Test (2012) 21:605–634
DOI 10.1007/s11749-011-0265-z

ORIGINAL PAPER

Monitoring changes in the error distribution of autoregressive models based on Fourier methods

Zdeněk Hlávka · Marie Hušková · Claudia Kirch · Simos G. Meintanis

Received: 7 December 2010 / Accepted: 25 July 2011 / Published online: 8 September 2011

Abstract We develop a procedure for monitoring changes in the error distribution of autoregressive time series while controlling the overall size of the sequential test. The proposed procedure, unlike standard procedures which are also referred to, utilizes the empirical characteristic function of properly estimated residuals. The limit behavior of the test statistic is investigated under the null hypothesis as well as under alternatives. Since the asymptotic null distribution contains unknown parameters, a bootstrap procedure is proposed in order to actually perform the test and corresponding results on the finite–sample performance of the new method are presented. As it turns out the procedure is not only able to detect distributional changes but also changes in the regression coefficient.

Keywords Empirical characteristic function · Change point analysis

Mathematics Subject Classification (2000) 62M10 · 62G10 · 62G20

Z. Hlávka · M. Hušková
Faculty of Mathematics and Physics, Department of Statistics, Charles University in Prague, Sokolovská 83, 186 75 Praha 8, Czech Republic

Z. Hlávka
e-mail: hlavka@karlin.mff.cuni.cz

M. Hušková
e-mail: huskova@karlin.mff.cuni.cz

C. Kirch
Institute of Stochastics, Karlsruhe Institute of Technology, Kaiserstr. 89, 76133 Karlsruhe, Germany
e-mail: claudia.kirch@kit.edu

S.G. Meintanis (✉)
Department of Economics, National and Kapodistrian University of Athens, 8 Pesmazoglou Street, 105 59 Athens, Greece
e-mail: simosmei@econ.uoa.gr

1 Introduction

Change-point analysis for distributional changes with i.i.d. observations and the study of structural breaks in the parameters of time series has received wide attention; see for instance Yao (1990), Horváth (1993), Bai (1993), Davis et al. (1995), Einmahl and McKeague (2003), Hušková et al. (2007, 2008), and Gombay and Serban (2009). For a full-book treatment on theoretical and methodological issues of change-point analysis the reader is referred to Csörgő and Horváth (1997).

On the other hand works on structural breaks due to a change in the distribution of a time series are relatively few. Hušková and Meintanis (2006a, 2006b) develop detection procedures for distributional changes with i.i.d. observations. In this paper we extend their results in two ways. First, the observations need no longer be independent. Instead we assume a linear autoregressive structure. Second, we operate within the framework of on-line monitoring analysis whereby data are not observed at once but arrive in a sequential manner—one by one. Then, following each new observation we would like to know whether our model is still capable of explaining the current observations. This type of procedure plays an increasingly important role in applications as data sets are often collected automatically or without significant costs. Examples include financial data sets, e.g. in risk management (Andreou and Ghysels 2006) or in CAPM models (Aue et al. 2010), as well as medical data sets, e.g. when monitoring intensive care patients (Fried and Imhoff 2004). More applications can be found in other areas of applied statistics. The consideration of such data sets leads to sequential statistical analysis which is also called online monitoring.

To fix the model, let $\{X_j,\ j = p+1, \ldots, n\}$ be an AR(p) process defined by the equation

$$X_j = \boldsymbol{\beta}^T \boldsymbol{X}_{j-1} + \varepsilon_j, \tag{1.1}$$

where $\boldsymbol{X}_{j-1} = (X_{j-1}, \ldots, X_{j-p})^T$, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ is an unknown regression parameter. In (1.1) the errors ε_j are independent, each having a corresponding distribution function F_j with mean zero and finite variance. Also the AR process is assumed to be causal, i.e., the characteristic polynomial $P(z) = 1 - \beta_1 z - \cdots - \beta_p z^p$, is assumed to satisfy $P(z) \neq 0$, $\forall |z| \leq 1$.

The idea in the sequential testing methods which we consider is as follows: We suppose that there exists a historic or training data set $X_1, \ldots, X_T$, with no change, i.e., following (1.1) with $F_1 = \cdots = F_T$. Practically, this is the data set based on which we estimate the appropriate parameters. In particular, we postulate model (1.1) with no change in distribution and estimate β as well as the distribution F_1 of ε_1 from $X_1, \ldots, X_T$. Then we start monitoring, i.e., observing the data $X_{T+1}, X_{T+2}, \ldots$ sequentially. After each new observation we decide whether there is evidence of a change and in this case we terminate the monitoring procedure and decide for the alternative. Otherwise we continue monitoring.

Here we monitor for a change in the distributional aspect of the errors ε_j, i.e., we wish to test the hypotheses

$$\begin{aligned} &\mathbb{H}_0 : F_j = F_0, \quad j = T+1, T+2, \ldots \ vs. \\ &\mathbb{H}_1 : F_{T+j} = F_0, \quad j \leq T + j_0;\ F_{T+j} = F^0 \neq F_0,\ j > T + j_0, \end{aligned} \tag{1.2}$$

of a change in the distribution F_j, where F_0, F^0, and the time of a change $j_0 \geq 1$, are considered unknown. As it turns out, however, the monitoring schemes developed for these distributional changes are also able to detect changes in the regression coefficient.

In view of the fact that the errors are unobserved, typically one computes the residuals

$$\hat{\varepsilon}_j = X_j - \hat{\boldsymbol{\beta}}_T^T \mathbf{X}_{j-1}^T, \tag{1.3}$$

from (1.1) by using some standard estimator $\hat{\boldsymbol{\beta}}_T := \hat{\boldsymbol{\beta}}_T(X_1, \dots, X_T)$ of $\boldsymbol{\beta}$ such as the least squares estimator, based only on the training data set $X_1, \dots, X_T$, and fulfilling

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) = O_P(1), \quad \text{as } T \to \infty.$$

Note that for asymptotic considerations we let the length of the historic data set T go to infinity. Hence the estimation of the model parameters from the historic data set improves. The total number of observations, however, is random (and possibly infinite) as we stop monitoring as soon as we can reject.

Based on the estimated residuals in (1.3), several monitoring schemes may be devised, each corresponding to a standard goodness-of-fit statistic. Traditional goodness-of-fit tests, however, make use of the empirical distribution function (EDF) of these residuals, whereas here we utilize the empirical characteristic function (ECF) of the residuals. This approach was employed by Hušková and Meintanis (2006a, 2006b) in order to test for distributional changes of independent observations in an off-line setting and was found to have a satisfactory performance. For earlier attempts to utilize the ECF in the context of testing with time series the reader is referred to Hong (1999), Epps (1988, 1987).

In particular, the Fourier formulation which we advocate here utilizes the Cramér–von-Mises type statistics

$$T_{\text{CF}}(j, \gamma) = \rho_{j,T}(\gamma) \int_{-\infty}^{\infty} \left|\hat{\phi}_{T,T+j}(u) - \hat{\phi}_{p,T}(u)\right|^2 w(u)\, du, \tag{1.4}$$

where $0 < \gamma \leq 1$, and

$$\hat{\phi}_{j_1,j_2}(u) = \frac{1}{j_2 - j_1} \sum_{t=j_1+1}^{j_2} e^{iu\hat{\varepsilon}_t}$$

is the ECF of the residuals. In (1.4), $\rho_{j,T}(\gamma)$ denotes a weight function needed to control (asymptotically) the probability α of type-I error for the sequential test procedure, while $w(u)$ is an extra weight function introduced to smooth out the periodic components of the ECF.

We reject the null hypothesis when for the first time $T_{\text{CF}}(j, \gamma) \geq c_\alpha$ for an appropriately chosen critical value c_α. In this case we stop monitoring. Otherwise we continue. The associated stopping rule is given by

$$\tau(T) = \begin{cases} \inf\{1 \leq j < L_T : T_{\text{CF}}(j, \gamma) \geq c_\alpha\}, \\ \infty, \quad \text{if } T_{\text{CF}}(j, \gamma) < c_\alpha \text{ for all } 1 \leq j < L_T. \end{cases}$$

We shall distinguish between open-end procedures where $L_T = \infty$ and closed-end procedures where $L_T = \lfloor NT \rfloor + 1$ for some $N > 0$. (Note that the closed-end procedures are sometimes also called curtailed or truncated.)

As in classical hypothesis testing, our aim is to control the overall value of α, i.e.,

$$\lim_{T\to\infty} P_{\mathbb{H}_0}\bigl(\tau(T) < \infty\bigr) = \alpha. \tag{1.5}$$

In this context Theorem 2.1 below shows how to choose the critical values so that (1.5) holds, i.e., so that the procedure has asymptotic size α. On the other hand, Theorems 2.2 and 2.3 show that this monitoring procedure detects a large class of alternatives with probability one asymptotically, i.e.,

$$\lim_{T\to\infty} P_{\mathbb{H}_1}\bigl(\tau(T) < \infty\bigr) = 1. \tag{1.6}$$

Thinking of the monitoring procedure in terms of classical statistics yields the following test statistic

$$\mathrm{CF}_T(\gamma) = \mathrm{CF}_T(\hat{\varepsilon}_{p+1}, \hat{\varepsilon}_{p+2}, \dots; \gamma) := \sup_{1\le j<L_T} T_{\mathrm{CF}}(j,\gamma), \tag{1.7}$$

which is only used to obtain asymptotics, whereas the actual calculation is performed sequentially as already explained above.

We have already pointed out that one can also develop related procedures based on empirical distribution functions. To this end, denote by $\hat{F}_{T,T+j}(z)$ and $\hat{F}_{p,T}(z)$ the EDF based on $\hat{\varepsilon}_{T+1}, \dots, \hat{\varepsilon}_{T+j}$ and $\hat{\varepsilon}_{p+1}, \dots, \hat{\varepsilon}_T$, respectively. Lee et al. (2009) proposed Kolmogorov–Smirnov-type statistic defined by

$$\mathrm{KS}_T^{\mathrm{Lee}}(\zeta) = \sup_{1\le j<L_T} d_{j,T}(\zeta) \sup_z \bigl|\hat{F}_{T,T+j}(z) - \hat{F}_{p,T}(z)\bigr|, \tag{1.8}$$

where $d_{j,T}(\zeta) = \sqrt{T}(T/(T+j))^{\zeta} j/(T+j)$ for some $\zeta > 0$ are weights and showed that the limit null distribution of such test statistic is an asymptotically distribution-free functional of a two-dimensional Gaussian process. The advantage is that this limit distribution, unlike the one we obtained for our procedure (cf. Theorem 2.1), does not depend on the unknown error distribution. However, and in order to obtain limit properties additional assumptions on the smoothness of the error distribution are needed. Our preliminary simulation results, not reported in this paper, were in a good agreement with the simulation study performed in Lee et al. (2009): the KS-test statistic with Lee's weights seems to reliably detect the change in the error distribution only when an early and very large change occurs in a situation with large number of observations.

Finally we note that along the line of our test procedure based on $\mathrm{CF}_T(\gamma)$ one can also develop a KS-type test statistics with $d_{j,T}(\zeta)$ in (1.8) replaced by $\rho_{j,T}^{1/2}(\gamma)$, where $0 < \gamma \le 1$ as in (1.4), i.e.,

$$\mathrm{KS}_T(\gamma) = \sup_{1\le j<L_T} \rho_{j,T}^{1/2}(\gamma) \sup_z \bigl|\hat{F}_{T,T+j}(z) - \hat{F}_{p,T}(z)\bigr|. \tag{1.9}$$

Certain comparisons based on simulations using the test statistic $\mathrm{KS}_T(\gamma)$ are reported in Sect. 4.

2 Asymptotic results

Here we present and discuss results on the asymptotic distribution of the test statistic $\mathrm{CF}_T(\gamma)$ defined in (1.7) both under the null hypothesis and under some alternatives. The latter results lead to consistency in the sense of (1.5) as well as (1.6). Note that we suppress the weight parameter γ, and write CF_T for simplicity.

Recall that we work with the sequence $\{X_j,\ j = p+1, \ldots\}$ following the model:

$$X_j = \boldsymbol{\beta}^T \boldsymbol{X}_{j-1} + \varepsilon_j, \quad j = p+1, \ldots,$$

where $\boldsymbol{X}_{j-1} = (X_{j-1}, \ldots, X_{j-p})^T$, $\boldsymbol{\beta}$ is an unknown p-regression parameter, and $\varepsilon_j,\ j = 1, \ldots$ are innovations which under the null hypothesis satisfy the following assumptions:

(A.1) $\{\varepsilon_j,\ j = 0, \pm 1, \ldots\}$ are i.i.d. random variables with common distribution function F_0 having zero mean, positive variance and $\mathsf{E}\,|\varepsilon_j|^4 < \infty$.

(A.2) The initial values $X_1, \ldots, X_p$ are independent of $\varepsilon_{p+1}, \ldots;\ \beta_p \neq 0$, and the roots of the polynomial $t^p - \beta_1 t^{p-1} - \cdots - \beta_p$ are less than one in absolute value.

(A.3) The vector $\boldsymbol{X}_{p+1} = (X_p, \ldots, X_1)^T$ of initial observations may be written as

$$\boldsymbol{X}_{p+1} = \sum_{j=0}^{\infty} \boldsymbol{B}^j \boldsymbol{e}_{p+1-j}\,, \tag{2.1}$$

where

$$\boldsymbol{B} = \begin{pmatrix} \beta_1, \ldots, \beta_p \\ \boldsymbol{I}_{p-1} \quad \boldsymbol{O} \end{pmatrix} \quad \text{and} \quad \boldsymbol{e}_k = (\varepsilon_k, 0, \ldots, 0)^T, \tag{2.2}$$

with $\boldsymbol{I}_{p-1}$ denoting the $(p-1)$-dimensional unit matrix.

It is also assumed that the estimator $\hat{\boldsymbol{\beta}}_T = \hat{\boldsymbol{\beta}}_T(X_1, \ldots, X_T)$ of $\boldsymbol{\beta}$ and the weight function $w(\cdot)$ satisfy

(A.4)

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) = O_P(1), \quad T \to \infty.$$

(A.5) $w(t),\ t \in R^1$, is a symmetric function such that

$$\int t^4 w(t)\, dt < \infty.$$

Theorem 2.1 *Let* $\{X_t\}$ *follow model* (1.1) *and let Assumptions* (A.1)–(A.5) *be satisfied. Then under the null hypothesis of no change the following asymptotic results hold for the test statistic* CF_T *in* (1.7).

(a) *For the open-end procedure* (*i.e.*, $L_T = \infty$) *and as* $T \to \infty$,

$$\mathrm{CF}_T \xrightarrow{\mathcal{D}} \sup_{0<t<1} \left| t^\gamma \left(\int w(u)\,du - \mathsf{E} \int \cos\big((\varepsilon_1 - \varepsilon_2)u\big) w(u)\,du \right) + \sum_{q=1}^{\infty} \lambda_q \frac{W_q^2(t) - t}{t^{1-\gamma}} \right|,$$

where E *denotes the expectation with respect to the innovations* $\varepsilon_1, \varepsilon_2$,

$$\rho_{j,T}(\gamma) = T \left(\frac{j}{T+j} \right)^{1+\gamma}, \quad 0 < \gamma \le 1, \tag{2.3}$$

$W_q(\cdot)$ *are independent Wiener processes and* λ_q *are square-summable eigenvalues which depend on the unknown underlying distribution function* F_0 (*refer to* (6.4)).

(b) *For the closed-end procedure* (*i.e.*, $L_T = \lfloor NT \rfloor + 1$) *and as* $T \to \infty$,

$$\mathrm{CF}_T \xrightarrow{\mathcal{D}} \sup_{0<t<N} c_w(t) \left| t(1+t) \left(\int w(u)\,du - \mathsf{E} \int \cos\big((\varepsilon_1 - \varepsilon_2)u\big) w(u)\,du \right) + \sum_{q=1}^{\infty} \lambda_q \big(W_{q,1}(t) - t W_{q,2}(1)\big)^2 - t(1+t) \right|,$$

where $\rho_{j,T}(\gamma) = \frac{j^2}{T} c_w(\frac{j}{T})$ *and* $c_w(t) \ge 0$ *is continuous on* $(0, N]$ *such that there exists* $0 \le \alpha < 1$ *with* $\lim_{t\to 0} t^\alpha c_w(t) < \infty$, *and* $W_{q,1}(\cdot)$, $W_{q,2}(\cdot)$ *are independent Wiener processes and* λ_q *are as in* (*a*).

Theorem 2.1 shows that the limit distribution depends on unknown quantities that are determined by the unknown distribution of the innovations ε_j and consequently does not provide an approximation for critical values of the CF-statistic. Therefore a bootstrap procedure suitable for the above sequential setup is useful for practical applications and will be discussed in Sect. 3 below. Also note that the conditions on the weight function for the closed-end procedure include in particular the weight functions as given for the open-end procedure, and that the limit distribution of CF_T is the same if we replace the residuals $\hat{\varepsilon}_j$ by ε_j (see Lemma 6.3).

Remark 1 The proposed procedure based on CF_T defined in (1.7) can easily be extended to different models such as regression models or ARMA-sequences among others. The key to the proofs is to be able to estimate the residuals ε_j by $\hat{\varepsilon}_j$, such that the limit distribution of the resulting test procedure does not change.

Next we have a look at the asymptotic behavior of CF_T under a class of alternatives. In particular, Assumption (A.1) is replaced by the following assumption:

(B.1) $\{\varepsilon_j,\ j = 0, \pm 1, \ldots\}$ are independent random variables with zero mean, positive variance and finite moment $\mathsf{E}|\varepsilon_j|^4 < \infty$ and having the distribution function F_0 for $j \le T + j_0$ and F^0 for $j > T + j_0$, for some $j_0 \ge 1$, $F_0 \ne F^0$.

Theorem 2.2 *Let* $\{X_t\}$ *follow model* (1.1) *and let Assumptions* (A.2)–(A.5) *and* (B.1) *be satisfied, i.e., a change of the error distribution takes place. Let also* $\rho_{j,T}(\gamma)$ *be defined as in* (2.3). *Then for the open-end procedure and as* $T \to \infty$,

$$\mathrm{CF}_T \to \infty, \quad \textit{in probability}.$$

Moreover, if $j_0 = \lfloor T t_0 \rfloor$ *with some* $t_0 \geq 0$, *then as* $T \to \infty$,

$$\mathrm{CF}_T / T \to^P \sup_{t_0 < t < \infty} \left(\frac{t}{1+t} \right)^{1+\gamma} \left(\frac{t - t_0}{t} \right)^2 \times \int \left| \varphi_0(u) - \varphi^0(u) \right|^2 w(u)\, du,$$

where $\varphi_0(t)$ *and* $\varphi^0(t)$ *are characteristic functions before and after the change, respectively. The assertion remains true for closed-end procedures if* $t_0 < N$ *and where the* sup *is taken over the set* $t_0 < t \leq N$.

The above theorem shows that our test procedure detects distributional changes of the type described by (1.2), as required. However, it will be shown below that the test procedure has also some power with respect to changes in the autoregressive coefficient. So one should also apply a test for a change in the autoregressive parameter, which, however, does not have power against distributional changes, in order to distinguish between the two (Hušková et al. 2007, 2008).

Consider

$$X_j = \boldsymbol{\beta}^T \boldsymbol{X}_{j-1} + \boldsymbol{\delta}^T \boldsymbol{X}_{j-1} I\{j > T + j_0\} + \varepsilon_j, \quad j \geq 1, \tag{2.4}$$

where $\boldsymbol{\delta} \neq \mathbf{0}$ and $j_0 \geq 1$ are both unknown, and all other symbols are as in model (1.1). As in Hušková et al. (2007) we also assume:

(B.2) The observations $X_{p+1}, \ldots$ follow the model (2.4) with $j_0 = \lfloor T t_0 \rfloor$, $t_0 \geq 0$; $X_1, \ldots, X_p$ are independent of $\varepsilon_{p+1}, \ldots, \varepsilon_T, \ldots$; $\beta_p \neq 0$, and the roots of the polynomial $t^p - \beta_1 t^{p-1} - \cdots - \beta_p$ are less than one in absolute value. In addition, $\beta_p + \delta_p \neq 0$, and the roots of $t^p - (\beta_1 + \delta_1) t^{p-1} - \cdots - (\beta_p + \delta_p)$ are also less than one in absolute value, for $\boldsymbol{\delta} \neq \mathbf{0}$ fixed.

Theorem 2.3 *Let model* (2.4) *fulfill* (A.1), (B.2), (A.3)–(A.5), $j_0 = \lfloor T t_0 \rfloor$ *for some* $t_0 \geq 0$, *i.e., a change in the regression coefficient takes place and let* $\rho_{j,T}(\gamma)$ *be defined as in* (2.3). *Then, for open-end procedures and as* $T \to \infty$,

$$\mathrm{CF}_T / T \to^P \sup_{t_0 < t < \infty} \left(\frac{t}{t+1} \right)^{1+\gamma} \left(\frac{t - t_0}{t} \right)^2 \times \int \left| \varphi_0(u) \big(\varphi_X(u) - 1 \big) \right|^2 w(u)\, du,$$

where $\varphi_X(u)$ *is the characteristic function of* $\sum_{j=1}^p \delta_j Z_{q-j}$ *with* $\{Z_q\}_q$ *being an* $AR(p)$ *process with parameters* $\boldsymbol{\beta} + \boldsymbol{\delta}$. *The assertion remains true for closed-end procedures if* $t_0 < N$ *and where the* sup *is taken over the set* $t_0 < t \leq N$.

3 Bootstrap procedures

In order to apply the tests we need critical values, and the standard approach would be to use the quantiles of the asymptotic distribution. However, from Theorem 2.1 it is clear that this is not feasible here as the limit distribution depends on too many unknown parameters. Therefore, we will apply resampling methods to approximate the null distribution.

The simplest approach is a classical bootstrap based on the estimated residuals of the training data: Let $U_T(p+1), \ldots, U_T(\tilde{L}_T)$ be i.i.d. uniform on $p+1, \ldots, T$, and independent of $\{X_t\}$, where we choose $\tilde{L}_T = L_T - 1$ in case of the closed-end procedure and $\tilde{L}_T/T \to \infty$ in case of the open-end procedure. Let

$$\varepsilon^*(t) = \hat{\varepsilon}_{U_T(t)},$$

with $\hat{\varepsilon}_j$ as in (1.3).

The bootstrap critical value $c_\alpha(X_1, \ldots, X_T)$ is chosen minimal such that

$$P_T^*\big(\mathrm{CF}_T\big(\varepsilon^*(1), \ldots, \varepsilon^*(\tilde{L}_T); \gamma\big) \le c_\alpha(X_1, \ldots, X_T)\big) \ge 1 - \alpha,$$

where $P_T^*(\cdot) = P(\cdot \mid X_1, \ldots, X_T)$. We can easily simulate the above conditional distribution by drawing B random realizations of $\{U_T(\cdot)\}$.

The above bootstrap scheme only uses the training sample $X_1, \ldots, X_T$. Therefore the following theorem holds under assumptions on the training set only, and no additional assumptions on the data after monitoring starts are needed.

Theorem 3.1 *If $X_1, \ldots, X_T$ follow model* (1.1) *fulfilling assumptions* (A.1)–(A.5), *then*

$$c_\alpha(X_1, \ldots, X_T) \xrightarrow{P} c_\alpha,$$

where c_α is the α-quantile of the asymptotic distribution in Theorem 2.1.

The above theorem shows that a test using critical values based on the bootstrap approximation has asymptotic size α and asymptotic power one under the alternatives in Theorems 2.2 and 2.3. In addition, the bootstrap test is asymptotically equivalent to the large-sample test which, however, is not feasible as c_α depends on too many unknown parameters; see Theorem 2.1.

Usually a bootstrap procedure is not optimal in case of smaller training data sets. This is not surprising as we create a data set of length L_T from a data set of length T which is much smaller than L_T. In the simpler location setting Kirch (2008) showed via simulations that this in fact leads to a loss of power. Nevertheless, adaptations of the above bootstrap schemes, including observations obtained during monitoring, are possible (cf. Kirch 2008 for the location model and Hušková and Kirch 2011 for a change in the regression coefficient).

4 Simulation study

In the previous sections we derived monitoring procedures with an asymptotic level α and asymptotic power 1 for a large class of alternatives. In this section we conduct a small simulation study to see how the test behaves for small samples. A more detailed simulation study including also some extensions and variations of the procedure is in preparation.

From (1.4) we have by straightforward calculations that the test statistic admits the following representation:

$$T_{\mathrm{CF}}(j,\gamma)=\rho_{j,T}(\gamma)\Bigg[\frac{1}{j^2}\sum_{t,s=T+1}^{T+j}h_w(\hat{\varepsilon}_{ts})+\frac{1}{(T-p)^2}\sum_{t,s=p+1}^{T}h_w(\hat{\varepsilon}_{ts})-\frac{2}{j(T-p)}\sum_{t=T+1}^{T+j}\sum_{s=p+1}^{T}h_w(\hat{\varepsilon}_{ts})\Bigg], \tag{4.1}$$

where $\rho_{j,T}(\gamma)$ is defined in (2.3), $\hat{\varepsilon}_{ts}=\hat{\varepsilon}_t-\hat{\varepsilon}_s$, and

$$h_w(x)=\int_{-\infty}^{\infty}\cos(ux)w(u)\,du. \tag{4.2}$$

The choice of the weight function $w(\cdot)$ is primarily guided by the desire to render the integral in (4.2) and the test statistic in (4.1) in closed-form. If this is accomplished then the procedure can be easily implemented since one can obtain from $T_{\mathrm{CF}}(j,\gamma)$ the next term $T_{\mathrm{CF}}(j+1,\gamma)$, by means of a simple recursion. There are several choices for $w(\cdot)$ that serve the purpose of computational simplicity, the most popular being exponential-type functions of the form $w(u)=\exp(-a|u|^b)$, for $a>0$ and $b=1$ or $b=2$. In the simulations we use $w(u)=w_a(u)=\exp(-a|u|)$, with several values of $a>0$. Also, the weight function for the sequential procedure is given by (2.3) for $\gamma=0.1$ and $\gamma=1$. We use a historic data set of $T=50$ and $T=200$ and a monitoring length NT with $N=4$. The residuals $\hat{\varepsilon}_t$ in (4.1) are obtained by using a standard least squares estimator calculated from the training sample. For the calculation of the bootstrap distribution 500 random bootstrap samples have been used.

We have already remarked in Sect. 1 that the KS-statistic (1.8) did not work well in the setup of our simulation study. Therefore, in order to provide also a comparison with some procedure based on EDF, we include simulation results for the KS-type test defined by (1.9). We have used bootstrap critical values obtained by the bootstrap scheme described in Sect. 3. We are not aware of any theoretical results concerning the application of bootstrap on the sequential test statistic (1.9) but the behavior of the bootstrap approximation in our simulation study was "very reasonable" and we have decided to include these results as an illustration comparing an EDF- and ECF-based procedure.

The training sample is always an AR(1) process (1.1) with the regression parameter $\boldsymbol{\beta}=\beta_0=0.5$ and normally distributed error terms with standard deviation $sd(\varepsilon_i)=\sigma_0=1$. The symbols $\beta^0=\beta_0+\delta$ and σ^0 denote the value of the same parameters after the change at time j_0.

We consider the following types of change:

– Change in the regression coefficient.
– Change in scale.
– Change from a normal distribution to a Student t-distribution with 4 degrees of freedom.
– Change from a normal distribution to a χ^2-distribution with 4 degrees of freedom.

Note that both the Student t-distribution and the χ^2-distribution were centered and standardized.

As usual we assess the quality of the tests by (i) the actually achieved level of the test as well as the achieved power. However, in a sequential setup it is also of interest to know (ii) how fast a change is detected by the proposed procedure.

To visualize these properties we use the following.

(i) *Achieved Size-Power Curves (ASP)* Each blue line corresponds to a specific combination of (γ, a) and to the null hypothesis, and shows on the y-axis the actual (observed) level, for a nominal size given by the x-axis. Likewise, each red line corresponds to a specific combination of (γ, a) and to one specific alternative, and shows on the y-axis the size-corrected power, i.e., the empirical power of the test corresponding to a true (observed) level α, where α is given by the x-axis. These plots are based on 1 000 repetitions of the procedure.

(ii) *Estimated density of the run length (EDR)* The run length is the point in time at which the null hypothesis is rejected. In the plots it is calculated for a true size 5% test (not a nominal one). This was done in order to obtain comparable plots for all procedures without having to take size-differences into account, which obviously have an important influence on the run length. The vertical dotted line indicates where the monitoring starts. This is a lower bound for the run length but—due to artifacts of the kernel density estimation procedure—it can happen that the estimated density is positive there. The vertical solid line indicates the position of the change. Note that the density does not integrate to one since it also attains positive mass corresponding to the samples for which the null hypothesis was not rejected.

In Fig. 1 we consider the influence of the shape of the weight function given by (2.3), with respect to the value of γ. Some typical plots are shown for an early as well as a late change. From these plots it becomes clear that $\gamma = 0.1$ detects early changes better than the procedure based on $\gamma = 1$, but at the cost of having a high probability of falsely detecting late changes before they occur. In addition we observe a power loss for late changes (and moderate monitoring horizon). This behavior is well known in sequential change-point analysis and has already been reported in different settings (cf. e.g. Horváth et al. 2004; note that a γ close to 0 in our setting corresponds to a γ close to $1/2$ in their setting, while $\gamma = 1$ in our setting corresponds to $\gamma = 0$ in their setting due to a different normalization). In Fig. 1 the plots are only given for $a = 1$, but other values of a lead to similar results.

With Figs. 2 and 3 we assess the influence of the parameter a, which determines the shape of the weight function $w_a(u)$ of the CF_T statistic. To this end we fix the value of γ at $\gamma = 1$. Also we compare the new procedure with the Kolmogorov–Smirnov-type sequential test defined by (1.9). In Fig. 2, plots include a change in the AR-parameter as well as in the scale of the error distribution. From these plots

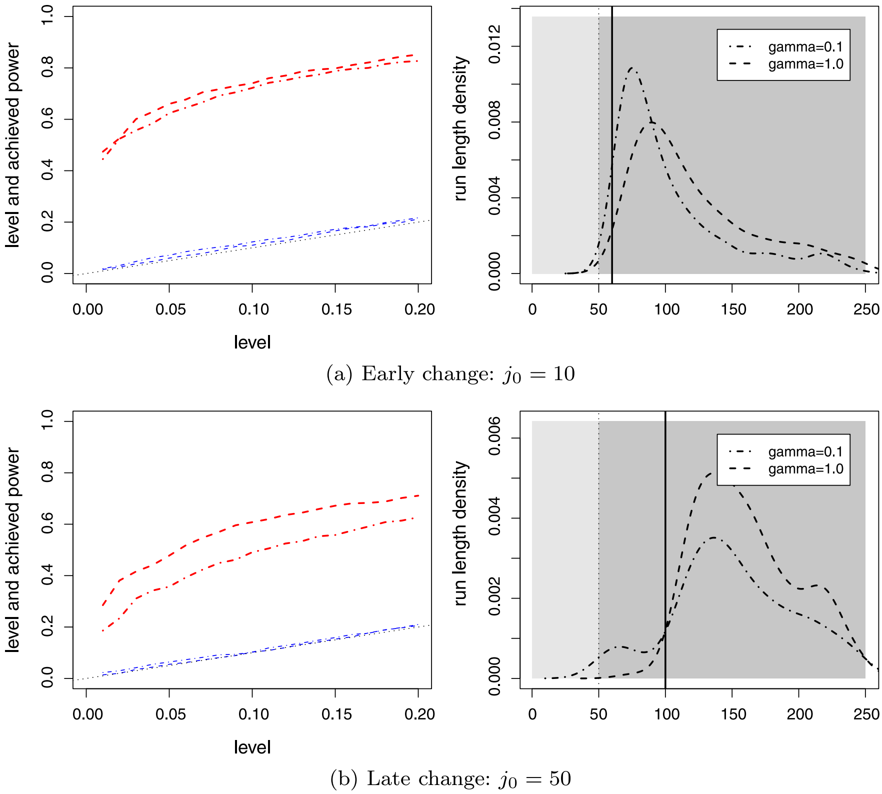
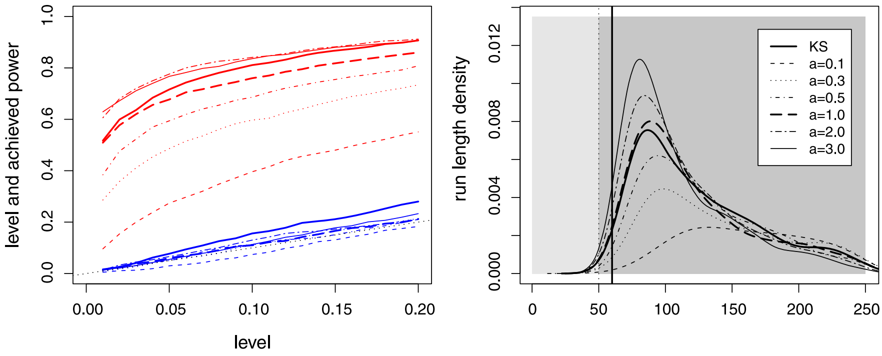


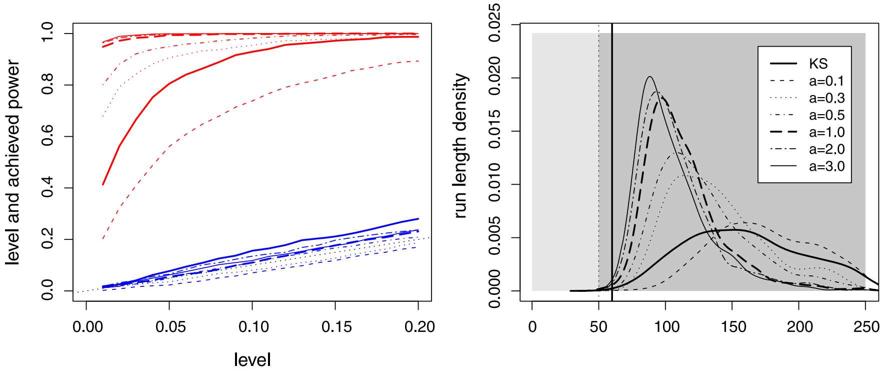
Fig. 1 ASP and EDR plots for the CF_T test with normal errors; change in the regression parameter from $\beta_0 = 0.5$ to $\beta^0 = 0.9$ ($\gamma = 0.1$ and $\gamma = 1.0$, $\sigma_0 = \sigma^0 = 1$, $a = 1$, $T = 50$, $N = 4$)

it becomes clear that intermediate values of a hold the level best. However, larger values lead to more powerful procedures both in terms of overall detection rate as well as detection delay while still having a reasonable size. They also yield better results than the Kolmogorov–Smirnov-type test especially in the situation of a change in scale parameter.

The plots in Fig. 3 include changes in the distribution of the errors from a normal distribution to a t_4-distribution, as well as from a normal distribution to a χ^2-distribution with 4 degrees of freedom. Corresponding results for the Kolmogorov–Smirnov-type test are also included. According to Fig. 3(a), i.e., for $T = 50$ and a change from normal to t_4, both procedures have a very low power for small samples (but are unbiased). Apparently, the difference between the two distributions is not large enough to be detectable at a satisfying power with a historic sample of size only $T = 50$ at hand. However, this is not due to the change-point setting or the sequential nature of the test, since even in a simple (non-sequential) two-sample situation both tests have a very low power in distinguishing these two distributions; for a nice review of two-sample tests, including the Kolmogorov–Smirnov test, the reader is referred to Dufour and Farhat (2002). This statement is easily verified by a small simulation study. For example, with two equally sized samples of length 50 we obtain an empiri-



(a) Change in the AR-parameter: $\beta^0 = 0.9$



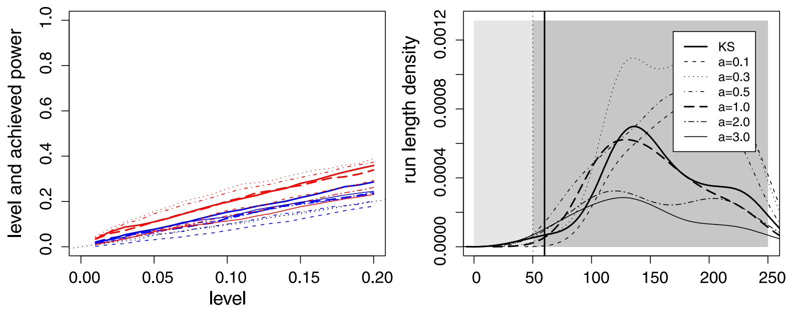
(b) Change in the scale parameter: $\sigma^0 = 2$

Fig. 2 Dependency on a, fixed $\gamma = 1$, $j_0 = 10$, $T = 50$, $N = 4$, $\beta_0 = 0.5$, $\sigma_0 = 1$, normal errors: ASP—as well as EDR-Plots

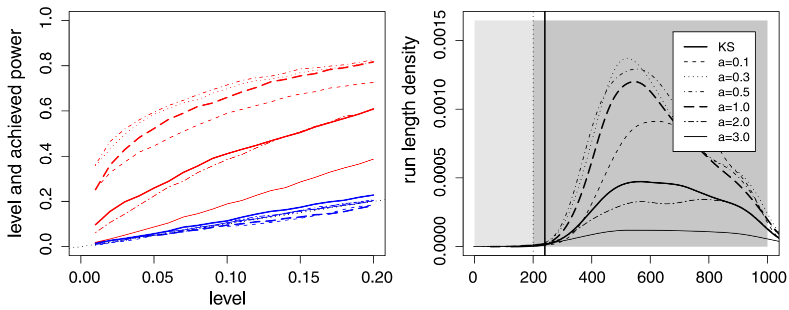
cal power (calculated from 500 simulations with 1000 bootstrap replicates) of 0.1 for the CF_T-test with $a = 1$, and a corresponding power of 0.072 for the KS-test, each at nominal level 5%.

The size results depicted in Fig. 3 are reasonable for all values of a, but best for intermediate values of this parameter. Naturally, for the larger historic data set $T = 200$ (Fig. 3(b)), the power increases and it becomes clear that in this situation small values of a yield best results in terms of power and detection delay. Furthermore, our procedure clearly outperforms the Kolmogorov–Smirnov-type test. In Figs. 3(c) and (d) analogous pictures for a change from a normal to a χ^2-distribution with 4 degrees of freedom can be found. The power is somewhat higher than for the change to a t_4-distribution, but the general conclusions remain the same.

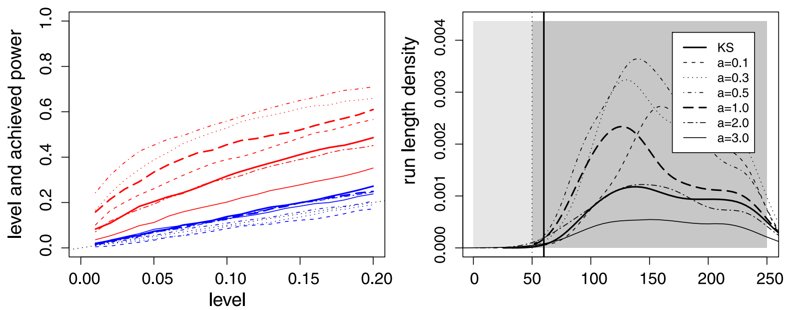
As far as the choice of the weight parameter a is concerned it is clear from the simulation results that the value of a has some effect on the power properties of the proposed test. In order to motivate the discussion we shall remain within the context of the weight function $w(t) = e^{-a|t|}$, but similar considerations hold for other weight functions, and the issue of the proper choice of the weight function resembles that of choosing the kernel in non-parametric density estimation: The choice of the



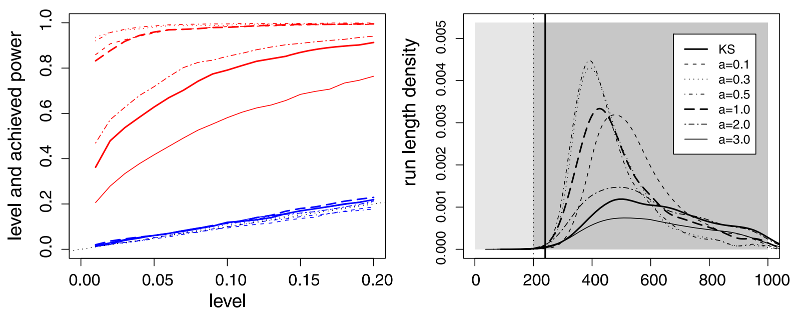
(a) Change from normal to t_4: $T = 50$



(b) Change from normal to t_4: $T = 200$



(c) Change from normal to χ^2_4: $T = 50$



(d) Change from normal to χ^2_4: $T = 200$

Fig. 3 Dependency on a, fixed $\gamma = 1$, $j_0 = 1/5T$, $N = 4$, $\beta_0 = \beta^0 = 0.5$, $\sigma_0 = \sigma^0 = 1$: ASP—as well as EDR-Plots kernel function (resp. weight function) matters less than the choice of the value of the bandwidth (resp. of the weight parameter a). In this connection note that choosing a large value of a causes the weight function to decay rapidly, and consequently the test statistic is dominated, at least asymptotically, by the behavior of the CF around zero. Since, however, the tail behavior of a distribution is reflected on the behavior of its CF around zero, putting most of the weight around $t = 0$ should render the test statistic powerful against distributions with markedly different tail characteristics. On the other hand, by 'moving' to a smaller value of a we also put weight in differences in the middle sections of a pair of distributions (which is expected to compensate for some loss of diagnostic power against distributions with great differences in the tails). Of course these comments are of qualitative nature and in order to be more specific one should have in mind concrete alternative directions away from a specific null hypothesis, which is clearly not possible in our nonparametric context. But even within a fully parametric setting the problem of choosing a 'good' value for a is far from trivial and the reader is referred to Epps (1999) and Tenreiro (2009) for a power analysis in the context of testing for normality against specific alternatives. Otherwise, proper values of a should be determined empirically via Monte Carlo simulation of the behavior of the test under a wide variety of alternatives. As an example and based on our simulation results we could suggest a value in the vicinity of $a = 1$ as a compromise choice with good overall power properties for the test statistic.

5 Conclusion

We propose goodness-of-fit procedures for the error distribution of AR models in the sequential set-up. The new tests utilize an L^2-type discrepancy measure between a couple of empirical characteristic functions (ECFs) of the residuals; the first ECF includes the residuals computed from a training data set with no change in the distribution of random errors, while the other ECF is based on the residuals after this training data set has been observed. Asymptotic results are provided both under the null hypothesis of no change, as well as under alternative hypotheses. The latter results imply the consistency of the proposed test in the case of a change in the error distribution, but also in the case of a change in the parameter of the underlying AR-model. Additionally, a bootstrap procedure is proposed which is straightforward to apply, thereby circumventing the drawback that the asymptotic null distribution of the test statistic is parametric in nature. A simulation study supports our asymptotic results, by reporting empirical level close to the nominal size even for small samples, and percentage of rejection under alternatives which suggests that the new test is able to detect distributional changes as well as changes in the autoregression parameter. Extra simulation results include favorable comparisons with a Kolmogorov–Smirnov-type test.

6 Proofs

We start with the proofs of some auxiliary lemmas. D will denote a positive generic constant.

Lemma 6.1 *Let Assumptions* (A.2)–(A.3) *and either* (A.1) (*null hypothesis*) *or* (B.1) (*change in distribution*) *or* (A.1) *and* (B.2) (*change in regression coefficient*) *be satisfied. Then for an arbitrary* $\kappa > 0$ *there exists* $A > 0$ *such that for* $\eta > 3/2$

$$\text{(a)}\quad P\left(\max_{1\le k\le Q} k^{-\eta}\left\|\sum_{j=T+1}^{T+k} \boldsymbol{X}_{q,j-1}\right\|^2 \ge A\right) \le \kappa,$$

$$\text{(b)}\quad P\left(\max_{1\le k\le Q} k^{-\eta}\left\|\sum_{j=T+1}^{T+k} \left(\boldsymbol{X}_{q,j-1}\boldsymbol{X}_{q,j-1}^T - E\left(\boldsymbol{X}_{q,j-1}\boldsymbol{X}_{q,j-1}^T\right)\right)\right\|^2 \ge A\right) \le \kappa,$$

$$\text{(c)}\quad \max_{k\ge\sqrt{T}} \int \left|\frac{1}{k}\sum_{j=T+1}^{T+k}\left(g(t\epsilon_j) - Eg(t\epsilon_j)\right)\right|^2 w(t)\,dt = o_P(1),$$

$$\text{(d)}\quad \max_{k\ge\sqrt{T}} \int \left|\frac{1}{k}\sum_{j=T+k^\circ+1}^{T+k}\left(g\left(t\boldsymbol{\delta}^T\boldsymbol{X}_{q,j}\right) - Eg\left(t\boldsymbol{\delta}^T\boldsymbol{X}_{q,j}\right)\right)\right|^2 w(t)\,dt = o_P(1),$$

where $\boldsymbol{X}_{q,j} = (X_j,\ldots,X_{j-q})^T$, *for any fixed* q*-dimensional vector* $\boldsymbol{\delta}$, *any bounded function* g *with bounded first derivative*, *and any positive integer* Q.

Proof We start with the proof if either (A.1) or (B.2) holds. Assertion (b) is given in Lemma 4.2 in Hušková et al. (2007) for $q = p$, the assertion for $q \ne p$ and (a) follows analogously. The key are the following moment bounds given in Corollary 4.1 in Hušková et al. (2007) in addition to some Hájek–Rényi type inequalities. For some $\rho \in (0, 1)$ it holds

$$\begin{aligned} &\mathrm{E}|X_{j-v}X_{j-s}| \le D\rho^{|v-s|}, \quad 1 \le v,\ s \le j,\ j \ge p,\\ &\left|\operatorname{cov}\left(X_{j-v}X_{j-s}, X_{j+h-v}X_{j+h-s}^T\right)\right| \le D\rho^{2|h|},\\ &\quad h \ge 0,\ 1 \le v,\ s \le p,\ j \ge p. \end{aligned} \tag{6.1}$$

To prove (c) first note that due to the boundedness of g it holds for any $K > 0$ that

$$\begin{aligned} &\int \left|\frac{1}{k}\sum_{j=T+1}^{T+k}\left(g(t\epsilon_j) - \mathrm{E}g(t\epsilon_j)\right)\right|^2 w(t)\,dt\\ &\quad\le \int w(t)\,dt \sup_{|t|\le K}\left|\frac{1}{k}\sum_{j=T+1}^{T+k}\left(g(t\epsilon_j) - \mathrm{E}g(t\epsilon_j)\right)\right|^2\\ &\qquad + \sup_{x\in\mathbb{R}} g^2(x)\int_{|t|>K} w(t)\,dt, \end{aligned}$$

where the second summand becomes arbitrarily small for K large enough. Concerning the first summand we apply a uniform law of large numbers for stationary and ergodic sequences by Ranga Rao (1962). Here, the sequence is even i.i.d. up to $T + k_0$

and starting at $T+k_0$ and the condition in Ranga Rao (1962) $\mathrm{E}\sup_{|t|\le K}|g(t\epsilon_0)|<\infty$ is fulfilled due to boundedness of g. We get

$$\sup_{k\ge\sqrt{T}}\sup_{|t|\le K}\left|\frac{1}{k}\sum_{j=T+1}^{T+k}\big(g(t\epsilon_j)-\mathrm{E}g(t\epsilon_j)\big)\right|^2$$

$$\stackrel{\mathcal{D}}{=}\sup_{k\ge\sqrt{T}}\sup_{|t|\le K}\left|\frac{1}{k}\sum_{j=1}^{k}\big(g(t\epsilon_j)-\mathrm{E}g(t\epsilon_j)\big)\right|^2\to 0 \quad \text{a.s.}$$

for $T\to\infty$. This yields the assertion.

For the proof of (d) first consider $\widetilde{X}_j=(\boldsymbol{\beta}^T+\boldsymbol{\delta}^T)\widetilde{\mathbf{X}}_{j-1}+\varepsilon_j,\ j\ge T+k_0$. Further assume that $\widetilde{\mathbf{X}}_{T+k_0-1}$ fulfills (2.1) with β_l replaced by $\beta_l+\delta_l$ and all ε_l following F^0. By (B.2) $\widetilde{\mathbf{X}}$ is stationary and ergodic and the same arguments as in (c) lead to the assertion with X_j replaced by $\widetilde{X}_j$. Similar arguments as in Sect. 4 in Hušková et al. (2007) yield

$$|X_j-\widetilde{X}_j|\le C\rho^{j-T-k_0}\|\mathbf{X}_{T+k_0-1}-\widetilde{\mathbf{X}}_{T+k_0-1}\| \tag{6.2}$$

for some $C>0,\ 0<\rho<1$. Since $\|\mathbf{X}_{T+k_0-1}-\widetilde{\mathbf{X}}_{T+k_0-1}\|=O_P(1)$ uniformly in T,k_0 we get by the mean-value theorem and the fact that the first derivative of g is bounded

$$\sup_{k\ge\sqrt{T}}\sup_{|t|\le K}\left|\frac{1}{k}\sum_{j=T+k^\circ+1}^{T+k}\big(g\big(t\boldsymbol{\delta}^T\boldsymbol{X}_{q,j}\big)-\mathrm{E}g\big(t\boldsymbol{\delta}^T\boldsymbol{X}_{q,j}\big)\big)\right|^2$$

$$=o_P(1)+O_P(1)\|\delta\|K\sup_{k\ge\sqrt{T}}\frac{1}{k}\sum_{j=T+k_0}^{T+k}\rho^{j-T-k_0}=o_P(1).$$

□

Lemma 6.2 *Under the assumptions of Theorem 2.1, respectively, of Theorem 2.2 it holds for the open-end as well as closed-end procedure that*

$$\sup_{1\le j<L_T}\big|T_{\mathrm{CF}}(j,\gamma)-T_{\mathrm{CF}}(j,\gamma;\varepsilon_1,\varepsilon_2\ldots)\big|=o_P(1),$$

where $T_{\mathrm{CF}}(j,\gamma;\varepsilon_1,\varepsilon_2\ldots)$ denotes the test statistic (1.4) with $\hat{\varepsilon}_i$ replaced by ε_i.

Proof We will study the differences $\hat{C}_k(t)-C_k(t)$ and $\hat{S}_k(t)-S_k(t)$ where

$$C_k(t)=\frac{1}{k}\sum_{j=T+1}^{T+k}\cos(t\varepsilon_j)-\frac{1}{T}\sum_{j=1+p}^{T}\cos(t\varepsilon_j),$$

$S_k(t)$ is defined analogously with $\cos(\cdot)$ replaced by $\sin(\cdot)$, and $\hat{C}_k(t)$, $\hat{S}_k(t)$ with ε_j replaced by $\hat{\varepsilon}_j$. By the Taylor expansion, $\cos(t\hat{\varepsilon}_j)=\cos(t\varepsilon_j)-t(\hat{\varepsilon}_j-\varepsilon_j)\sin(t\varepsilon_j)+R_{jC}(t)$, where $R_{jC}(t)$ is a remainder term. Then $\hat{C}_k(t)$ can be decomposed:

$$\hat{C}_k(t)-C_k(t)=\hat{C}_{k1}(t)+\hat{C}_{k2}(t)$$

with

$$\hat{C}_{k1}(t) = -\left(\frac{1}{k}\sum_{j=T+1}^{T+k} t(\hat{\varepsilon}_j - \varepsilon_j)\sin(t\varepsilon_j) - \frac{1}{T}\sum_{j=1+p}^{T} t(\hat{\varepsilon}_j - \varepsilon_j)\sin(t\varepsilon_j)\right),$$

$$\hat{C}_{k2}(t) = \frac{1}{k}\sum_{j=T+1}^{T+k} R_{jC}(t) - \frac{1}{T}\sum_{j=1+p}^{T} R_{jC}(t).$$

Since $|R_{jC}(t)| \le Dt^2(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_T)^T \boldsymbol{X}_{j-1}\boldsymbol{X}_{j-1}^T(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_T)$ for some positive $D > 0$ we also have

$$\left|\hat{C}_{k2}(t)\right|^2 \le Dt^4\left((\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_T)^T\left(\frac{1}{k}\sum_{j=T+1}^{T+k}\boldsymbol{X}_{j-1}\boldsymbol{X}_{j-1}^T + \frac{1}{T}\sum_{j=p+1}^{T}\boldsymbol{X}_{j-1}\boldsymbol{X}_{j-1}^T\right)\right.$$
$$\left.\times(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_T)\right)^2.$$

Recall that by the Cauchy–Schwarz inequality $x^T Ax \le \|x\|^2\,\|A\|_F \le Dx^T Ax \le \|x\|^2\,\|A\|$, where $\|\cdot\|_F$ denotes the Frobenius norm and $\|\cdot\|$ the Euclidean norm. Hence we get by (A.4) and Lemma 6.1 that

$$\left|\hat{C}_{k2}(t)\right|^2 \le Dt^4\|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_T\|^4\left(\left\|\frac{1}{k}\sum_{j=T+1}^{T+k}\boldsymbol{X}_{j-1}\boldsymbol{X}_{j-1}^T\right\|^2 + \left\|\frac{1}{T}\sum_{j=p+1}^{T}\boldsymbol{X}_{j-1}\boldsymbol{X}_{j-1}^T\right\|^2\right)$$
$$= O_P(1)\frac{t^4}{T^2} \tag{6.3}$$

uniformly in k. Hence by (A.5)

$$\max_{1\le k<\infty} T\left(\frac{k}{T+k}\right)^{1+\gamma}\int\left|\hat{C}_{k2}(t)\right|^2 w(t)\,dt = O_P(1)\frac{1}{T}\int t^4 w(t)\,dt = o_P(1).$$

Next we show the negligibility of $\hat{C}_{k1}(t)$. We use the decomposition

$$\hat{C}_{k1}(t) = (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_T)^T\left(\frac{t}{k}A_{k1}(t) - \frac{t}{T}B_1(t) + \frac{t}{k}A_{k2}(t) - \frac{t}{T}B_2(t)\right),$$

where

$$A_{k1}(t) = \sum_{j=T+1}^{T+k}\boldsymbol{X}_{j-1}\big(-\sin(t\varepsilon_j) + \mathrm{E}\sin(t\varepsilon_j)\big),$$

$$A_{k2}(t) = -\mathrm{E}\big(\sin(t\varepsilon_1)\big)\sum_{j=T+1}^{T+k}\boldsymbol{X}_{j-1}$$

and $B_1(t)$, $B_2(t)$ are defined analogously with the sum taken from $p+1$ to T.

By (A.4) one obtains

$$\left|\hat{C}_{k1}(t)\right|^2 = O_P(1)\frac{1}{T}\left\|\left(\frac{t}{k}A_{k1}(t) - \frac{t}{T}B_1(t) + \frac{t}{k}A_{k2}(t) - \frac{t}{T}B_2(t)\right)\right\|^2$$

uniformly in k and t. $\{A_{k1}(t), \sigma(X_1,\dots,X_{T+k}), k \geq 1\}$ is a martingale for each $t \in R^1$ with $\mathrm{E}A_{k1}(t) = 0$, var $A_{k1}(t) \leq Dk$. Consequently

$$\left\{\int t^2\left\|A_{k1}(t)\right\|^2 w(t)\,dt,\ \sigma(X_1,\dots,X_{T+k}),\ k \geq 1\right\}$$

is a nonnegative submartingale with

$$\mathrm{E}\int t^2\left\|A_{k1}(t)\right\|^2 w(t)\,dt \leq Dk.$$

By Chows inequality (cf. e.g. Chow and Teicher 1997, Sect. 7.4, Theorem 8) it holds for a nonnegative submartingale S_k, $v_l \geq v_{l+1} \geq 0$ and $\lambda > 0$

$$\lambda P\Big(\max_{1\leq l\leq n} v_l S_l > \lambda\Big) \leq \sum_{l=2}^{n} v_l \mathrm{E}(S_l - S_{l-1}) + v_1 \mathrm{E}S_1$$
$$= \sum_{l=1}^{n-1}(v_l - v_{l+1})\mathrm{E}S_l + v_n \mathrm{E}S_n.$$

If $v_n \mathrm{E}S_n \to 0$ as $n \to \infty$ we get

$$\lambda P\Big(\max_{1\leq l\leq n} v_l S_l > \lambda\Big) \leq \sum_{l\geq 1}(v_l - v_{l+1})\mathrm{E}S_l.$$

From this we can conclude with $v_l = T^{-1-\gamma}l^{\gamma-1}$ for $l = 1,\dots,T$ and $v_l = l^{-2}$ for $l > T$ ($D > 0$ is a generic constant which may change from line to line)

$$\lambda P\left(\max_{1\leq k<\infty}\left(\frac{k}{k+T}\right)^{1+\gamma}\frac{1}{k^2}\int t^2\left\|A_{k1}(t)\right\|^2 w(t)\,dt > \lambda\right)$$
$$\leq \lambda P\left(\max_{1\leq k<\infty} v_l \int t^2\left\|A_{k1}(t)\right\|^2 w(t)\,dt > \lambda\right)$$
$$\leq DT^{-1-\gamma}\sum_{k=1}^{T}k^{\gamma-1} + \sum_{k\geq T}k^{-2} \leq DT^{-1} = o(1),$$

where the last line follows from the mean-value theorem.

By Lemma 6.1

$$\max_{1\leq k<\infty}\left\|\sum_{j=T+1}^{T+k}\boldsymbol{X}_{j-1}\right\|^2\frac{1}{k^2}\left(\frac{k}{T+k}\right)^{1+\gamma} = o_P(1),$$

which implies

$$\max_{1\le k<\infty}\left(\frac{k}{k+T}\right)^{1+\gamma}\frac{1}{k^2}\int t^2\left\|A_{k2}(t)\right\|^2 w(t)\,dt = o_P(1).$$

Similar relations hold true also for $B_1(t)$ and $B_2(t)$. It is in fact an easier situation since the random part does not depend on k.

Combining the above arguments we obtain

$$\max_{1\le k<\infty} T\left(\frac{k}{T+k}\right)^{1+\gamma}\int\left\|\hat{C}_{k1}(t)\right\|^2 w(t)\,dt = o_P(1).$$

Analogous argument for $\hat{S}_k(t)-S_k(t)$ complete the proof for the open-end procedure, the result for the closed-end procedure is obtained analogously. □

Lemma 6.3 *The assertions of Theorem 2.1 hold under the same assumptions if one replaces* $\mathrm{CF}_T(\hat{\varepsilon}_{p+1},\hat{\varepsilon}_{p+2},\ldots;\gamma)$ *by* $\mathrm{CF}_T(\varepsilon_{p+1},\varepsilon_{p+2},\ldots;\gamma)$.

Proof The proof is very close to the proof of Theorem A in Hušková and Meintanis (2006a). Therefore we only give a sketch here. Let $h(x,y)=\int\cos(u(x-y))w(u)\,du$ and $\widetilde{h}(x,y)=h(x,y)-\mathrm{E}h(x,\varepsilon_1)-\mathrm{E}h(\varepsilon_2,y)+\mathrm{E}h(\varepsilon_1,\varepsilon_2)$. Analogous to the decomposition (18) in Hušková and Meintanis (2006a) we get

$$\int_{-\infty}^{\infty}\left|\widetilde{\phi}_{T,T+k}(u)-\widetilde{\phi}_{p,T}(u)\right|^2 w(u)\,du = A_{k1}+A_{k2}+A_{k3},$$

where

$$A_{k1}=\frac{T+k}{kT}\left(\int w(u)\,du-\mathrm{E}h(\varepsilon_1,\varepsilon_2)\right),$$

$$A_{k2}=\frac{1}{k^2}\sum_{v_1=T+1}^{T+k}\sum_{\substack{v_2=T+1\\ v_2\neq v_1}}^{T+k}\widetilde{h}(\varepsilon_{v_1},\varepsilon_{v_2})+\frac{1}{T^2}\sum_{v_1=1}^{T}\sum_{\substack{v_2=1\\ v_2\neq v_1}}^{T}\widetilde{h}(\varepsilon_{v_1},\varepsilon_{v_2})$$

$$-\frac{2}{Tk}\sum_{v_1=1}^{T}\sum_{v_2=T+1}^{T+k}\widetilde{h}(\varepsilon_{v_1},\varepsilon_{v_2}),$$

$$A_{k3}=-\frac{2}{k^2}\sum_{v=T+1}^{T+k}\left(\mathrm{E}\big(h(\varepsilon_v,\varepsilon_0)\,|\,\varepsilon_v\big)-\mathrm{E}h(\varepsilon_1,\varepsilon_2)\right)$$

$$-\frac{2}{T^2}\sum_{v=1}^{T}\left(\mathrm{E}\big(h(\varepsilon_v,\varepsilon_0)\,|\,\varepsilon_v\big)-\mathrm{E}h(\varepsilon_1,\varepsilon_2)\right),$$

where $\varepsilon_0,\varepsilon_1,\ldots$ i.i.d. Note that $Z_v=\mathrm{E}(h(\varepsilon_v,\varepsilon_0)\,|\,\varepsilon_v)-\mathrm{E}h(\varepsilon_1,\varepsilon_2)$ are i.i.d. with zero mean and finite variance, hence an application of the Hájek–Rényi inequality e.g. for the first term of A_{k3} and the open-end procedure as $T \to \infty$ for arbitrary $\eta > 0$ and $0 < \gamma \le 1$ yields

$$P\left(\max_{k\ge 1} \frac{T}{T+k}\left(\frac{k}{k+T}\right)^{\gamma} \frac{1}{k}\left|\sum_{j=T+1}^{T+k} Z_j\right| \ge \eta\right)$$
$$\le C\eta^{-2}\left(\frac{1}{T^{2\gamma}}\sum_{k=1}^{T}\frac{1}{k^{2-2\gamma}} + \sum_{k\ge T}\frac{1}{k^2}\right) \to 0$$

for some $C > 0$. A similar expression holds for the second term of A_{k3} and in case of the closed-end procedure. This shows that A_{k3} is asymptotically negligible.

We investigate A_{k2} now. Analogously to Hušková and Meintanis (2006a) there exist orthonormal functions $g_j(\cdot)$ and eigenvalues λ_j such that

$$\begin{aligned}
&\widetilde{h}(x,y) \overset{L^2_\varepsilon}{=} \sum_{j=1}^{\infty} \lambda_j g_j(x) g_j(y), \quad \text{i.e.,} \\
&\lim_{L\to\infty} \mathrm{E}\left(\widetilde{h}(\varepsilon_1,\varepsilon_2) - \sum_{j=1}^{L}\lambda_j g_j(\varepsilon_1) g_j(\varepsilon_2)\right)^2 = 0, \\
&\mathrm{E}g_j(\varepsilon_1) = 0, \qquad \mathrm{E}g_j^2(\varepsilon_1) = 1, \qquad \mathrm{E}g_j(\varepsilon_1)g_k(\varepsilon_2) = 0, \quad j \ne k, \\
&\mathrm{E}\widetilde{h}^2(\varepsilon_1,\varepsilon_2) = \sum_{j=1}^{\infty}\lambda_j^2 < \infty.
\end{aligned} \tag{6.4}$$

Let $\widetilde{h}_L(x,y) = \sum_{s=1}^{L} \lambda_s g_s(x) g_s(y)$ and $A_{k2}(L)$ defined as A_{k2} with $\widetilde{h}$ replaced by $\widetilde{h}_L$. As in Hušková and Meintanis (2006a)

$$S_k := \sum_{1\le i<j\le k} \left(\widetilde{h}(\varepsilon_i,\varepsilon_j) - \widetilde{h}_L(\varepsilon_i,\varepsilon_j)\right)$$

is a martingale and therefore we get by the Hájek–Rényi inequality for the open-end procedure and the first term of $A_{k2} - A_{k2}(L)$ for all $\eta > 0$ and some $C > 0$

$$P\left(\max_{k\ge 1}\frac{T}{T+k}\left(\frac{k}{k+T}\right)^{\gamma}\frac{1}{k}|S_k| \ge \eta\right)$$
$$\le \frac{1}{\eta^2}\sum_{k\ge 1}\frac{1}{k^2}\mathrm{E}\left(\widetilde{h}(\varepsilon_1,\varepsilon_2) - \widetilde{h}_L(\varepsilon_1,\varepsilon_2)\right)^2 \le \frac{C}{\eta^2}\sum_{j=L+1}^{\infty}\lambda_j^2 \overset{L\to\infty}{\longrightarrow} 0$$

and a similar expression for the closed-end procedure and the other terms. Hence for any $\eta_1, \eta_2 > 0$ and all $L \ge L_0$ (for some L_0) it holds

$$P\left(\max_{k\ge T}\rho_{k,T}(\gamma)\left|A_{k2} - A_{k2}(L)\right| \ge \eta_1\right) \le \eta_2.$$

It holds

$$A_{k2}(L) = \sum_{q=1}^{L} \lambda_q \left(\frac{T}{k^2} B_1^2(q,k) - \frac{T+k}{Tk} - B_2(q,k) \right),$$

$$B_1(q,k) = \frac{1}{\sqrt{T}} \sum_{v=T+1}^{T+k} \left(g_q(\varepsilon_v) - \frac{1}{T} \sum_{j=1}^{T} g_q(\varepsilon_j) \right),$$

$$B_2(q,k) = \frac{1}{k^2} \sum_{v=T+1}^{T+k} \left(g_q^2(\varepsilon_v) - 1 \right) + \frac{1}{T^2} \sum_{v=1}^{T} \left(g_q^2(\varepsilon_v) - 1 \right).$$

By the strong law of large numbers we get (as $\max_{k \le \log T} \frac{\rho_{k,T}}{k} \to 0$)

$$\max_k \rho_{k,T}(\gamma) B_2(q,k) \to 0 \quad \text{a.s.}$$

Concerning $B_1(q,k)$ we first note that it is sufficient also in case of the open-end procedure to consider the supremum up to DT for some D large enough.

To this end, note that by the Hájek–Rényi inequality for any $\eta > 0$ and some $C > 0$

$$P\left(\max_{k \ge DT} \left(\frac{k}{k+T} \right)^{1/2+\gamma/2} \frac{T^{1/2}}{k} \left| \sum_{v=T+1}^{T+k} g_q(e_v) \right| \ge \eta \right)$$
$$\le \frac{C}{\eta^2} T \sum_{k \ge DT} \frac{1}{k^2} + \frac{C}{D\eta^2} \le \frac{2C}{\eta^2} \frac{1}{D},$$

which becomes arbitrarily small for D large enough, hence it is sufficient to consider the maximum up to DT even in case of the open-end procedure.

Similarly, using again the Hájek–Rényi inequality one can see that the maximum over $k \le \delta T$ is negligible for δ small enough (in case of the open-end procedure) ($\gamma \ne 0$):

$$P\left(\max_{k \le \delta T} \sqrt{\frac{T}{k^2} \left(\frac{k}{T+k} \right)^{1+\gamma}} \left| \sum_{v=T+1}^{T+k} g_q(e_v) \right| \ge \eta \right)$$
$$\le \frac{C}{\eta^2} T \sum_{k \le \delta T} \frac{1}{k^{1-\gamma}(T+k)^{1+\gamma}} \le \frac{C}{-\eta^2} T^{-\gamma} \sum_{k \le \delta T} \frac{1}{k^{1-\gamma}} \le \frac{C}{\eta^2} \delta^{\gamma},$$

which becomes arbitrarily small for $\delta \to 0$. The result for the closed-end procedure is obtained similarly. Now, we can use the functional limit theorem and get (noting that $\{\frac{1}{\sqrt{T}} \sum_{v=T+1}^{T+k} g_q(e_v) : k\}$ and $\frac{1}{T} \sum_{j=1}^{T} g_q(e_j)$ are independent)

$$\max_{\delta T \le k < DT} \rho_{k,T}(\gamma) \sum_{q=1}^{L} \lambda_q \left(\frac{T}{k^2} B_1^2(q,k) - \frac{T+k}{Tk} \right)$$

$$= \max_{\delta T \le k < DT} \rho_{k,T}(\gamma) \sum_{q=1}^{L} \lambda_q \left(\frac{T}{k^2} \left(W_{q,1}\left(\frac{k}{T}\right) - \frac{k}{T} W_{q,2}(1) \right)^2 - \frac{T+k}{Tk} \right) + o_P(1),$$

where $W_{q,1}(\cdot)$, $W_{q,2}(\cdot)$, $q = 1, \ldots, L$ are independent Wiener processes. For the open-end procedure we still need to note that

$$\frac{T^2}{k^2} \left(\frac{k}{T+k} \right)^{1+\gamma} = \left(\frac{T}{T+k} \right)^2 \left(\frac{T+k}{k} \right)^{1-\gamma}.$$

Similar arguments as before yield that this has the same asymptotic distribution as the complete maximum over $1 \le k < \infty$ for the open-end procedure, respectively, $1 \le k \le NT$ for the closed-end procedure. In the proof of Theorem 2.1. Horváth et al. (2004) show that ($t \approx k/T$)

$$\sup_{k \ge 1} \frac{|W_{q,1}(\frac{k}{T}) - \frac{k}{T} W_{q,2}(1)|}{\frac{T+k}{T} (\frac{k}{T+k})^{(1-\gamma)/2}} \xrightarrow{\mathcal{D}} \sup_{t \ge 0} \frac{|W_{q,1}(t) - t W_{q,2}(1)|}{(1+t)(\frac{t}{1+t})^{(1-\gamma)/2}}.$$

After an index transformation $l = t/(1+t)$ they further obtain

$$\max_{t \ge 0} \frac{|W_{q,1}(t) - t W_{q,2}(1)|}{(1+t)(\frac{t}{1+t})^{(1-\gamma)/2}} \stackrel{D}{=} \sup_{0 \le l \le 1} \frac{|W_q(l)|}{l^{(1-\gamma)/2}},$$

where $W_q(\cdot)$ are independent Wiener processes.

Similar arguments as above yield that it is asymptotically equivalent to consider the complete sum over $q \ge 1$.

Taking the negligibility of A_{k3} into account analogous arguments as above give the limit behavior as given in Theorem 2.1 for the joint supremum $\sup_{k \ge 1} |A_{k1} + A_{k2} + A_{k3}|$, thus completing the proof. □

Proof of Theorem 2.1 It follows immediately from Lemmas 6.2 and 6.3. □

Proof of Theorem 2.2 By Lemma 6.2 it is sufficient to consider the statistic CF_T, where $\hat{\varepsilon}_t$ are replaced by ε_t. By Lemma 6.1 ($g = \cos, \sin$) it follows

$$\max_{1 \le k < \infty} \left(\frac{k}{T+k} \right)^{1+\gamma} \int \left| \frac{1}{k} \sum_{j=T+1}^{T+k} \big(\exp(it\varepsilon_j) - \mathrm{E}\big(\exp(it\varepsilon_j)\big)\big) \right|^2 w(t)\, dt = o_P(1),$$

$$\max_{1 \le k < \infty} \left(\frac{k}{T+k} \right)^{1+\gamma} \int \left| \frac{1}{T} \sum_{j=1}^{T} \big(\exp(it\varepsilon_j) - \mathrm{E}\big(\exp(it\varepsilon_j)\big)\big) \right|^2 w(t)\, dt = o_P(1).$$

From this we can conclude immediately that $\mathrm{CF}_T = O_P(1/T)$ as well as the limit distribution of CF_T/T in case of $k_0 = \lfloor t_0 T \rfloor$. □

The following lemma is needed to prove Theorem 2.3.

Lemma 6.4 *Under the assumptions of Theorem 2.3 it holds for the open-end procedure*

$$\begin{aligned}&\big(\mathrm{CF}_T(\hat{\varepsilon}_{p+1},\hat{\varepsilon}_{p+2},\dots;\gamma)-\mathrm{CF}_T(\varepsilon_{p+1},\varepsilon_{p+2},\dots;\gamma)\big)/T\\&\quad\xrightarrow{P}\sup_{t_0<t<\infty}\left(\frac{t}{t+1}\right)^{1+\gamma}\left(\frac{t-t_0}{t}\right)^2\int\big|\varphi_0(u)\big(\varphi_X(u)-1\big)\big|^2w(u)\,du.\end{aligned}\tag{6.5}$$

For the closed-end procedure an analogous assertion holds where the supremum is taken over $t_0<t<N$.

Proof We follow the lines of the proof of Lemma 6.2. We study the differences

$$\hat{C}_k(t)-C_k^A(t),\quad \hat{S}_k(t)-S_k^A(t),$$

where $\hat{C}_k(t)$, $\hat{S}_k(t)$ are as in the proof of Lemma 6.2,

$$C_k^A(t)=\frac{1}{k}\sum_{j=T+1}^{T+k}\cos\big(t\big(\varepsilon_j+\boldsymbol{\delta}^T\boldsymbol{X}_{j-1}I\{j>T+k_0\}\big)\big)-\frac{1}{T}\sum_{j=1+p}^{T}\cos(t\varepsilon_j),$$

and $S_k^A(t)$ is defined analogously with $\cos(\cdot)$ replaced by $\sin(\cdot)$. By a Taylor expansion

$$\begin{aligned}&\cos(t\hat{\varepsilon}_j)-\cos\big(t\big(\varepsilon_j+\boldsymbol{\delta}^T\boldsymbol{X}_{j-1}I\{j>T+k_0\}\big)\big)\\&\quad=-t\big(\hat{\varepsilon}_j-\varepsilon_j-\boldsymbol{\delta}^T\boldsymbol{X}_{j-1}I\{j>T+k_0\}\big)\sin\big(t\big(\varepsilon_j+\boldsymbol{\delta}^T\boldsymbol{X}_{j-1}I\{j>T+k_0\}\big)\big)\\&\qquad+R_{jC}^A(t),\end{aligned}$$

where $R_{jC}^A(t)$ is a remainder term. Then $\hat{C}_k(t)$ can be decomposed as follows:

$$\hat{C}_k(t)=C_k^A(t)+\hat{C}_{k1}^A(t)+\frac{1}{k}\sum_{j=T+1}^{T+k}R_{jC}^A(t)-\frac{1}{T}\sum_{j=1+p}^{T}R_{jC}(t)$$

with $R_{jC}(t)$ as in the proof of Lemma 6.2 and

$$\begin{aligned}\hat{C}_{k1}^A(t)=-\Bigg(&\frac{1}{k}\sum_{j=T+1}^{T+k}t\big(\hat{\varepsilon}_j-\big(\varepsilon_j+\boldsymbol{\delta}^T\boldsymbol{X}_{j-1}I\{j>T+k_0\}\big)\big)\\&\times\sin\big(t\big(\varepsilon_j+\boldsymbol{\delta}^T\boldsymbol{X}_{j-1}I\{j>T+k_0\}\big)\big)\\&-\frac{1}{T}\sum_{j=1+p}^{T}t(\hat{\varepsilon}_j-\varepsilon_j)\sin(t\varepsilon_j)\Bigg).\end{aligned}$$

Similarly as in the proof of Lemma 6.2 we get

$$\max_{1\le k<\infty}\left(\frac{k}{T+k}\right)^{1+\gamma}\int\left|\frac{1}{k}\sum_{j=T+1}^{T+k}R_{jC}^A(t)-\frac{1}{T}\sum_{j=1+p}^{T}R_{jC}(t)\right|^2 w(t)\,dt=o_P(1).$$

Concerning $\hat{C}_{k1}^A(t)$ we use the decomposition

$$\hat{C}_{k1}^A(t)=(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_T)^T\left(\frac{t}{k}A_{k1}^A(t)-\frac{t}{T}B_1(t)+\frac{t}{k}A_{k2}^A(t)-\frac{t}{T}B_2(t)\right),$$

where $B_1(t)$, $B_2(t)$ are as in the proof of Lemma 6.2 and

$$\begin{aligned}A_{k1}^A(t)=\sum_{j=T+1}^{T+k}&\boldsymbol{X}_{j-1}\big(-\sin\big(t\big(\varepsilon_j+\boldsymbol{\delta}^T\boldsymbol{X}_{j-1}I\{j>T+k_0\}\big)\big)\\&+\mathrm{E}\big(\sin\big(t\big(\varepsilon_j+\boldsymbol{\delta}^T\boldsymbol{X}_{j-1}I\{j>T+k_0\}\big)\big)|\boldsymbol{X}_{j-1}|\big)\big),\\A_{k2}^A(t)=-\sum_{j=T+1}^{T+k}&\boldsymbol{X}_{j-1}\mathrm{E}\big(\sin\big(t\big(\varepsilon_j+\boldsymbol{\delta}^T\boldsymbol{X}_{j-1}I\{j>T+k_0\}\big)\big)|\boldsymbol{X}_{j-1}|\big).\end{aligned}$$

Negligiblity of $B_1(t)$, $B_2(t)$ follows from the proof of Lemma 6.2 and of $A_{k1}^A(t)$ analogously. The term $A_{k2}^A(t)$ has to be treated more carefully. Notice that by Jensen's inequality, by Lemma 6.1, and the fact that $\mathrm{E}X_{j-l}^2\le C$ for some $C>0$:

$$\begin{aligned}&\max_{1\le k<\infty}\left(\frac{k}{k+T}\right)^{1+\gamma}\frac{1}{k^2}\int t^2\big\|A_{k2}^A(t)\big\|^2 w(t)\,dt\\&\quad\le\int t^2 w(t)\,dt\sum_{l=1}^{p}\max_{1\le k<\infty}\frac{1}{k}\sum_{j=T+1}^{T+k}X_{j-l}^2=O_P(1).\end{aligned}$$

Combining the above arguments we have

$$\begin{aligned}&\max_{1\le k<\infty}\left(\frac{k}{T+k}\right)^{1+\gamma}\int\big\|\hat{C}_k(t)-C_k^A(t)\big\|^2 w(t)\,dt\\&\quad=O_P\big(T^{-1}\big)=o_P(1).\end{aligned}$$

Analogous arguments hold for $\hat{S}_k(t)-S_k^A(t)$ as well as the closed-end procedure.

Noticing that

$$\mathrm{E}\big(C_k^A(t)+iS_k^A(t)\big)=\varphi_0(t)\big(\varphi_X(t)-1\big)\frac{\max(0,k-k_0)}{k}$$

it remains to show

$$\max_{1\le k<\infty}\left(\frac{k}{T+k}\right)^{1+\gamma}\int\big|C_k^A(t)-\mathrm{E}C_k^A(t)\big|^2 w(t)\,dt=o_P(1)$$

which follows by Lemma 6.1 ($g=\cos,\sin$). □

Proof of Theorem 2.3 To prove our theorem it suffices to show that, $T \to \infty$,

$$\mathrm{CF}_T(\varepsilon_{p+1}, \varepsilon_{p+2}, \dots; \gamma)/T = o_P(1) \tag{6.6}$$

in addition to (6.5). The latter follows immediately from Lemma 6.4, while assertion (6.6) follows from Lemma 6.3. □

Proof of Theorem 3.1 The proof follows along the line of the proof of Theorem 2.1 but certain parts are more delicate since we work now with a triangular array.

Denote by $\mathrm{E}^*, \mathrm{var}^*, P^*$ etc. expectation, variance and probability w.r.t. $\varepsilon^*(p)$, $\varepsilon^*(p+1), \dots$ given $X_1, X_2, \dots$, i.e., for example $\mathrm{E}^*(\cdot) = \mathrm{E}(\cdot|X_1, X_2, \dots)$.

We have to study conditional limit behavior of

$$\begin{aligned}\mathrm{CF}_T^* &= \mathrm{CF}_T\bigl(\varepsilon^*(p+1), \varepsilon^*(p+2), \dots; \gamma\bigr)\\ &= \max_{1\le k\le TN} T\left(\frac{k}{T+k}\right)^{1+\gamma} \int_{-\infty}^{\infty} \bigl|\varphi^*_{T,T+k}(u) - \varphi^*_{p,T}(u)\bigr|^2 w(u)\,du,\end{aligned}$$

where

$$\varphi^*_{T,T+k}(u) = \frac{1}{k}\sum_{j=T+1}^{T+k} \exp\bigl\{iu\varepsilon^*(j)\bigr\}, \quad u \in R^1,\ k \ge 1$$

and an analogous expression for $\varphi^*_{p,T}$. First, we show that we can replace φ^* by

$$\widetilde{\varphi}^*_{T,T+k}(u) = \frac{1}{k}\sum_{j=T+1}^{T+k} \exp\{iu\varepsilon_{U_T(j)}\}$$

and an analogous expression for $\widetilde{\varphi}^*_{p,T}(u)$, $\widetilde{\mathrm{CF}}^*_T = \mathrm{CF}_T(\varepsilon_{U_T(p+1)}, \varepsilon_{U_T(p+2)}, \dots; \gamma)$. Precisely, we will show that for any $\eta_1, \eta_2 > 0$ it holds for T large enough

$$P^*\bigl(\bigl|\mathrm{CF}^*_T - \widetilde{\mathrm{CF}}^*_T\bigr| \ge \eta_1\bigr) \le \eta_2 + o_P(1).$$

The proof is essentially analogous to the proof of Lemma 6.2 and will only be sketched. The decompositions remain true but $\mathbf{X}_j$ has to be replaced by $\mathbf{X}_{U_T(j)}$ and $\varepsilon_{U_T(j)}$. In the notation we indicate this by $*$, e.g. $\hat{C}^*_{k1}(t)$.

By an application of the Hájek–Rényi inequality it holds for any $D > 0$

$$\begin{aligned}&P^*\left(\max_k \left\| \frac{1}{k}\sum_{j=T+1}^{T+k} \mathbf{X}_{U_T(j-1)}\mathbf{X}^T_{U_T(j-1)} - \mathrm{E}^*\mathbf{X}_{U_T(1)}\mathbf{X}^T_{U_T(1)}\right\| \ge D\right)\\ &\le \frac{1}{D^2}\,\mathrm{var}^*\bigl\|\mathbf{X}_{U_T(p+1)}\mathbf{X}^T_{U_T(p+1)}\bigr\| \sum_{k\ge 1}\frac{1}{k^2}\\ &\le \frac{1}{D^2} C \frac{1}{T-p}\sum_{j=p+1}^{T} \bigl\|\mathbf{X}_j\mathbf{X}_j^T\bigr\|^2 \le \frac{1}{D^2}\bigl(C + o_P(1)\bigr),\end{aligned}$$

where $C > 0$ is now and in the following a generic constant which can change from line to line. This leads similarly by the analog of (6.3) to

$$P^*\left(\max_{1\le k<\infty} T\left(\frac{k}{T+k}\right)^{1+\gamma} \int \left|\hat{C}^*_{k2}(t)\right|^2 w(t)\,dt \ge \eta\right) \le \varepsilon + o_P(1) \tag{6.7}$$

for any $\eta, \varepsilon > 0$. The decomposition of $\hat{C}^*_{k1}(t)$ is slightly different than in the proof of Lemma 6.2. Let

$$\hat{C}^*_{k1}(t) = (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_T)^T \left(\frac{t}{k}\hat{C}^*_{k11}(t) - \frac{t}{T}\hat{C}^*_{k12}(t)\right),$$

where

$$\hat{C}^*_{k11}(t) = -\sum_{j=T+1}^{T+k} \big(\boldsymbol{X}_{U_T(j)-1}\sin(t\varepsilon_{U_T(j)}) - \mathrm{E}^*\big(\boldsymbol{X}_{U_T(j)-1}\sin(t\varepsilon_{U_T(j)})\big)\big),$$

and $\hat{C}^*_{k12}(t)$ is defined analogously with the sum taken from $p+1$ to T. The assertion follows analogously to the assertion for $\hat{C}_{k1}(t)$ in the proof of Lemma 6.2 proving (6.7).

The remainder of the proof is close to the proof of Lemma 6.3, so we only sketch the differences. Equation (6.7) shows that it is sufficient to study

$$\int_{-\infty}^{\infty} \left|\tilde{\varphi}^*_{T,T+k}(u) - \tilde{\varphi}^*_{p,T}(u)\right|^2 w(u)\,du = A_{k1} + A^*_{k2} + A^*_{k3} + A^*_{k4},$$

where A_{k1} is as in the proof of Lemma 6.3 and

$$\begin{aligned}
A^*_{k2} &= \frac{1}{k^2}\sum_{v_1=T+1}^{T+k}\sum_{\substack{v_2=T+1\\ v_2\ne v_1}}^{T+k} \widetilde{h}(\varepsilon_{U_T(v_1)}, \varepsilon_{U_T(v_2)}) + \frac{1}{T^2}\sum_{v_1=1}^{T}\sum_{\substack{v_2=1\\ v_2\ne v_1}}^{T} \widetilde{h}(\varepsilon_{U_T(v_1)}, \varepsilon_{U_T(v_2)}) \\
&\quad - \frac{2}{kT}\sum_{v_1=1}^{T}\sum_{v_2=1+T}^{T+k} \widetilde{h}(\varepsilon_{U_T(v_1)}, \varepsilon_{U_T(v_2)}),\\
A^*_{k3} &= -\frac{2}{k^2}\sum_{v=T+1}^{T+k}\big(\mathrm{E}\big(h(\varepsilon_{U_T(v)}, \varepsilon_0)|\varepsilon_{U_T(v)}\big) - \mathrm{E}^*\mathrm{E}\big(h(\varepsilon_{U_T(v)}, \varepsilon_0)|\varepsilon_{U_T(v)}\big)\big)\\
&\quad - \frac{2}{T^2}\sum_{v=p+1}^{T}\big(\mathrm{E}\big(h(\varepsilon_{U_T(v)}, \varepsilon_0)|\varepsilon_{U_T(v)}\big) - \mathrm{E}^*\mathrm{E}\big(h(\varepsilon_{U_T(v)}, \varepsilon_0)|\varepsilon_{U_T(v)}\big)\big),\\
A^*_{k4} &= -2\,\frac{k+T}{kT}\big(\mathrm{E}^*\mathrm{E}\big(h(\varepsilon_{U_T(p+1)}, \varepsilon_0)|\varepsilon_{U_T(p+1)}\big) - \mathrm{E}h(\varepsilon_1, \varepsilon_2)\big).
\end{aligned}$$

As for A_{k3} in the proof of Lemma 6.3 we obtain by an application of the Hájek–Rényi inequality for any $\eta > 0$

$$P^*\left(\max_{k\geq 1} T\left(\frac{k}{T+k}\right)^{1+\gamma} |A^*_{k3}| \geq \eta\right) = o_P(1).$$

Furthermore it holds

$$\begin{aligned}&\max_{k\geq 1} T\left(\frac{k}{T+k}\right)^{1+\gamma} \left|A^*_{k4}\right| \\ &\quad\leq \frac{1}{T-p}\left|\sum_{j=p+1}^{T}\left(\mathrm{E}\big(h(\varepsilon_j,\varepsilon_0)|\varepsilon_j\big) - \mathrm{E}h(\varepsilon_1,\varepsilon_2)\right)\right| = o_P(1).\end{aligned}$$

Define

$$S^*_k := \sum_{1\leq i<j\leq k}\left(\widetilde{h}(\varepsilon_{U_T(i)},\varepsilon_{U_T(j)}) - \widetilde{h}_L(\varepsilon_{U_T(i)},\varepsilon_{U_T(j)})\right).$$

As $T\to\infty$ an application of the Markov inequality yields

$$\mathrm{E}^*S^*_k = \frac{k(k-1)}{2}\frac{1}{T^2}\sum_{v_1=1}^{T}\sum_{v_2=1}^{T}\left(\widetilde{h}(\varepsilon_{v_1},\varepsilon_{v_2}) - \widetilde{h}_L(\varepsilon_{v_1},\varepsilon_{v_2})\right) = O_P(1)\frac{k^2}{T}\sqrt{\sum_{j\geq L}\lambda_j^2}$$

for some C by Lemma A in Serfling (1980), p. 183, which shows that

$$\sup_{k\geq 1}\frac{T}{T+k}\left(\frac{k}{k+T}\right)^{\gamma}\frac{1}{k}\left|\mathrm{E}^*S^*_k\right|$$

becomes sufficiently small for L large enough. $S^*_k - \mathrm{E}^*S^*_k$ can be expressed as a linear combinations of martingales (cf. Serfling 1980, pp. 178–179), therefore the Hájek–Rényi inequality yields, as $T\to\infty$,

$$\begin{aligned}&P^*\left(\max_{k\geq 1}\frac{T}{T+k}\left(\frac{k}{k+T}\right)^{\gamma}\frac{1}{k}\left|S^*_k - \mathrm{E}^*S^*_k\right| \geq A\right) \\ &\quad\leq \frac{C}{A^2}\sum_{k\geq 1}\frac{1}{k^2}\frac{1}{T^2}\sum_{j=1}^{T}\sum_{v=1}^{T}\left(\widetilde{h}(\varepsilon_j,\varepsilon_v) - \widetilde{h}_L(\varepsilon_j,\varepsilon_v)\right)^2 \\ &\quad= \frac{C}{A^2}\sum_{j=L+1}^{\infty}\lambda_j^2 + o_P(1)\end{aligned}$$

for each L, for the open-end procedure and the first term of $A_{k2}(\widetilde{h}) - A_{k2}(\widetilde{h}_L)$ for all $A > 0$ and some $C > 0$. The right hand side becomes arbitrarily small for L large enough. Hence for any $A_1, A_2 > 0$ and all $L \geq L_0$ (for some L_0) T large, it holds

$$P\left(\max_{k\geq T}\rho_{k,T}(\gamma)\left|A^*_{k2} - A^*_{k2}(L)\right| \geq A_1\right) \leq A_2,$$

where

$$A^*_{k2}(L) = \sum_{q=1}^{L} \lambda_q \left(\frac{T}{k^2} B_1^{*2}(q,k) - \frac{T+k}{Tk} - B_2^*(q,k) \right),$$

$$B_1^*(q,k) = \frac{1}{\sqrt{T}} \sum_{v=T+1}^{T+k} \left(g_q(\varepsilon_{U_T(v)}) - \frac{1}{T} \sum_{j=1}^{T} g_q\left(\varepsilon_{U_T(j)}\right) \right),$$

$$B_2^*(q,k) = \frac{1}{k^2} \sum_{v=T+1}^{T+k} \left(g_q^2(\varepsilon_{U_T(v)}) - 1 \right) + \frac{1}{T^2} \sum_{v=1}^{T} \left(g_q^2(\varepsilon_{U_T(j)}) - 1 \right).$$

We start with the term $B_2^*(q,k)$. First note that for $r > 1$

$$\mathrm{E}^* \left| g_q^2(\varepsilon_{U_T(1)}) - 1 \right|^2 = T^{r-1} \frac{1}{T^r} \sum_{v=1}^{T} \left(g_q^2(\varepsilon_v) - 1 \right)^2 = o_P\left(T^{r-1}\right) \tag{6.8}$$

by Theorem 5.2.3 (i)(α) in Chow and Teicher (1997) since

$$\sum_{j\geq 1} P\left(\left(g_q^2(\varepsilon_v) - 1 \right)^r \geq j^r \right) = \sum_{j \geq 1} P\left(\left| g_q^2(\varepsilon_v) - 1 \right| \geq j \right) = \mathrm{E} \left| g_q^2(\varepsilon_v) - 1 \right| < \infty.$$

According to Shorack and Wellner (1986), inequality 4 (p. 858) in addition to the von Bahr Esseen inequality (Shorack and Wellner 1986, p. 858) it holds for i.i.d. random variables with mean 0 for $1 < r \leq 2$

$$\mathrm{E} \max_{1 \leq k \leq T} \left(|S_k|^r \right) \leq c T\, \mathrm{E} |X_1|^r$$

for some $c > 0$, Theorem 1.1 in Fazekas and Klesov (2000) then gives for $b_1 \geq \cdots \geq b_T > 0$

$$\mathrm{E} \left(\max_{1 \leq k \leq T} b_k |S_k| \right)^r \leq c\, \mathrm{E} |X_1|^r \sum_{k=1}^{T} b_k^r$$

for some $c > 0$. We have to distinguish the cases $k \leq T$ and $k \geq T$. By the above inequality for $1 < r < \min(2, 1/(1-\gamma))$, i.e., $0 \leq r(1-\gamma) < 1$, and (6.8) it holds

$$\mathrm{E}^* \left(\max_{k \leq T} \frac{T}{T+k} \left(\frac{1}{T+k} \right)^{\gamma} \frac{1}{k^{1-\gamma}} \left| \sum_{v=T+1}^{T+k} \left(g_q^2(\varepsilon_{U_T(v)}) - 1 \right) \right| \right)$$

$$\leq T^{-r\gamma} \sum_{k=1}^{T} \frac{1}{k^{r(1-\gamma)}}\, \mathrm{E}^* \left| g_q^2(\varepsilon_{U_T(1)}) - 1 \right|^2 = o_P(1).$$

Similarly for $r = 2$

$$\mathrm{E}^*\left(\max_{k>T} T\left(\frac{k}{T+k}\right)^{1+\gamma}\frac{1}{k^2}\left|\sum_{v=T+1}^{T+k}\left(g_q^2(\varepsilon_{U_T(v)})-1\right)\right|\right)$$
$$\leq T^2\sum_{k>T}\frac{1}{k^4}\,\mathrm{E}^*\left|g_q^2(\varepsilon_{U_T(1)})-1\right|^2 = o_P(1).$$

An application of the Markov inequality yields the negligibility of $B_2^*(q,k)$.

It remains to show that the process $\{(B_1^*(1,\lfloor Ts\rfloor),\dots,B_1^*(L,\lfloor Ts\rfloor));\ s\in[0,A]\}$ converges weakly (conditionally) to a Wiener process for all $A>0$. The convergence of the finite-dimensional distributions follows e.g. from Singh (1981), Theorem 1, while tightness follows by Theorem 15.6 in Billingsley (1968) and the finiteness of the second moments. We can then conclude as in the proof of Lemma 6.3. □

Acknowledgements The work of Z. Hlávka was partially supported by MSM 0021620839, the work of M. Hušková was partially supported by the grants 201/09/0755 and MSM 0021620839, and the work of S. Meintanis by grant number KA: 70/4/7658 of the program 'Kapodistrias' of the special account for research of the National and Kapodistrian University of Athens. The position of C. Kirch was financed by the Stifterverband für die Deutsche Wissenschaft by funds of the Claussen-Simon-trust and the research supported by the DFG grant KI 1443/2-1.

References

Andreou E, Ghysels E (2006) Monitoring disruptions in financial markets. J Econom 135(1–2):77–124

Aue A, Hoermann S, Horváth L, Hušková M (2010) Sequential testing for the stability of portfolio betas. Econom Theory (accepted)

Bai J (1993) On the partial sums of residuals in autoregressive and moving average models. J Time Ser Anal 14(3):247–260

Billingsley P (1968) Convergence of probability measures. Wiley, New York

Chow YS, Teicher H (1997) Probability theory: independence, interchangeability, martingales, 3rd edn. Springer texts in statistics. Springer, New York

Csörgő M, Horváth L (1997) Limit theorems in change-point analysis. Wiley, Chichester

Davis RA, Huang DW, Yao YC (1995) Testing for a change in the parameter values and order of an autoregressive model. Ann Stat 23(1):282–304

Dufour JM, Farhat A (2002) Exact nonparametric two-sample homogeneity tests. In: Goodness-of-fit tests and model validity (Paris, 2000). Stat. ind. technol. Birkhäuser, Boston, pp 435–448

Einmahl JHJ, McKeague IW (2003) Empirical likelihood based hypothesis testing. Bernoulli 9(2):267–290

Epps TW (1987) Testing that a stationary time series is Gaussian. Ann Stat 15(4):1683–1698

Epps TW (1988) Testing that a Gaussian process is stationary. Ann Stat 16(4):1667–1683

Epps TW (1999) Limiting behavior of the ICF test for normality under Gram-Charlier alternatives. Stat Probab Lett 42(2):175–184

Fazekas I, Klesov O (2000) A general approach to the strong laws of large numbers. Teor Veroâtn Ee Primen 45(3):568–583

Fried R, Imhoff M (2004) On the online detection of monotonic trends in time series. Biom J 46(1):90–102

Gombay E, Serban D (2009) Monitoring parameter change in AR(p) time series models. J Multivar Anal 100(4):715–725

Hong Y (1999) Hypothesis testing in time series via the empirical characteristic function: a generalized spectral density approach. J Am Stat Assoc 94(448):1201–1220

Horváth L (1993) Change in autoregressive processes. Stoch Process Appl 44(2):221–242

Horváth L, Hušková M, Kokoszka P, Steinebach J (2004) Monitoring changes in linear models. J Stat Plan Inference 126(1):225–251

Hušková M, Kirch C (2011) Bootstrapping sequential change-point tests for linear regression. Metrika. doi:10.1007/s001-0347-7

Hušková M, Meintanis SG (2006a) Change point analysis based on empirical characteristic functions. Metrika 63(2):145–168

Hušková M, Meintanis SG (2006b) Change-point analysis based on empirical characteristic functions of ranks. Seq Anal 25(4):421–436

Hušková M, Prášková Z, Steinebach J (2007) On the detection of changes in autoregressive time series. I. Asymptotics. J Stat Plan Inference 137(4):1243–1259

Hušková M, Kirch C, Prášková Z, Steinebach J (2008) On the detection of changes in autoregressive time series. II. Resampling procedures. J Stat Plan Inference 138(6):1697–1721

Kirch C (2008) Bootstrapping sequential change-point tests. Seq Anal 27(3):330–349

Lee S, Lee Y, Na O (2009) Monitoring distributional changes in autoregressive models. Commun Stat, Theory Methods 38(16–17):2969–2982

Ranga Rao R (1962) Relations between weak and uniform convergence of measures with applications. Ann Math Stat 33:659–680

Serfling RJ (1980) Approximation theorems of mathematical statistics. Wiley series in probability and mathematical statistics. Wiley, New York

Shorack GR, Wellner JA (1986) Empirical processes with applications to statistics. Wiley series in probability and mathematical statistics. Wiley, New York

Singh K (1981) On the asymptotic accuracy of Efron's bootstrap. Ann Stat 9(6):1187–1195

Tenreiro C (2009) On the choice of the smoothing parameter for the BHEP goodness-of-fit test. Comput Stat Data Anal 53(4):1038–1053

Yao YC (1990) On the asymptotic behavior of a class of nonparametric tests for a change-point problem. Stat Probab Lett 9(2):173–177