



**HAL**  
open science

# Regenerative block empirical likelihood for Markov chains

Hugo Harari-Kermadec

► **To cite this version:**

Hugo Harari-Kermadec. Regenerative block empirical likelihood for Markov chains. *Journal of Non-parametric Statistics*, 2011, 23 (3), p. 781-802, GNST-2010-05-08.R2. hal-00566140

**HAL Id: hal-00566140**

**<https://hal.science/hal-00566140>**

Submitted on 15 Feb 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## RESEARCH ARTICLE

### Regenerative block empirical likelihood for Markov chains

Hugo Harari-Kermadec

*ENS-Cachan & Université Paris 1, SAMM*  
(Received 00 Month 200x; in final form 00 Month 200x)

Empirical likelihood is a powerful semi-parametric method increasingly investigated in the literature. However, most authors essentially focus on an i.i.d. setting. In the case of dependent data, the classical empirical likelihood method cannot be directly applied on the data but rather on blocks of consecutive data catching the dependence structure. Generalization of empirical likelihood based on the construction of blocks of increasing nonrandom length have been proposed for time series satisfying mixing conditions. Following some recent developments in the bootstrap literature, we propose a generalization for a large class of Markov chains, based on small blocks of various lengths. Our approach makes use of the regenerative structure of Markov chains, which allows us to construct blocks which are almost independent (independent in the atomic case). We obtain the asymptotic validity of the method for positive recurrent Markov chains and present some simulation results.

**Keywords:** Nummelin splitting technique, time series, Empirical Likelihood  
**MSC codes:** 62G05, 62F35, 62F40

Hugo Harari-Kermadec  
Dpt Eco-gestion, ENS-CACHAN  
61, avenue du Président Wilson  
94235 Cachan cedex, France

phone : +331 47 40 20 95  
fax : +331 47 40 24 60  
email : hugo.harari@ens-cachan.fr

## 1. Introduction

Empirical Likelihood (EL), introduced by Owen (1988), is a powerful semi-parametric method. It can be used in a very general setting and leads to effective estimation, tests and confidence intervals. This method shares many good properties with the conventional parametric log-likelihood ratio: both statistics have  $\chi^2$  limiting distribution and are Bartlett correctable, meaning that the error can be reduced from  $\mathcal{O}(n^{-1})$  to  $\mathcal{O}(n^{-2})$  by a simple adjustment. An additional property of EL is that the corresponding confidence intervals and tests do not rely on an estimator of the variance. This last property is specially noticeable for dependent data, since estimating the variance is then a challenging issue.

Owen's framework has been intensively studied in the 90's (see Owen 2001, for an overview), leading to many generalizations and applications, but mainly for an i.i.d. setting. Some adaptations of EL for dependent data have been introduced, such as the Block Empirical Likelihood (BEL) of Kitamura (1997) for weakly dependent processes or the subject-wise and elementwise empirical likelihoods of Wang et al. (2010) for longitudinal data. BEL is inspired by similarities with the bootstrap methodology. Kitamura proposed to apply the empirical likelihood framework not directly on the data but on blocks of consecutive data, to catch the dependence structure. This idea, known as *Block Bootstrap* (BB) or blocking technique (in the probabilistic literature, see Doukhan and Ango Nze 2004, for references) goes back to Kunsch (1989) in the bootstrap literature and has been intensively exploited in this field (see Lahiri 2003, for a survey). However, the BB performance has been questioned, see Götze and Kunsch (1996) and Horowitz (2003). Indeed it is known that the blocking technique distorts the dependence structure of the data generating process and its performance strongly relies on the choice of the block size. From a theoretical point of view, the assumptions used to prove the validity of BB and BEL are generally strong: it is assumed that the process is stationary and satisfies some strong-mixing properties (some non-stationary processes can nevertheless be handled, see Synowiecki (2007) for example). In addition to having a precise control of the coverage probability of the confidence intervals, we have to assume that the strong mixing coefficients are exponentially decreasing (see Kitamura 1997; Lahiri 2003). Moreover, the choice of the tuning parameter (the block size) may be quite difficult from a practical point of view.

In this paper, we focus on generalizing empirical likelihood to Markov chains. Questioning the restriction implied by the Markovian setting is a natural issue. It should be mentioned that homogeneous Markov chain models cover a huge number of time series models. In particular, a Markov chain can always be written in a nonparametric way:  $X_i = h(X_{i-1}, \dots, X_{i-p}, \varepsilon_i)$ , where  $(\varepsilon_i)_{i \geq 0}$  is i.i.d. with density  $f$  and, for  $i > 0$ ,  $\varepsilon_i$  is independent of  $(X_k)_{0 \leq k < i}$  (see Kallenberg 2002). Note that both  $h$  and  $f$  are unknown functions. Such representations explain why, provided that  $p$  is large enough, any process of length  $n$  can be generated by a Markov chain, see Knight (1975). Note also that a Markov chain may not be necessarily strong-mixing. For instance, the simple linear model  $X_i = \frac{1}{2}(X_{i-1} + \varepsilon_i)$  with  $\mathbb{P}(\varepsilon_i = 1) = \mathbb{P}(\varepsilon_i = 0) = \frac{1}{2}$  is not strong-mixing (see Doukhan and Ango Nze 2004, for results on dependence in Econometrics). Doukhan and Ango Nze (2004) gives many classical econometric models that can be seen as Markovian: ARMA, ARCH and GARCH processes, bilinear and threshold models.

Our approach is also inspired by some recent developments in the bootstrap literature on Markov chains: instead of choosing blocks of constant length, we use the Markov chain structure to choose some adequate cutting times and then we obtain blocks of

various lengths. This construction, introduced in Bertail and Cl emen on (2004), catches the dependence structure. It is originally based on the existence of an atom for the chain i.e. an accessible set on which the transition kernel is constant (see Meyn and Tweedie 2009, chapter 1.5). The existence of an atom allows us to cut the chain into regeneration blocks, separated from each other by a visit to the atom. These blocks (of random lengths) are independent by the strong Markov property. Once these blocks are obtained, the *Regenerative Block-Bootstrap (RBB)* consists in resampling the data blocks to build new regenerative processes. The rate of convergence of the pivotal statistic obtained by resampling these blocks ( $\mathcal{O}(n^{-1+\varepsilon})$ ) is better than the one obtained for the Block Bootstrap ( $\mathcal{O}(n^{-3/4})$ ) and is close to the classical rate  $\mathcal{O}(n^{-1})$  obtained in the i.i.d. case, see G otze and Kunsch (1996) and Lahiri (2003).

These improvements suggest that a version of the empirical likelihood (EL) method based on such blocks could yield improved results in comparison to the method presented in Kitamura (1997). Indeed it is known that EL enjoys somehow the same properties in terms of accuracy as the bootstrap but without any Monte-Carlo step. The main idea is to consider the renewal blocks as independent observations and to follow the empirical likelihood method. Such a program is made possible by transforming the original problem based on moments under the stationary distribution into an equivalent problem under the distribution of the observable blocks (via Kac's Theorem). The advantages of the method proposed in this paper are at least twofold: first the construction of the blocks is automatic and entirely determined by the data: it leads to a unique version of the empirical likelihood program. Second there is not need to ensure stationarity nor any strong mixing condition to obtain a better coverage probability for the corresponding confidence regions.

Assuming that the chain is atomic is a strong restriction of this method. This hypothesis holds for discrete Markov chains and queuing (or storage) systems returning to a stable state (for instance the empty queue): see chapter 2.4 of Meyn and Tweedie (2009). However this method can be extended to the more general case of Harris chains. Indeed, any chain having some recurrent properties can be extended to a chain possessing an atom which then enjoys some regenerative properties. Nummelin gives an explicit construction of such an extension that we recall in Section 4 (see Athreya and Ney 1978; Nummelin 1978). In Bertail and Cl emen on (2006), an extension of the RBB procedure to general Harris chains based on the Nummelin' splitting technique is proposed (*the Approximate Regenerative Block-Bootstrap, ARBB*). One purpose of this paper is to prove that these approximatively regenerative blocks can also be used in the framework of empirical likelihood and lead to consistent results.

The outline of the paper is the following. In Section 2, notations are set out and key concepts of the Markov atomic chain theory are recalled. In Section 3, we present how to construct regenerative data blocks and confidence regions based on these blocks. We give the main properties of the corresponding asymptotic statistics. In Section 4 the Nummelin splitting technique is shortly recalled and a framework to adapt the regenerative empirical likelihood method to general Harris chains is proposed. We essentially obtain consistent results but also briefly discuss test and higher order properties. In Section 5, we present some moderate sample size simulations.

## 2. Preliminary statement

### 2.1. Framework

For the sake of simplicity, we use the same notations as Bertail and Cléménçon (2004) when possible. We consider a chain  $X = (X_i)_{i \in \mathbb{N}}$  on a state space  $(E, \mathcal{E})$ , with initial distribution  $\nu$  and transition probability  $\Pi$ . For a set  $B \in \mathcal{E}$  and  $i \in \mathbb{N}$ , we thus denote

$$X_0 \sim \nu \text{ and } \mathbb{P}(X_i \in B \mid X_0, \dots, X_{i-1}) = \Pi(X_{i-1}, B) \text{ a.s. .}$$

Recurrence properties will be important in the following. An irreducible chain is said *positive recurrent* when it admits an invariant probability :

$$\exists \mu \text{ probability measure on } E, \mu \Pi = \mu, \text{ where } \mu \Pi(\cdot) = \int_{x \in E} \mu(dx) \Pi(x, \cdot).$$

We assume that the chain is aperiodic (i.e.  $X$  is not cyclic) and that there exists a measure  $\psi$  such as the chain is  $\psi$ -irreducible. This simply means that for any starting state  $x$  in  $E$  and any set  $A$  of positive  $\psi$ -measure, the chain visits  $A$  with probability 1. A  $\psi$ -irreducible chain is said *Harris recurrent* if every measurable set with positive  $\psi$ -measure visited once is visited infinitely often with probability 1.

In what follows,  $\mathbb{P}_\nu$  and  $\mathbb{P}_x$  (for  $x$  in  $E$ ) denote respectively the probability measure when  $X_0 \sim \nu$  and  $X_0 = x$ . The indicator function of an event  $\mathcal{A}$  is denoted by  $\mathbb{1}_{\mathcal{A}}$ . The corresponding expectations are denoted  $\mathbb{E}_\nu(\cdot)$ ,  $\mathbb{E}_x[\cdot]$  and  $\mathbb{E}_{\mathcal{A}}[\cdot]$ . For further details and traditional properties of Markov chains, we refer to Revuz (1984) or Meyn and Tweedie (2009).

Notice that the chain  $X$  is not supposed stationary (since  $\nu$  may differ from  $\mu$ ) nor strong-mixing. To simplify the exposition, we do not treat in this paper the fully non-stationary case corresponding to null recurrence. Results in that direction may be found in Tjostheim (1990).

### 2.2. Atomic Markov chains

Assume that the chain is  $\psi$ -irreducible and possesses an accessible atom, i.e. a set  $A$  with  $\psi(A) > 0$  such that  $\Pi(x, \cdot) = \Pi(y, \cdot)$  for all  $x, y$  in  $A$ . The class of atomic Markov chains contains not only chains defined on a countable state space but also many specific Markov models used to study queuing systems and stock models (see Asmussen 2003, for models involved in queuing theory). In the discrete case, any recurrent state is an accessible atom: the choice of the atom is thus left to the statistician who can for instance use the most visited point. In many other situations the atom is determined by the structure of the model (for a random walk on  $\mathbb{R}^+$ , with continuous increment, 0 is the only possible atom).

Denote by  $\tau_A = \tau_A(1) = \inf \{k \geq 1, X_k \in A\}$  the hitting time of the atom  $A$  (the first visit) and, for  $j \geq 2$ , denote by  $\tau_A(j) = \inf \{k > \tau_A(j-1), X_k \in A\}$  the successive return times to  $A$ . The sequence  $(\tau_A(j))_{j \geq 1}$  defines the successive times at which the chain forgets its past, called *regeneration times*. Indeed, the transition probability being constant on the atom,  $X_{\tau_A+1}$  only depends on the information that  $X_{\tau_A}$  is in  $A$  and not any more on the actual value of  $X_{\tau_A}$  itself.

For any initial distribution  $\nu$ , the sample path of the chain may be divided into blocks

of random length corresponding to consecutive visits to  $A$ :

$$B_j = (X_{\tau_A(j)+1}, \dots, X_{\tau_A(j+1)}).$$

The sequence of blocks  $(B_j)_{1 \leq j < \infty}$  is then i.i.d. by the strong Markov property (Meyn and Tweedie 2009, page 73). Notice that the block  $B_0 = (X_1, \dots, X_{\tau_A})$  is independent of the other blocks, but does not have the same distribution, because it depends on the initial distribution  $\nu$ .

Let  $m : E \times \mathbb{R}^p \rightarrow \mathbb{R}^r$  be a measurable function and  $\theta_0$  be the true value of some parameter  $\theta \in \mathbb{R}^p$  of the chain, given by an estimating equation on the invariant measure  $\mu$ :

$$\mathbb{E}_\mu[m(X, \theta_0)] = 0. \quad (1)$$

Dimensions are of importance: the number of constraints  $r$  must be at least equal to the number of parameters  $p$ , for identification reasons. The estimation of the mean is a just-identified case ( $r = p$ ):  $\theta_0 = \mathbb{E}_\mu[X]$  and  $m(X, \theta) = X - \theta$ .

In this framework, Kac's Theorem, stated below (Meyn and Tweedie 2009, Theorem 10.2.2) allows us to write functionals of the stationary distribution  $\mu$  as functionals of the distribution of a regenerative block.

**Theorem 2.1 Kac** *Let  $X$  be an aperiodic,  $\psi$ -irreducible Markov chain with an accessible atom  $A$ .  $X$  is positive recurrent if and only if  $\mathbb{E}_A[\tau_A] < \infty$ . In such a case,  $X$  admits an unique invariant probability distribution  $\mu$ , the Pitman's occupation measure given by*

$$\mu(F) = \mathbb{E}_A \left[ \sum_{i=1}^{\tau_A} \mathbb{1}_{X_i \in F} \right] / \mathbb{E}_A[\tau_A], \text{ for all } F \in \mathcal{E}.$$

In the following we denote

$$M(B_j, \theta) = \sum_{i=\tau_A(j)+1}^{\tau_A(j+1)} m(X_i, \theta)$$

so that we can rewrite the estimating equation (1) as:

$$\mathbb{E}_A[M(B_j, \theta_0)] = 0. \quad (2)$$

Kac's Theorem allows us to use the decomposition of the chain into independent blocks to obtain limit theorems for atomic chains. See for example Meyn and Tweedie (2009) for the Law of Large Numbers (LLN, page 415), Central Limit Theorem (CLT, page 416), Law of Iterated Logarithm (page 416), Bolthausen (1982) for the Berry-Esseen Theorem and Bertail and Cl  men  on (2004) for Edgeworth expansions. These results are established under hypotheses related to the distribution of the  $B_j$ 's :

**Return time conditions:**

$$\begin{aligned} \mathbf{HO}(\kappa) &: \mathbb{E}_A[\tau_A^\kappa] < \infty, \\ \mathbf{HO}(\kappa, \nu) &: \mathbb{E}_\nu[\tau_A^\kappa] < \infty, \end{aligned}$$

where  $\kappa > 0$  and  $\nu$  is the initial distribution of the chain. When the chain is stationary and strong mixing, these hypotheses can be related to the rate of decay of  $\alpha$ -mixing coefficients  $\alpha(p)$ , see Bolthausen (1982). In particular, the hypotheses are satisfied if  $\sum_{j \geq 1} j^\kappa \alpha(j) < \infty$ .

*Block-moment conditions:*

$$\mathbf{H1}(\kappa, m) : \mathbb{E}_A \left[ \left( \sum_{i=1}^{\tau_A} \|m(X_i, \theta_0)\| \right)^\kappa \right] < \infty,$$

$$\mathbf{H1}(\kappa, \nu, m) : \mathbb{E}_\nu \left[ \left( \sum_{i=1}^{\tau_A} \|m(X_i, \theta_0)\| \right)^\kappa \right] < \infty.$$

The assumptions on  $\nu$  allow to control the first block  $B_0$ . Equivalence of these assumptions with easily checkable drift conditions may be found in Meyn and Tweedie (2009), Appendix A.

### 3. The regenerative case

#### 3.1. Regenerative Block Empirical Likelihood algorithm

Let  $X_1, \dots, X_n$  be an observation of the chain  $X$ . If we assume that we know an atom  $A$  for the chain, the construction of the regenerative blocks is then trivial. Consider the empirical distribution of the blocks:

$$\mathbb{P}_{l_n} = \frac{1}{l_n} \sum_{j=1}^{l_n} \delta_{B_j},$$

where  $l_n$  is the number of complete regenerative blocks, and the multinomial distributions

$$\mathbb{Q} = \sum_{j=1}^{l_n} q_j \delta_{B_j}, \text{ with } 0 < q_j \text{ and } \sum_{j=1}^{l_n} q_j = 1,$$

dominated by  $\mathbb{P}_{l_n}$ . To obtain a confidence region, we will apply Owen (1990)'s method to the blocks  $B_j$ : we are going to minimize the Kullback discrepancy between  $\mathbb{Q}$  and  $\mathbb{P}_{l_n}$  under the condition (2). More precisely, the *Regenerative Block Empirical Likelihood* is defined in the next 4 steps:

**Algorithm 1** ReBEL - Regenerative Block Empirical Likelihood construction:

- (1) Count the number of visits to  $A$  up to time  $n$ :  $l_n + 1 = \sum_{i=1}^n \mathbb{1}_{X_i \in A}$ .
- (2) Divide the observed trajectory  $X^{(n)} = (X_1, \dots, X_n)$  into  $l_n + 2$  blocks corresponding to the pieces of the sample path between consecutive visits to the atom  $A$ ,

$$B_0 = (X_1, \dots, X_{\tau_A(1)}), B_1 = (X_{\tau_A(1)+1}, \dots, X_{\tau_A(2)}), \dots,$$

$$B_{l_n} = (X_{\tau_A(l_n)+1}, \dots, X_{\tau_A(l_n+1)}), B_{l_n+1}^{(n)} = (X_{\tau_A(l_n+1)+1}, \dots, X_n),$$

with the convention  $B_{l_n+1}^{(n)} = \emptyset$  when  $\tau_A(l_n + 1) = n$ .

- (3) Drop the first block  $B_0$  and the last one  $B_{l_n+1}^{(n)}$  (possibly empty when  $\tau_A(l_n + 1) = n$ ).
- (4) Evaluate the empirical log-likelihood ratio  $r_n(\theta)$  (practically on a grid of the set of interest):

$$r_n(\theta) = - \sup_{(q_1, \dots, q_{l_n})} \left\{ \log \left[ \prod_{j=1}^{l_n} l_n q_j \right] \middle| \sum_{j=1}^{l_n} q_j \cdot M(B_j, \theta) = 0, \sum_{j=1}^{l_n} q_j = 1 \right\}.$$

Using Lagrange arguments, this can be more easily calculated as

$$r_n(\theta) = \sup_{\lambda \in \mathbb{R}^p} \left\{ \sum_{j=1}^{l_n} \log [1 + \lambda' M(B_j, \theta)] \right\}.$$

**Remark 1** Small samples: Possibly, if the chain does not visit  $A$ ,  $l_n = -1$ . Of course the algorithm cannot be implemented and no confidence interval can be built. Actually, even when  $l_n \geq 0$ , the algorithm can be meaningless and at least a reasonable number of blocks are needed to build a confidence interval. In the positive recurrent case, it is known that  $l_n \sim n/\mathbb{E}_A[\tau_A]$  a.s. and the length of each block has expectation  $\mathbb{E}_A[\tau_A]$ . Many regenerations of the chain should then be observed as soon as  $n$  is significantly larger than  $\mathbb{E}_A[\tau_A]$ . Of course, the next results are asymptotic, for finite sample consideration on empirical likelihood methods (in the i.i.d. setting), refer to Bertail et al. (2008).

The next theorem states the asymptotic validity of ReBEL in the case  $r = p$  (just-identified case). For this, we introduce the ReBEL confidence region defined as follows:

$$C_{n,\alpha} = \left\{ \theta \in \mathbb{R}^p \mid 2 \cdot r_n(\theta) \leq F_{\chi_p^2}^{-1}(1 - \alpha) \right\},$$

where  $F_{\chi_p^2}$  is the distribution function of a  $\chi^2$  distribution with  $p$  degrees of freedom.

**Remark 2** Scaling factor: Block empirical likelihood methods usually need a scaling factor to compensate the eventual overlap of the blocks, denoted  $A_N^{-1}$  in Kitamura (1997), page 2089. The regenerative perspective used here forbid any overlap and therefore avoid such a factor.

**Theorem 3.1**: *Let  $\mu$  be the invariant measure of the chain, let  $\theta_0 \in \mathbb{R}^p$  be the parameter of interest, satisfying  $\mathbb{E}_\mu[m(X, \theta_0)] = 0$ . Assume*

$$\Sigma = \mathbb{E}_A[\tau_A]^{-1} \mathbb{E}_A[M(B, \theta_0)M(B, \theta_0)']$$

*is of full-rank. If **H0**(1,  $\nu$ ), **H0**(2) and **H1**(2,  $m$ ) hold, then*

$$2r_n(\theta_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi_p^2$$

*and therefore*

$$\mathbb{P}_\nu(\theta_0 \in C_{n,\alpha}) \xrightarrow[n \rightarrow \infty]{} 1 - \alpha.$$



The proof relies on the same arguments as the one for empirical likelihood based on i.i.d. data. This can be easily understood: our data, the regenerative blocks, are i.i.d. (Owen 1990, 2001). The only difference with the classical use of empirical likelihood is that the length of the data (i.e. the number of blocks) is a random value  $l_n$ . However, we have that  $n/l_n \rightarrow \mathbb{E}_A(\tau_A)$  a.s. (Meyn and Tweedie 2009, page 425). The proof is given in the appendix.

**Remark 3** Convergence rate: Let's make some very brief discussion on the rate of convergence of this method. Bertail and Cléménçon (2004) shows that the Edgeworth expansion of the mean standardized by the empirical variance holds up to  $\mathcal{O}_\nu(n^{-1})$  (in contrast to what is expected when considering a variance built on fixed length blocks). It follows from their result that

$$\mathbb{P}_\nu(2r_n(\theta_0) \leq u) = F_{\chi_p^2}^{-1}(u) + \mathcal{O}_\nu(n^{-1})$$

This is already (without Bartlett correction) better than the Bartlett corrected empirical likelihood when fixed length blocks are used (Kitamura 1997). Actually, we expect, in this atomic framework, that a Bartlett correction would lead to the same result as in the i.i.d. case:  $\mathcal{O}(n^{-2})$ . However, to prove this conjecture, we should establish an Edgeworth expansion for the likelihood ratio (which can be derived from the Edgeworth expansion for self-normalized sums) up to order  $\mathcal{O}(n^{-2})$  which is a very technical task. This is left for further work.

**Remark 4** Change of discrepancy: Empirical likelihood can be seen as a contrast method based on the Kullback discrepancy. To replace the Kullback discrepancy by some other discrepancy is an interesting problem which has led to some recent works in the i.i.d. case. Newey and Smith (2004) generalized empirical likelihood to the family of Cressie-Read discrepancies (see also Guggenberger and Smith 2005). The resulting methodology, *Generalized Empirical Likelihood*, is included in the empirical  $\varphi$ -discrepancy method introduced by Bertail et al. (2008) (see also Bertail et al. 2007; Kitamura 2006).

In the dependent case, it should be mentioned that the constant length blocks procedure has been studied in the case of empirical Euclidean likelihood by Lin and Zhang (2001). A method based on the Cressie-Read discrepancies for tilting time series data has been introduced by Hall and Yao (2003). Our proposal, stated here for the Kullback discrepancy only, is straightforwardly compatible with these generalizations (Cressie-Read and  $\varphi$ -discrepancy).

An important issue is the behavior of the empirical log-likelihood ratio under a local alternative, i.e. if the moment equation (1) is misspecified :  $\mathbb{E}_\mu[m(X, \theta_0)] = \delta/\sqrt{n}$ . The result states as follows.

**Theorem 3.2**: *Let  $\mu$  be the invariant measure of the chain, let  $\theta_0 \in \mathbb{R}^p$  be the parameter of interest, satisfying  $\mathbb{E}_\mu[m(X, \theta_0)] = \delta/\sqrt{n}$ . Assume that  $\Sigma$  is of full-rank. If **H0**(1,  $\nu$ ), **H0**(2) and **H1**(2,  $m$ ) hold, then the empirical log-likelihood ratio has an asymptotic noncentral chi-square distribution with  $p$  degrees of freedom and noncentrality parameter  $\delta'\Sigma^{-1}\delta$*

$$2r_n(\theta_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi_p'^2(\delta'\Sigma^{-1}\delta).$$

The proof is postponed to the appendix. It is a classical result that the log-likelihood

ratio is asymptotically noncentral chi-square and that the critical order is  $n^{-1/2}$ . The interesting quantity to study the efficiency of the method in this context is the noncentrality parameter. Newey (1985) gives the asymptotic distribution of the pivotal statistic based on optimally weighted GMM which is a standard tool for dependent data. Unfortunately, Newey's results are stated in a parametric context and it is therefore impossible to compare them with Theorem 3.2.

Nevertheless, ReBEL can easily be compared with the Continuously updated GMM (CUE-GMM) which is very close to the optimally weighted GMM. CUE-GMM estimators have been shown to coincide with empirical Euclidean likelihood (EEL), see Antoine et al. (2007). The difference between the EL and EEL being just a change of discrepancy (see Note 4), it is then straightforward to adapt the proof of Theorem 3.2 to the case of the EEL. The developments of the pivotal statistics coincide for the two first order and therefore they lead to the same asymptotic distribution in the case of misspecification. EL is thus as efficient as the optimally weighted GMM.

### 3.2. Estimation and the over-identified case

The properties of empirical likelihood proved by Qin and Lawless (1994) can be extended to our Markovian setting. In order to state the corresponding results respectively on estimation, confidence region under over-identification ( $r > p$ ) and hypotheses testing, we introduce the following additional assumptions. Assume that there exists a neighborhood  $V$  of  $\theta_0$  and a real positive function  $N$  with  $\mathbb{E}_\mu [N(X)] < \infty$ , such that:

- H2(a)**  $\partial m(x, \theta)/\partial \theta$  is continuous in  $\theta$  and bounded in norm by  $N(x)$  for  $\theta$  in  $V$ ,
- H2(b)**  $D = \mathbb{E}_\mu[\partial m(X, \theta_0)/\partial \theta]$  is of full rank,
- H2(c)**  $\partial^2 m(x, \theta)/\partial \theta \partial \theta'$  is continuous in  $\theta$  and bounded in norm by  $N(x)$  for  $\theta$  in  $V$ ,
- H2(d)**  $\|m(x, \theta)\|^3$  is bounded by  $N(x)$  on  $V$ .

Notice that **H2(d)** implies in particular the block moment condition **H1(3, m)** since by Kac's Theorem

$$\mathbb{E}_\mu [\|m(X, \theta)\|^3] = \frac{\mathbb{E}_A [\sum_{i=1}^{\tau_A} \|m(X_i, \theta)\|^3]}{\mathbb{E}_A [\tau_A]} \leq \frac{\mathbb{E}_A [\sum_{i=1}^{\tau_A} N(X_i)]}{\mathbb{E}_A [\tau_A]} = \mathbb{E}_\mu [N(X)] < \infty.$$

Empirical likelihood provides a natural way to estimate  $\theta_0$  in the i.i.d. case (Qin and Lawless 1994). This can be straightforwardly extended to Markov chains. The estimator is the maximum empirical likelihood estimator defined by

$$\tilde{\theta}_n = \arg \inf_{\theta \in \Theta} \{r_n(\theta)\}.$$

The next theorem shows that, under natural assumptions on  $m$  and  $\mu$ ,  $\tilde{\theta}_n$  is an asymptotically Gaussian estimator of  $\theta_0$ .

**Theorem 3.3:** *Assume that the hypotheses of Theorem 3.1 holds. Under the additional assumptions **H2(a)**, **H2(b)** and **H2(d)**,  $\tilde{\theta}_n$  is a consistent estimator of  $\theta_0$ . If in addition **H2(c)** holds, then  $\tilde{\theta}_n$  is asymptotically Gaussian:*

$$\sqrt{n}(\tilde{\theta}_n - \theta_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}\left(0, (D' \Sigma^{-1} D)^{-1}\right).$$

Notice that both  $D$  and  $\Sigma$  can be easily estimated by empirical sums over the blocks. The corresponding estimator for  $(D'\Sigma^{-1}D)^{-1}$  is straightforwardly convergent by the LLN for Markov chains.

**Remark 5** Asymptotic covariance matrix : Our asymptotic covariance matrix  $(D'\Sigma^{-1}D)^{-1}$  is to be compared with the asymptotic covariance matrix  $V_\theta$  of Kitamura (1997)'s estimator, which coincide with the asymptotic covariance matrix of the optimally weighted GMM estimator. Both matrix are very similar:  $V_\theta = (D'\mathcal{S}^{-1}D)^{-1}$ , where  $\mathcal{S}$  is the counterpart of our  $\Sigma$  for weakly dependent processes:

$$\mathcal{S} = \lim_{n \rightarrow \infty} n^{-1} \left( \sum_{i=1}^n m(X_i, \theta_0) \right) \left( \sum_{i=1}^n m(X_i, \theta_0) \right)'$$

For a process being both weakly dependent and Markovian (and in particular in the i.i.d. case),  $\mathcal{S} = \Sigma$  and therefore  $V_\theta = (D'\Sigma^{-1}D)^{-1}$ .

The case of over-identification ( $r > p$ ) is an important feature, specially for econometric applications. In such a case, the statistic  $2r_n(\tilde{\theta}_n)$  may be considered to test the moment equation (1):

**Theorem 3.4:** *Under the assumptions of Theorem 3.3, if the moment equation (1) holds, then we have*

$$2r_n(\tilde{\theta}_n) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi_{r-p}^2.$$

We now turn to a theorem equivalent to Theorem 3.1. In the over-identified case, the likelihood ratio statistic used to test  $\theta = \theta_0$  must be corrected. We now define

$$W_{1,n}(\theta) = 2r_n(\theta) - 2r_n(\tilde{\theta}_n).$$

The ReBEL confidence region of nominal level  $1 - \alpha$  in the over-identified case is now given by

$$C_{n,\alpha}^1 = \left\{ \theta \in \mathbb{R}^p \mid W_{1,n}(\theta) \leq F_{\chi_p^2}^{-1}(1 - \alpha) \right\}.$$

**Theorem 3.5:** *Under the assumptions of Theorem 3.3, the likelihood ratio statistic for  $\theta = \theta_0$  is asymptotically  $\chi_p^2$ :*

$$W_{1,n}(\theta_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi_p^2$$

and  $C_{n,\alpha}^1$  is then an asymptotic confidence region of nominal level  $1 - \alpha$ .

To test a sub-vector of the parameter, we can also build the corresponding empirical likelihood ratio (Guggenberger and Smith 2005; Kitamura 1997; Kitamura et al. 2004; Qin and Lawless 1994). Let  $\theta = (\gamma, \beta)'$  be in  $\mathbb{R}^q \times \mathbb{R}^{p-q}$ , where  $\gamma \in \mathbb{R}^q$  is the parameter of interest and  $\beta \in \mathbb{R}^{p-q}$  is a nuisance parameter. Assume that the true value of the

parameter of interest is  $\gamma_0$ . The empirical likelihood ratio statistic in this case becomes

$$W_{2,n}(\gamma) = 2 \cdot \left( \inf_{\beta} r_n((\gamma, \beta)') - \inf_{\theta} r_n(\theta) \right) = 2 \cdot \left( \inf_{\beta} r_n((\gamma, \beta)') - r_n(\tilde{\theta}_n) \right),$$

and the empirical likelihood confidence region is given by

$$C_{n,\alpha}^2 = \left\{ \gamma \in \mathbb{R}^q \mid W_{2,n}(\gamma) \leq F_{\chi_q^2}^{-1}(1 - \alpha) \right\}.$$

**Theorem 3.6:** *Under the assumptions of Theorem 3.3,*

$$W_{2,n}(\gamma_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi_q^2$$

and  $C_{n,\alpha}^2$  is then an asymptotic confidence region of nominal level  $1 - \alpha$ .

## 4. The case of general Harris chains

### 4.1. Algorithm

As explained in the introduction, the splitting technique introduced in Nummelin (1978) allows us to extend our algorithm to general Harris recurrent chains. The idea is to extend the original chain to a “virtual” chain with an atom. The splitting technique relies on the crucial notion of *small set*. Additional definitions are needed: a set  $S \in \mathcal{E}$  is said to be *small* if there exist  $\delta > 0$ , a positive integer  $q$  and a probability measure  $\Phi$  supported by  $S$  such that, for all  $x \in S$ ,  $A \in \mathcal{E}$ ,

$$\Pi^q(x, A) \geq \delta \Phi(A), \tag{3}$$

$\Pi^q$  being the  $q$ -th iterate of the transition probability  $\Pi$ . Note that an accessible small set always exists for  $\psi$ -irreducible chains (Jain and Jamison 1967).

In the case  $q > 1$ , a first step is typically to reduce the order to 1 by stacking lagged values (an example is given in section 5.2). Nevertheless, this complicates the exposition and the demonstrations since the resulting transition probability has no density and since the splitting technique leads to 1-dependence instead of independence. See Bertail et al. (2009) and Adamczak (2008) on that issues. For simplicity, we assume in the following that  $q = 1$  and that  $\Phi$  has a density  $\phi$  with respect to some reference measure  $\lambda(\cdot)$ .

The idea to construct the split chain  $\tilde{X} = (X, W)$  is the following:

- if  $X_i \notin S$ , generate (conditionally to  $X_i$ )  $W_i$  as a Bernoulli random value, with probability  $\delta$ .
- if  $X_i \in S$ , generate (conditionally to  $X_i$ )  $W_i$  as a Bernoulli random value, with probability  $\delta \phi(X_{i+1})/p(X_i, X_{i+1})$ ,

where  $p$  is the transition density of the chain  $X$  with respect to  $\lambda$ . This construction essentially relies on the fact that under the minorization condition (3),  $\Pi(x, A)$  may be written on  $S$  as a mixture:  $\Pi(x, A) = (1 - \delta)(\Pi(x, A) - \delta \Phi(A))/(1 - \delta) + \delta \Phi(A)$ , which

is constant (independent of the starting point  $x$ ) when the second component is picked (see Bertail and Cléménçon 2006, for details).

When constructed this way, the split chain is an atomic Markov chain, with marginal distribution equal to the original distribution of  $X$  (see Meyn and Tweedie 2009, page 427). The atom is then  $A = S \times \{1\}$ . In practice, we will only need to know when the split chain visits the atom, i.e. we only need to simulate  $W_i$  when  $X_i \in S$ .

Those visits to the atom are therefore the date of regeneration of the chain, and the number of visits acts as a sample size. In practice, the choice of the small set is then decisive for the performance of the algorithm. A balance needs to be achieved: if  $S$  were chosen too large, it would be visited very often, but the minorization condition (3) would likely be poor and therefore  $\delta$  would be small. This would lead to many realization  $W_i = 0$  and few  $W_i = 1$ . Most of the visits of  $X_i$  to the small set would then be wasted since they would not give a regeneration time. This balance is not a curse: it gives a natural data-driven tuning of the small set and prevent from the difficulties rising in the choice of kernel bandwidth for example. For a discussion on the practical choice of the small set, see Bertail and Cléménçon (2006).

The return time conditions are now defined as uniform moment condition over the small set:

$$\mathbf{H0}(S, \kappa) : \sup_{x \in S} \mathbb{E}_x[\tau_S^\kappa] < \infty,$$

$$\mathbf{H0}(S, \kappa, \nu) : \mathbb{E}_\nu[\tau_S^\kappa] < \infty.$$

The Block-moment conditions become

$$\mathbf{H1}(S, \kappa, m) : \sup_{x \in S} \mathbb{E}_x \left[ \left( \sum_{i=1}^{\tau_S} \|m(X_i, \theta_0)\| \right)^\kappa \right] < \infty,$$

$$\mathbf{H1}(S, \kappa, \nu, m) : \mathbb{E}_\nu \left[ \left( \sum_{i=1}^{\tau_S} \|m(X_i, \theta_0)\| \right)^\kappa \right] < \infty.$$

Unfortunately, the Nummelin technique involves the transition density of the chain, which is of course unknown in a nonparametric approach. An approximation  $p_n$  of this density can however be computed easily by using standard kernel methods. This leads us to the following version of the empirical likelihood program.

**Algorithm 1** Approximate regenerative block EL construction:

- (1) Find an estimator  $p_n$  of the transition density (for instance a Nadaraya-Watson estimator).
- (2) Choose a small set  $S$  and a density  $\phi$  on  $S$  and evaluate  $\delta = \min_{x, y \in S} \left\{ \frac{p_n(x, y)}{\phi(y)} \right\}$ .
- (3) When  $X$  visits  $S$ , generate  $\widehat{W}_i$  as a Bernoulli with parameter  $\delta \phi(X_{i+1}) / p_n(X_i, X_{i+1})$ . If  $\widehat{W}_i = 1$ , the approximate split chain  $(X_i, \widehat{W}_i) = \widehat{X}_i$  visits the atom  $A = S \times \{1\}$  and  $i$  is an approximate regenerative time. These times define the approximate return times  $\widehat{\tau}_A(j)$ .
- (4) Count the number of visits to  $A$  up to time  $n$ :  $\widehat{l}_n + 1 = \sum_{i=1}^n \mathbb{1}_{\widehat{X}_i \in A}$ .
- (5) Divide the observed trajectory  $X^{(n)} = (X_1, \dots, X_n)$  into  $\widehat{l}_n + 2$  blocks corresponding to the pieces of the sample path between approximate return times to the atom

A,

$$\begin{aligned}\widehat{B}_0 &= (X_1, \dots, X_{\widehat{\tau}_A(1)}), \widehat{B}_1 = (X_{\widehat{\tau}_A(1)+1}, \dots, X_{\widehat{\tau}_A(2)}), \dots, \\ \widehat{B}_{\widehat{l}_n} &= (X_{\widehat{\tau}_A(\widehat{l}_n)+1}, \dots, X_{\widehat{\tau}_A(\widehat{l}_n+1)}), \widehat{B}_{\widehat{l}_n+1}^{(n)} = (X_{\widehat{\tau}_A(\widehat{l}_n+1)+1}, \dots, X_n),\end{aligned}$$

with the convention  $\widehat{B}_{\widehat{l}_n+1}^{(n)} = \emptyset$  when  $\widehat{\tau}_A(\widehat{l}_n + 1) = n$ .

(6) Drop the first block  $\widehat{B}_0$ , and the last one  $\widehat{B}_{\widehat{l}_n+1}^{(n)}$  (possibly empty when  $\widehat{\tau}_A(\widehat{l}_n + 1) = n$ ).

(7) Define

$$M(\widehat{B}_j, \theta) = \sum_{i=\widehat{\tau}_A(j)+1}^{\widehat{\tau}_A(j+1)} m(X_i, \theta).$$

Evaluate the empirical log-likelihood ratio  $\widehat{r}_n(\theta)$  (practically on a grid of the set of interest):

$$\widehat{r}_n(\theta) = - \sup_{(q_1, \dots, q_{\widehat{l}_n})} \left\{ \log \left[ \prod_{j=1}^{\widehat{l}_n} \widehat{l}_n q_j \right] \middle| \sum_{j=1}^{\widehat{l}_n} q_j \cdot M(\widehat{B}_j, \theta) = 0, \sum_{j=1}^{\widehat{l}_n} q_j = 1 \right\}.$$

Using Lagrange arguments, this can be more easily calculated as

$$\widehat{r}_n(\theta) = \sup_{\lambda \in \mathbb{R}^p} \left\{ \sum_{j=1}^{\widehat{l}_n} \log \left[ 1 + \lambda' M(\widehat{B}_j, \theta) \right] \right\}.$$

## 4.2. Main theorem

The practical use of this algorithm crucially relies on the preliminary computation of a consistent estimator of the transition density. We thus consider some conditions on the uniform consistency of the density estimator  $p_n$ . These assumptions are satisfied for the usual kernel or wavelets estimators of the transition density.

**H3** For a sequence of nonnegative real numbers  $(\alpha_n)_{n \in \mathbb{N}}$  converging to 0 as  $n \rightarrow \infty$ ,  $p(x, y)$  is estimated by  $p_n(x, y)$  at the rate  $\alpha_n$  for the mean square error when error is measured by the  $L^\infty$  loss over  $S \times S$ :

$$\mathbb{E}_\nu \left[ \sup_{(x, y) \in S \times S} |p_n(x, x') - p(x, x')|^2 \right] = \mathcal{O}_\nu(\alpha_n), \text{ as } n \rightarrow \infty.$$

**H4** The minorizing probability  $\Phi$  is such that  $\inf_{x \in S} \phi(x) > 0$ .

**H5** The densities  $p$  and  $p_n$  are bounded over  $S^2$  and  $\inf_{x, y \in S} p_n(x, y)/\phi(y) > 0$ .

Since the choice of  $\Phi$  is left to the statistician, we can use for instance the uniform distribution over  $S$ , even if it may not be optimal to do so. In such a case, **H4** is

automatically satisfied. Similarly, it is not difficult to construct an estimator  $p_n$  satisfying the constraints of **H5**.

Results of the previous section can then be extended to Harris chains:

**Theorem 4.1:** *Let  $\mu$  be the invariant measure of the chain, and  $\theta_0 \in \mathbb{R}^p$  be the parameter of interest, satisfying  $\mathbb{E}_\mu[m(X, \theta_0)] = 0$ . Consider  $A = S \times \{1\}$  an atom of the split chain,  $\tau_A$  the hitting time of  $A$  and  $B = (X_1, \dots, X_{\tau_A})$ . Assume the hypotheses **H3**, **H4** and **H5**, and suppose that  $\mathbb{E}_A[M(B, \theta_0)M(B, \theta_0)']$  is of full rank.*

(a) *If **H0**( $S, 4, \nu$ ) and **H0**( $S, 2$ ) holds as well as **H1**( $S, 4, \nu, m$ ) and **H1**( $S, 2, m$ ), then we have in the just-identified case ( $r = p$ ):*

$$2\hat{r}_n(\theta_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi_p^2$$

and therefore

$$\hat{C}_{n,\alpha} = \left\{ \theta \in \mathbb{R}^p \mid 2 \cdot \hat{r}_n(\theta) \leq F_{\chi_p^2}^{-1}(1 - \alpha) \right\}.$$

is an asymptotic confidence region of level  $1 - \alpha$ .

(b) *Under the additional assumptions **H2**(a), **H2**(b) and **H2**(d),*

$$\hat{\theta} = \arg \inf_{\theta \in \Theta} \{ \hat{r}_n(\theta) \}$$

is a consistent estimator of  $\theta_0$ . If in addition **H2**(c) holds, then  $\sqrt{n}(\hat{\theta} - \theta_0)$  is asymptotically normal.

(c) *In the case of over-identification ( $r > p$ ), we have:*

$$\widehat{W}_{1,n}(\theta_0) = 2\hat{r}_n(\theta_0) - 2\hat{r}_n(\hat{\theta}) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi_p^2$$

and

$$\widehat{C}_{n,\alpha}^1 = \left\{ \theta \in \mathbb{R}^p \mid \widehat{W}_{1,n}(\theta) \leq F_{\chi_p^2}^{-1}(1 - \alpha) \right\},$$

is an asymptotic confidence region of level  $1 - \alpha$ . The moment equation (1) can be tested by using the following convergence in law:

$$2\hat{r}_n(\hat{\theta}) \xrightarrow[n \rightarrow \infty]{\text{under (1)}} \chi_{r-p}^2.$$

(d) *Let  $\theta = (\gamma, \beta)'$ , where  $\gamma \in \mathbb{R}^q$  and  $\beta \in \mathbb{R}^{p-q}$ . Under the hypotheses  $\gamma = \gamma_0$ ,*

$$\widehat{W}_{2,n}(\gamma_0) = 2 \inf_{\beta} \hat{r}_n((\gamma_0, \beta)') - 2\hat{r}_n(\hat{\theta}) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi_q^2$$

and then

$$\widehat{C}_{n,\alpha}^2 = \left\{ \gamma \in \mathbb{R}^q \mid \widehat{W}_{2,n}(\gamma) \leq F_{\chi_q^2}^{-1}(1 - \alpha) \right\},$$

is an asymptotic confidence region of level  $1 - \alpha$  for the parameter of interest  $\gamma$ .

## 5. Some simulation results

### 5.1. Illustrative example

We introduce here an example in order to illustrate the method in a very simple setting and compare ReBEL with BEL (Kitamura 1997) in a situation that favor none.

We consider a AR(1), which is also a Markov chain of order 1, defined as follows:

$$X_0 = 0, \quad \varepsilon_i \sim \mathcal{U}([- \sqrt{12}, \sqrt{12}]) \text{ and } X_i = 0.9X_{i-1} + \varepsilon_i,$$

$\varepsilon_t$  being i.i.d. uniformly distributed random variables of mean 0 and variance 1. Since  $0.9 < 1$ , the chain is recurrent. The parameter of interest is the mean of the chain,  $\mu = 0$ .

We choose in the simulations a small set  $S = [-a; a]$ . On each simulation,  $a$  is chosen from a small grid in order to maximize the number of regeneration times as explained in section 4.1. ReBEL algorithm is then implemented and coverage probabilities are calculated for the nominative level 95%.

We also compute BEL coverage probabilities, using non-overlapping blocks of constant length the integer part of  $n^{1/3}$ , where  $n$  is the data set length.

We get the following results, for 10 000 replications:

table A.5 should approximatively here

On these simulations, ReBEL seems to be better fitted. This may be due to the fact that ReBEL' small set length is data driven whereas BEL's blocks length is constant over the replications. In the following section, we set the small set once for all the replication.

### 5.2. Estimation of the threshold crossing rate of a TGARCH

The aim of this section is to show that ReBEL can be adapted to complex data and can outperform competing methods. Some applications of empirical likelihood to dependent data have been carried out, such as Li and Wang (2003) on Stanford Heart Transplant data or Owen (2001) on bristlecone pine tree rings. In his book, Owen motivates his use of empirical likelihood to study the tree rings data set by its asymmetry: "we could not capture such asymmetry in an AR model with normally distributed errors" (Owen 2001, page 168).

To motivate the use of empirical likelihood, we propose here to generate data sets with strong asymmetry properties to illustrate the applicability of the method. For this, we consider a family of models introduced to study financial data, the TGARCH (Rabemananjara and Zakoïan 1993). This model has been designed to handle non symmetric data, such as stock return series in presence of asymmetry in the volatility. We think in particular to applications on modeling electricity prices series (Cornec and Harari-Kermadec 2008). These series are very hard to model because of their very asymmetric behavior and because of the presence of very sharp peaks alternating with periods of low volatility. Application of ReBEL to these series seems to be promising, see Cornec and Harari-Kermadec (2008).



The data generating process is the following:

$$\begin{cases} X_i = 0.97X_{i-1} + \varepsilon_i & \text{with } X_0 = 0, \\ \varepsilon_i = \sigma_i \nu_i & \text{with } \nu_i \sim \text{NID}(0, 1), \\ \sigma_i = 1 + 0.5|\varepsilon_{i-1}| + 0.4\varepsilon_{i-1}^+ & \text{with } \varepsilon_0 = 0. \end{cases}$$

where the  $\nu_i$  are standard normal random values independent of all other random variables and  $x^+$  is the positive part of  $x$ :  $x^+ = \max\{0, x\}$ . Of course, in the following, this generating mechanism is considered unknown. Retrieving the underlying mechanism by just looking at the data is a difficult task and this motivates the use of a non parametric approach in this context.

It is straightforward that  $(X, \varepsilon)$  is a Markov chain of order 1. As  $\varepsilon_{i-1} = X_{i-1} - 0.97X_{i-2}$ , it is immediate that  $X$  is a Markov chain of order 2. ReBEL algorithm can then be applied to  $\mathbf{X}_i = (X_i, X_{i-1})$ , which is a Markov chain of order 1.

In practice, the order  $k$  of the Markov chain is unknown and is therefore to be estimated. We propose the following heuristic procedure to estimate the order:

- (1) Suppose  $k = 1$ .
- (2) Build the block according to Algorithm 1.
- (3) Evaluate the moment condition over the blocks:  $Y_j = M(\hat{B}_j, \theta)$ .
- (4) Perform a test of independence (or at least of non correlation) of the  $(Y_1, \dots, Y_{l_n-1})$ , for example by testing the nullity of  $\rho$  given by  $Y_i = \rho Y_{i-1} + \nu_i$ . Other tests may be considered as well, such as tests based on kernel estimators of the density.
- (5) If the independence (or non correlation) is rejected, set  $k = k + 1$  and restart at point 2.

In order to apply Kitamura (1997)'s Block Empirical Likelihood (BEL),  $X$  must be weakly dependent. As the sum of the coefficients of  $|\varepsilon_{i-1}|$  and  $\varepsilon_{i-1}^+$  is smaller than 1, the volatility of the data generating process is contracting. Therefore one can easily check the weak dependence of the process.

### 5.3. Confidence intervals

We are interested in estimating the probability of crossing a high threshold. This is an interesting problem because of the asymmetry of the data and a problem of practical interest for electricity prices. Indeed, production means are only profitable above some level. The probability of crossing the profitability threshold is therefore essential to estimate. The parameter of interest is defined here as:

$$\theta_0 = \mathbb{E}_\mu [\mathbb{1}_{\{X_i \geq 10\}}] = \mathbb{P}_\mu (X_i \geq 10).$$

and its value (estimated on a simulated data set of size  $10^6$ ) is  $\theta_0 = 0.1479$ . A first advantage of ReBEL is that such a parameter, defined with respect to the underlying invariant measure  $\mu$ , is naturally handled by this method, whereas no unbiased estimating equation is available for BEL.

We simulate a data set of length 1000 and perform a test to estimate the order of the chain. We build an estimator  $p_n$  of the transition density  $p$  based on Gaussian kernels. The hypothesis  $q = 1$  is rejected whereas  $q = 2$  is not. As the chain is then considered 2-dimensional, we consider a small set of the form  $S^2$  where  $S$  is an interval. The

interval  $S$  has been chosen empirically to maximize the number of blocks and is equal to  $[-1.3; 4.7]$ . It is set once and for all and is not updated at each replication. On the graphic corresponding to one simulation,  $\mathbf{X}$  is in the small set  $S^2$  when the trajectory of  $X$  is in between the 2 plain black lines  $y = -1.3$  and  $y = 4.7$  for two consecutive times. For  $i$  such that  $\mathbf{X}_i$  visits  $S^2$ , we generate a Bernoulli  $B_i$  as in Algorithm 1, and if  $B_i = 1$ ,  $i$  is a approximate renewal time. On the simulation,  $S$  is visited 231 times, leading to 18 renewal times, marked by a vertical green line.

Figure A1 should be approximately here

The block length adapts to the local behavior of the chain: regions of low volatility lead to small blocks (between 500 and 700) whereas regions with high values lead to larger blocks (like the 142-484 block). It can be noticed that high values concentrate in few blocks, because the dependence is well captured by Algorithm 1. BEL procedure leads to constant length block which cannot adapt to the dependence structure. As suggested by Hall et al. (1995), the BEL blocks used in the following are of length  $n^{1/3} = 10$  and then the chain is divided into 100 non overlapping blocks. The overlapping block perform poorly and won't be considered in the following.

Now that we have ReBEL approximately regenerative blocks, we can apply Theorem 4.1(a) to obtain a confidence interval for  $\theta$ . We give a BEL confidence interval as well for comparison. We also consider two simpler methods as references for the performances of ReBEL: the simple sample mean  $mean = n^{-1} \sum_i \mathbb{1}_{\{X_i \geq 10\}}$  and the mean over the regenerative blocks

$$trunc = \frac{\sum_{k=1}^{l_n} \sum_{i=\hat{\tau}(k)+1}^{\hat{\tau}(k+1)} \mathbb{1}_{\{X_i \geq 10\}}}{\sum_{k=1}^{l_n} (\hat{\tau}(k+1) - \hat{\tau}(k))} = \frac{\sum_{k=1}^{l_n} \sum_{i=\hat{\tau}(k)+1}^{\hat{\tau}(k+1)} \mathbb{1}_{\{X_i \geq 10\}}}{\hat{\tau}(l_n + 1) - \hat{\tau}(1)}.$$

The simple mean do not deal with the dependence and we expect it to perform poorly. The second reference method *trunc* uses the splitting technique in its expression, but in practice it only differs from *mean* by the fact that it discards the first and last blocks.

An important point here is that to build confidence intervals with these two methods, an estimator of the variance is needed. In fact, if these estimators seem much simpler than BEL and ReBEL, the difficulty is mainly transferred to the estimation of their variances. This issue is difficult in a general dependence setting. In the applications, we used a bootstrap estimator of the variance of *mean* and *trunc* according to Götze and Kunsch (1996).

Having in mind that difficulty, it is important to stress that ReBEL and BEL confidence intervals do not rely on an estimation of the variance of the estimator. This property is well-known for methods based on empirical likelihood that automatically estimate a variance at each point of the confidence interval, see for example the Continuously updated GMM Antoine et al. (2007). Additional results on the self-normalized properties of these methods have been investigated in Bertail et al. (2008).

Figure A2 should be approximately here

*Mean* and BEL estimators and confidence intervals appear biased to the right. This is most likely due to the effect of data from the first and last blocks, discarded by ReBEL and *trunc*. It can be noticed that BEL blocks being more numerous, the confidence interval is tighter for BEL than for ReBEL.

To compare the considered methods, we also compute coverage probabilities and type-II errors (which is equivalent to power in terms of test) of confidence intervals with nominal level 95%. To test the behavior under the alternative, we evaluate the statistics at the erroneous points  $\theta = \theta_0 + 5/\sqrt{n}$  and  $\theta = \theta_0 + 10/\sqrt{n}$  and check if the null hypotheses if rejected or not.

The 2 000 simulation results are summarized in Table A2, for  $n = 1000, 5000$  and  $10000$ .

Table A2 should be approximately here

Globally, ReBEL's coverage probabilities are better than BEL's, whereas its type-II error are bigger. This is coherent with Figure A2: ReBEL confidence interval leads to better coverage probabilities but is larger than BEL's (and therefore type-II errors are bigger for ReBEL). *Mean* and *trunc* perform well for  $n = 1000$  but show some limits for  $n = 5000$  and  $10000$ . It seems that ReBEL is the only method converging to the nominal level 95%.

Coverage probabilities at other nominal level can also be investigated, and we make a Monte-Carlo experiment (10 000 repetitions) in order to confirm the adequacy to the asymptotic distribution achieved by the ReBEL algorithm. Data sets length are 10 000.

Figure A3 should be approximately here

Figure A3 shows the adequacy of the log likelihood to the asymptotic distribution given by Theorem 4.1. The QQ-plots is almost linear and is close to the  $45^\circ$  line.

## 6. Conclusion

This paper propose an alternative point of view on dependent data sets and a corresponding semi-parametric methodology. Random length blocks allow to adapt to the dependence structure of the data. We have shown that ReBEL enjoys desirable properties corresponding to that of optimal reference methods for strong-mixing series. Simulations indicate that our algorithm at least competes with Kitamura's BEL when both methods can be applied.

This method seems to be a promising tool to handle dependent data when classical parametric models do not perform well, for example in presence of asymmetry and non normality of the innovations.

## References

- Adamczak, R. (2008), "A tail inequality for suprema of unbounded empirical processes with applications to Markov chains," *Electron. J. Probab.*, 13, 1000–1034.
- Antoine, B., Bonnal, H., and Renault, E. (2007), "On the efficient use of the informational content of estimating equations: Implied probabilities and Euclidean empirical likelihood," *Journal of Econometrics*, 138, 461–487.
- Asmussen, S., *Applied Probabilities and Queues*, Springer, New York (2003).
- Athreya, K., and Ney, P. (1978), "A new approach to the limit theory of recurrent Markov chains," *Trans. Amer. Math. Soc.*, 245, 493–501.
- Bertail, P., and Cléménçon, S. (2004), "Approximate Regenerative Block-Bootstrap for

- Markov Chains: second-order properties,” in *Compstat 2004 Proceedings*, Physica Verlag.
- Bertail, P., and Cl emen on, S. (2004), “Edgeworth expansions for suitably normalized sample mean statistics of atomic Markov chains,” *Prob. Th. Rel. Fields*, 130, 388–414.
- Bertail, P., and Cl emen on, S. (2006), “Regeneration-based statistics for Harris recurrent Markov chains,” in *Dependence in Probability and Statistics*, eds. P. Bertail, P. Doukhan and P. Soulier, Vol. 187 of *Lecture Notes in Statistics*, Springer.
- Bertail, P., and Cl emen on, S. (2006), “Regenerative Block Bootstrap for Markov Chains,” *Bernoulli*, 12, 689–712.
- Bertail, P., Cl emen on, S., and Tressou, J. (2009), “Renewal approach to U-Statistics for Markovian data,” in *41 emes Journ ees de Statistique, SFdS, Bordeaux*, <http://hal.archives-ouvertes.fr/inria-00386724/>.
- Bertail, P., Gautherat, E., and Harari-Kermadec, H. (2008), “Exponential bounds for multivariate self-normalized sums,” *Electronic communication in probability*, 13, 628–640.
- Bertail, P., Harari-Kermadec, H., and Ravaille, D. (2007), “ $\varphi$ -Divergence empirique et vraisemblance empirique g eneralis ee,” *Annales d’ conomie et de Statistique*, pp. 131–157.
- Bolthausen, E. (1982), “The Berry-Esseen Theorem for strongly mixing Harris recurrent Markov Chains,” *Z. Wahr. Verw. Gebiete*, 60, 283–289.
- Cornec, M., and Harari-Kermadec, H. (2008), “Modeling spot electricity prices with regenerative blocks,” in *IASTED ASM*.
- Doukhan, P., and Ango Nze, P. (2004), “Weak dependence, models and applications to econometrics,” *Econometric Theory*, 20, 995–1045.
- Guggenberger, P., and Smith, R.J. (2005), “Generalized empirical likelihood estimators and tests under weak, partial and strong identification,” *Econometric Theory*, 21, 667–709.
- G otze, F., and Kunsch, H.R. (1996), “Second order correctness of the blockwise bootstrap for stationary observations,” *Annals of Statistics*, 24, 1914–1933.
- Hall, P., Horowitz, J., and Jing, B.Y. (1995), “On blocking rules for the bootstrap with dependent data,” *Biometrika*, 82, 561–574.
- Hall, P., and Yao, Q. (2003), “Data tilting for time series,” *Journal of the Royal Statistical Society, Series B*, 65, 425–442.
- Horowitz, J. (2003), “The Bootstrap in Econometrics,” *Statistical Science*, 18, 211–218.
- Jain, J., and Jamison, B. (1967), “Contributions to Doebelin’s theory of Markov processes,” *Z. Wahrsch. Verw. Geb.*, 8, 9–40.
- Kallenberg, O., *Foundations of modern probability*, Springer-Verlag, New York (2002).
- Kitamura, Y. (1997), “Empirical likelihood methods with weakly dependent processes,” *Annals of Statistics*, 25, 2084–2102.
- Kitamura, Y., “Empirical likelihood methods in Econometrics: theory and practice.” Cowles Foundation discussion paper n 1569. Available at SSRN: <http://ssrn.com/abstract=917901> (2006).
- Kitamura, Y., Tripathi, G., and Ahn, H. (2004), “Empirical likelihood-based inference in conditional moment restriction models,” *Econometrica*, 72, 1667–1714.
- Knight, F. (1975), “A Predictive View of Continuous Time Processes,” *Annals of Probability*, 3, 573–596.
- Kunsch, H.R. (1989), “The jackknife and the bootstrap for general stationary observations,” *Annals of Statistics*, 17, 1217–1241.
- Lahiri, S.N., *Resampling methods for dependent Data*, Springer (2003).

- Li, G., and Wang, Q.H. (2003), “EMPIRICAL LIKELIHOOD REGRESSION ANALYSIS FOR RIGHT CENSORED DATA,” *Statistica Sinica*, 13, 51–68.
- Lin, L., and Zhang, R. (2001), “Blockwise empirical Euclidean likelihood for weakly dependent processes,” *Statistics and Probability Letters*, 53, 143–152.
- Meyn, S.P., and Tweedie, R.L., *Markov Chains and Stochastic Stability*, Cambridge university press (2009).
- Newey, W.K. (1985), “Generalized method of moments specification testing,” *Journal of Econometrics*, 29, 229–256.
- Newey, W.K., and Smith, R.J. (2004), “Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators,” *Econometrica*, 72, 219–255.
- Nummelin, E. (1978), “A splitting technique for Harris recurrent chains,” *Z. Wahrsch. Verw. Gebiete*, 43, 309–318.
- Owen, A.B. (1988), “Empirical Likelihood Ratio Confidence intervals for a Single Functional,” *Biometrika*, 75, 237–249.
- Owen, A.B. (1990), “Empirical likelihood ratio confidence regions,” *Annals of Statistics*, 18, 90–120.
- Owen, A.B., *Empirical Likelihood*, Chapman and Hall/CRC, Boca Raton (2001).
- Qin, Y.S., and Lawless, J. (1994), “Empirical likelihood and General Estimating Equations,” *Annals of Statistics*, 22, 300–325.
- Rabemananjara, R., and Zakoïan, J.M. (1993), “Threshold Arch Models and Asymmetries in Volatility,” *Journal of Applied Econometrics*, 8, 31–49.
- Revuz, D., *Markov Chains*, North-Holland (1984).
- Synowiecki, R. (2007), “Consistency and application of moving block bootstrap for non-stationary time series with periodic and almost periodic structure,” *Bernoulli*, 13, 1151–1178.
- Teicher, H., and Chow, Y.S., *Probability Theory: Independence, Interchangeability, Martingales*, Springer-Verlag, New York (1988).
- Tjøstheim, D. (1990), “Non-linear time series and Markov chains,” *Advances in Applied Probability*, 22, 587–611.
- Wang, S., Qian, L., and Carroll, R.J. (2010), “Generalized empirical likelihood methods for analyzing longitudinal data,” *Biometrika*, 97, 79–93.

## Appendix A. Proofs

### A.1. Lemmas for the atomic case

Denote  $Y_j = M(B_j, \theta_0)$ ,  $\bar{Y} = 1/l_n \sum_{j=1}^{l_n} Y_j$  and define

$$S_{l_n}^2 = 1/l_n \sum_{j=1}^{l_n} M(B_j, \theta_0)M(B_j, \theta_0)' = 1/l_n \sum_{j=1}^{l_n} Y_j Y_j' \text{ and } S_{l_n}^{-2} = (S_{l_n}^2)^{-1}.$$

To demonstrate Theorem 3.1, we need 2 technical lemmas.

**Lemma A.1:** *Assume that  $\mathbb{E}_A[M(B, \theta_0)M(B, \theta_0)']$  exists and is full-rank, with ordered eigenvalues  $\sigma_p \geq \dots \geq \sigma_1 > 0$ . Then, assuming  $\mathbf{HO}(1, \nu)$  and  $\mathbf{HO}(1)$ , we have*

$$S_{l_n}^2 \rightarrow_{\nu} \mathbb{E}_A[M(B, \theta_0)M(B, \theta_0)'].$$

Therefore, for all  $u \in \mathbb{R}^p$  with  $\|u\| = 1$ ,

$$\sigma_1 + o_\nu(1) \leq u' S_{l_n}^2 u \leq \sigma_p + o_\nu(1).$$

**A.1.0.1. Proof:** The convergence of  $S_{l_n}^2$  is a LLN for the sum of a random number of random variables, and is a straightforward corollary of the Theorem 6 of (Teicher and Chow 1988, chapter 5.2, page 131).

**Lemma A.2:** Assuming **H0**(1,  $\nu$ ), **H0**(2) and **H1**(2,  $m$ ), we have

$$\max_{1 \leq j \leq l_n} \|Y_j\| = o_\nu(n^{1/2}).$$

**A.1.0.2. Proof:** By **H1**(2,  $m$ ),

$$\mathbb{E}_A \left[ \left( \sum_{i=1}^{\tau_1} \|m(X_i, \theta_0)\| \right)^2 \right] < \infty,$$

and then,

$$\mathbb{E}_A[\|Y_1\|^2] = \mathbb{E}_A \left[ \left\| \sum_{i=1}^{\tau_1} m(X_i, \theta_0) \right\|^2 \right] < \infty.$$

By Lemma A.1 of Antoine et al. (2007), the maximum of  $n$  i.i.d. real-valued random variables with finite variance is  $o(n^{1/2})$ . Let  $Z_n$  be the maximum of  $n$  independent copies of  $\|Y_1\|$ ,  $Z_n$  is then such as  $Z_n = o_\nu(n^{1/2})$ . As  $l_n$  is smaller than  $n$ ,  $\max_{1 \leq j \leq l_n} \|Y_j\|$  is bounded by  $Z_n$  and therefore,  $\max_{1 \leq j \leq l_n} \|Y_j\| = o_\nu(n^{1/2})$ .

## A.2. Proof of Theorem 3.1

The likelihood ratio statistic  $r_n(\theta_0)$  is the supremum over  $\lambda \in \mathbb{R}^p$  of  $\sum_{j=1}^{l_n} \log(1 + \lambda' Y_j)$ . The first order condition at the supremum  $\lambda_n$  is then:

$$1/l_n \sum_{j=1}^{l_n} \frac{Y_j}{1 + \lambda_n' Y_j} = 0. \tag{A1}$$

Multiplying by  $\lambda_n$  and using  $1/(1+x) = 1 - x/(1+x)$ , we have

$$1/l_n \sum_{j=1}^{l_n} (\lambda_n' Y_j) \left( 1 - \frac{\lambda_n' Y_j}{1 + \lambda_n' Y_j} \right) = 0, \text{ and then } \lambda_n' \bar{Y} = 1/l_n \sum_{j=1}^{l_n} \frac{\lambda_n' Y_j Y_j' \lambda_n}{1 + \lambda_n' Y_j}.$$

Now we may bound the denominators  $1 + \lambda'_n Y_j$  by  $1 + \|\lambda_n\| \max_j \|Y_j\|$  and then

$$\lambda'_n \bar{Y} = 1/l_n \sum_{j=1}^{l_n} \frac{\lambda'_n Y_j Y'_j \lambda_n}{1 + \lambda'_n Y_j} \geq \frac{\lambda'_n S_{l_n}^2 \lambda_n}{(1 + \|\lambda_n\| \max_j \|Y_j\|)}$$

Multiply both sides by the denominator,  $\lambda'_n \bar{Y} (1 + \|\lambda_n\| \max_j \|Y_j\|) \geq \lambda'_n S_{l_n}^2 \lambda_n$  or

$$\lambda'_n \bar{Y} \geq \lambda'_n S_{l_n}^2 \lambda_n - \|\lambda_n\| \max_j \|Y_j\| \lambda'_n \bar{Y}.$$

Dividing by  $\|\lambda_n\|$  and setting  $u = \lambda_n / \|\lambda_n\|$ , we have

$$u' \bar{Y} \geq \|\lambda_n\| \left[ u' S_{l_n}^2 u - \max_j \|Y_j\| u' \bar{Y} \right]. \quad (\text{A2})$$

Now we control the terms between the square brackets. First, by Lemma A.1,  $u' S_{l_n}^2 u$  is bounded between  $\sigma_1 + o_\nu(1)$  and  $\sigma_p + o_\nu(1)$ . Second, by Lemma A.2,  $\max_j \|Y_j\| = o_\nu(n^{1/2})$ . Third, the CLT applied to the  $Y_j$ 's gives  $\bar{Y} = \mathcal{O}_\nu(n^{-1/2})$ . Then, inequality (A2) gives

$$\mathcal{O}_\nu(n^{-1/2}) \geq \|\lambda_n\| \left[ u' S_{l_n}^2 u - o_\nu(n^{1/2}) \mathcal{O}_\nu(n^{-1/2}) \right] = \|\lambda_n\| (u' S_{l_n}^2 u + o_\nu(1)),$$

and  $\|\lambda_n\|$  is then  $\mathcal{O}_\nu(n^{-1/2})$ .

Using the first order condition (A1) as well as the equality  $1/(1+x) = 1-x+x^2/(1+x)$ , we get

$$0 = 1/l_n \sum_{j=1}^{l_n} Y_j \left( 1 - \lambda'_n Y_j + \frac{(\lambda'_n Y_j)^2}{1 + \lambda'_n Y_j} \right) = \bar{Y} - S_{l_n}^2 \lambda_n + 1/l_n \sum_{j=1}^{l_n} \frac{Y_j (\lambda'_n Y_j)^2}{1 + \lambda'_n Y_j}.$$

The last term is  $o_\nu(n^{-1/2})$  by Lemma A.2 of Antoine et al. (2007) and then  $\lambda_n = S_{l_n}^{-2} \bar{Y} + o_\nu(n^{-1/2})$ .

Now, developing the log up to the second order,

$$2r_n(\theta_0) = 2 \sum_{j=1}^{l_n} \log(1 + \lambda'_n Y_j) = 2l_n \lambda'_n \bar{Y} - l_n \lambda'_n S_{l_n}^2 \lambda_n + 2 \sum_{j=1}^{l_n} \eta_j,$$

where the  $\eta_i$  are such that, for some positive  $B$  and with probability tending to 1,  $|\eta_j| \leq B |\lambda'_n Y_j|^3$ . Since, by Lemma A.2,  $\max_j \|Y_j\| = o_\nu(n^{1/2})$ ,

$$\sum_{j=1}^{l_n} \|Y_j\|^3 \leq n \max_j \|Y_j\| \left( \frac{1}{l_n} \sum_{j=1}^{l_n} \|Y_j\|^2 \right) = n o_\nu(n^{1/2}) \mathcal{O}_\nu(1) = o_\nu(n^{3/2})$$

from which we find

$$2 \sum_{j=1}^{l_n} \eta_j \leq B \|\lambda_n\|^3 \sum_{j=1}^{l_n} \|Y_j\|^3 = \mathcal{O}_\nu(n^{-3/2}) o_\nu(n^{3/2}) = o_\nu(1).$$

Finally,

$$2r_n(\theta_0) = 2l_n\lambda'_n\bar{Y} - l_n\lambda'_n S_{l_n}^2 \lambda_n + o_\nu(1) = l_n\bar{Y}' S_{l_n}^{-2}\bar{Y} + o_\nu(1) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi_p^2.$$

This concludes the proof of Theorem 3.1.

### A.3. Proof of Theorem 3.2

We keep the notations of the previous subsection. Note that instead of  $\mathbb{E}_A[Y_j] = 0$ , we have

$$\mathbb{E}_A[Y_j] = \delta \frac{\mathbb{E}_A[\tau_A]}{\sqrt{n}} \sim \delta \sqrt{\frac{\mathbb{E}_A[\tau_A]}{l_n}},$$

because  $l_n \sim n/\mathbb{E}_A[\tau_A]$ . The beginning of the proof is similar to that of Theorem 3.1: the misspecification is not significant at first order:  $\bar{Y}$  remains  $\mathcal{O}_\nu(n^{-1/2})$ . We obtain:

$$2r_n(\theta_0) = l_n\bar{Y}' S_{l_n}^{-2}\bar{Y} + o_\nu(1).$$

$\bar{Y} - \delta \sqrt{\mathbb{E}_A[\tau_A]/l_n}$  is asymptotically Gaussian with variance  $\mathbb{E}_A[M(B, \theta_0)M(B, \theta_0)']$ , which is the limit in probability of  $S_{l_n}^{-2}$ . Therefore,

$$2r_n(\theta_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi_p'^2(\delta' \Sigma^{-1} \delta).$$

### A.4. Proof of Theorem 3.3

In order to prove Theorem 3.3, we use a result established by Qin and Lawless (1994).

**Lemma A.3** Qin & Lawless, 1994 *Let  $Z, Z_1, \dots, Z_n \sim F$  be i.i.d. observations in  $\mathbb{R}^d$  and consider a function  $g : \mathbb{R}^d \times \mathbb{R}^p \rightarrow \mathbb{R}^r$  such that  $\mathbb{E}_F[g(Z, \theta_0)] = 0$ . Suppose that the following hypotheses hold:*

- (1)  $\mathbb{E}_F[g(Z, \theta_0)g'(Z, \theta_0)]$  is positive definite,
- (2)  $\partial g(z, \theta)/\partial \theta$  is continuous and bounded in norm by an integrable function  $G(z)$  in a neighborhood  $V$  of  $\theta_0$ ,
- (3)  $\|g(z, \theta)\|^3$  is bounded by  $G(z)$  on  $V$ ,
- (4) the rank of  $\mathbb{E}_F[\partial g(Z, \theta_0)/\partial \theta]$  is  $p$ ,
- (5)  $\frac{\partial^2 g(z, \theta)}{\partial \theta \partial \theta'}$  is continuous and bounded by  $G(z)$  on  $V$ .

Then, the maximum empirical likelihood estimator  $\tilde{\theta}_n$  is a consistent estimator and  $\sqrt{n}(\tilde{\theta}_n - \theta_0)$  is asymptotically normal with mean zero.

Set

$$Z = B_1 = (X_{\tau_A(1)+1}, \dots, X_{\tau_A(2)}) \in \bigcup_{n \in \mathbb{N}} \mathbb{R}^n$$



and  $g(Z, \theta) = M(B_1, \theta)$ . Expectation under  $F$  is then replaced by  $\mathbb{E}_A$ . Theorem 3.3 is a straightforward application of the Lemma A.3 as soon as the assumptions hold.

By assumption,  $\mathbb{E}_A[M(B, \theta_0)M(B, \theta_0)']$  is of full rank. This implies (1).

By **H2(a)**, there is a neighborhood  $V$  of  $\theta_0$  and a function  $N$  such that, for all  $i$  between  $\tau_A + 1$  and  $\tau_A(2)$ ,  $\partial m(X_i, \theta)/\partial \theta$  is continuous on  $V$  and bounded in norm by  $N(X_i)$ .  $\partial M(B_1, \theta)/\partial \theta$  is then continuous as a sum of continuous functions and is bounded for  $\theta$  in  $V$  by  $L(B_1) = \sum_{i=\tau_A(1)+1}^{\tau_A(2)} N(X_i)$ . Since  $N$  is such that  $\mathbb{E}_\mu[N(X)] < \infty$ , we have by Kac's Theorem,

$$\mathbb{E}_A \left[ \sum_{i=\tau_A(1)+1}^{\tau_A(2)} N(X_i) \right] / \mathbb{E}_A[\tau_A] = \mathbb{E}_A[L(B_1)] / \mathbb{E}_A[\tau_A] < \infty.$$

The bounding function  $L(B_1)$  is then integrable. This gives assumption (2). Assumption (5) is derived from **H2(c)** by the same arguments.

By **H2(d)**,  $\|m(X_i, \theta)\|^3$  is bounded by  $N(X_i)$  for  $\theta$  in  $V$ , and then

$$\|M(B_1, \theta)\|^3 \leq \sum_{i=\tau_A(1)+1}^{\tau_A(2)} \|m(X_i, \theta)\|^3 \leq \sum_{i=\tau_A(1)+1}^{\tau_A(2)} N(X_i) = L(B_1).$$

Thus,  $\|M(B_1, \theta)\|^3$  is also bounded by  $L(B_1)$  for  $\theta$  in  $V$ , and hypotheses (3) follows.

By Kac's Theorem,

$$\mathbb{E}_A[\tau_A]^{-1} \mathbb{E}_A[\partial M(B_1, \theta_0)/\partial \theta] = \mathbb{E}_\mu[\partial m(X_i, \theta_0)/\partial \theta],$$

which is supposed to be of full rank by **H2(b)**. Thus  $\mathbb{E}_A[\partial M(B_1, \theta_0)/\partial \theta]$  is of full rank and this gives assumption (4). This concludes the proof of Theorem 3.3.

Under the same hypotheses, Theorem 2 and Corollaries 4 and 5 of Qin and Lawless (1994) hold. They give respectively our Theorems 3.5, 3.4 and 3.6.

### A.5. Proof of Theorem 4.1

Suppose that we know the real transition density  $p$ . The chain can then be split with the Nummelin technique as above. We get an atomic chain  $\tilde{X}$ . Let's denote by  $B_j$  the blocks obtained from this chain. The Theorem (3.1) can then be applied to  $Y_j = M(B_j, \theta_0)$ .

Unfortunately  $p$  is unknown and then we can not use the  $Y_j$ . Instead, we have the vectors  $\hat{Y}_j = M(\hat{B}_j, \theta_0)$ , built on approximatively regenerative blocks. To prove the Theorem 4.1, we essentially need to control the difference between the two statistics  $\bar{Y} = \frac{1}{l_n} \sum_{j=1}^{l_n} Y_j$  and  $\hat{Y} = \frac{1}{\hat{l}_n} \sum_{j=1}^{\hat{l}_n} \hat{Y}_j$ . This can be done by using Lemmas (5.2) and (5.3) in Bertail and Cléménçon (2006): under **H0**( $S, 4, \nu$ ), we get

$$\left| \frac{\hat{l}_n}{n} - \frac{l_n}{n} \right| = \mathcal{O}_\nu(\alpha_n^{1/2}) \tag{A3}$$

and under  $\mathbf{H1}(S, 4, \nu, m)$  and  $\mathbf{H1}(S, 2, m)$ ,

$$\left\| \frac{\hat{l}_n}{n} \widehat{Y} - \frac{l_n}{n} \overline{Y} \right\| = \left\| \frac{1}{n} \sum_{j=1}^{\hat{l}_n} \widehat{Y}_j - \frac{1}{n} \sum_{j=1}^{l_n} Y_j \right\| = \mathcal{O}_\nu(n^{-1} \alpha_n^{1/2}).$$

With some straightforward calculus, we have

$$\left\| \widehat{Y} - \overline{Y} \right\| \leq \frac{n}{\hat{l}_n} \left\| \frac{\hat{l}_n}{n} \widehat{Y} - \frac{l_n}{n} \overline{Y} \right\| + \left| \frac{l_n}{\hat{l}_n} - 1 \right| \|\overline{Y}\|. \quad (\text{A4})$$

Since

$$\left| l_n - n/\mathbb{E}_A[\tau_A] \right| \xrightarrow[n \rightarrow \infty]{a.s.} 0,$$

equation (A3) gives

$$\frac{n}{\hat{l}_n} = \frac{n}{l_n} \left( 1 + \frac{n}{l_n} \frac{\hat{l}_n - l_n}{n} \right)^{-1} = \mathcal{O}_\nu(\mathbb{E}_A[\tau_A]) \left( 1 + \mathcal{O}_\nu(\mathbb{E}_A[\tau_A]) \mathcal{O}_\nu(\alpha_n^{1/2}) \right)^{-1} = \mathcal{O}_\nu(\mathbb{E}_A[\tau_A])$$

and

$$\left| \frac{l_n}{\hat{l}_n} - 1 \right| = \frac{n}{\hat{l}_n} \left| \frac{\hat{l}_n - l_n}{n} \right| = \mathcal{O}_\nu(\mathbb{E}_A[\tau_A]) \mathcal{O}_\nu(\alpha_n^{1/2}) = \mathcal{O}_\nu(\alpha_n^{1/2}).$$

From this and equation (A4), we deduce:

$$\left\| \widehat{Y} - \overline{Y} \right\| \leq \mathcal{O}_\nu(\mathbb{E}_A[\tau_A]) \mathcal{O}_\nu(n^{-1} \alpha_n^{1/2}) + \mathcal{O}_\nu(\alpha_n^{1/2}) \mathcal{O}_\nu(n^{-1/2}) = \mathcal{O}_\nu(\alpha_n^{1/2} n^{-1/2}). \quad (\text{A5})$$

Therefore

$$n^{1/2} \widehat{Y} = n^{1/2} \overline{Y} + n^{1/2} (\overline{Y} - \widehat{Y}) = n^{1/2} \overline{Y} + \mathcal{O}_\nu(\alpha_n^{1/2}).$$

Using this and the CLT for the  $Y_i$ , we show that  $n^{1/2} \widehat{Y}$  is asymptotically Gaussian.

The same kind of arguments give a control on the difference between empirical variances. Consider

$$\widehat{S}_{\hat{l}_n}^2 = \sum_{j=1}^{\hat{l}_n} \widehat{Y}_j \widehat{Y}_j' \text{ and } \widehat{S}_{\hat{l}_n}^{-2} = (\widehat{S}_{\hat{l}_n}^2)^{-1}.$$

By Lemma (5.3) of Bertail and Cléménçon (2006) we have, under  $\mathbf{H1}(S, 4, \nu, m)$  and

$\mathbf{H1}(S, 2, m)$ ,  $\left\| \frac{\hat{l}_n}{n} \widehat{S}_{\hat{l}_n}^2 - \frac{l_n}{n} S_{l_n}^2 \right\| = \mathcal{O}_\nu(\alpha_n)$ , and then

$$\left\| \widehat{S}_{\hat{l}_n}^2 - S_{l_n}^2 \right\| \leq \frac{n}{\hat{l}_n} \left\| \frac{\hat{l}_n}{n} \widehat{S}_{\hat{l}_n}^2 - \frac{l_n}{n} S_{l_n}^2 \right\| + \left| \frac{l_n}{\hat{l}_n} - 1 \right| \|S_{l_n}^2\| = \mathcal{O}_\nu(\alpha_n) + \mathcal{O}_\nu(\alpha_n^{1/2}) = o_\nu(1). \quad (\text{A6})$$

The proof of Theorem (3.1) is then also valid for the approximated blocks  $\widehat{B}_j$  and reduce to the study of the square of a self-normalized sum based on the pseudo-blocks. We have  $\hat{r}_n(\theta_0) = \sup_{\lambda \in \mathbb{R}^p} \left\{ \sum_{j=1}^{\hat{l}_n} \log \left[ 1 + \lambda' \widehat{Y}_j \right] \right\}$ . Let  $\hat{\lambda}_n = -\widehat{S}_{\hat{l}_n}^{-2} \widehat{Y}$  +  $o_\nu(n^{-1/2})$  be the optimum value of  $\lambda$ , we have

$$2\hat{r}_n(\theta_0) = -2\hat{l}_n \hat{\lambda}_n' \widehat{Y} - \sum_{j=1}^{\hat{l}_n} (\hat{\lambda}_n' \widehat{Y}_j)^2 + o_\nu(1) = \hat{l}_n \widehat{Y}' \widehat{S}_{\hat{l}_n}^{-2} \widehat{Y} + o_\nu(1).$$

Using the controls given by equations (A5) and (A6), we get

$$2\hat{r}_n(\theta_0) = [l_n + \mathcal{O}_\nu(n\alpha_n^{1/2})] \times \left[ \overline{Y}' + \mathcal{O}_\nu \left( \sqrt{\frac{\alpha_n}{n}} \right) \right] \times [S_{l_n}^{-2} + o_\nu(1)] \times \left[ \overline{Y} + \mathcal{O}_\nu \left( \sqrt{\frac{\alpha_n}{n}} \right) \right] + o_\nu(1).$$

Developing this product, the main term is  $l_n \overline{Y}' S_{l_n}^{-2} \overline{Y} \sim_\nu 2r_n(\theta_0)$  and all other terms are  $o_\nu(1)$ , yielding

$$2\hat{r}_n(\theta_0) = l_n \overline{Y}' S_{l_n}^{-2} \overline{Y} + o_\nu(1) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi_p^2.$$

Results (b), (c) and (d) can be derived from the atomic case by using the same arguments.

$n$	ReBEL	BEL
250	0.92	0.82
500	0.94	0.88
1000	0.94	0.91

Table A1. Cover rates of confidence intervals for the mean of an AR(1). Comparison of ReBEL and BEL for different data set lengths. Nominal level is 0.95 .

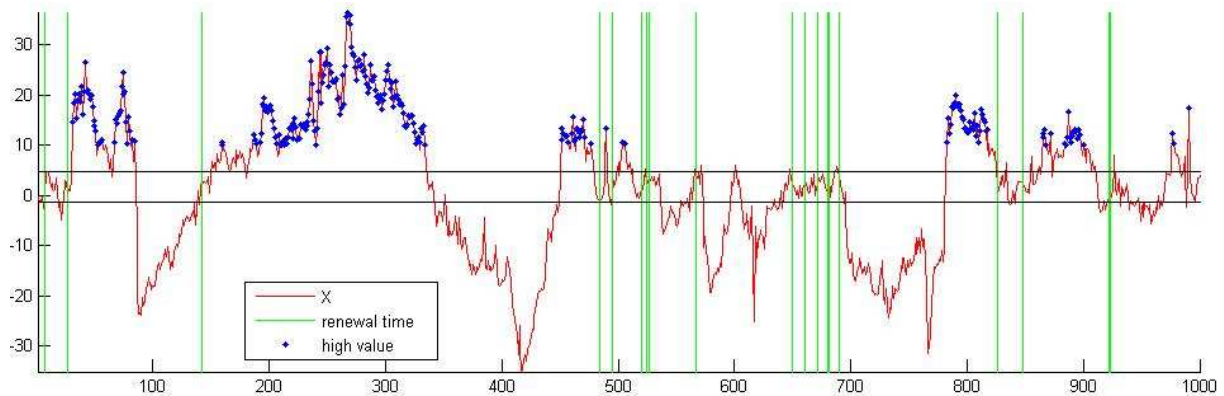


Figure A1. The plain curve is a chain of length 1000. The horizontal lines limit the small set. The 18 renewal times are marked by vertical lines. High values are marked by a dot.

$n$	$\theta = \theta_0$				$\theta = \theta_0 + 5/\sqrt{n}$				$\theta = \theta_0 + 10/\sqrt{n}$			
	ReBEL	BEL	mean	trunc	ReBEL	BEL	mean	trunc	ReBEL	BEL	mean	trunc
1000	54	55	58	58	24	13	14	20	07	02	02	06
5000	88	67	74	74	52	27	30	29	12	01	01	02
10000	92	70	76	77	59	31	34	33	11	02	02	02

Table A2. Coverage probabilities and type-II errors (percent) under the null and two alternatives, for ReBEL against BEL, and 2 reference methods. Nominal level is 95%

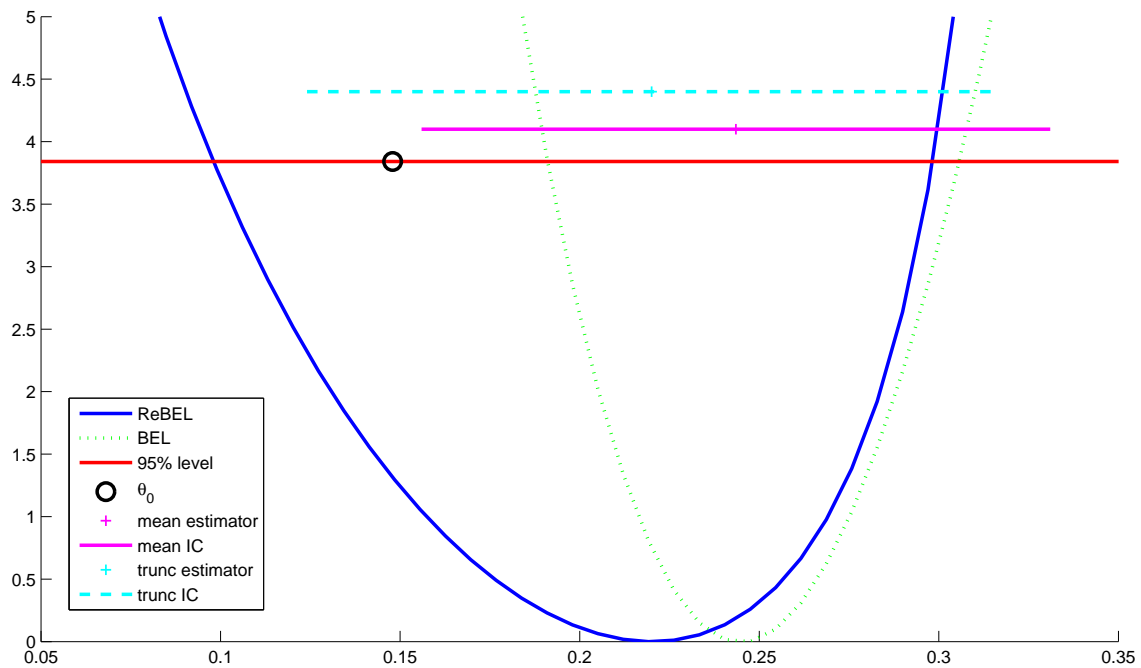


Figure A2. The plain curve gives the ReBEL likelihood, whereas the dotted curve shows the BEL likelihood. The red horizontal line marks the 95% level and  $\theta_0$  is marked by a circle on that line. The *mean* CI is the magenta segment whereas the *trunc* CI is the larger dotted cyan segment.

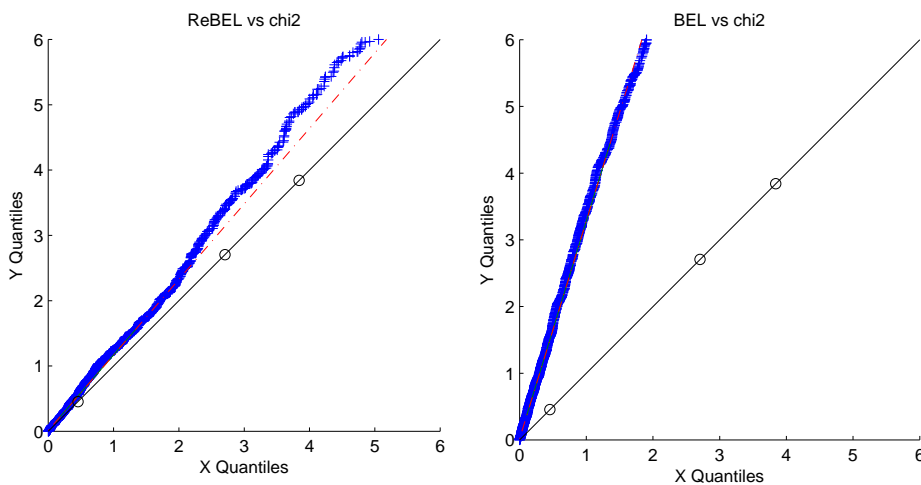


Figure A3. QQ-plots of 10 000 Monte-Carlo repetitions of ReBEL statistic versus  $\chi_1^2$  quantiles. The solid reference line is the 45° line. The reference circles on that line mark the 50%, 90% and 95% levels. Data set length is  $n = 10\,000$ .