehat{\ell}_t(\theta) = Y_t \log \widehat{\ell}_\theta^t - \widehat{\ell}_\theta^t$  with  $\widehat{f}_\theta^t = f_\theta(Y_{t-1}, Y_{t-2}, \dots, Y_1, 0, \dots)$ . So, for any  $t \in \{1, \dots, t_1^*\}$ ,  $\widehat{f}_\theta^t = f_\theta(Y_{t-1}^{(1)}, Y_{t-2}^{(1)}, \dots, Y_1^{(1)}, 0, \dots)$ . Then  $\theta \mapsto \widehat{L}(T_{1,t_1^*}, \theta)$  is the likelihood function of the stationary process  $(Y_t^{(1)})_{t \in \mathbb{Z}}$  computed on  $\{1, \dots, t_1^*\}$ . According to Theorem 3.1, it holds that  $\widehat{\theta}_n(T_{1,t_1^*}) \xrightarrow[n \rightarrow +\infty]{a.s.} \theta_1^*$ . Recall that

$$\begin{aligned} \widehat{\Sigma}_n(u_n) &= \frac{1}{2} \left[ \left( \frac{1}{u_n} \sum_{t=1}^{u_n} \frac{1}{\widehat{f}_\theta^t} \left( \frac{\partial}{\partial \theta} \widehat{f}_\theta^t \right) \left( \frac{\partial}{\partial \theta} \widehat{f}_\theta^t \right)' \right) \Big|_{\theta = \widehat{\theta}_n(T_{1,u_n})} \right. \\ &\quad \left. + \left( \frac{1}{n-u_n} \sum_{t=u_n+1}^n \frac{1}{\widehat{f}_\theta^t} \left( \frac{\partial}{\partial \theta} \widehat{f}_\theta^t \right) \left( \frac{\partial}{\partial \theta} \widehat{f}_\theta^t \right)' \right) \Big|_{\theta = \widehat{\theta}_n(T_{u_n+1,n})} \right]. \end{aligned}$$

Denote  $l_{t,1}(\theta) = Y_t^{(1)} \log f_\theta^{t,1} - f_\theta^{t,1}$  where

$$f_\theta^{t,1} = f_\theta(Y_{t-1}^{(1)}, Y_{t-2}^{(1)}, \dots, Y_1^{(1)}, 0 \dots).$$

We have,  $\widehat{\theta}_n(T_{1,u_n}) \xrightarrow[n \rightarrow +\infty]{a.s.} \theta_1^*$  and, from Lemma 7.2, it holds that

$$\frac{1}{u_n} \sum_{t=1}^{u_n} \frac{1}{\widehat{f}_\theta^t} \left( \frac{\partial}{\partial \theta} \widehat{f}_\theta^t \right) \left( \frac{\partial}{\partial \theta} \widehat{f}_\theta^t \right)' \xrightarrow[n \rightarrow +\infty]{a.s.} \Sigma^{(1)}$$

where  $\Sigma^{(1)} = E \left[ \frac{1}{f_{\theta_1^*}^{0,1}} \left( \frac{\partial}{\partial \theta} f_{\theta_1^*}^{0,1} \right) \left( \frac{\partial}{\partial \theta} f_{\theta_1^*}^{0,1} \right)' \right]$ . Also, we have

$$\widehat{L}_n(T_{t_1^*+1,n}, \theta) = \sum_{t=t_1^*+1}^n \widehat{\ell}_t(\theta) \text{ where } \widehat{\ell}_t(\theta) = Y_t \log \widehat{f}_\theta^t - \widehat{f}_\theta^t,$$

with

$$\widehat{f}_\theta^t = f_\theta(Y_{t-1}, Y_{t-2}, \dots, Y_1, 0, \dots) = f_\theta(Y_{t-1}^{(1)}, Y_{t-2}^{(1)}, \dots, Y_1^{(1)}, 0, \dots).$$

Define

$$\widehat{L}_{n,2}(T_{t_1^*+1,n}, \theta) = \sum_{t=t_1^*+1}^n \widehat{\ell}_{t,2}(\theta), \text{ where } \widehat{\ell}_{t,2}(\theta) = Y_t^{(2)} \log \widehat{f}_\theta^{t,2} - \widehat{f}_\theta^{t,2},$$

with  $\widehat{f}_\theta^{t,2} = f_\theta(Y_{t-1}^{(2)}, Y_{t-2}^{(2)}, \dots, Y_1^{(2)}, 0, \dots)$ .

Remarks that the difference between  $\widehat{f}_\theta^t$  and  $\widehat{f}_\theta^{t,2}$  lies on the dependence with the past.  $\widehat{f}_\theta^t$  can contain  $Y_{t-1}^{(1)}$ , but not  $\widehat{f}_\theta^{t,2}$ .  $\theta \mapsto \widehat{L}_{n,2}(T_{t_1^*+1,n}, \theta)$  is the approximated likelihood of the stationary model after change. By Theorem 3.1, it holds that

$$\widehat{\theta}_n^{(2)}(T_{t_1^*+1,n}) = \operatorname{argmax}_{\theta \in \Theta} \widehat{L}_{n,2}(T_{t_1^*+1,n}, \theta) \xrightarrow[+\rightarrow +\infty]{a.s.} \theta_2^*.$$

Let us show that

$$\frac{1}{n - t_1^*} \left\| \widehat{L}_n(T_{t_1^*+1,n}, \theta) - \widehat{L}_{n,2}(T_{t_1^*+1,n}, \theta) \right\|_{\Theta} \xrightarrow[n \rightarrow +\infty]{a.s.} 0.$$

According to

$$\begin{aligned} \frac{1}{n - t_1^*} \left\| \widehat{L}_n(T_{t_1^*+1,n}, \theta) - \widehat{L}_{n,2}(T_{t_1^*+1,n}, \theta) \right\|_{\Theta} &\leq \\ &\frac{1}{n - t_1^*} \sum_{k=1}^{n-t_1^*} \left\| \widehat{\ell}_{t_1^*+k}(\theta) - \widehat{\ell}_{t_1^*+k,2}(\theta) \right\|_{\Theta}, \end{aligned}$$

by using again Kounias and Weng (1969) [29], it suffices to show that

$$\sum_{k \geq 1} \frac{1}{k} E \left\| \widehat{\ell}_{t_1^*+k}(\theta), \widehat{\ell}_{t_1^*+k,2}(\theta) \right\|_{\Theta} = \sum_{k \geq 1} \frac{1}{t - t_1^*} E \left\| \widehat{\ell}_t(\theta) - \widehat{\ell}_{t,2}(\theta) \right\|_{\Theta} < \infty.$$

For  $t \geq t_1^* + 1$ , we have

$$\begin{aligned} E \left[ \left\| \widehat{\ell}_t(\theta) - \widehat{\ell}_{t,2}(\theta) \right\|_{\Theta} \right] &= E \left[ \left\| Y_t^{(2)} \log \widehat{f}_{\theta}^t - \widehat{f}_{\theta}^t - Y_t^{(2)} \log \widehat{f}_{\theta}^{t,2} + \widehat{f}_{\theta}^{t,2} \right\|_{\Theta} \right] \\ &\leq E \left[ \left| Y_t^{(2)} \right| \left\| \log \widehat{f}_{\theta}^t - \log \widehat{f}_{\theta}^{t,2} \right\|_{\Theta} + \left\| \widehat{f}_{\theta}^t - \widehat{f}_{\theta}^{t,2} \right\|_{\Theta} \right]. \end{aligned}$$

We can show, as in the proof of Theorem 3.1, that  $\left\| \log \widehat{f}_{\theta}^t - \log \widehat{f}_{\theta}^{t,2} \right\|_{\Theta} \leq \frac{1}{\underline{c}} \left\| \widehat{f}_{\theta}^t - \widehat{f}_{\theta}^{t,2} \right\|_{\Theta}$ . Hence

$$\begin{aligned} E \left[ \left\| \widehat{\ell}_t(\theta) - \widehat{\ell}_{t,2}(\theta) \right\|_{\Theta} \right] &\leq E \left[ \left( \frac{|Y_t^{(2)}|}{\underline{c}} + 1 \right) \left\| \widehat{f}_{\theta}^t - \widehat{f}_{\theta}^{t,2} \right\|_{\Theta} \right] \\ &\leq \left( E \left[ \left( \frac{|Y_t^{(2)}|}{\underline{c}} + 1 \right)^2 \right] \right)^{1/2} \times \left( E \left\| \widehat{f}_{\theta}^t - \widehat{f}_{\theta}^{t,2} \right\|_{\Theta}^2 \right)^{1/2} \\ &\leq C \left( E \left\| \widehat{f}_{\theta}^t - \widehat{f}_{\theta}^{t,2} \right\|_{\Theta}^2 \right)^{1/2}. \end{aligned}$$

But, for  $t \geq t_1^* + 1$ , we have

$$\begin{aligned} &\left\| \widehat{f}_{\theta}^t - \widehat{f}_{\theta}^{t,2} \right\|_{\Theta} \\ &= \left\| f_{\theta}(Y_{t-1}^{(2)}, \dots, Y_{t_1^*+1}^{(2)}, Y_{t_1^*}^{(1)}, \dots, Y_1^{(1)}, 0, \dots) - f_{\theta}(Y_{t-1}^{(2)}, \dots, Y_{t_1^*+1}^{(2)}, 0, \dots) \right\| \\ &\leq \sum_{j=t-t_1^*+1}^t \alpha_j^{(0)}(\Theta) |Y_{t-j}^{(1)}|. \end{aligned}$$

Thus, by using Minkowski's inequality, it holds that

$$\begin{aligned} E \left[ \left\| \widehat{\ell}_t(\theta) - \widehat{\ell}_{t,2}(\theta) \right\|_{\Theta} \right] &\leq C \left( E \left\| \widehat{f}_{\theta}^t - \widehat{f}_{\theta}^{t,2} \right\|_{\Theta}^2 \right)^{1/2} \\ &\leq C \left( E \left[ \left( \sum_{j=t-t_1^*+1}^t \alpha_j^{(0)}(\Theta) |Y_{t-j}^{(1)}| \right)^2 \right] \right)^{1/2} \\ &\leq \sum_{j=t-t_1^*+1}^t \alpha_j^{(0)}(\Theta) \left( E \left[ |Y_{t-j}^{(1)}|^2 \right] \right)^{1/2} \\ &\leq C \sum_{j=t-t_1^*+1}^t \alpha_j^{(0)}(\Theta). \end{aligned}$$

Thus, it comes that

$$\sum_{t \geq t_1^*+1} \frac{1}{t-t_1^*} \left\| \widehat{\ell}_t(\theta) - \widehat{\ell}_{t,2}(\theta) \right\|_{\Theta} \leq C \sum_{t \geq t_1^*+1} \frac{1}{t-t_1^*} \sum_{j=t-t_1^*+1}^t \alpha_j^{(0)}(\Theta)$$

$$\leq \sum_{t \geq t_1^*+1} \sum_{j=t-t_1^*+1}^t \frac{1}{t-t_1^*} \alpha_j^{(0)}(\Theta).$$

Set  $l = t - t_1^*$ , we have

$$\begin{aligned} \sum_{t \geq t_1^*+1} \sum_{j=t-t_1^*+1}^t \frac{1}{t-t_1^*} \alpha_j^{(0)}(\Theta) &= \sum_{l \geq 1} \sum_{j=l+1}^{l+t_1^*} \frac{1}{l} \alpha_j^{(0)}(\Theta) \leq \sum_{l \geq 1} \sum_{j=l}^{l+t_1^*} \frac{1}{l} \alpha_j^{(0)}(\Theta) \\ &\leq \sum_{j=1}^{t_1^*} \sum_{l=1}^j \frac{1}{l} \alpha_j^{(0)}(\Theta) + \sum_{j=t_1^*+1}^{+\infty} \sum_{l=j-t_1^*}^j \frac{1}{l} \alpha_j^{(0)}(\Theta) \\ &\leq C \sum_{j=1}^{t_1^*} \alpha_j^{(0)}(\Theta) \log j + \sum_{j=t_1^*+1}^{+\infty} \alpha_j^{(0)}(\Theta) \sum_{l=1}^j \frac{1}{l} \\ &\leq C \sum_{j \geq 1} \alpha_j^{(0)}(\Theta) \log j < +\infty. \end{aligned}$$

So, we have

$$\sum_{t \geq t_1^*+1} \frac{1}{t-t_1^*} \left\| \widehat{\ell}_t(\theta) - \widehat{\ell}_{t,2}(\theta) \right\|_{\Theta} \leq C < \infty \text{ a.s.}$$

Hence,

$$\frac{1}{n-t_1^*} \left\| \widehat{L}_n(T_{t_1^*,n}, \theta) - \widehat{L}_{n,2}(T_{t_1^*,n}, \theta) \right\|_{\Theta} \xrightarrow[n \rightarrow +\infty]{a.s.} 0.$$

According to the proof of Theorem 3.1,

$$\frac{1}{n-t_1^*} \left\| \widehat{L}_{n,2}(T_{t_1^*,n}, \theta) - E \ell_{0,2}(\theta) \right\|_{\Theta} \xrightarrow[n \rightarrow +\infty]{a.s.} 0$$

where

$$l_{t,2}(\theta) = Y_t^{(2)} \log f_{\theta}^{t,2} - f_{\theta}^{t,2}, \text{ with } f_{\theta}^{t,2} = f_{\theta}(Y_{t-1}^{(2)}, Y_{t-2}^{(2)}, \dots).$$

Moreover, the function  $\theta \mapsto E(l_{0,2}(\theta))$  has a unique maximum at  $\theta_2^*$ . This is enough to conclude that  $\widehat{\theta}_n(T_{t_1^*+1,n}) \xrightarrow[n \rightarrow +\infty]{a.s.} \theta_2^*$ . To complete the proof of this part of Theorem 4.2, remarks that the two matrices in the definition of  $\widehat{\Sigma}_n(u_n)$  are positive semi-definite (by definition) and the first one converges a.s. to  $\Sigma^{(1)}$  which is positive definite. Thus, for  $n$  large enough, we have

$$\begin{aligned} \widehat{C}_n &= \max_{v_n \leq k \leq n-v_n} \widehat{C}_{n,k} \geq \widehat{C}_{n,t_1^*} \\ &\geq \frac{1}{q \binom{t_1^*}{n}} \frac{t_1^{*2} (n-t_1^*)^2}{n^3} \left( \widehat{\theta}_n(T_{1,t_1^*}) - \widehat{\theta}_n(T_{t_1^*+1,n}) \right)'. \end{aligned}$$

$$\begin{aligned}
& \left[ \frac{1}{2} \left( \frac{1}{u_n} \sum_{t=1}^{u_n} \frac{1}{\widehat{f}_\theta^t} \left( \frac{\partial \widehat{f}_\theta^t}{\partial \theta} \right) \left( \frac{\partial \widehat{f}_\theta^t}{\partial \theta} \right)' \right) \Big|_{\theta=\widehat{\theta}_n(T_{1,u_n})} \right] \left( \widehat{\theta}_n(T_{1,t_1^*}) - \widehat{\theta}_n(T_{t_1^*+1,n}) \right) \\
& \geq \frac{1}{\sup_{0 < \tau < 1} q(\tau)} n(\tau_1^*(1 - \tau_1^*))^2 \left( \widehat{\theta}_n(T_{1,t_1^*}) - \widehat{\theta}_n(T_{t_1^*+1,n}) \right)' \\
& \left[ \frac{1}{2} \left( \frac{1}{u_n} \sum_{t=1}^{u_n} \frac{1}{\widehat{f}_\theta^t} \left( \frac{\partial \widehat{f}_\theta^t}{\partial \theta} \right) \left( \frac{\partial \widehat{f}_\theta^t}{\partial \theta} \right)' \right) \Big|_{\theta=\widehat{\theta}_n(T_{1,u_n})} \right] \left( \widehat{\theta}_n(T_{1,t_1^*}) - \widehat{\theta}_n(T_{t_1^*+1,n}) \right) \\
& \geq C \times n \left( \widehat{\theta}_n(T_{1,t_1^*}) - \widehat{\theta}_n(T_{t_1^*+1,n}) \right) \\
& \left[ \frac{1}{2} \left( \frac{1}{u_n} \sum_{t=1}^{u_n} \frac{1}{\widehat{f}_\theta^t} \left( \frac{\partial \widehat{f}_\theta^t}{\partial \theta} \right) \left( \frac{\partial \widehat{f}_\theta^t}{\partial \theta} \right)' \right) \Big|_{\theta=\widehat{\theta}_n(T_{1,u_n})} \right] \left( \widehat{\theta}_n(T_{1,t_1^*}) - \widehat{\theta}_n(T_{t_1^*+1,n}) \right) \\
& \xrightarrow[n \rightarrow +\infty]{a.s.} +\infty.
\end{aligned}$$

This holds because  $\widehat{\theta}_n(T_{1,t_1^*}) - \widehat{\theta}_n(T_{t_1^*+1,n}) \xrightarrow[n \rightarrow +\infty]{a.s.} \theta_1^* - \theta_2^* \neq 0$ , and

$$\frac{1}{u_n} \sum_{t=1}^{u_n} \frac{1}{\widehat{f}_\theta^t} \left( \frac{\partial}{\partial \theta} \widehat{f}_\theta^t \right) \left( \frac{\partial}{\partial \theta} \widehat{f}_\theta^t \right)' \xrightarrow[n \rightarrow +\infty]{a.s.} \Sigma^{(1)},$$

which is positive definite.

This completes the first part of the proof of Theorem 4.2.

2. It goes along the same line as in the proof of Theorem 2 of [27], by using the approximation of likelihood as above. ■

## Acknowledgements

The authors thank Konstantinos Fokianos for his kind support, especially for his help in the real data application study.

## References

- [1] BARDET, J.-M. AND WINTENBERGER, O., Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes. *Ann. Statist.* 37 (2009), 2730–2759. [MR2541445](#)
- [2] BARDET, J.-M., KENGNE, W. AND WINTENBERGER, O., Detecting multiple change-points in general causal time series using penalized quasi-likelihood. *Electronic Journal of Statistics* 6 (2012), 435–477. [MR2988415](#)
- [3] BERKES, I., HORVÁTH, L., KOKOSZKA, P.S. AND SHAO, Q.-M., On discriminating between long-range dependence and changes in mean. *Ann. Statist.* 34 (2006), 1140–1165. [MR2278354](#)
- [4] BILLINGSLEY, P., *Convergence of Probability Measures*. John Wiley & Sons Inc., New York (1968). [MR0233396](#)

- [5] BRÄNNÄS, K. AND SHAHIDUZZAMAN QUORESHI, A.M.M., Integer-valued moving average modelling of the number of transactions in stocks. *Applied Financial Economics* 20 (2010), 1429–1440.
- [6] CSÖRGO, M., CSÖRGO, S., HORVÁTH, L. AND MASON, D.M., Weighted empirical and quantile processes. *The Annals of Probability* 14 (1986), 31–85. [MR0815960](#)
- [7] CSÖRGO, M. AND HORVÁTH, L., *Weighted Approximations in Probability and Statistics*. Wiley Chichester (1993). [MR1215046](#)
- [8] DAVIS, R.A., DUNSMUIR, W. AND STREETT, S., Maximum likelihood estimation for an observation driven model for Poisson counts. *Methodol. Comput. Appl. Probab.* 7 (2005), 149–159. [MR2189786](#)
- [9] DAVIS, R.A. AND LIU, H., Theory and Inference for a Class of Observation-Driven Models with Application to Time Series of Counts. *Preprint, arXiv:1204.3915*.
- [10] DAVIS, R.A. AND WU, R., A negative binomial model for time series of counts. *Biometrika* 96 (2009), 735–749. [MR2538769](#)
- [11] DOUC, R., DOUKHAN, P. AND MOULINES, E., Ergodicity of observation-driven time series models and consistency of the maximum likelihood estimator. *Stochastic Process. Appl.* 123 (2013), 2620–2647. [MR3054539](#)
- [12] DOUKHAN, P. AND WINTENBERGER, O., Weakly dependent chains with infinite memory. *Stochastic Process. Appl.* 118 (2008), 1997–2013. [MR2462284](#)
- [13] DOUKHAN, P., FOKIANOS, K. AND TJØSTHEIM, D., On weak dependence conditions for Poisson autoregressions. *Statist. and Probab. Letters* 82 (2012), 942–948. [MR2910041](#)
- [14] DOUKHAN, P., FOKIANOS, K. AND TJØSTHEIM, D., Correction to “On weak dependence conditions for Poisson autoregressions” [Statist. Probab. Lett. 82 (2012) 942–948]. *Statist. and Probab. Letters* 83 (2013), 1926–1927. [MR2910041](#)
- [15] FERLAND, R., LATOUR, A. AND ORAICHI, D., Integer-valued GARCH process. *J. Time Ser. Anal.* 27 (2006), 923–942. [MR2328548](#)
- [16] FOKIANOS, K. AND FRIED, R., Interventions in INGARCH processes. *J. Time Ser. Anal.* 31 (2010), 210–225. [MR2676152](#)
- [17] FOKIANOS, K. AND FRIED, R., Interventions in log-linear Poisson autoregression. *Statistical Modelling* 12 (2012), 1–24. [MR3179504](#)
- [18] FOKIANOS, K. AND NEUMANN, M., A goodness-of-fit test for Poisson count processes. *Electronic Journal of Statistics* 7 (2013), 793–819. [MR3040560](#)
- [19] FOKIANOS, K., RAHBK, A. AND TJØSTHEIM, D., Poisson autoregression. *Journal of the American Statistical Association* 104 (2009), 1430–1439. [MR2596998](#)
- [20] FOKIANOS, K. AND TJØSTHEIM, D., Nonlinear Poisson autoregression. *Ann. Inst. Stat. Math.* 64 (2012), 1205–1225. [MR2981620](#)
- [21] FRANKE, J., KIRCH, C. AND TADJUIDJE KAMGAING, J., Changepoints in times series of counts. *J. Time Ser. Anal.* 33 (2012), 757–770. [MR2969909](#)
- [22] HAIRER, M. AND MATTINGLY, J., Ergodicity of the 2d navier-stokes equations with degenerate stochastic forcings. *Ann. Math.* 164 (2006), 993–1032. [MR2259251](#)

- [23] HELD, L., HÖHLE, M. AND HOFMANN, M., A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Statistical Modelling* 5 (2005), 187–199. [MR2210732](#)
- [24] INCLÁN, C. AND TIAO, G.C., Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association* 89 (1994), 913–923. [MR1294735](#)
- [25] KANG, J. AND LEE, S., Parameter change test for random coefficient integer-valued autoregressive processes with application to polio data analysis. *Journal of Time Series Analysis* 30(2) (2009), 239–258. [MR2508343](#)
- [26] KEDEM, B. AND FOKIANOS, K., *Regression Models for Time Series Analysis*. Hoboken, Wiley, NJ (2002). [MR1933755](#)
- [27] KENGNE W., Testing for parameter constancy in general causal time-series models. *J. Time Ser. Anal.* 33 (2012), 503–518. [MR2915099](#)
- [28] KIERFER, J., K-sample analogues of the Kolmogorov-Smirnov and Cramér-v.Mises tests. *Ann. Math. Statist* 30 (1959), 420–447. [MR0102882](#)
- [29] KOUNIAS, E.G. AND WENG, T.-S., An inequality and almost sure convergence. *Annals of Mathematical Statistics* 40 (1969), 1091–1093. [MR0245058](#)
- [30] LAMBERT, D., Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 34 (1992), 1–14.
- [31] NEUMANN, M., Absolute regularity and ergodicity of Poisson count processes. *Bernoulli* 17 (2011), 1268–1284. [MR2854772](#)
- [32] RABEMANANJARA, R. AND ZAKOÏAN, J.M., Threshold ARCH models and asymmetries in volatility. *Journal of Applied Econometrics* 8 (1993), 31–49.
- [33] STRAUMANN, D. AND MIKOSCH, T., Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *Ann. Statist.* 34 (2006), 2449–2495. [MR2291507](#)
- [34] TJØSTHEIM, D., Some recent theory for autoregressive count time series. *TEST* 21 (2012), 413–438. [MR2983224](#)
- [35] WEISS, C.H., Modelling time series of counts with overdispersion. *Stat. Methods Appl.* 18 (2009), 507–519. [MR2551338](#)
- [36] ZAKOÏAN, J.-M., Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* 18 (1994), 931–955.