Change point detection in autoregressive models with no moment assumptions

Fumiya Akashi
Waseda University
Department of Applied Mathematics
169-8555, Tokyo, Japan

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany

Yan Liu
Waseda University
Department of Applied Mathematics
169-8555, Tokyo, Japan

Abstract

In this paper we consider the problem of detecting a change in the parameters of an autoregressive process, where the moments of the innovation process do not necessarily exist. An empirical likelihood ratio test for the existence of a change point is proposed and its asymptotic properties are studied. In contrast to other work on change point tests using empirical likelihood, we do not assume knowledge of the location of the change point. In particular, we prove that the maximizer of the empirical likelihood is a consistent estimator for the parameters of the autoregressive model in the case of no change point and derive the limiting distribution of the corresponding test statistic under the null hypothesis. We also establish consistency of the new test. A nice feature of the method consists in the fact that the resulting test is asymptotically distribution free and does not require an estimate of the long run variance. The asymptotic properties of the test are investigated by means of a small simulation study, which demonstrates good finite sample properties of the proposed method.

Keywords and Phrases: Empirical likelihood, change point analysis, infinite variance, autoregressive processes
AMS Subject Classification: 62M10, 62G10, 62G35

1 Introduction

The problem of detecting structural breaks in time series has been studied for a long time. Since the seminal work of Page (1954, 1955), who proposed a sequential scheme for identifying changes in the mean of a sequence of independent random variables, numerous authors have worked on this problem. A large part of the literature concentrates on CUSUM tests, which are nonparametric by design [see Aue and Horváth (2013) for a recent review and some important references]. Other authors make distributional assumptions to construct tests for structural breaks. For example, Gombay and Horváth (1990) suggested a likelihood ratio procedure to test for a change in the mean and extensions of this method can be found in the monograph of Csörgö and Horváth (1997) and the reference therein. An important problem in this context is the detection of changes in the parameters of an autoregressive process and we refer to the work of Andrews (1993), Bai (1993, 1994), Davis et al. (1995), Lee et al. (2003) and Berkes et al. (2011) among others who proposed CUSUM-type and likelihood ratio tests.
In practice, however, the distribution of random variables is rarely known and its misspecification may result in an invalid analysis using likelihood ratio methods. One seminal method to treat the likelihood ratio empirically has been investigated by Owen (1988), Qin and Lawless (1994) in a general context and extended by Chuang and Chan (2002) to estimate and test parameters in an autoregressive model. In change point analysis the empirical likelihood approach can be viewed as a compromise between the completely parametric likelihood ratio and nonparametric CUSUM method. Baragona et al. (2013) used this concept to construct a test for change-points and showed that in the case where the location of the break points is known,

the limiting distribution of the corresponding test statistic is a chi-square distribution. Ciuperca and Salloum (2015) considered the change point problem in a non-linear model with independent data without assuming knowledge of its location and derived an extreme value distribution as limit distribution of the empirical likelihood ratio test statistic. These findings are similar in spirit to the meanwhile classical results in Csörgö and Horváth (1997), who considered the likelihood ratio test.

The purpose of the present paper is to investigate an empirical likelihood test for a change in the parameters of an autoregressive process with infinite variance (more precisely we do not assume the existence of any moments). Our work is motivated by the fact that in many fields, such as electrical engineering, hydrology, finance and physical systems, one often observes "heavy-tailed" data [see Nolan (2015) or Samoradnitsky and Taqqu (1994) among many others]. To deal with such data, many authors have developed L_1-based methods. For example, Chen et al. (2008) constructed a robust test for a linear hypothesis of the parameters based on least absolute deviation. Ling (2005) and Pan et al. (2007) proposed self-weighted least absolute deviation-based estimators for (parametric) time series models with an infinite variance innovation process and show the asymptotic normality of the estimators. However, the limit distribution of the L_1-based statistics usually contains the unknown probability density of the innovation process, which is difficult to estimate. For example, Ling (2005) and Pan et al. (2007) used kernel density estimators for this purpose, but the choice of the corresponding bandwidth is not clear and often depends on users.

To circumvent problems of this type in the context of change point analysis, we combine in this paper quantile regression and empirical likelihood methods. As a remarkable feature, the asymptotic distribution of the proposed test statistic does not involve unknown quantities of the model even if we consider autoregressive models with an infinite variance in the innovation process. We would also like to emphasize that the nonparametric CUSUM tests proposed by Bai (1993, 1994) for detecting structural breaks in the parameters of an autoregressive process assume the existence of the variance of the innovations. However, an alternative to the method proposed here are CUSUM tests based on quantile regression, which has been recently considered by Qu (2008), Su and Xiao (2008) and Zhou et al. (2015) among others.

The remaining part of this paper is organized as follows. In Section 2, we introduce the model, the testing problem and the so-called self-weighted empirical likelihood ratio test statistic. Our main results are given in Section 3, where we derive the limit distribution of the proposed test statistic and prove consistency. The finite sample properties of the proposed test are investigated in Section 4 by means of a simulation study. We also compare the test proposed in this paper with the CUSUM test using quantile regression [see Qu (2008)]. While the empirical likelihood based test suggested here is competitive with the CUSUM test using quantile regression when the innovation process is Gaussian, it performs remarkably better than the CUSUM test of Qu (2008) if the innovation process has heavy tails. Moreover, the new test is robust with respect non-stationarity even when the process is nearly a unit root process. Finally, rigorous proofs of the results relegated to Section 5.

2 Change point tests using empirical likelihood

Throughout this paper the following notations and symbols are used. The set of all integers and real numbers are denoted as $\mathbb{Z}$ and $\mathbb{R}$, respectively. For any sequence of random vectors $\{A_n : n \geq 1\}$ we denote by

$$A_n \xrightarrow{\mathcal{P}} A \quad \text{and} \quad A_n \xrightarrow{\mathcal{L}} A$$

convergence in probability and law to a random vector A, respectively. The transpose of a matrix M is denoted by M', and $\|M\| = \{\text{tr}(M'M)\}^{1/2}$ is the Frobenius norm. We denote the i-dimensional zero vector, the $j \times k$ zero matrix and the $l \times l$ identity matrix by 0_i, $O_{j\times k}$ and $I_{l\times l}$, respectively.

Consider the autoregressive model of order p (AR(p) model) defined by

$$y_t = X_{t-1}'\beta + e_t, \tag{2.1}$$

where $X_{t-1} = (y_{t-1}, \dots, y_{t-p})'$ and $\beta \in \mathbb{R}^p$ and assume that the innovation process $\{e_t : t \in \mathbb{Z}\}$ is a sequence of independent and identically distributed (i.i.d.) random

variables with vanishing median. Let $\{y_{1-p}, \ldots, y_n\}$ be an observed stretch from the model (2.1) for $\beta = \beta_0$, where $\beta_0 = (\beta_1, \ldots, \beta_p)'$ denotes the "true" parameter.
This paper focuses on a posteriori type change point problem for the parameters in the AR(p) process (2.1). More precisely, we consider the model

$$y_t = \begin{cases} X_{t-1}'\theta_1 + e_t & (1 \le t \le k^*) \\ X_{t-1}'\theta_2 + e_t & (k^* + 1 \le t \le n) \end{cases}$$

for some vector $\theta_1, \theta_2 \in \mathbb{R}^p$, where $k^* \in \{1, \ldots, n\}$ is the unknown time point of the change. The testing problem for a change point in the autoregressive process can then be formulated by the following hypotheses:

$$H_0 : \theta_1 = \theta_2 = \beta_0 \quad \text{against} \quad H_1 : \theta_1 \neq \theta_2. \tag{2.2}$$

Note that we neither assume knowledge of the change point k^* (if the null hypothesis is not true) nor of the true value $\beta_0 \in \mathbb{R}^p$ (if the null hypothesis holds).

For the testing problem (2.2), we construct an empirical likelihood ratio (ELR) test. To be precise, let $\mathbb{I}$ denote the indicator function. As the median of e_t is zero, the moment condition

$$\mathbb{E}\Big[\Big\{\frac{1}{2} - \mathbb{I}(y_t - X_{t-1}'\beta_0 \le 0)\Big\}a^*(X_{t-1})\Big] = 0_m \tag{2.3}$$

holds under the null hypothesis H_0 in (2.2), where $a^*(X_{t-1})$ is any m-dimensional measurable function of X_{t-1} independent of e_t. Motivated by the moment conditions (2.3), we first introduce the self-weighted moment function

$$g(\mathscr{Y}_t^p, \beta) := \Big\{\frac{1}{2} - \mathbb{I}\left(y_t - X_{t-1}'\beta \le 0\right)\Big\}a^*(X_{t-1}) \quad (t = 1, \ldots, n),$$

where $\mathscr{Y}_t^p = (y_t, \ldots, y_{t-p})$ and $a^*(X_{t-1}) = w_{t-1}a(X_{t-1})$, $a(x) = (x', \varphi(x)')'$ is an $m = (p+q)$-dimensional function, φ a q-dimensional function, $w_{t-1} = w(y_{t-1}, \ldots, y_{t-p})$ a self-weight and w some positive weight function. We can choose the weight function w and φ arbitrarily provided that Assumption 3.2 in Section 3 holds. In particular, we can use $a(x) = x$, which corresponds to the case $q = 0$ (see also Section 4).
Note that under the null hypothesis H_0, we have that $\mathbb{E}[g(\mathscr{Y}_t^p, \beta_0)] = 0_m$ for all $t = 1, \ldots, n$. Let $r_{n,k}$ be $(v_1, \ldots, v_k, v_{k+1}, \ldots, v_n)'$ be a vector in the unit cube $[0,1]^n$,

then the empirical likelihood (EL), for $\beta = \theta_1$ before the change point $k \in \{1, \ldots, n\}$ and $\beta = \theta_2$ after the change point, is defined by

$$L_{n,k}(\theta_1, \theta_2) := \sup\Big\{ \Big(\prod_{i=1}^{k} v_i \Big) \Big(\prod_{j=k+1}^{n} v_j \Big) : r_{n,k} \in \mathcal{P}_{n,k} \cap \mathcal{M}_{n,k}(\theta_1, \theta_2) \Big\},$$

where $\mathcal{P}_{n,k}$ and $\mathcal{M}_{n,k}(\theta_1, \theta_2)$ are subsets of the cube $[0,1]^n$ defined as

$$\mathcal{P}_{n,k} := \Big\{ r_{n,k} \in [0,1]^n : \sum_{i=1}^{k} v_i = \sum_{j=k+1}^{n} v_j = 1 \Big\}$$

and

$$\mathcal{M}_{n,k}(\theta_1, \theta_2) := \Big\{ r_{n,k} \in [0,1]^n : \sum_{i=1}^{k} v_i g(\mathscr{Y}_i^p, \theta_1) = \sum_{j=k+1}^{n} v_j g(\mathscr{Y}_j^p, \theta_2) = 0_m \Big\}.$$

Note that the unconstrained maximum EL is represented as

$$L_{n,k,E} := \sup\Big\{ \prod_{i=1}^{n} v_i : r_{n,k} \in \mathcal{P}_{n,k} \Big\} = k^{-k}(n-k)^{-(n-k)},$$

and hence, the logarithm of the empirical likelihood ratio (ELR) statistic is given by

$$\begin{aligned} l_{n,k}(\theta_1, \theta_2) &:= -\log \frac{L_{n,k}(\theta_1, \theta_2)}{L_{n,k,E}} \\ &= -\log \sup\Big\{ \Big(\prod_{i=1}^{k} k v_i \Big) \Big(\prod_{j=k+1}^{n} (n-k) v_j \Big) : r_{n,k} \in \mathcal{P}_{n,k} \cap \mathcal{M}_{n,k}(\theta_1, \theta_2) \Big\} \\ &= \Big[\sum_{i=1}^{k} \log\big\{1 - \lambda' g(\mathscr{Y}_i^p, \theta_1)\big\} + \sum_{j=k+1}^{n} \log\big\{1 - \eta' g(\mathscr{Y}_j^p, \theta_2)\big\} \Big], \end{aligned} \quad (2.4)$$

where (2.4) is obtained by the Lagrange multiplier method and the multipliers λ, $\eta \in \mathbb{R}^m$ satisfy

$$\sum_{i=1}^{k} \frac{g(\mathscr{Y}_i^p, \theta_1)}{1 - \lambda' g(\mathscr{Y}_i^p, \theta_1)} = \sum_{j=k+1}^{n} \frac{g(\mathscr{Y}_j^p, \theta_2)}{1 - \eta' g(\mathscr{Y}_j^p, \theta_2)} = 0_m.$$

We finally define the test statistic for the change point problem (2.2). Since the maximum ELR under H_0 is given by

$$P_{n,k} := \sup_{\beta \in \mathcal{B}} \{-l_{n,k}(\beta, \beta)\},$$

one may define the ELR test statistic by

$$T_n := 2 \max_{\lfloor r_1 n \rfloor \leq k \leq \lfloor r_2 n \rfloor} P_{n,k}, \tag{2.5}$$

where $0 < r_1 < r_2 < 1$ for fixed constants. Note that we do not consider the maximum of $\{P_{n,k} \mid k = 1, \ldots, n\}$ as $P_{n,k}$ can not be estimated accurately for small and large values of k (see Theorem 3.1 in Section 3 for more details). The asymptotic properties of a weighted version of this statistic are investigated in the following section.

Remark 2.1. The approach presented here can be naturally extended to the general τ-quantile regression models. To be precise, suppose that

$$Q_y(\tau \mid X_{t-1}) = \inf\{y : P(y_t < y \mid X_{t-1}) \geq \tau\}$$

denotes the τth-quantile of y_t conditional on X_{t-1} and assume that $Q_y(\tau \mid X_{t-1}) = \beta(\tau)' X_{t-1}$. The moment condition

$$\mathbb{E}[g^{(\tau)}(\mathscr{Y}_t^p, \beta_0(\tau))] = 0_m$$

still holds under the null hypothesis H_0, if we define

$$g^{(\tau)}(\mathscr{Y}_t^p, \beta(\tau)) := \psi_\tau(y_t - \beta(\tau)' X_{t-1}) a^*(X_{t-1})$$

and $\psi_\tau(u) := \{\tau - \mathbb{I}(u \leq 0)\}$.

Remark 2.2. The method can also be extended to develop change point analysis based on the generalized empirical likelihood (GEL). A GEL test statistic for the change point problem (2.2) can be defined by

$$l^\rho_{n,k}(\theta_1, \theta_2) = 2\Big[\sup_{\lambda \in \mathbb{R}^m} \sum_{i=1}^{k} \rho\left\{\lambda' g(\mathscr{Y}_i^p, \theta_1)\right\} + \sup_{\eta \in \mathbb{R}^m} \sum_{j=k+1}^{n} \rho\left\{\eta' g(\mathscr{Y}_j^p, \theta_2)\right\}\Big],$$

where ρ is a real-valued, concave, twice differentiable function defined on an open interval of the real line that contains the point 0 with $\rho'(0) = \rho''(0) = 1$. Typical examples for the choice of ρ are given by $\rho(\nu) = -\log(1-\nu)$ and

$$\rho(\nu) = \frac{(1+c\nu)^{(c+1)/c} - 1}{c+1}. \tag{2.6}$$

Using Lagrangian multipliers, it is easy to see that the choice $\rho(\nu) = -\log(1-\nu)$ yields the empirical likelihood method discussed so far. The class associated with (2.6) is called the Cressie-Read family [see Cressie and Read (1984)].

3 Main results

In this section we state our main results. Throughout this paper, let F and f denote the distribution function and the probability density function of e_t, respectively. We impose the following assumptions.

Assumption 3.1.

(i) $\beta_0 \in \mathrm{Int}(\mathcal{B})$, where the parameter space $\mathcal{B}$ is a compact set in $\mathbb{R}^p$ with non-empty interior.

(ii) $1 - \beta_1 z - \cdots - \beta_p z^p \neq 0$ for $|z| \leq 1$ and $\beta \in \mathcal{B}$.

(iii) The median of e_t is zero.

(iv) The distribution function F of e_t is continuous and differentiable at the point 0 with positive derivative $F'(0) = f(0)$.

Assumption 3.2. $\mathbb{E}[(w_{t-1} + w_{t-1}^2)(\|a(X_{t-1})\|^2 + \|a(X_{t-1})\|^3)] < \infty$.

Assumption 3.3. The matrix $\mathbb{E}[g(\mathscr{Y}_t^p, \beta_0) g(\mathscr{Y}_t^p, \beta_0)']$ is positive definite.

Assumption 3.4.

(i) There exists a constant $\gamma > 2$ such that $\mathbb{E}[\|a^*(X_{t-1})\|^\gamma] < \infty$.

(ii) Let $v_t := \mathrm{sign}(e_t)a^*(X_{t-1})$. Then the sequence $\{v_t : t \in \mathbb{Z}\}$ is strong mixing with mixing coefficients α_l that satisfy $\sum_{l=1}^{\infty} \alpha_l^{1-2/\gamma} < \infty$.

The *maximum EL estimator* $\hat{\beta}_{n,k}$ is defined by

$$-l_{n,k}(\hat{\beta}_{n,k}, \hat{\beta}_{n,k}) = \sup_{\beta \in \mathcal{B}} \{-l_{n,k}(\beta, \beta)\} ,$$

and the consistency with corresponding rate of convergence of this statistic are given in the following theorem.

Theorem 3.1. *Suppose that Assumptions 3.1-3.4 hold and define $k^* := rn$ for some $r \in (0,1)$. Then, under the null hypothesis H_0, we have, as $n \to \infty$,*

$$\hat{\beta}_{n,k^*} - \beta_0 = O_p\left(n^{-1/2}\right).$$

As seen from Theorem 3.1, T_n is not accurate for small k and $n-k$ as the result does not hold if $k/n = o(1)$ or $(n-k)/n = o(1)$. In addition, the ELR statistic is not computable for small k and $n-k$. For this reason, we consider in the following discussion the trimmed and weighted-version of EL ratio test statistic, defined by

$$\tilde{T}_n := 2 \max_{k_{1n} \le k \le k_{2n}} h\left(\frac{k}{n}\right) P_{n,k}, \tag{3.1}$$

where h is a given weight function, $k_{1n} := r_1 n$, $k_{2n} := r_2 n$ and $0 < r_1 < r_2 < 1$. If $\tilde{T}_n$ takes a significant large value, we have enough reason to reject the null hypothesis H_0 of no change point. We also need a further assumption to control a remainder terms in the stochastic expansion of $\tilde{T}_n$.

Assumption 3.5. $\sup_{0<r<1} h(r)^2 < \infty$.

With this additional assumption the limit distribution of the test statistic (3.1) can be derived in the following theorem.

Theorem 3.2. *Suppose that Assumptions 3.1-3.5 hold. Then, under the null hypothesis H_0 of no change point*

$$\tilde{T}_n \xrightarrow{\mathcal{L}} T := \sup_{r_1 \le r \le r_2} \left\{ r^{-1}(1-r)^{-1} h(r) \left\| B(r) - rB(1) \right\|^2 + h(r) B(1)' Q B(1) \right\} \tag{3.2}$$

as $n \to \infty$. Here $\{B(r) : r \in [0,1]\}$ is an m-dimensional vector of independent Brownian motions and the matrix Q is defined by

$$Q = I_{m\times m} - \Omega^{-1/2} G \Sigma G' \Omega^{-1/2}, \tag{3.3}$$

where $A^{1/2}$ denotes the square root of a nonnegative definite matrix A, $G = G(\beta_0) = \partial g(\beta_0)/\partial \beta'$, $\Sigma = (G'\Omega^{-1}G)^{-1}$ and

$$\Omega := \mathbb{E}[g(\mathscr{Y}_t^p, \beta_0) g(\mathscr{Y}_t^p, \beta_0)'] = \frac{1}{4}\mathbb{E}[a^*(X_{t-1}) a^*(X_{t-1})']. \tag{3.4}$$

A test for the hypotheses in (2.2) is now easily obtained by rejecting the null hypothesis in (2.2) whenever

$$\tilde{T}_n > q_{1-\alpha}, \tag{3.5}$$

where $q_{1-\alpha}$ is the $(1-\alpha)$-quantile of the distribution of the random variable T defined on the right-hand side of equation (3.2) (using an appropriate estimate of the matrix Q).

Theorem 3.3. *Suppose that Assumptions 3.1-3.5 and the alternative $H_1 : \theta_1 \neq \theta_2$ hold. Then we have*

$$\tilde{T}_n \xrightarrow{\mathcal{P}} \infty$$

as $n \to \infty$.

Theorem 3.3 shows that the power of the test (3.5) approaches 1 at any fixed alternative. In other words, the test is consistent.

4 Finite sample properties

In this section, we illustrate the finite sample properties of the ELR test (3.5) for the hypothesis (2.2) by means of small simulation study. For this purpose we consider the AR(1) model

$$y_t = \beta y_{t-1} + e_t,$$

where the coefficient β satisfies

$$\beta = \begin{cases} \theta_1 & (t = 1, \dots, k^*) \\ \theta_2 & (t = k^* + 1, \dots, n) \end{cases} .$$

For the calculation of the ELR statistic $\tilde{T}_n$ in (3.1), we use the functions $a(x) = x$ and $h(r) = r(1-r)$ throughout this section. Following Ling (2005), the self-weights are chosen as

$$w_{t-1} = \begin{cases} 1 & (d_{t-1} = 0) \\ (c/d_{t-1})^3 & (d_{t-1} \neq 0) \end{cases} ,$$

where $d_{t-1} = |y_{t-1}|\mathbb{I}(|y_{t-1}| > c)$ and c is the 95%-quantile of the sample $\{y_0, y_1, \dots, y_n\}$. The trimming parameters in the definition of the statistic $\tilde{T}_n$ are chosen as $r_{1n} = 0.1$ and $r_{2n} = 0.9$. The critical value in (3.5) is obtained as the empirical 95% quantile of the Monte-Carlo samples

$$\left\{ \max_{k_{1n} \le k \le k_{2n}} \left(B^{(l)}(k/n) - (k/n) B^{(l)}(1) \right)^2 : l = 1, \dots, 1000 \right\},$$

where $B^{(1)}(\cdot), \dots, B^{(1000)}(\cdot)$ are independent standard Brownian motions (note that in this case, the matrix in (3.3) is given by $Q = 0$).

In Figures 1-3, we display the rejection probabilities of the ELR test (3.5) for the hypothesis (2.2), where the nominal level is chosen as $\alpha = 0.05$. The horizontal and vertical axes show, respectively, the values of θ_2 and the rejection rate of the hypothesis $H_0 : \theta_1 = \theta_2$ at this point (θ_1 is fixed as 0.3). The sample sizes are given by $n = 100, 200$ and 400 and the distribution of the innovation process is a standard normal distribution (Figure 1), a t-distribution with 2 degrees of freedom (Figure 2) and a Cauchy distribution (Figure 3). We also consider two values of the parameter r in the definition of the change point $k^* = rn$, that is $r = 0.5$ and $r = 0.8$.

We observe that for small sample sizes, the test is slightly conservative and that the approximation of the nominal level improves with increasing sample size. The alternatives are rejected with reasonable probabilities, where the power is larger in the case $r = 0.5$ than for $r = 0.8$. A comparison of the different distributions in Figures 1-3 shows that the power is lower for standard normal distributed innovations, while an error process with a Cauchy distribution yields the largest rejection probabilities. Other simulations show a similar picture, and the results are omitted for the sake of brevity.

Figure 1: *Simulated rejection probabilities of the ELR test* (3.5) *in the AR(1) model with normal distributed innovations.*

(a) $\theta_1 = 0.3$, $r = 0.5$ (b) $\theta_1 = 0.3$, $r = 0.8$

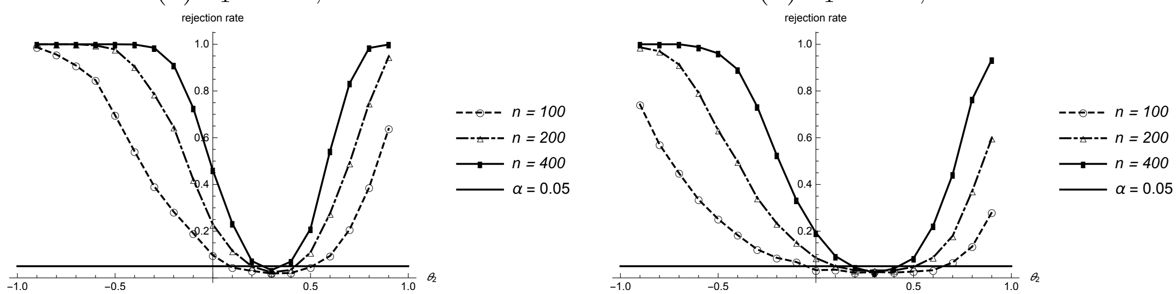


Figure 2: *Simulated rejection probabilities of the ELR test* (3.5) *in the AR(1) model with t-distributed innovations.*

(a) $\theta_1 = 0.3$, $r = 0.5$ (b) $\theta_1 = 0.3$, $r = 0.8$

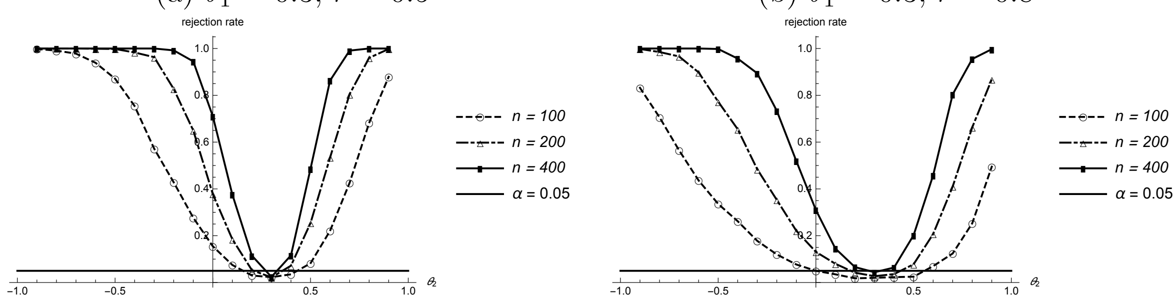
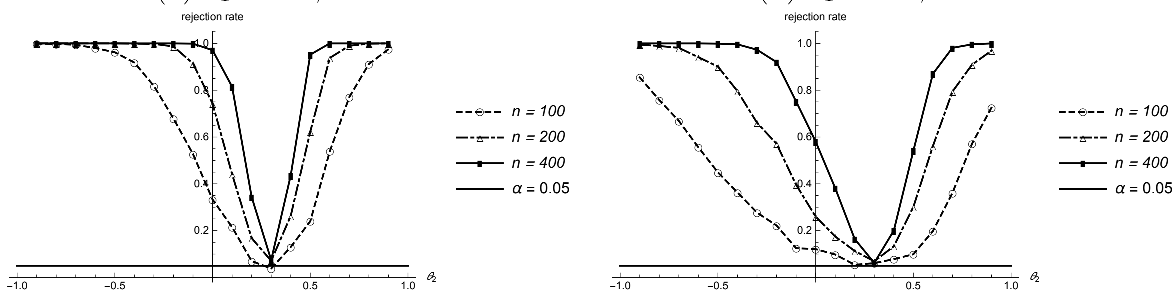


Figure 3: *Simulated rejection probabilities of the ELR test* (3.5) *in the AR(1) model with Cauchy distributed innovations.*

(a) $\theta_1 = 0.3$, $r = 0.5$ (b) $\theta_1 = 0.3$, $r = 0.8$



In the second part of this section we compare the new test defined by (3.5) with the CUSUM test in Qu (2008) which uses quantile regression. The test statistic for the median in Qu (2008) is defined by

$$\mathrm{SQ}_{0.5} = \sup_{\lambda \in [0,1]} \| H_{\lambda,n}(\hat{\beta}) - \lambda H_{1,n}(\hat{\beta}) \|, \tag{4.1}$$

where $\| \cdot \|$ is the sup norm, $\hat{\beta}$ is the median regressor,

$$H_{\lambda,n} = (\mathcal{X}'\mathcal{X})^{-1/2} \sum_{i=1}^{[\lambda n]} |y_t - X_{t-1}'\hat{\beta}| X_{t-1},$$

and the matrix $\mathcal{X}$ is given by $\mathcal{X} = (X_1, \ldots, X_n)'$. In Figures 4-6, we display the rejection probabilities of the test based on the statistic T_n in (2.5), $\tilde{T}_n$ in (3.1) and $\mathrm{SQ}_{0.5}$ in (4.1) for the hypothesis (2.2), where the nominal level is chosen as $\alpha = 0.05$. The horizontal and vertical axes show, respectively, the values of θ_2 and the rejection rate of the hypothesis $H : \theta_1 = \theta_2$ at this point (θ_1 is fixed as 0.3). The distribution of the innovation process is a standard normal distribution (Figure 4), a t-distribution with 2 degree of freedom (Figure 5) and a Cauchy distribution (Figure 6) and the sample sizes are given by $n = 100$, 200 and 400 in each case. Again we consider two different locations for the change point k^* corresponding to the values $r = 0.5$ and $r = 0.8$.

We observe that all tests derived from the three statistics T_n in (2.5) (corresponding to the weight function $h(r) \equiv 1$), $\tilde{T}_n$ in (3.1) (corresponding to the weight function $h(r) = r(1-r)$) and $\mathrm{SQ}_{0.5}$ in (4.1) are slightly conservative and that the approximation of the nominal level improves with increasing sample size [see Figure 4-6 for the value $\theta_2 = \theta_1 = 0.3$]. The approximation is usually more accurate for $r = 0.5$.
Next we compare the power of the different tests (i.e. $\theta_2 \neq \theta_1 = 0.3$) for different distributions of the innovations. In the case of Gaussian innovations all tests shows a similar behavior (see Figure 4) and only if the case $n = 200$ and $r = 0.8$ the ELR test based on the (unweighted) statistic T_n shows a better performance as the tests based on $\tilde{T}_n$ and $\mathrm{SQ}_{0.5}$. Moreover, for Gaussian innovations all three tests show a remarkable robustness against non-stationarity, that is $|\theta_2| = 1$.
In Figure 5 we display corresponding results for t_2-distributed innovations. The differences in the approximation of the nominal level are negligible ($\theta_2 = \theta_1 = 0.3$). If $r = 0.5$ we do not observe substantial differences in the power between the three tests (independently of the sample size). On the other hand, if $r = 0.8$ the tests based on ELR statistics $\tilde{T}_n$ and T_n yield larger rejection probabilities than the test $\mathrm{SQ}_{0.5}$ (see the right part of Figure Figure 5). Interestingly the unweighted test based on T_n shows a better performance than the test based on $\tilde{T}_n$ in these cases. Again, all tests are robust with respect to non-stationarity.
Finally, in Figure 6 we display the rejection probabilities of the three tests for Cauchy distributed innovations, where we again do not observe differences in the approximation of the nominal level ($\theta_2 = \theta_1 = 0.3$). On the other hand the differences in power between the tests based on ELR and quantile regression are remarkable. In all cases the ELR tests based on T_n and $\tilde{T}_n$ have substantially more power than the test based on $\mathrm{SQ}_{0.5}$. The ELR test based on the unweighted statistic T_n shows a better performance than the ELR test based on $\tilde{T}_n$. This superiority is less pronounced in the case $r = 0.5$ but clearly visible for $r = 0.8$. Finally, in contrast to the test based on $SQ_{0.5}$ the ELR tests based on T_n and $\tilde{T}_n$ are robust against non-stationarity (i.e. $|\theta_2| = 1$) for Cauchy distributed innovations and clearly detect a change in the parameters in these cases.

Figure 4: *Simulated rejection probabilities of various change point tests based on the statistics T_n, $\tilde{T}_n$ and $SQ_{0.5}$ defined in* (2.5), (3.1) *and* (4.1), *respectively. The model is given by an AR(1) model with normal distributed innovations.*

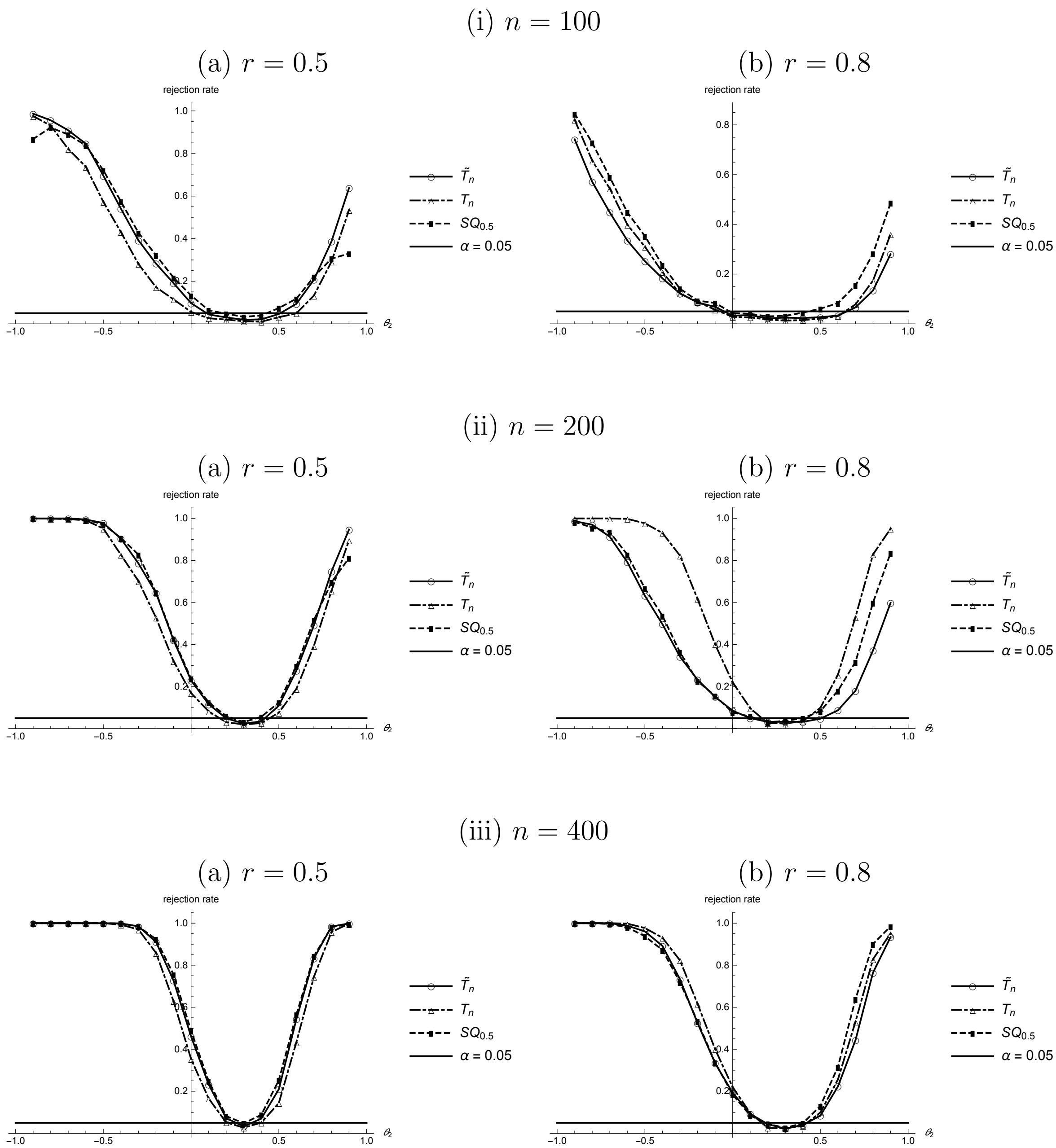


Figure 5: *Simulated rejection probabilities of various change point tests based on the statistics T_n, $\tilde{T}_n$ and $SQ_{0.5}$ defined in* (2.5), (3.1) *and* (4.1), *respectively. The model is given by an AR(1) model with t_2-distributed innovations.*

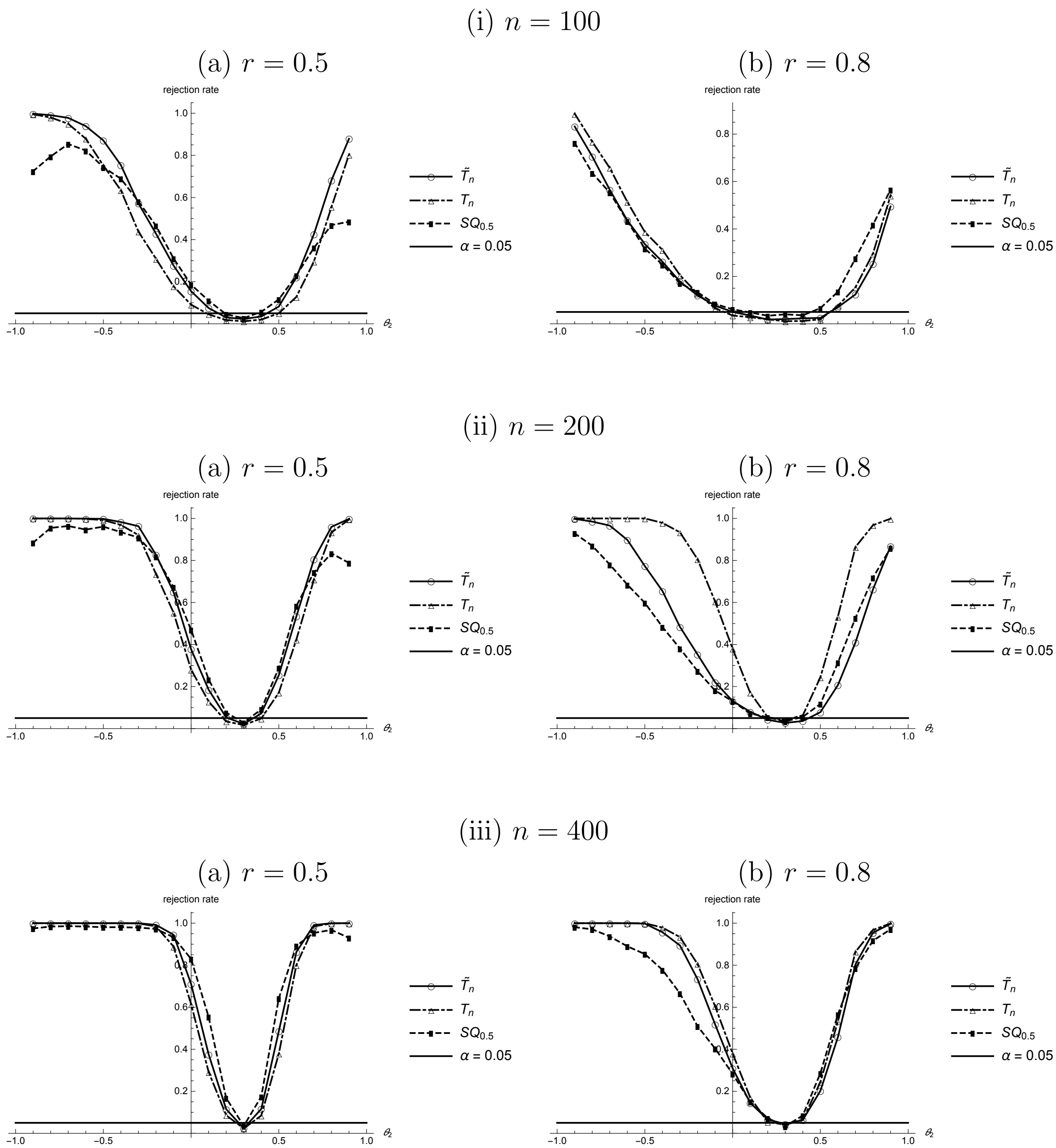
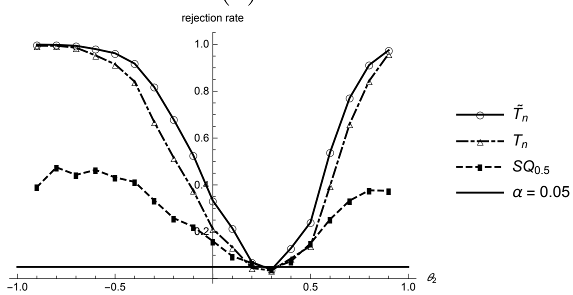


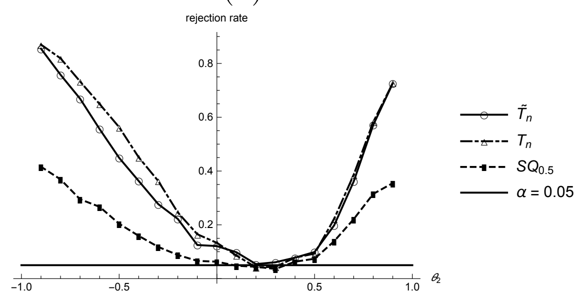
Figure 6: *Simulated rejection probabilities of various change point tests based on the statistics T_n, $\tilde{T}_n$ and $SQ_{0.5}$ defined in* (2.5), (3.1) *and* (4.1), *respectively. The model is given by an AR(1) model with Cauchy distributed innovations.*

(i) $n = 100$

(a) $r = 0.5$

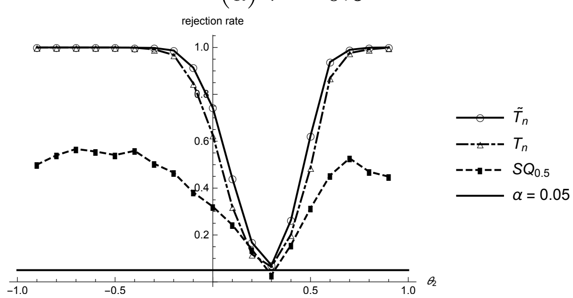


(b) $r = 0.8$

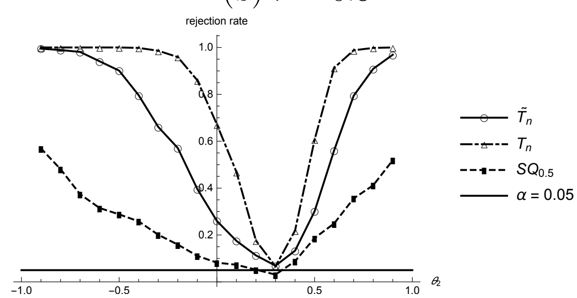


(ii) $n = 200$

(a) $r = 0.5$

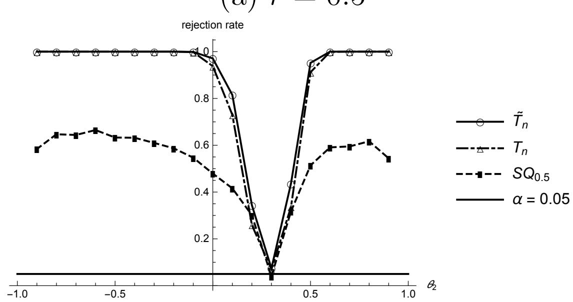


(b) $r = 0.8$

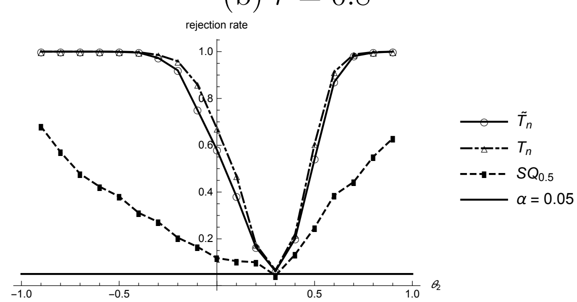


(iii) $n = 400$

(a) $r = 0.5$



(b) $r = 0.8$



5 Proofs

This section gives rigorous proofs of all results in this paper. In what follows, C will denote a generic positive constant that varies in different places. "with probability approaching one" will be abbreviated as w.p.a.1. Moreover, we use the following notations throughout this section:

$$
\begin{aligned}
&g_i(\beta) = g(\mathscr{Y}_i^p, \beta), \quad g(\beta) = \mathbb{E}[g(\mathscr{Y}_i^p, \beta)],\\
&\hat{P}_k^1(\beta, \lambda) = \frac{1}{k}\sum_{i=1}^{k} \log\{1 - \lambda' g_i(\beta)\},\\
&\hat{P}_{n,k}^2(\beta, \eta) = \frac{1}{n-k}\sum_{j=k+1}^{n} \log\{1 - \eta' g_j(\beta)\},\\
&\hat{\Lambda}_k^1(\beta) = \{\lambda \in \mathbb{R}^m : |\lambda' g_i(\beta)| < 1 \text{ for all } i = 1, \ldots, k\},\\
&\hat{\Lambda}_{n,k}^2(\beta) = \{\eta \in \mathbb{R}^m : |\eta' g_j(\beta)| < 1 \text{ for all } j = k+1, \ldots, n\},\\
&\hat{g}(\beta) = \frac{1}{n}\sum_{i=1}^{n} g(\mathscr{Y}_i^p, \beta) = \frac{1}{n}\sum_{i=1}^{n} g_i(\beta),\\
&\hat{g}_k^1(\beta) = \frac{1}{k}\sum_{i=1}^{k} g_i(\beta) \quad \text{and} \quad \hat{g}_{n,k}^2(\beta) = \frac{1}{n-k}\sum_{j=k+1}^{n} g_j(\beta).
\end{aligned}
$$

5.1 Proof of Theorem 3.1

We start proving several auxiliary results which are required in the proof of Theorem 3.1.

Lemma 5.1. *Suppose that Assumption 3.4 (i) holds. For* $1/\gamma < \zeta < 1/2$, *let*

$$\Lambda_{n,k} = \{(\lambda, \eta) \in \mathbb{R}^{2m} : \|\lambda\| \le Ck^{-\zeta}, \quad \|\eta\| \le C(n-k)^{-\zeta}\}.$$

Then, as $n \to \infty$, *we have*

$$\sup_{\beta \in \mathcal{B}, \lambda \in \Lambda_{n,k^*}} \max_{1 \le i \le k^*} |\lambda' g_i(\beta)| \xrightarrow{\mathcal{P}} 0, \qquad \sup_{\beta \in \mathcal{B}, \eta \in \Lambda_{n,k^*}} \max_{k^*+1 \le j \le n} |\eta' g_j(\beta)| \xrightarrow{\mathcal{P}} 0.$$

Also, $\Lambda_{n,k^*} \subset \hat{\Lambda}_{k^*}^1(\beta) \times \hat{\Lambda}_{n,k^*}^2(\beta)$ *for all* $\beta \in \mathcal{B}$ *w.p.a.1.*

Proof. Let $b_i = \sup_{\beta\in\mathcal{B}} \|g_i(\beta)\|$. By Assumption 3.4 (i), we can choose $\gamma > 2$ such that $K = \mathbb{E}[b_1^\gamma]^{1/\gamma}$ is finite. Then, for any $\delta > 0$, we can define $M(\delta) = K/\delta^{1/\gamma}$ and obtain

$$P\Big(\max_{1\le i\le k^*} b_i \ge M(\delta)k^{*1/\gamma}\Big) \le \sum_{i=1}^{k^*} P\left(b_i \ge M(\delta)k^{*1/\gamma}\right) = \sum_{i=1}^{k^*} P\left(b_i^\gamma \ge M(\delta)^\gamma k^*\right)$$
$$\le \sum_{i=1}^{k^*} \frac{\mathbb{E}[b_i^\gamma]}{M(\delta)^\gamma k^*} = \delta.$$

Consequently, $\max_i b_i = O_p(k^{*1/\gamma})$ and by the Cauchy-Schwartz inequality we have

$$\sup_{\beta\in\mathcal{B},\lambda\in\Lambda_{n,k^*}} \max_{1\le i\le k^*} |\lambda' g_i(\beta)| \le \sup_{\lambda\in\Lambda_{n,k^*}} \|\lambda\| \max_{1\le i\le k^*} b_i = O_p(k^{*-\zeta+1/\gamma}),$$

which implies

$$\sup_{\beta\in\mathcal{B},\lambda\in\Lambda_{n,k^*}} \max_{1\le i\le k^*} |\lambda' g_i(\beta)| \xrightarrow{\mathcal{P}} 0.$$

Similarly, it follows that

$$\sup_{\beta\in\mathcal{B},\eta\in\Lambda_{n,k^*}} \max_{k^*+1\le j\le n} |\eta' g_j(\beta)| \xrightarrow{\mathcal{P}} 0.$$

Therefore, $\Lambda_{n,k^*} \subset \hat{\Lambda}^1_{k^*}(\beta)\times\hat{\Lambda}^2_{n,k^*}(\beta)$ for all $\beta\in\mathcal{B}$ w.p.a.1, which completes the proof of Lemma 5.1. □

Lemma 5.2. *Suppose that Assumptions 3.1 – 3.4 hold, and there exists a sequence* $\{\overline{\beta}_{n,k^*}\} \subset \mathcal{B}$ *such that*

$$\overline{\beta}_{n,k^*} \xrightarrow{\mathcal{P}} \beta_0,\ \hat{g}^1_{k^*}(\overline{\beta}_{n,k^*}) = O_p(k^{*-1/2}) \text{ and } \hat{g}^2_{n,k^*}(\overline{\beta}_{n,k^*}) = O_p((n-k^*)^{-1/2})$$

as $n\to\infty$. *Denote* $\overline{\beta}_{n,k^*}$ *by* $\overline{\beta}$. *Then, under* H_0,

$$\overline{\lambda} := \arg\max_{\lambda\in\hat{\Lambda}^1_{k^*}(\overline{\beta})} \hat{P}^1_{k^*}(\overline{\beta},\lambda) \quad \text{and} \quad \overline{\eta} := \arg\max_{\eta\in\hat{\Lambda}^2_{n,k^*}(\overline{\beta})} \hat{P}^2_{n,k^*}(\overline{\beta},\eta)$$

exist w.p.a.1. Moreover, as $n\to\infty$ *we have*

$$\begin{aligned} \overline{\lambda} &= O_p(k^{*-1/2}),\quad \overline{\eta} = O_p((n-k^*)^{-1/2}),\\ \hat{P}^1_{k^*}(\overline{\beta},\overline{\lambda}) &= O_p(k^{*-1}),\quad \hat{P}^2_{n,k^*}(\overline{\beta},\overline{\eta}) = O_p((n-k^*)^{-1}). \end{aligned}$$

Proof. We only show the statement for $\overline{\lambda}$, the corresponding statement for $\overline{\eta}$ follows by similar arguments. Since Λ_{n,k^*} is a closed set, it follows that

$$\check{\lambda} := \arg \max_{\lambda \in \Lambda_{n,k^*}} \hat{P}^1_{k^*}(\overline{\beta}, \lambda)$$

exists (note that $\hat{P}_{k^*}(\overline{\beta}, \lambda)$ is a concave function of λ). From Lemma 5.1 it follows that $\hat{P}^1_{k^*}(\overline{\beta}, \lambda)$ is continuously twice differentiable with respect to λ w.p.a.1. By a Taylor expansion at $\lambda = 0_m$, there exists a point $\dot{\lambda}$ on the line joining $\check{\lambda}$ and 0_m such that

$$\begin{aligned} 0 = \hat{P}^1_{k^*}(\overline{\beta}, 0_m) &\leq \hat{P}^1_{k^*}(\overline{\beta}, \check{\lambda}) \\ &= -\check{\lambda}' \hat{g}^1_{k^*}(\overline{\beta}) + \frac{1}{2}\check{\lambda}'\Big[\frac{1}{k^*}\sum_{i=1}^{k^*} \rho^1_i(\dot{\lambda}) g_i(\overline{\beta}) g_i(\overline{\beta})'\Big]\check{\lambda}, \end{aligned} \tag{5.1}$$

where $\rho^1_i(\lambda) = -1/(1 - \lambda' g_i(\overline{\beta}))^2$. Note that the definition of $g_i(\beta)$ implies

$$g_i(\beta) g_i(\beta)' = \frac{1}{4} a^*(X_{i-1}) a^*(X_{i-1})'$$

for any $\beta \in \mathcal{B}$. By Lemma 5.1 we have $\rho^1_i(\dot{\lambda}) \geq -C$ uniformly with respect to i w.p.a.1. Furthermore, the ergodicity of $\{X_t : t \in \mathbb{Z}\}$ implies that the random variable

$$\hat{\Omega}^1_{k^*} := (4k^*)^{-1} \sum_{i=1}^{k^*} a^*(X_{i-1}) a^*(X_{i-1})'$$

converges to Ω in probability. Hence the minimum eigenvalue of $\hat{\Omega}^1_{k^*}$ is bounded away from 0 w.p.a.1. and we obtain

$$\begin{aligned} -\check{\lambda}' \hat{g}^1_{k^*}(\overline{\beta}) + \frac{1}{2}\check{\lambda}'\Big[\frac{1}{k^*}\sum_{i=1}^{k^*} \rho^1_i(\dot{\lambda}) g_i(\overline{\beta}) g_i(\overline{\beta})'\Big]\check{\lambda} &\leq \|\check{\lambda}\| \, \|\hat{g}^1_{k^*}(\overline{\beta})\| - \frac{C}{2}\check{\lambda}\hat{\Omega}^1_{k^*}\check{\lambda} \\ &\leq \|\check{\lambda}\| \|\hat{g}^1_{k^*}(\overline{\beta})\| - C\|\check{\lambda}\|^2 \end{aligned} \tag{5.2}$$

w.p.a.1. Dividing both sides of (5.2) by $\|\check{\lambda}\|$, we get

$$\|\check{\lambda}\| = O_p(k^{*-1/2}) = o_p(k^{*-\zeta}),$$

and hence $\check{\lambda} \in \mathrm{Int}(\hat{\Lambda}^1_{k^*})$ w.p.a.1. Again by Lemma 5.1, the concavity of $\hat{P}^1_{k^*}(\overline{\beta}, \lambda)$ and the convexity of $\hat{\Lambda}^1_{k^*}(\overline{\beta})$, it follows that $\overline{\lambda} = \check{\lambda}$ exists w.p.a.1 and $\overline{\lambda} = O_p(k^{*-1/2})$. These results also imply that $\hat{P}^1_{k^*}(\overline{\beta}, \overline{\lambda}) = O_p(k^{*-1})$. By similar arguments, we can show the corresponding results for $\overline{\eta}$ and $\hat{P}^2_{n,k^*}(\overline{\beta}, \overline{\eta})$. □

Next, let us consider the estimator $\hat{\beta}_{n,k}$ of Theorem 3.1. Recall that $\hat{\beta}_{n,k}$ is the minimizer of

$$l_{n,k}(\beta, \beta) = k \sup_{\lambda \in \hat{\Lambda}^1_k(\beta)} \hat{P}^1_k(\beta, \lambda) + (n-k) \sup_{\eta \in \hat{\Lambda}^2_{n,k}(\beta)} \hat{P}^2_{n,k}(\beta, \eta).$$

Let us define

$$\hat{P}_{n,k}(\beta, \lambda, \eta) := k\hat{P}^1_k(\beta, \lambda) + (n-k)\hat{P}^2_{n,k}(\beta, \eta) \tag{5.3}$$

and

$$\hat{\lambda}_{n,k} := \arg \max_{\lambda \in \hat{\Lambda}^1_k(\hat{\beta}_{n,k})} \hat{P}^1_k(\hat{\beta}_{n,k}, \lambda), \quad \hat{\eta}_{n,k} := \arg \max_{\eta \in \hat{\Lambda}^2_{n,k}(\hat{\beta}_{n,k})} \hat{P}^2_{n,k}(\hat{\beta}_{n,k}, \eta). \tag{5.4}$$

Lemma 5.3. *Suppose that Assumptions 3.1 – 3.4 hold. Then, under the null hypothesis H_0 of no change point we have*

$$\hat{g}^1_{k^*}(\hat{\beta}_{n,k^*}) = O_p(k^{*-1/2}), \quad \hat{g}^2_{n,k^*}(\hat{\beta}_{n,k^*}) = O_p((n-k^*)^{-1/2})$$

as $n \to \infty$.

Proof. Define $\hat{g}^l_{n,k} := \hat{g}^l(\hat{\beta}_{n,k})$ for $l = 1, 2$,

$$\tilde{\lambda}_{n,k} := -k^{-1/2}\hat{g}^1_{n,k}/\|\hat{g}^1_{n,k}\|, \quad \tilde{\eta}_{n,k} := -(n-k)^{-1/2}\hat{g}^2_{n,k}/\|\hat{g}^2_{n,k}\|, \tag{5.5}$$

then it follows from (5.4) that

$$\hat{P}^1_k(\hat{\beta}_{n,k}, \tilde{\lambda}_{n,k}) \le \hat{P}^1_k(\hat{\beta}_{n,k}, \hat{\lambda}_{n,k}) \quad \text{and} \quad \hat{P}^2_{n,k}(\hat{\beta}_{n,k}, \tilde{\eta}_{n,k}) \le \hat{P}^2_{n,k}(\hat{\beta}_{n,k}, \hat{\eta}_{n,k}),$$

which implies the inequality

$$\hat{P}_{n,k}(\hat{\beta}_{n,k}, \tilde{\lambda}_{n,k}, \tilde{\eta}_{n,k}) \le \hat{P}_{n,k}(\hat{\beta}_{n,k}, \hat{\lambda}_{n,k}, \hat{\eta}_{n,k}). \tag{5.6}$$

By similar arguments as used in (5.1) and (5.2) we have

$$\hat{P}_{n,k^*}(\hat{\beta}_{n,k^*}, \tilde{\lambda}_{n,k^*}, \tilde{\eta}_{n,k^*}) \geq k^{*1/2}\|\hat{\hat{g}}^1_{n,k^*}\| + (n-k^*)^{1/2}\|\hat{\hat{g}}^2_{n,k^*}\| - c_0 \tag{5.7}$$

w.p.a.1, where c_0 is the same constant as in the proof of Lemma 5.2. On the other hand, we have the following inequality:

$$\begin{aligned}\hat{P}_{n,k}(\hat{\beta}_{n,k}, \hat{\lambda}_{n,k}, \hat{\eta}_{n,k}) &= \inf_{\beta\in\mathcal{B}} \sup_{\lambda\in\hat{\Lambda}^1_k(\hat{\beta}_{n,k}),\ \eta\in\hat{\Lambda}^2_{n,k}(\hat{\beta}_{n,k})} \hat{P}_{n,k}(\beta,\lambda,\eta) \\ &\leq \sup_{\lambda\in\hat{\Lambda}^1_k(\beta_0),\ \eta\in\hat{\Lambda}^2_{n,k}(\beta_0)} \hat{P}_{n,k}(\beta_0,\lambda,\eta) \\ &\leq k \sup_{\lambda\in\hat{\Lambda}^1_k(\beta_0)} \hat{P}^1_k(\beta_0,\lambda) + (n-k) \sup_{\eta\in\hat{\Lambda}^2_{n,k}(\beta_0)} \hat{P}^2_{n,k}(\beta_0,\eta). \end{aligned}\tag{5.8}$$

Applying Lemma 5.2 with $\overline{\beta}_{n,k^*} = \beta_0$ yields

$$\sup_{\lambda\in\hat{\Lambda}^1_{k^*}(\beta_0)} \hat{P}^1_{k^*}(\beta_0,\lambda) = O_p(k^{*-1}), \qquad \sup_{\eta\in\hat{\Lambda}^2_{n,k^*}(\beta_0)} \hat{P}^2_{n,k^*}(\beta_0,\eta) = O_p((n-k^*)^{-1}), \tag{5.9}$$

and from (5.8) and (5.9), we get

$$\hat{P}_{n,k^*}(\hat{\beta}_{n,k^*}, \hat{\lambda}_{n,k^*}, \hat{\eta}_{n,k^*}) = O_p(1). \tag{5.10}$$

Finally, from (5.6), (5.7) and (5.10), we have

$$-c_0 \leq -c_0 + k^{*1/2}\|\hat{\hat{g}}^1_{n,k^*}\| + (n-k^*)^{1/2}\|\hat{\hat{g}}^2_{n,k^*}\| \leq \hat{P}_{n,k^*}(\hat{\beta}_{n,k^*}, \hat{\lambda}_{n,k^*}, \hat{\eta}_{n,k^*}) = O_p(1),$$

which implies

$$\|\hat{\hat{g}}^1_{n,k^*}\| = \|\hat{g}^1_{k^*}(\hat{\beta}_{n,k^*})\| = O_p(k^{*-1/2}) \quad \text{and} \quad \|\hat{\hat{g}}^2_{n,k^*}\| = \|\hat{g}^2_{n,k^*}(\hat{\beta}_{n,k^*})\| = O_p((n-k^*)^{-1/2}),$$

establishing the assertion of Lemma 5.3. □

Proof. [Proof of Theorem 3.1] By Lemma 5.3 we have $\hat{g}(\hat{\beta}_{n,k^*}) = o_p(1)$. Then, it follows from the triangular inequality and uniform law of large numbers that

$$\|g(\hat{\beta}_{n,k^*})\| \leq \|g(\hat{\beta}_{n,k^*}) - \hat{g}(\hat{\beta}_{n,k^*})\| + \|\hat{g}(\hat{\beta}_{n,k^*})\|$$

$$\leq \sup_{\beta \in \mathcal{B}} \|g(\beta) - \hat{g}(\beta)\| + \|\hat{g}(\hat{\beta}_{n,k^*})\| = o_p(1).$$

Since $g(\beta)$ has a unique zero at β_0, the function $\|g(\beta)\|$ must be bounded away from zero outside any neighborhood of β_0. Therefore, $\hat{\beta}_{n,k^*}$ must be inside any neighborhood of β_0 w.p.a.1. and therefore, $\hat{\beta}_{n,k^*} \xrightarrow{\mathcal{P}} \beta_0$.
Next, we show that $\hat{\beta}_{n,k^*} - \beta_0 = O_p(n^{-1/2})$. As $k^* = rn$, by Lemma 5.3, we have

$$\hat{g}(\hat{\beta}_{n,k^*}) = n^{-1} \left\{ k^* \hat{g}^1_{k^*}(\hat{\beta}_{n,k^*}) + (n - k^*) \hat{g}^2_{n,k^*}(\hat{\beta}_{n,k^*}) \right\} = O_p(n^{-1/2})$$

and the central limit theorem implies

$$\hat{g}(\beta_0) = O_p \left[n^{-1} \left\{ k^{*1/2} + (n - k^*)^{1/2} \right\} \right] = O_p(n^{1/2}).$$

Further,

$$\|\hat{g}(\hat{\beta}_{n,k^*}) - \hat{g}(\beta_0) - g(\hat{\beta}_{n,k^*})\| \leq (1 + \sqrt{n}\|\hat{\beta}_{n,k^*} - \beta_0\|) o_p(n^{-1/2}), \tag{5.11}$$

which yields

$$\begin{aligned}\|g(\hat{\beta}_{n,k^*})\| &\leq \|\hat{g}(\hat{\beta}_{n,k^*}) - \hat{g}(\beta_0) - g(\hat{\beta}_{n,k^*})\| + \|\hat{g}(\hat{\beta}_{n,k^*})\| + \|\hat{g}(\beta_0)\| \\ &= (1 + \sqrt{n}\|\hat{\beta}_{n,k^*} - \beta_0\|) o_p(n^{-1/2}) + O_p \left[n^{-1} \left\{ k^{*1/2} + (n - k^*)^{1/2} \right\} \right].\end{aligned}$$

Moreover, similar arguments as given in Newey and McFadden (1994) on page 2191, the differentiability of $\|g(\beta)\|$ and the estimate $\|g(\hat{\beta}_n)\| \geq C\|\hat{\beta}_n - \beta_0\|$ w.p.a.1. show that

$$\|\hat{\beta}_{n,k^*} - \beta_0\| = (1 + \sqrt{n}\|\hat{\beta}_{n,k^*} - \beta_0\|) o_p(n^{-1/2}) + O_p \left[n^{-1} \left\{ k^{*1/2} + (n - k^*)^{1/2} \right\} \right],$$

and hence

$$\{1 + o_p(1)\}\|\hat{\beta}_{n,k^*} - \beta_0\| = o_p(n^{-1/2}) + O_p \left[n^{-1} \left\{ k^{*1/2} + (n - k^*)^{1/2} \right\} \right]. \tag{5.12}$$

If $k^* = rn$ the right-hand side of (5.12) is of order $O_p(n^{-1/2})$, which completes the proof of Theorem 3.1.

5.2 Proof of Theorem 3.2

We first show that $\hat{P}_{n,k^*}(\beta,\lambda,\eta)$ in (5.3) is well approximated by some function near its optima using a similar reasoning as in Parente and Smith (2011). For this purpose let us define

$$\hat{L}_k^1(\beta,\lambda) = \{-G(\beta-\beta_0) - \hat{g}_k^1(\beta_0)\}'\lambda - \frac{1}{2}\lambda'\Omega\lambda,$$
$$\hat{L}_{n,k}^2(\beta,\eta) = \{-G(\beta-\beta_0) - \hat{g}_{n,k}^2(\beta_0)\}'\eta - \frac{1}{2}\eta'\Omega\eta$$

and

$$\hat{L}_{n,k}(\beta,\lambda,\eta) := k\hat{L}_k^1(\beta,\lambda) + (n-k)\hat{L}_{n,k}^2(\beta,\eta).$$

Furthermore, hereafter redefine

$$\tilde{\beta}_{n,k} := \arg\min_{\beta\in\mathcal{B}} \sup_{\lambda\in\mathbb{R}^m,\eta\in\mathbb{R}^m} \hat{L}_{n,k}(\beta,\lambda,\eta),$$
$$\tilde{\lambda}_{n,k} := \arg\max_{\lambda\in\mathbb{R}^m} \hat{L}_k^1(\tilde{\beta},\lambda) \quad \text{and} \quad \tilde{\eta}_{n,k} := \arg\max_{\eta\in\mathbb{R}^m} \hat{L}_{n,k}^2(\tilde{\beta},\eta).$$

Lemma 5.4. *Suppose that Assumptions 3.1-3.4 hold. Then, under H_0,*

$$\hat{P}_{n,k^*}(\hat{\beta}_{n,k^*},\hat{\lambda}_{n,k^*},\hat{\eta}_{n,k^*}) = \hat{L}_{n,k^*}(\tilde{\beta}_{n,k^*},\tilde{\lambda}_{n,k^*},\tilde{\eta}_{n,k^*}) + o_p(1)$$

as $n\to\infty$.

Proof. It is sufficient to show that

(i) $\hat{P}_{n,k^*}(\hat{\beta}_{n,k^*},\hat{\lambda}_{n,k^*},\hat{\eta}_{n,k^*}) - \hat{L}_{n,k^*}(\hat{\beta}_{n,k^*},\hat{\lambda}_{n,k^*},\hat{\eta}_{n,k^*}) = o_p(1)$,

(ii) $\hat{L}_{n,k^*}(\hat{\beta}_{n,k^*},\hat{\lambda}_{n,k^*},\hat{\eta}_{n,k^*}) - \hat{L}_{n,k^*}(\tilde{\beta}_{n,k^*},\hat{\lambda}_{n,k^*},\hat{\eta}_{n,k^*}) = o_p(1)$,

(iii) $\hat{L}_{n,k^*}(\tilde{\beta}_{n,k^*},\hat{\lambda}_{n,k^*},\hat{\eta}_{n,k^*}) - \hat{L}_{n,k^*}(\tilde{\beta}_{n,k^*},\tilde{\lambda}_{n,k^*},\tilde{\eta}_{n,k^*}) = o_p(1)$.

For a proof of (i) we note that a Taylor expansion leads to

$$\hat{P}_k^1(\hat{\beta}_{n,k},\hat{\lambda}_{n,k}) = -\hat{\lambda}_{n,k}'\hat{\hat{g}}_{n,k}^1 + \frac{1}{2}\hat{\lambda}_{n,k}'\Big[\frac{1}{k}\sum_{i=1}^{k}\rho_i^1(\ddot{\lambda})a^*(X_{i-1})a^*(X_{i-1})'\Big]\hat{\lambda}_{n,k},$$

where $\ddot{\lambda}$ is on the line joining the points $\hat{\lambda}_{n,k}$ and 0_m. Observing the definition of $\hat{L}^1(\hat{\beta}_{n,k}, \hat{\lambda}_{n,k})$ this yields the estimate

$$\begin{aligned}\left|\hat{P}_k^1(\hat{\beta}_{n,k}, \hat{\lambda}_{n,k}) - \hat{L}^1(\hat{\beta}_{n,k}, \hat{\lambda}_{n,k})\right| \ &\leq \ \left| -\left(\hat{\hat{g}}_{n,k}^1 - \hat{g}_k^1(\beta_0) - G(\hat{\beta} - \beta_0)\right)' \hat{\lambda}_{n,k}\right| \\ &+ \ \left|\frac{1}{2}\hat{\lambda}'_{n,k}\Big[\frac{1}{k}\sum_{i=1}^{k} \dot{\rho}_i^1 a^*(X_{i-1})a^*(X_{i-1})' + \Omega\Big]\hat{\lambda}_{n,k}\right|. \end{aligned} \tag{5.13}$$

Since $\hat{\beta}_{n,k^*} \xrightarrow{\mathcal{P}} \beta_0$ by Theorem 3.1, we can take $\overline{\beta}_{n,k^*} = \hat{\beta}_{n,k^*}$ in Lemma 5.2, and obtain $\hat{\lambda}_{n,k^*} = O_p(n^{-1/2})$. Then, recalling (5.11), the first term in (5.13) (where k is replaced by k^*) becomes

$$\begin{aligned}&\left| -\left(\hat{\hat{g}}_{n,k^*}^1 - \hat{g}_{k^*}^1(\beta_0) - G(\hat{\beta}_{n,k^*} - \beta_0)\right)' \hat{\lambda}_{n,k^*}\right| \\ &\leq \left\{\left\|\hat{\hat{g}}_{n,k^*}^1 - \hat{g}_{k^*}^1(\beta_0) - g(\hat{\beta}_{n,k^*})\right\| + \left\|g(\hat{\beta}_{n,k^*}) - G(\hat{\beta}_{n,k^*} - \beta_0)\right\|\right\} \left\|\hat{\lambda}_{n,k^*}\right\| \\ &= \left\{ \left(1 + \sqrt{n}\left\|\hat{\beta}_{n,k^*} - \beta_0\right\|\right) o_p(n^{-1/2}) + O_p\Big(\left\|\hat{\beta}_{n,k^*} - \beta_0\right\|^2 \Big)\right\} O_p(n^{-1/2}) \\ &= o_p(n^{-1}).\end{aligned}$$

Moreover, the second term in (5.13) is of order $o_p(k^{*-1})$. Hence, we get

$$\left|\hat{P}_{k^*}^1(\hat{\beta}_{n,k^*}, \hat{\lambda}_{n,k^*}) - \hat{L}^1(\hat{\beta}_{n,k}, \hat{\lambda}_{n,k^*})\right| = o_p(k^{*-1})$$

and similarly

$$\left|\hat{P}_{n,k^*}^2(\hat{\beta}_{n,k^*}, \hat{\eta}_{n,k^*}) - \hat{L}^2(\hat{\beta}_{n,k^*}, \hat{\eta}_{n,k^*})\right| = o_p((n-k^*)^{-1}).$$

Combining these estimates yields

$$\hat{P}_{n,k^*}(\hat{\beta}_{n,k^*}, \hat{\lambda}_{n,k^*}, \hat{\eta}_{n,k^*}) - \hat{L}_{n,k^*}(\hat{\beta}_{n,k^*}, \hat{\lambda}_{n,k^*}, \hat{\eta}_{n,k^*}) = o_p(1),$$

which is the statement (i).

For a proof of (ii) we first show

$$\left|\hat{P}_{n,k^*}(\tilde{\beta}_{n,k^*}, \hat{\lambda}_{n,k^*}, \hat{\eta}_{n,k^*}) - \hat{L}_{n,k^*}(\tilde{\beta}_{n,k^*}, \hat{\lambda}_{n,k^*}, \hat{\eta}_{n,k^*})\right| = o_p(1).$$

Note that the function $\hat{L}_{n,k}(\beta, \lambda, \eta)$ is smooth in β, λ and η. Then, the first order conditions for an interior global maximum

$$0_p = \frac{\partial \hat{L}_{n,k}(\beta, \lambda, \eta)}{\partial \beta} = -G'\left\{k\lambda + (n-k)\eta\right\},$$

$$0_m = \frac{\partial \hat{L}_{n,k}(\beta,\lambda,\eta)}{\partial \lambda} = -k\left\{G(\beta-\beta_0) + \hat{g}^1_k(\beta_0) + \Omega\lambda\right\},$$
$$0_m = \frac{\partial \hat{L}_{n,k}(\beta,\lambda,\eta)}{\partial \eta} = -(n-k)\left\{G(\beta-\beta_0) + \hat{g}^2_{n,k}(\beta_0) + \Omega\eta\right\}$$

are satisfied for the point $(\beta',\lambda',\eta') = (\tilde{\beta}'_{n,k}, \tilde{\lambda}'_{n,k}, \tilde{\eta}'_{n,k})$. These conditions can be rewritten in matrix form as

$$\begin{pmatrix} O_{p\times p} & G' & G' \\ G & k^{-1}\Omega & O_{m\times m} \\ G & O_{m\times m} & (n-k)^{-1}\Omega \end{pmatrix} \begin{pmatrix} \tilde{\beta}_{n,k} - \beta_0 \\ k\tilde{\lambda}_{n,k} \\ (n-k)\tilde{\eta}_{n,k} \end{pmatrix} + \begin{pmatrix} 0_p \\ \hat{g}^1_k(\beta_0) \\ \hat{g}^2_{n,k}(\beta_0) \end{pmatrix} = 0_{p+2m}. \quad (5.14)$$

With the notations

$$\begin{aligned} \Sigma &:= (G'\Omega^{-1}G)^{-1}, \quad H := \Omega^{-1}G\Sigma, \\ P^1_k &:= \Omega^{-1} - \frac{k}{n}H\Sigma^{-1}H', \quad P^2_{n,k} := \Omega^{-1} - \frac{n-k}{n}H\Sigma^{-1}H', \end{aligned}$$

the system (5.14) is equivalent to

$$\begin{aligned} &\begin{pmatrix} \tilde{\beta}_{n,k} - \beta_0 \\ k\tilde{\lambda}_{n,k} \\ (n-k)\tilde{\eta}_{n,k} \end{pmatrix} \\ &= \begin{pmatrix} n^{-1}\Sigma & -kn^{-1}H' & -(n-k)n^{-1}H' \\ -kn^{-1}H & -kP^1_k & k(n-k)n^{-1}H\Sigma^{-1}H' \\ -(n-k)n^{-1}H & k(n-k)n^{-1}H\Sigma^{-1}H' & -(n-k)P^2_{n,k} \end{pmatrix} \begin{pmatrix} 0_p \\ \hat{g}^1_k(\beta_0) \\ \hat{g}^2_{n,k}(\beta_0) \end{pmatrix} \\ &= \begin{pmatrix} -H'\hat{g}(\beta_0) \\ -k\left\{\Omega^{-1}\hat{g}^1_k(\beta_0) - H\Sigma^{-1}H'\hat{g}(\beta_0)\right\} \\ -(n-k)\left\{\Omega^{-1}\hat{g}^2_{n,k}(\beta_0) - H\Sigma^{-1}H'\hat{g}(\beta_0)\right\} \end{pmatrix}. \end{aligned} \quad (5.15)$$

Consequently, $\tilde{\beta}_{n,k^*} - \beta_0$, $\tilde{\lambda}_{n,k^*}$ and $\tilde{\eta}_{n,k^*}$ are of order $O_p(n^{-1/2})$, $O_p(k^{*-1/2})$ and $O_p((n-k^*)^{-1/2})$, respectively. Therefore, by the same arguments as given in the proof of (i), it follows that

$$|\hat{P}_{n,k^*}(\tilde{\beta}_{n,k^*}, \hat{\lambda}_{n,k^*}, \hat{\eta}_{n,k^*}) - \hat{L}_{n,k^*}(\tilde{\beta}_{n,k^*}, \hat{\lambda}_{n,k^*}, \hat{\eta}_{n,k^*})| = o_p(1).$$

This relationship and the fact that $(\hat{\beta}'_{n,k^*}, \hat{\lambda}'_{n,k^*}, \hat{\eta}'_{n,k^*})'$ and $(\tilde{\beta}'_{n,k^*}, \tilde{\lambda}'_{n,k^*}, \tilde{\eta}'_{n,k^*})'$ are the saddle points of the functions $\hat{P}_{n,k^*}(\beta, \lambda, \eta)$ and $\hat{L}_{n,k^*}(\beta, \lambda, \eta)$, respectively, imply that

$$\begin{aligned} \hat{L}_{n,k^*}(\hat{\beta}_{n,k^*}, \hat{\lambda}_{n,k^*}, \hat{\eta}_{n,k^*}) &= \hat{P}_{n,k^*}(\hat{\beta}_{n,k^*}, \hat{\lambda}_{n,k^*}, \hat{\eta}_{n,k^*}) + o_p(1) \\ &\le \hat{P}_{n,k^*}(\tilde{\beta}_{n,k^*}, \hat{\lambda}_{n,k^*}, \hat{\eta}_{n,k^*}) + o_p(1) \\ &= \hat{L}_{n,k^*}(\tilde{\beta}_{n,k^*}, \hat{\lambda}_{n,k^*}, \hat{\eta}_{n,k^*}) + o_p(1). \end{aligned} \tag{5.16}$$

On the other hand,

$$\begin{aligned} \hat{L}_{n,k^*}(\tilde{\beta}_{n,k^*}, \hat{\lambda}_{n,k^*}, \hat{\eta}_{n,k^*}) &\le \hat{L}_{n,k^*}(\tilde{\beta}_{n,k^*}, \tilde{\lambda}_{n,k^*}, \tilde{\eta}_{n,k^*}) \\ &\le \hat{L}_{n,k^*}(\hat{\beta}_{n,k^*}, \tilde{\lambda}_{n,k^*}, \tilde{\eta}_{n,k^*}) \\ &= \hat{P}_{n,k^*}(\hat{\beta}_{n,k^*}, \tilde{\lambda}_{n,k^*}, \tilde{\eta}_{n,k^*}) + o_p(1) \\ &\le \hat{P}_{n,k^*}(\hat{\beta}_{n,k^*}, \hat{\lambda}_{n,k^*}, \hat{\eta}_{n,k^*}) + o_p(1) \\ &= \hat{L}_{n,k^*}(\hat{\beta}_{n,k^*}, \hat{\lambda}_{n,k^*}, \hat{\eta}_{n,k^*}) + o_p(1). \end{aligned} \tag{5.17}$$

Thus, (5.16) and (5.17) lead to

$$\hat{L}_{n,k^*}(\hat{\beta}_{n,k^*}, \hat{\lambda}_{n,k^*}, \hat{\eta}_{n,k^*}) - \hat{L}_{n,k^*}(\tilde{\beta}_{n,k^*}, \hat{\lambda}_{n,k^*}, \hat{\eta}_{n,k^*}) = o_p(1).$$

Finally, we can prove (iii) by similar arguments that

$$\begin{aligned} \hat{L}_{n,k^*}(\tilde{\beta}_{n,k^*}, \tilde{\lambda}_{n,k^*}, \tilde{\eta}_{n,k^*}) &\le \hat{L}_{n,k^*}(\hat{\beta}_{n,k^*}, \tilde{\lambda}_{n,k^*}, \tilde{\eta}_{n,k^*}) \\ &= \hat{P}_{n,k^*}(\hat{\beta}_{n,k^*}, \tilde{\lambda}_{n,k^*}, \tilde{\eta}_{n,k^*}) + o_p(1) \\ &\le \hat{P}_{n,k^*}(\hat{\beta}_{n,k^*}, \hat{\lambda}_{n,k^*}, \hat{\eta}_{n,k^*}) + o_p(1) \\ &\le \hat{P}_{n,k^*}(\tilde{\beta}_{n,k^*}, \hat{\lambda}_{n,k^*}, \hat{\eta}_{n,k^*}) + o_p(1) \\ &= \hat{L}_{n,k^*}(\tilde{\beta}_{n,k^*}, \hat{\lambda}_{n,k^*}, \hat{\eta}_{n,k^*}) + o_p(1) \end{aligned}$$

and

$$\hat{L}_{n,k^*}(\tilde{\beta}_{n,k^*}, \hat{\lambda}_{n,k^*}, \hat{\eta}_{n,k^*}) \le \hat{L}_{n,k^*}(\tilde{\beta}_{n,k^*}, \tilde{\lambda}_{n,k^*}, \tilde{\eta}_{n,k^*}).$$

Consequently, $\hat{L}_{n,k^*}(\tilde{\beta}_{n,k^*}, \hat{\lambda}_{n,k^*}, \hat{\eta}_{n,k^*}) = \hat{L}_{n,k^*}(\tilde{\beta}_{n,k^*}, \tilde{\lambda}_{n,k^*}, \tilde{\eta}_{n,k^*}) + o_p(1)$, which implies (iii). □

Proof. [Proof of Theorem 3.2] By (5.3), (5.4), Lemma 5.4 and (5.15) it follows that

$$\begin{aligned}
\sup_{\beta\in\mathcal{B}}\{-l_{n,k^*}(\beta,\beta)\} &= \hat{P}_{n,k^*}(\hat{\beta}_{n,k^*},\hat{\lambda}_{n,k^*},\hat{\eta}_{n,k^*}) \\
&= \hat{L}_{n,k^*}(\tilde{\beta}_{n,k^*},\tilde{\lambda}_{n,k^*},\tilde{\eta}_{n,k^*}) + R_{n,k^*} \\
&= \frac{k^*}{2}\tilde{\lambda}'_{n,k^*}\Omega\tilde{\lambda}_{n,k^*} + \frac{n-k^*}{2}\tilde{\eta}'_{n,k^*}\Omega\tilde{\eta}_{n,k^*} + R_{n,k^*} + o_p(1) \\
&= \frac{k^*}{2}\hat{g}^1_{k^*}(\beta_0)'\Omega^{-1}\hat{g}^1_{k^*}(\beta_0) + \frac{n-k^*}{2}\hat{g}^2_{n,k^*}(\beta_0)'\Omega^{-1}\hat{g}^2_{n,k^*}(\beta_0) \\
&\quad - \frac{n}{2}\hat{g}(\beta_0)'H\Sigma^{-1}H'\hat{g}(\beta_0) + R_{n,k^*} + o_p(1) \\
&= \hat{M}_{n,k^*} + R_{n,k^*} + o_p(1), \qquad (5.18)
\end{aligned}$$

where

$$\begin{aligned}
\hat{M}_{n,k} &= \frac{\left\|\hat{W}_n(k/n) - (k/n)\hat{W}_n(1)\right\|^2}{2\phi(k/n)} + \frac{\hat{W}_n(1)'Q\hat{W}_n(1)}{2}, \\
\hat{W}_n(r) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{[rn]}\Omega^{-1/2}g(\mathscr{Y}_t^p,\beta_0), \\
R_{n,k} &= \hat{P}_{n,k}(\hat{\beta}_{n,k},\hat{\lambda}_{n,k},\hat{\eta}_{n,k}) - \hat{L}_{n,k}(\tilde{\beta}_{n,k},\tilde{\lambda}_{n,k},\tilde{\eta}_{n,k}),
\end{aligned}$$

$\phi(u) = u(1-u)$ and $[x]$ denotes the integer part of real number x. As shown in Lemma 5.4,

$$\max_{k_{1n}\le k^*\le k_{2n}}|R_{n,k^*}| = \sup_{r_1\le r\le r_2}|R_{n,rn}| = o_p(1).$$

Second, from Assumption 3.4 and Lemma 2.2 in Phillips (1987), it follows that

$$\left\{c'\hat{W}_n(r) : r\in[0,1]\right\} \xrightarrow{\mathcal{L}} \{c'B(r) : r\in[0,1]\},$$

for any vector $c\in\mathbb{R}^m$, where $\{B(r) : r\in[0,1]\}$ is an m-dimensional standard Brownian motion. Hence, the Cramér-Wold device and the continuous mapping theorem lead to

$$\begin{aligned}
\tilde{T}_n &= 2\max_{k_{1n}\le k\le k_{2n}}\left\{h\Big(\frac{k}{n}\Big)\hat{M}_{n,k}\right\} \\
&= \sup_{k_{1n}/n\le r\le k_{2n}/n}\left\{\frac{h(k/n)}{\phi(k/n)}\left\|\hat{W}_n(r) - ([rn]/n)\hat{W}_n(1)\right\|^2 + h([rn]/n)\hat{W}_n(1)'Q\hat{W}_n(1)\right\}
\end{aligned}$$

$$\xrightarrow{\mathcal{L}} \sup_{r_1 \le r \le r_2} \Big\{ \frac{h(r)}{\phi(r)} \|B(r) - rB(1)\|^2 + h(r)B(1)'QB(1) \Big\}.$$

5.3 Proof of Theorem 3.3

Proof. Without loss of generality, suppose that $\theta_2 \neq \beta_0$. This implies that there exist a neighborhood $U(\beta_0)$ of β_0 and a neighborhood $U(\theta_2)$ of θ_2 such that

$$U(\beta_0) \cap U(\theta_2) = \emptyset.$$

Under the alternative it follows that $\hat{\beta}_{n,k^*} \notin U(\beta_0)$ or $\hat{\beta}_{n,k^*} \notin U(\theta_2)$. Note that $\mathbb{E}[g(\mathcal{Y}_t^p, \theta_2)] \neq 0$ for $1 \le t \le k^*$ and $\mathbb{E}[g(\mathcal{Y}_t^p, \beta_0)] \neq 0$ for $k^* + 1 \le t \le n$. From a uniform law of large numbers, $\hat{g}^1_{k^*}(\hat{\beta}_{n,k^*})$ or $\hat{g}^2_{n,k^*}(\hat{\beta}_{n,k^*})$ is outside a neighborhood of 0 for any sufficiently large n.
Now, if we consider $\hat{g}^1_k(\hat{\beta}_{n,k^*})$ instead of $\hat{g}^1_k(\beta_0)$ and $\hat{g}^2_{n,k}(\hat{\beta}_{n,k^*})$ instead of $\hat{g}^2_{n,k}(\beta_0)$ in (5.14), we find, as in (5.18), that $\sup_{\beta \in \mathcal{B}} \{-l_{n,k^*}(\beta, \beta)\}$ can be approximated by

$$\frac{k^*}{2} \hat{g}^1_{k^*}(\hat{\beta}_{n,k^*})' \Omega^{-1} \hat{g}^1_{k^*}(\hat{\beta}_{n,k^*}) + \frac{n-k^*}{2} \hat{g}^2_{n,k^*}(\hat{\beta}_{n,k^*})' \Omega^{-1} \hat{g}^2_{n,k^*}(\hat{\beta}_{n,k^*}) \\ - \frac{n}{2} \hat{g}(\hat{\beta}_{n,k^*})' H \Sigma^{-1} H' \hat{g}(\hat{\beta}_{n,k^*}) + R_{n,k^*} + o_p(1).$$

This time, however, we have

$$\frac{k^*}{2} \hat{g}^1_{k^*}(\hat{\beta}_{n,k^*})' \Omega^{-1} \hat{g}^1_{k^*}(\hat{\beta}_{n,k^*}) + \frac{n-k^*}{2} \hat{g}^2_{n,k^*}(\hat{\beta}_{n,k^*})' \Omega^{-1} \hat{g}^2_{n,k^*}(\hat{\beta}_{n,k^*}) \to \infty,$$

since $\hat{g}^1_{k^*}(\hat{\beta}_{n,k^*})' \Omega^{-1} \hat{g}^1_{k^*}(\hat{\beta}_{n,k^*}) + \hat{g}^2_{n,k^*}(\hat{\beta}_{n,k^*})' \Omega^{-1} \hat{g}^2_{n,k^*}(\hat{\beta}_{n,k^*}) > 0$ for any sufficiently large n. This completes the proof of Theorem 3.3.

Acknowledgements. The authors would like to thank Martina Stein who typed this manuscript with considerable technical expertise. The work of authors was supported by JSPS Grant-in-Aid for Young Scientists (B) (16K16022), Waseda University Grant for Special Research Projects (2016S-063) and the Deutsche Forschungsgemeinschaft (SFB 823: Statistik nichtlinearer dynamischer Prozesse, Teilprojekt A1 and C1).

References

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica: Journal of the Econometric Society 61*(4), 821–856.

Aue, A. and L. Horváth (2013). Structural breaks in time series. *Journal of Time Series Analysis 34*(1), 1–16.

Bai, J. (1993). On the partial sums of residuals in autoregressive and moving average models. *Journal of Time Series Analysis 14*, 247–260.

Bai, J. (1994). Convergence of the sequential empirical processes of residuals in ARMA models. *Annals of Statistics 22*(4), 2051–2061.

Baragona, R., F. Battaglia, D. Cucina, et al. (2013). Empirical likelihood for break detection in time series. *Electronic Journal of Statistics 7*, 3089–3123.

Berkes, I., Horváth, S. Ling, and J. Schauer (2011). Testing for structural change of AR model to threshold AR model. *Journal of Time Series Analysis 32*(5), 547–565.

Chen, K., Z. Ying, H. Zhang, and L. Zhao (2008). Analysis of least absolute deviation. *Biometrika 95*(1), 107–122.

Chuang, C.-S. and N. H. Chan (2002). Empirical likelihood for autoregressive models, with applications to unstable time series. *Statistica Sinica 12*, 387–407.

Ciuperca, G. and Z. Salloum (2015). Empirical likelihood test in a posteriori change-point nonlinear model. *Metrika 78*, 919–952.

Cressie, N. and T. R. Read (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society. Series B (Methodological) 46*(3), 440–464.

Csörgö, M. and L. Horváth (1997). *Limit Theorems in Change-Point Analysis*. John Wiley.

Davis, R. A., D. Huang, and Y.-C. Yao (1995). Testing for a change in the parameter values and order of an autoregressive model. *Annals of Statistics 23*(1), 282–304.

Gombay, E. and L. Horváth (1990). Asymptotic distributions of maximum likelihood tests for change in the mean. *Biometrika 77*(2), 411–414.

Lee, S., J. Ha, O. Na, and S. Na (2003). The cusum test for parameter change in time series models. *Scandinavian Journal of Statistics 30*(4), 781–796.

Ling, S. (2005). Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67*(3), 381–393.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing.

Handbook of Econometrics 4, 2111–2245.

Nolan, J. P. (2015). *Stable Distributions - Models for Heavy Tailed Data*. Boston: Birkhauser. In progress, Chapter 1 online at academic2.american.edu/∼jpnolan.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika 75*(2), 237–249.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika 41*(1/2), 100–115.

Page, E. S. (1955). Control charts with warning lines. *Biometrika 42(1-2)*, 243–257.

Pan, J., H. Wang, and Q. Yao (2007). Weighted least absolute deviations estimation for ARMA models with infinite variance. *Econometric Theory 23*(05), 852–879.

Parente, P. M. and R. J. Smith (2011). GEL methods for nonsmooth moment indicators. *Econometric Theory 27*(01), 74–113.

Phillips, P. C. (1987). Time series regression with a unit root. *Econometrica 55*(2), 277–301.

Qin, J. and J. Lawless (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics 22*(1), 300–325.

Qu, Z. (2008). Testing for structural change in regression quantiles. *Journal of Econometrics 146*(1), 170–184.

Samoradnitsky, G. and M. S. Taqqu (1994). *Stable non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, Volume 1. CRC Press.

Su, L. and Z. Xiao (2008). Testing for parameter stability in quantile regression models. *Statistics & Probability Letters 78*(16), 2768–2775.

Zhou, M., H. J. Wang, and Y. Tang (2015). Sequential change point detection in linear quantile regression models. *Statistics & Probability Letters 100*, 98–103.