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

Zdeněk Hlávka[*], Marie Hušková[†], Claudia Kirch[‡], Simos G. Meintanis[§]

6. December 2010

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

AMS Subject Classification 2000: 62M10, 62G10, 62G20

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

[*]Charles University in Prague, Faculty of Mathematics and Physics, Department of Statistics, Sokolovská 83, 186 75 Praha 8, Czech Republic, hlavka@karlin.mff.cuni.cz

[†]Charles University in Prague, Faculty of Mathematics and Physics, Department of Statistics, Sokolovská 83, 186 75 Praha 8, Czech Republic, huskova@karlin.mff.cuni.cz

[‡]Karlsruhe Institute of Technology, Institute of Stochastics, Kaiserstr. 89, 76133 Karlsruhe, Germany, claudia.kirch@kit.edu

[§]National and Kapodistrian University of Athens, Department of Economics, 8 Pesmazoglou Street, 105 59 Athens, Greece, simosmei@econ.uoa.gr

McKeague (2003), Hušková et al. (2007), Hušková et al. (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,b) develop detection procedures for the distributional changes in 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 $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^T$ 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 stationary i.e., the characteristic polynomial $P(z) = 1 - \beta_1 z - ... - \beta_p z^p$, 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 $F_1 = ... = F_T$. 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. observe 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 distributional aspect of the errors ε_j, i.e. we wish to test the hypothesis

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

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. It turns out however that 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

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

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

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

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. However, the total number of observations 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,b) in order to test for distributional changes of independent observations in an offline 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) and Epps (1987).

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

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

where

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

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

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

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

We shall distinguish between open-end procedures where $L_T = \infty$ and closed-end procedures where $L_T = \lfloor NT \rfloor + 1$ for some $N > 0$.

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

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

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

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

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

$$CF_T(\gamma) = CF_T(\widehat{\varepsilon}_{p+1}, \widehat{\varepsilon}_{p+2}, \ldots; \gamma) := \sup_{1\le j<L_T} T_{CF}(j,\gamma) \tag{1.7}$$

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

We have pointed out that one can also develop related procedures based on empirical distribution functions. To this end, denote by $\widehat{F}_{T,T+j}(z)$ and $\widehat{F}_{p,T}(z)$ the EDF based on $\widehat{\varepsilon}_{T+1}, \ldots, \widehat{\varepsilon}_{T+j}$ and $\widehat{\varepsilon}_{p+1}, \ldots, \widehat{\varepsilon}_T$, respectively. The corresponding Kolmogorov-Smirnov statistic is then defined by

$$KS_T(\gamma) = \sup_{1\le j<L_T} d_{j,T}(\gamma) \sup_z |\widehat{F}_{T,T+j}(z) - \widehat{F}_{p,T}(z)| \tag{1.8}$$

$\gamma \in (0, 1]$. The choice of $d_{j,T}(\gamma)$ corresponding to our choice (in the open-end procedure) is

$$d_{j,T}(\gamma) = \sqrt{T}\Big(\frac{j}{T+j}\Big)^{(1+\gamma)/2}.$$

The respective limit null distribution of $KS_T(\gamma)$ is an asymptotically distribution free functional of a two-dimensional Gaussian process. The advantage is that this limit distribution unlike the one we obtain for our procedure (cf. Theorem 2.1) does not depend on the unknown error distribution. One can then either use a bootstrap test similar to ours, or construct tests based on (simulated) asymptotic critical values. However, in order to obtain these results (more precisely to obtain the analogue of our Lemma 6.2) additional assumptions on the smoothness of the error distribution are needed.

Motivated by Bai (1994) in the nonsequential case, Lee et al. (2009) study procedures related to $KS_T(\gamma)$. However, they use the different standardization $d_{j,T}(\gamma) = \sqrt{T}\left(\frac{T}{T+j}\right)^a$ for some $a > 0$. EDF–based Cramér-von-Mises type test statistics can be developed along the same line, but they also have the disadvantage that additional assumptions on the smoothness of the error distribution are needed. More details on these type of statistics including some comparative simulations can be found in Hlávka et al. (2010).

2 Asymptotic results

Here we present and discuss results on asymptotic distribution of the test statistic $CF_T(\gamma)$ defined in (1.7) both under the null hypothesis and under some alternatives leading 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.

2 Asymptotic results

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

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

where $\boldsymbol{X}_{j-1} = (X_{j-1},\dots,X_{j-p})^T$, $\boldsymbol{\beta}$ is an unknown p-regression parameter, ε_j, $j = 1,\dots$ are innovations that fulfill under the null hypothesis the assumptions:

(A.1) $\{\varepsilon_j,\ j = 0,\pm 1,\dots\}$ 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 $X_1,\dots,X_p$ are independent of $\varepsilon_{p+1},\dots,\varepsilon_n$; $\beta_p \neq 0$, and the roots of the polynomial $t^p - \beta_1 t^{p-1} - \ldots - \beta_p$ are less than one in absolute value.

(A.3) The vector $\boldsymbol{X}_{p+1} = (X_p,\dots,X_1)^T$ of initial observations satisfies

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

where

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

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

(A.4) $\widehat{\boldsymbol{\beta}}_T$ is an estimator of $\boldsymbol{\beta}$ based on $X_1,\dots,X_T$ with properties

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

(A.5) Let $w(t)$, $t \in R^1$, be a symmetric function with properties

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

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

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

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

where

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

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

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

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

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

The theorem shows that the limit distribution depends on unknown quantities and does not provide an approximation for critical values of the CF–statistic. Therefore a bootstrap suitable for the above sequential setup is useful for practical applications and will be discussed in Section 3 below. Also note that the conditions on the weight function for the open-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 $\widehat{\varepsilon}_j$ by ε_j.

Remark 2.1. The proposed procedure based on CF_T defined in (1.7) can easily be extended to different residual-based models such as regression models or ARMA-sequences among others. The key to the proofs is to be able to estimate the residuals ε_t by $\widehat{\varepsilon}_t$, 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) $\{\varepsilon_j,\, j = 0, \pm 1, \dots\}$ are independent random variables having zero mean, positive variance and finite moment $\mathsf{E}\,|\varepsilon_j|^4 < \infty$ and having the distribution function F_0 for $j \leq T + j_0$ and F^0 for $j > T + j_0$, for some $j_0 \geq 0$, $F_0 \neq F^0$.

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. Then for the open-ended procedure, as $T \to \infty$*

$$CF_T \to \infty, \quad \text{in} \quad \text{probability}.$$

Moreover, if $j_0 = \lfloor Tt_0 \rfloor$ with some $t_0 \geq 0$ then, as $T \to \infty$

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

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 as required. However, the test procedure has also some power with respect to changes in the autoregressive coefficient. So, one should apply a test for a change in the autoregressive parameter as well, which does not have power against distributional changes, in order to distinguish between the two. 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, all other symbols are as in model (1.1). As in Hušková et al. (2007) we assume:

(B.2) The observations $X_{p+1}, \ldots,$ follow the above model with $j_0 = \lfloor Tt_0 \rfloor$, $t_0 \geq 0$; $X_1, ..., X_p$ are independent of $\varepsilon_{p+1}, \ldots, \varepsilon_T, \ldots$; $\beta_p \neq 0$, the roots of the polynomial $t^p - \beta_1 t^{p-1} - \ldots - \beta_p$ are less than one in absolute value, $\beta_p + \delta_p \neq 0$, and the roots of $t^p - (\beta_1 + \delta_1)t^{p-1} - \ldots - (\beta_p + \delta_p)$ are also less than one in absolute value, $\boldsymbol{\delta} \neq \mathbf{0}$ fixed.

Theorem 2.3. *Let model (2.4) fulfill (A.1), (B.2), (A.3)–(A.5), $j_0 = \lfloor Tt_0 \rfloor$ for some $t_0 \geq 0$, i.e. a change in the regression coefficient takes place. Then, for open-end procedures, as $T \to \infty$,*

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

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) 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. The standard approach is 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 $p+1, \ldots, T$ 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

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

with $\widehat{\varepsilon}_j$ as in (1.3). Note that for the LS–estimator $\widehat{\boldsymbol{\beta}}_T$ it holds that $\sum_{j=p+1}^T \widehat{\varepsilon}_j = 0$, so that the bootstrap errors have mean zero (conditionally), and therefore no extra centering is required.

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

$$P_T^* \left(CF_T(\varepsilon^*(1), \ldots, \varepsilon^*(\tilde{L}_T)) \leq c_\alpha(X_1, \ldots, X_T) \right) \geq 1 - \alpha.$$

where $P_T^*(\cdot) = P(\cdot | X_1, \ldots, X_T)$. 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, no additional assumptions on the data after monitoring starts is 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 the bootstrap test has asymptotic size α and asymptotic power one under the alternatives in Theorem 2.2 and 2.3. Asymptotically it is equivalent to the asymptotic test which is not feasible as c_α depends on too many unknown parameters.

The above bootstrap procedure is not optimal in case of smaller training data sets, which is not surprising as we create a data set of length L_T from a dataset of length T which is much smaller than L_T. In the simpler location setting Kirch (2008) showed in simulations that this in fact leads to a loss of power. Adaptations of the above bootstrap schemes including observations obtained during monitoring are possible (cf. Kirch (2008) for the location model as well as Hušková and Kirch (2010) for a change in the regression coefficient), but these modifications lead to a substantial increase in the technical difficulty of our proofs and will therefore not be considered here. More details as well as some simulations can be found in Hlávka et al. (2010).

4 Simulation study

In the previous sections we derived monitoring procedures with an asymptotic overall 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 can be found in Hlávka et al. (2010).

As a weight function for the integral in (1.4) we use $w(u) = w_a(u) = \exp(-a|u|)$, with several values of $a > 0$. 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 $N = 4$. For the calculation of the bootstrap distribution 500 random bootstrap samples have been used.

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 occurring at time j_0.

We consider the following types of changes:

- 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 as well as 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 the first two properties we use:

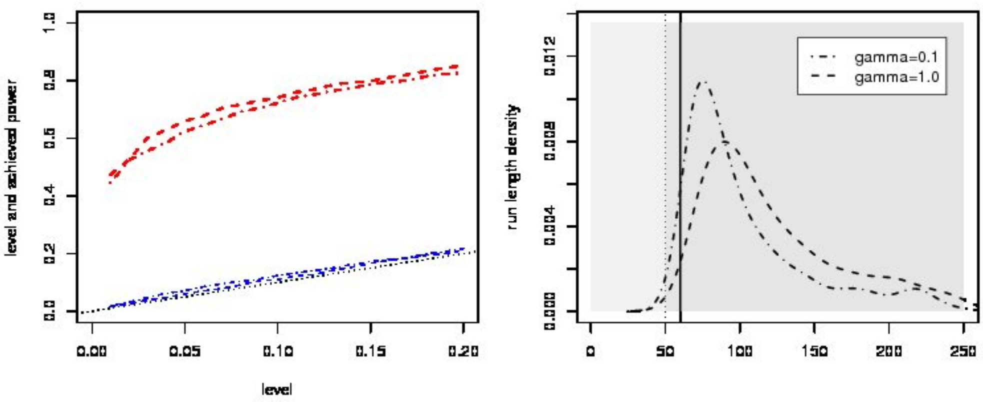
(i) Achieved Size-Power Curves (ASP)
The blue line corresponding to the null hypothesis shows the actual achieved level on the y-axis for a nominal one as given by the x-axis. The red line corresponding to one alternative shows the size-corrected power, i.e. the power of the test belonging to a true level α test where α is given by the x-axis. These plots are based on 1 000 repetitions of the procedure.

In order to visualize how fast changes are detected we show the:

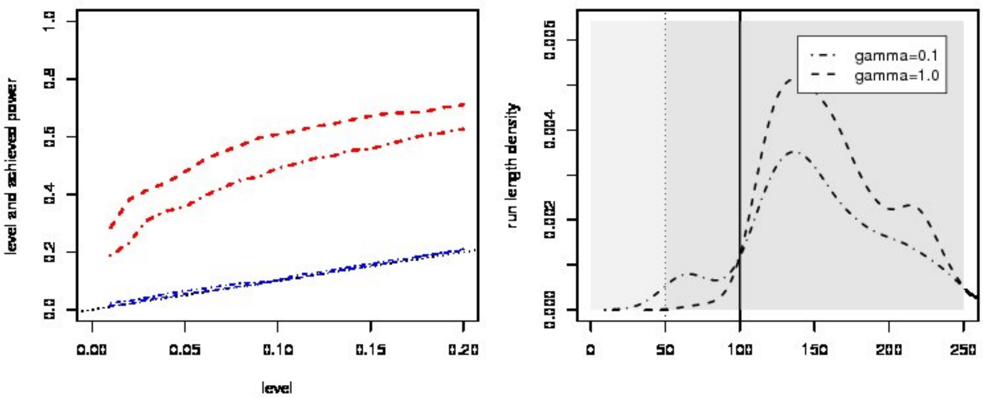
(ii) Estimated density of the run length (EDR)
The run length is the point in time at which the null hypothesis is rejected. In the plots it is calculated for a true size 5% test (not a nominal one). This is to obtain comparable plots for all procedures without having to take size-differences into account, 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 artefacts 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 Figure 4.1 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 Figure 4.1 the plots are only given for $a = 1$, but other values of a lead to similar results.

With Figure 4.2 and Figure 4.3 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 we fix the value of γ at $\gamma = 1$. Also we compare the new procedure with the Kolmogorov-Smirnov-type sequential test defined by (1.8). In Figure 4.2, plots include a change in the AR-parameter as well as in the scale of the error distribution. From these plots 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.



(a) Early change: $j_0 = 10$

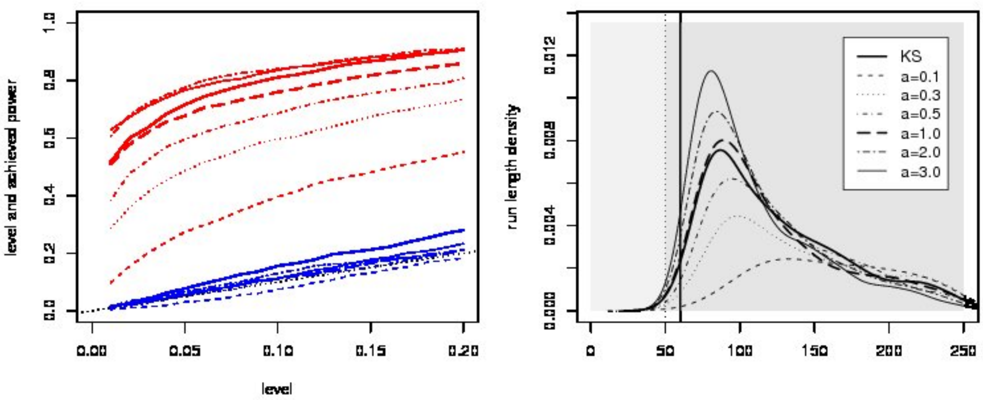


(b) Late change: $j_0 = 50$

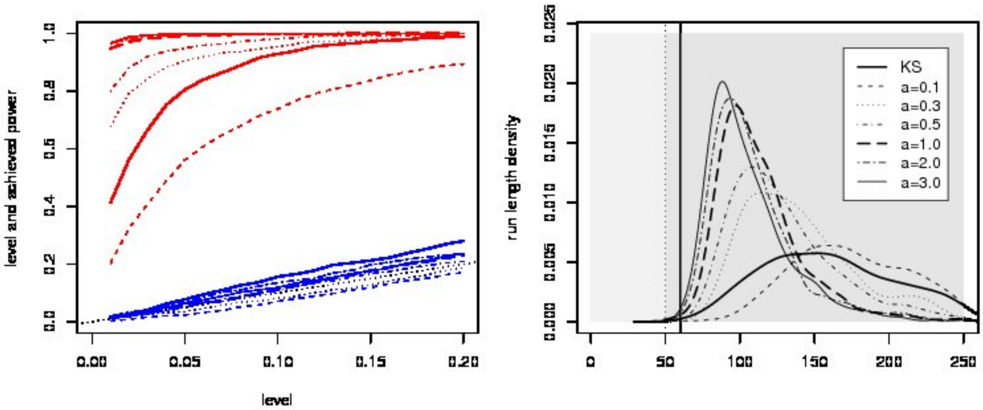
Figure 4.1: ASP and EDR plots for the CF_T test with normal errors; change in the regression parameter from $\beta_0 = 0.5$ to $\beta^0 = 0.9$ ($\gamma = 0.1$ and $\gamma = 1.0$, $\sigma_0 = \sigma^0 = 1$, $a = 1$, $T = 50$, $N = 4$)

The plots in Figure 4.3 include changes in the distribution of the errors: From a normal distribution, to a t_4–distribution as well as to a χ^2–distribution with 4 degrees of freedom. Corresponding results for the Kolmogorov-Smirnov-type test are also included. According to Figure 4.3 a), i.e. $T = 50$ and a change from normal to t_4, both procedures have a very low power for small samples (but are unbiased). Apparently, the differences between the two distributions is not large enough to be detectable at a satisfying power with only a historic sample of $T = 50$ 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; as an example, for two equally sized samples of length 50 we obtain an empirical 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 Figure 4.3 are reasonable for all values of a, but best for intermediate values of this parameter. Naturally, for the larger historic data set $T = 200$ (Figure 4.3 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



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



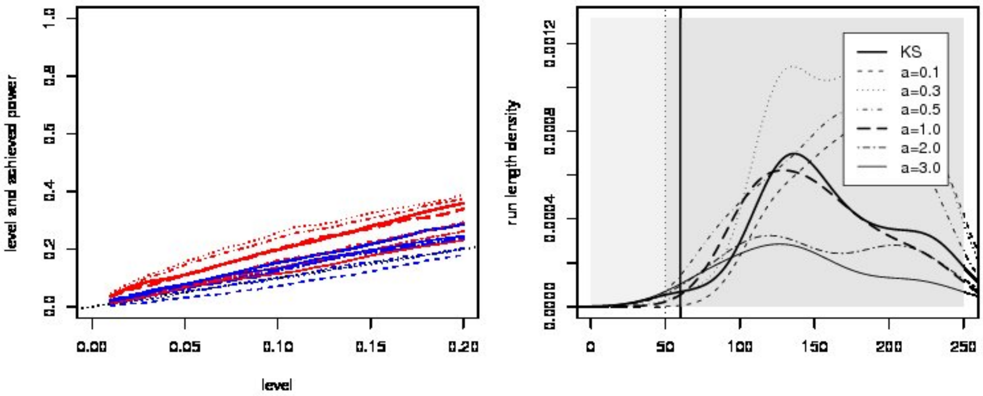
(b) Change in the scale parameter: $\sigma^0 = 2$

Figure 4.2: 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

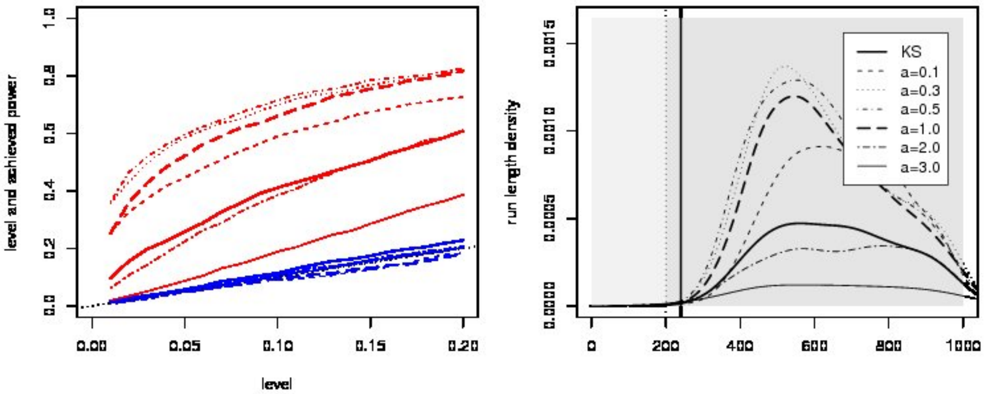
procedure clearly outperforms the Kolmogorov–Smirnov type test. In Figure 4.3 c) and d) analogous pictures for a change from a normal to a χ^2-distribution with 4 degrees of freedom can be found. The power is somewhat better than for the change to a t_4–distribution, but the general conclusions remain the same.

5 Conclusion

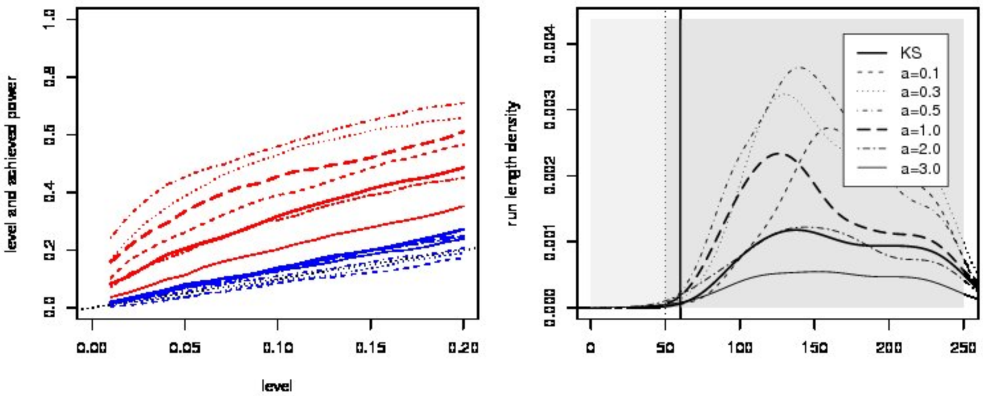
We propose goodness–of–fit procedures for the error distribution of AR models, in the sequential set–up. The new tests utilize an L^2–type discrepancy measure between a couple of empirical characteristic functions (ECFs) of the residuals; the first ECF includes the residuals computed from a training 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 chance 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



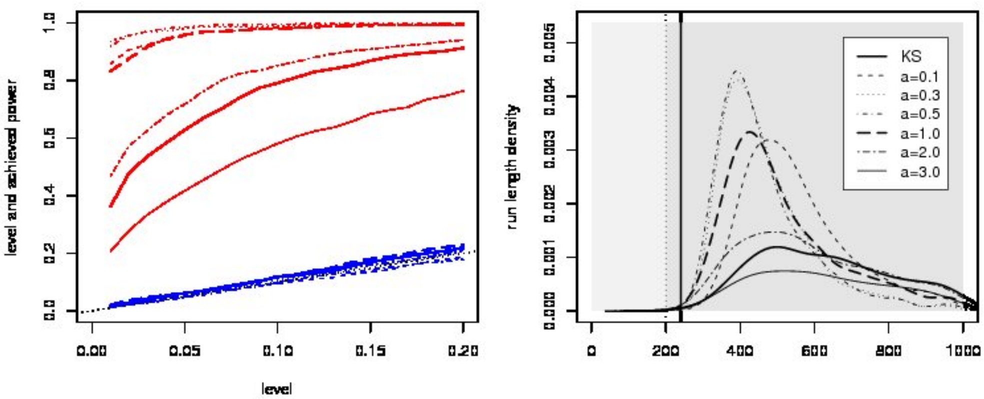
(a) Change from normal to t_4: $T = 50$



(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: $T = 200$

Figure 4.3: 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

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 $\eta > 3/2$*

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

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

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

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

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

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

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

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

$$\int\left|\frac{1}{k}\sum_{j=T+1}^{T+k}\big(g(t\epsilon_j)-\mathrm{E}g(t\epsilon_j)\big)\right|^2 w(t)\,dt$$

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

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

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

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

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 Section 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

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

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

■

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*

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

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

Proof. We will study the differences

$$\widehat{C}_k(t)-C_k(t),\quad \widehat{S}_k(t)-S_k(t)$$

where

$$\widehat{C}_k(t) = \frac{1}{k}\sum_{j=T+1}^{T+k}\cos(t\widehat{\varepsilon}_j) - \frac{1}{T}\sum_{j=1+p}^{T}\cos(t\widehat{\varepsilon}_j),$$
$$C_k(t) = \frac{1}{k}\sum_{j=T+1}^{T+k}\cos(t\varepsilon_j) - \frac{1}{T}\sum_{j=1+p}^{T}\cos(t\varepsilon_j),$$
$$\widehat{S}_k(t) = \frac{1}{k}\sum_{j=T+1}^{T+k}\sin(t\widehat{\varepsilon}_j) - \frac{1}{T}\sum_{j=1+p}^{T}\sin(t\widehat{\varepsilon}_j),$$
$$S_k(t) = \frac{1}{k}\sum_{j=T+1}^{T+k}\sin(t\varepsilon_j) - \frac{1}{T}\sum_{j=1+p}^{T}\sin(t\varepsilon_j).$$

By the Taylor expansion

$$\cos(t\widehat{\varepsilon}_j) = \cos(t\varepsilon_j) - t(\widehat{\varepsilon}_j - \varepsilon_j)\sin(t\varepsilon_j) + R_{jC}(t)$$

where $R_{jC}(t)$ is a remainder term. Then $\widehat{C}_k(t)$ can be decomposed:

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

with

$$\widehat{C}_{k1}(t) = -\Big(\frac{1}{k}\sum_{j=T+1}^{T+k} t(\widehat{\varepsilon}_j - \varepsilon_j)\sin(t\varepsilon_j) - \frac{1}{T}\sum_{j=1+p}^{T} t(\widehat{\varepsilon}_j - \varepsilon_j)\sin(t\varepsilon_j)\Big),$$
$$\widehat{C}_{k2}(t) = \frac{1}{k}\sum_{j=T+1}^{T+k} R_{jC}(t) - \frac{1}{T}\sum_{j=1+p}^{T} R_{jC}(t).$$

Since

$$|R_{jC}(t)| \leq Dt^2(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_T)^T \boldsymbol{X}_{j-1}\boldsymbol{X}_{j-1}^T(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_T)$$

for some positive $D > 0$ we also have

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

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

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

uniformly in k. Hence by (A.5)

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

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

$$\widehat{C}_{k1}(t) = (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_T)^T \Big(\frac{t}{k} \widehat{\boldsymbol{C}}_{k11}(t) + \frac{t}{T} \widehat{C}_{k12}(t) + \frac{t}{k} \widehat{C}_{k13}(t) + \frac{t}{T} \widehat{C}_{k14}(t) \Big)$$

where

$$\widehat{\boldsymbol{C}}_{k11}(t) = \sum_{j=T+1}^{T+k} \boldsymbol{X}_{j-1}(-\sin(t\varepsilon_j) + E\sin(t\varepsilon_j)),$$

$$\widehat{\boldsymbol{C}}_{k12}(t) = -\sum_{j=p+1}^{T} \boldsymbol{X}_{j-1}(-\sin(t\varepsilon_j) + E\sin(t\varepsilon_j)),$$

$$\widehat{\boldsymbol{C}}_{k13}(t) = -E(\sin(t\varepsilon_1)) \sum_{j=T+1}^{T+k} \boldsymbol{X}_{j-1},$$

$$\widehat{\boldsymbol{C}}_{k14}(t) = E(\sin(t\varepsilon_1)) \sum_{j=p+1}^{T} \boldsymbol{X}_{j-1}.$$

By (A.4) one obtains

$$|\widehat{C}_{k1}(t)|^2 = O_P(1) \frac{1}{T} \left\| \Big(\frac{t}{k} \widehat{\boldsymbol{C}}_{k11}(t) + \frac{t}{T} \widehat{\boldsymbol{C}}_{k12}(t) + \frac{t}{k} \widehat{\boldsymbol{C}}_{k13}(t) + \frac{t}{T} \widehat{\boldsymbol{C}}_{k14}(t) \Big) \right\|^2$$

uniformly in k and t. $\{\widehat{\boldsymbol{C}}_{k11}(t), \sigma(X_1, \ldots, X_{T+k}), k \geq 1\}$ is a martingale for each $t \in R^1$ with

$$\mathrm{E}\widehat{\boldsymbol{C}}_{k11}(t) = 0, \qquad \mathrm{var}\widehat{\boldsymbol{C}}_{k11}(t) \leq Dk.$$

Consequently

$$\left\{ \int t^2 \left\| \widehat{\boldsymbol{C}}_{k11}(t) \right\|^2 w(t)dt, \, \sigma(X_1, \ldots, X_{T+k}), \, k \geq 1 \right\}$$

is a nonnegative submartingale with

$$\mathrm{E} \int t^2 \left\| \widehat{\boldsymbol{C}}_{k11}(t) \right\|^2 w(t)dt \leq Dk.$$

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

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

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

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

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

$$
\begin{aligned}
&\lambda\, P\left(\max_{1\le k<\infty}\left(\frac{k}{k+T}\right)^{1+\gamma}\frac{1}{k^2}\int t^2\left\|\widehat{\boldsymbol{C}}_{k11}(t)\right\|^2 w(t)dt>\lambda\right)\\
&\le \lambda\, P\left(\max_{1\le k<\infty} v_l\int t^2\left\|\widehat{\boldsymbol{C}}_{k11}(t)\right\|^2 w(t)dt>\lambda\right)\\
&\le DT^{-1-\gamma}\sum_{k=1}^{T} k\left(k^{\gamma-1}-(k+1)^{\gamma-1}\right)+\sum_{k\ge T} k\left(k^{-2}-(k+1)^{-2}\right)\\
&\le DT^{-1-\gamma}\sum_{k=1}^{T} k^{\gamma-1}+\sum_{k\ge T} k^{-2}\le DT^{-1}=o(1).
\end{aligned}
$$

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

By Lemma 6.1

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

which implies

$$\max_{1\le k<\infty}\left(\frac{k}{k+T}\right)^{1+\gamma}\frac{1}{k^2}\int t^2||\widehat{\boldsymbol{C}}_{k13}(t)||^2 w(t)dt=o_P(1).$$

Similar relations hold true also for $\widehat{\boldsymbol{C}}_{k12}(t)$ and $\widehat{\boldsymbol{C}}_{k14}(t)$. It is in fact an easier situation since the random part does not depend on k.

Combining the above arguments we obtain

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

Analogous argument for $\widehat{S}_k(t) - S_k(t)$ 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* $CF_T(\widehat{\varepsilon}_{p+1}, \widehat{\varepsilon}_{p+2}, \ldots; \gamma)$ *by* $CF_T(\varepsilon_{p+1}, \varepsilon_{p+2}, \ldots; \gamma)$.

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

$$\int_{-\infty}^{\infty} \left|\widetilde{\phi}_{T,T+k}(u) - \widetilde{\phi}_{1,T}(u)\right|^2 w(u)\,du = A_{k1} + A_{k2} + A_{k3},$$

where

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

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

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

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

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

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

$$P\left(\max_{k\geq 1}\frac{T}{T+k}\left(\frac{k}{k+T}\right)^{\gamma}\frac{1}{k}\left|\sum_{j=T+1}^{T+k} Z_j\right| \geq \eta\right)$$

$$\leq C\eta^{-2}\Big(\frac{1}{T^{2\gamma}}\sum_{k=1}^{T}\frac{1}{k^{2-2\gamma}} + \sum_{k\geq T}\frac{1}{k^2}\Big) \to 0 \qquad (0<\gamma\leq 1).$$

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

$$\widetilde{h}(x,y) \overset{L^2_\varepsilon}{=} \sum_{j=1}^{\infty}\lambda_j g_j(x) g_j(y),$$

$$\text{i.e.}\quad \lim_{L\to\infty} E\left(\widetilde{h}(\varepsilon_1,\varepsilon_2) - \sum_{j=1}^{L}\lambda_j g_j(\varepsilon_1) g_j(\varepsilon_2)\right)^2 = 0,$$

$$\mathrm{E}g_j(\varepsilon_1) = 0, \quad \mathrm{E}g_j^2(\varepsilon_1) = 1, \quad \mathrm{E}g_j(\varepsilon_1)g_k(\varepsilon_2) = 0,\ j\neq k,$$

$$\mathrm{E}\widetilde{h}^2(\varepsilon_1,\varepsilon_2) = \sum_{j=1}^{\infty}\lambda_j^2 < \infty.$$

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)

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

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 $C > 0$

$$
\begin{aligned}
&P\left(\max_{k\geq 1} \frac{T}{T+k}\left(\frac{k}{k+T}\right)^{\gamma} \frac{1}{k}|S_k| \geq \eta\right) \\
&\leq \frac{1}{\eta^2}\sum_{k\geq 1} \frac{1}{k^2}\mathrm{E}\left(\widetilde{h}(\varepsilon_1,\varepsilon_2) - \widetilde{h}_L(\varepsilon_1,\varepsilon_2)\right)^2 \leq \frac{C}{\eta^2}\sum_{j=L+1}^{\infty} \lambda_j^2 \to 0 \qquad (L\to\infty)
\end{aligned}
$$

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

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

It holds

$$
\begin{aligned}
A_{k2}(L) &= \sum_{q=1}^{L} \lambda_q \left(\frac{T}{k^2}B_1^2(q,k) - \frac{T+k}{Tk} - B_2(q,k)\right),\\
B_1(q,k) &= \frac{1}{\sqrt{T}}\sum_{v=T+1}^{T+k}\left(g_q(\varepsilon_v) - \frac{1}{T}\sum_{j=1}^{T} g_q(\varepsilon_j)\right),\\
B_2(q,k) &= \frac{1}{k^2}\sum_{v=T+1}^{T+k}\left(g_q^2(\varepsilon_v)-1\right) + \frac{1}{T^2}\sum_{v=1}^{T}\left(g_q^2(\varepsilon_v)-1\right)
\end{aligned}
$$

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

$$
\max_{k} \rho_{k,T}(\gamma) B_2(q,k) \to 0 \qquad a.s.
$$

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 $C > 0$

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

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 $k \leq \delta T$ is negligible for δ small enough (in case of the open-end procedure) $(\gamma \neq 0)$:

$$
\begin{aligned}
&P\left(\max_{k\leq \delta T}\sqrt{\frac{T}{k^2}\left(\frac{k}{T+k}\right)^{1+\gamma}}\left|\sum_{v=T+1}^{T+k} g_q(e_v)\right| \geq \eta\right)\\
&\leq \frac{C}{\eta^2}T\sum_{k\leq\delta T}\frac{1}{k^{1-\gamma}(T+k)^{1+\gamma}} \leq \frac{C}{-\eta^2}T^{\gamma}\sum_{k\leq \delta T}\frac{1}{k^{1-\gamma}} \leq \frac{C}{\eta^2}\,\delta^{\gamma}
\end{aligned}
$$

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 ($t \approx k/T$)

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

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

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

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 $\widehat{\varepsilon}_t$ are replaced by ε_t. By Lemma 6.1 ($g = \cos, \sin$) it follows

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

From this we can conclude immediately that $CF_T = O_P(1/T)$ 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*

$$(CF_T(\widehat{\varepsilon}_{p+1},\widehat{\varepsilon}_{p+2},\ldots)-CF_T(\varepsilon_{p+1},\varepsilon_{p+2},\ldots))/T \\ \xrightarrow{P} \sup_{t_0<t<\infty}\Big(\frac{t}{t+1}\Big)^{1+\gamma}\Big(\frac{t-t_0}{t}\Big)^2\int|\varphi_0(u)(\varphi_X(u)-1)|^2w(u)du. \tag{6.4}$$

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 6.3. We study the differences

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

where $\widehat{C}_k(t)$, $\widehat{S}_k(t)$ are as in the proof of Lemma 6.3 and

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

$$S_k^A(t)=\frac{1}{k}\sum_{j=T+1}^{T+k}\sin(t(\varepsilon_j+\boldsymbol{\delta}^T\boldsymbol{X}_{j-1}I\{j>T+k_0\}))-\frac{1}{T}\sum_{j=1+p}^{T}\sin(t\varepsilon_j).$$

By a Taylor expansion

$$\cos(t\widehat{\varepsilon}_j)-\cos(t(\varepsilon_j+\boldsymbol{\delta}^T\boldsymbol{X}_{j-1}I\{j>T+k_0\})) \\ =-t(\widehat{\varepsilon}_j-\varepsilon_j-\boldsymbol{\delta}^T\boldsymbol{X}_{j-1}I\{j>T+k_0\})\sin(t(\varepsilon_j+\boldsymbol{\delta}^T\boldsymbol{X}_{j-1}I\{j>T+k_0\}))+R_{jC}^A(t)$$

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

$$\widehat{C}_k(t)=C_k^A(t)+\widehat{C}_{k1}^A(t)+\widehat{C}_{k2}^A(t)$$

with

$$\widehat{C}_{k1}^A(t)=-\Big(\frac{1}{k}\sum_{j=T+1}^{T+k}t(\widehat{\varepsilon}_j-(\varepsilon_j+\boldsymbol{\delta}^T\boldsymbol{X}_{j-1}I\{j>T+k_0\}))\sin(t(\varepsilon_j+\boldsymbol{\delta}^T\boldsymbol{X}_{j-1}I\{j>T+k_0\})) \\ -\frac{1}{T}\sum_{j=1+p}^{T}t(\widehat{\varepsilon}_j-\varepsilon_j)\sin(t\varepsilon_j)\Big),$$

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

$R_{jC}(t)$ as in the proof of Lemma 6.3. Similarly as in the proof of Lemma 6.3 we get

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

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

$$\widehat{C}_{k1}^A(t)=(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_T)^T\Big(\frac{t}{k}\widehat{\boldsymbol{C}}_{k11}^A(t)+\frac{t}{T}\widehat{\boldsymbol{C}}_{k12}(t)+\frac{t}{k}\widehat{\boldsymbol{C}}_{k13}^A(t)+\frac{t}{T}\widehat{\boldsymbol{C}}_{k14}(t)\Big),$$

where $\widehat{C}_{k12}(t)$, $\widehat{C}_{k14}(t)$ are as in the proof of Lemma 6.3 and

$$\widehat{\boldsymbol{C}}^A_{k11}(t) = \sum_{j=T+1}^{T+k} \boldsymbol{X}_{j-1}(-\sin(t(\varepsilon_j + \boldsymbol{\delta}^T \boldsymbol{X}_{j-1} I\{j > T + k_0\})) + \mathrm{E}(\sin(t(\varepsilon_j + \boldsymbol{\delta}^T \boldsymbol{X}_{j-1} I\{j > T + k_0\}))|\boldsymbol{X}_{j-1})),$$

$$\widehat{\boldsymbol{C}}^A_{k13}(t) = -\sum_{j=T+1}^{T+k} \boldsymbol{X}_{j-1}\mathrm{E}(\sin(t(\varepsilon_j + \boldsymbol{\delta}^T \boldsymbol{X}_{j-1} I\{j > T + k_0\}))|\boldsymbol{X}_{j-1})),$$

Negligiblity of $\widehat{\boldsymbol{C}}_{k12}(t), \widehat{\boldsymbol{C}}_{k14}(t)$ follows from the proof of Lemma 6.3 and of $\widehat{\boldsymbol{C}}^A_{k11}(t)$ analogously. The term $\widehat{\boldsymbol{C}}^A_{k13}(t)$ has to be treated more carefully. Notice that by Jenssens inequality

$$\begin{aligned}
&\max_{1 \le k < \infty} \left(\frac{k}{k+T}\right)^{1+\gamma} \frac{1}{k^2} \int t^2 \left\|\widehat{\boldsymbol{C}}^A_{k13}(t)\right\|^2 w(t)dt \\
&\le \int t^2 w(t)\,dt \max_{1 \le k < \infty} \frac{1}{k^2} \left(\sum_{j=T+1}^{T+k} \|\boldsymbol{X}_{j-1}\|\right)^2 \\
&\le \int t^2 w(t)\,dt \max_{1 \le k < \infty} \frac{1}{k} \sum_{j=T+1}^{T+k} \|\boldsymbol{X}_{j-1}\|^2 \\
&\le \int t^2 w(t)\,dt \sum_{l=1}^{p} \max_{1 \le k < \infty} \frac{1}{k} \sum_{j=T+1}^{T+k} X^2_{j-l} = O_P(1),
\end{aligned}$$

where the last line follows from Lemma 6.1 and the fact that $\mathrm{E}X^2_{j-l} \le C$ for some $C > 0$.

Combining the above arguments we have

$$\max_{1 \le k < \infty} \left(\frac{k}{T+k}\right)^{1+\gamma} \int \left\|\widehat{C}_k(t) - C^A_k(t)\right\|^2 w(t)dt = O_P(T^{-1}) = o_P(1).$$

Analogous arguments holds for $\widehat{S}_k(t) - S^A_k(t)$ as well as the closed-end procedure.

Noticing that

$$\mathrm{E}(C^A_k(t) + iS^A_k(t)) = \varphi_0(t)(\varphi_X(t) - 1)\frac{(k-k_0)_+}{k}$$

it remains to show

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

Since

$$\varepsilon_j + \boldsymbol{\delta}^T \boldsymbol{X}_{j-1} 1\{j > T + k_0\} = X_j - \boldsymbol{\beta}^T \boldsymbol{X}_{j-1} = \boldsymbol{\gamma}^T \boldsymbol{X}_{p+1,j-1}$$

for $\boldsymbol{\gamma} = (1, -\beta_1, \ldots, -\beta_p)^T$ it follows from Lemma 6.1 ($g = \cos, \sin$)

$$\begin{aligned}
&\max_{1 \le k < \infty} \left(\frac{k}{T+k}\right)^{1+\gamma} \int \left|\frac{1}{k} \sum_{j=T+k_0+1}^{T+k} \left(\exp\{it\boldsymbol{\gamma}^T \boldsymbol{X}_{p+1,j}\} - \mathrm{E}(\exp\{it\boldsymbol{\gamma}^T \boldsymbol{X}_{p+1,j}\})\right)\right|^2 w(t)dt \\
&= o_P(1).
\end{aligned}$$

■

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

$$CF_T(\varepsilon_{p+1}, \varepsilon_{p+2}, \ldots)/T = o_P(1) \tag{6.5}$$

in addition to (6.4). The latter follows immediately from Lemma 6.4, while assertion (6.5) 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. $\varepsilon^*(p), \varepsilon^*(p+1), \ldots$ given $X_1, X_2, \ldots,$, i.e. for example $\mathrm{E}^*(\cdot) = \mathrm{E}(\cdot|X_1, X_2, \ldots)$.

We have to study conditional limit behavior of

$$\begin{aligned} CF_T^* &= CF_T(\varepsilon^*(p+1), \varepsilon^*(p+2), \ldots) \\ &= \max_{1 \le k \le TN} T\Big(\frac{k}{T+k}\Big)^{1+\gamma} \int_{-\infty}^{\infty} \left|\varphi^*_{T,T+k}(u) - \varphi^*_{p,T}(u)\right|^2 w(u)\, du, \end{aligned}$$

where

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

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 $\widetilde{\varphi}^*_{p,T}(u)$, $\widetilde{CF}^*_T = CF_T(\varepsilon_{U_T(p+1)}, \varepsilon_{U_T(p+2)}, \ldots)$. Precisely, we will show that for any $\eta_1, \eta_2 > 0$ it holds for T large enough

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

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. $\widehat{C}^*_{k1}(t)$.

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

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

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

$$P^*\left(\max_{1\le k<\infty} T\left(\frac{k}{T+k}\right)^{1+\gamma}\int\left|\widehat{C}^*_{k2}(t)\right|^2 w(t)\,dt \ge \eta\right) \le \varepsilon + o_P(1) \tag{6.6}$$

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

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

where

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

$$\widehat{\boldsymbol{C}}^*_{k12}(t) = \sum_{j=p+1}^{T}\left(\boldsymbol{X}_{U_T(j)-1}\sin(t\varepsilon_{U_T(j)}) - E^*\left(\boldsymbol{X}_{U_T(j)-1}\sin(t\varepsilon_{U_T(j)})\right)\right).$$

The assertion follows analogously to the assertion for $\widehat{C}_{k1}(t)$ in the proof of Lemma 6.2 proving (6.6).

The remainder of the proof is close to the proof of Lemma 6.3, so we only scetch the differences. Equation (6.6) 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

$$A^*_{k2} = \frac{1}{k^2}\sum_{v_1=T+1}^{T+k}\sum_{\substack{v_2=T+1\\ v_2\ne v_1}}^{T+k}\widetilde{h}(\varepsilon_{U_T(v_1)}, \varepsilon_{U_T(v_2)}) + \frac{1}{T^2}\sum_{v_1=1}^{T}\sum_{\substack{v_2=1\\ v_2\ne v_1}}^{T}\widetilde{h}(\varepsilon_{U_T(v_1)}, \varepsilon_{U_T(v_2)}),$$
$$- \frac{2}{kT}\sum_{v_1=1}^{T}\sum_{v_2=1+T}^{T+k}\widetilde{h}(\varepsilon_{U_T(v_1)}, \varepsilon_{U_T(v_2)})$$

$$A^*_{k3} = -\frac{2}{k^2}\sum_{v=T+1}^{T+k}\left(\mathrm{E}(h(\varepsilon_{U_T(v)}, \varepsilon_0)|\varepsilon_{U_T(v)}) - \mathrm{E}^*\mathrm{E}\left(h(\varepsilon_{U_T(v)}, \varepsilon_0)|\varepsilon_{U_T(v)}\right)\right)$$
$$- \frac{2}{T^2}\sum_{v=p+1}^{T}\left(\mathrm{E}\left(h(\varepsilon_{U_T(v)}, \varepsilon_0)|\varepsilon_{U_T(v)}\right) - \mathrm{E}^*\mathrm{E}\left(h(\varepsilon_{U_T(v)}, \varepsilon_0)|\varepsilon_{U_T(v)}\right)\right)$$

$$A^*_{k4} = -2\,\frac{k+T}{kT}\left(\mathrm{E}^*\mathrm{E}\left(h(\varepsilon_{U_T(p+1)}, \varepsilon_0)|\varepsilon_{U_T(p+1)}\right) - Eh(\varepsilon_1, \varepsilon_2)\right).$$

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

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

Furthermore it holds

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

Define

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

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

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

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

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

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), p. 178-179), therefore the Hájek-Rényi inequality yields, as $T\to\infty$,

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

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

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

where

$$\begin{aligned}A_{k2}^*(L) &= \sum_{q=1}^{L}\lambda_q\left(\frac{T}{k^2}B_1^{*2}(q,k) - \frac{T+k}{Tk} - B_2^*(q,k)\right),\\ B_1^*(q,k) &= \frac{1}{\sqrt{T}}\sum_{v=T+1}^{T+k}\left(g_q(\varepsilon_{U_T(v)}) - \frac{1}{T}\sum_{j=1}^{T} g_q(\varepsilon_{U_T(j)}))\right),\\ B_2^*(q,k) &= \frac{1}{k^2}\sum_{v=T+1}^{T+k}\left(g_q^2(\varepsilon_{U_T(v)}) - 1\right) + \frac{1}{T^2}\sum_{v=1}^{T}\left(g_q^2(\varepsilon_{U_T(j)}) - 1\right).\end{aligned}$$

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

$$\mathrm{E}^*|g_q^2(\varepsilon_{U_T(1)}-1|^2 = T^{r-1}\frac{1}{T^r}\sum_{v=1}^{T}\left(g_q^2(\varepsilon_v)-1\right)^2 = o_P(T^{r-1}) \tag{6.7}$$

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 $1<r\leq 2$

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

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

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

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 (6.7) it holds

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

Similarly for $r=2$

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

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

It remains to show that the process $\{(B_1^*(1,\lfloor Ts\rfloor),\ldots,B_1^*(L,\lfloor Ts\rfloor));\ s\in[0,A]\}$ 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. ■

Acknowlegements

The work of Z. Hlávka was partially supported by MSM 0021620839, the work of M. Hušková was partially supported by the grants GACR 201/09/j006, 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

Elena Andreou and Eric Ghysels. Monitoring disruptions in financial markets. *J. Econometrics*, 135(1-2):77–124, 2006. ISSN 0304-4076.

A. Aue, S. Hoermann, L. Horváth, and M. Hušková. Sequential testing for the stability of portfolio betas. Submitted, 2010.

Jushan Bai. On the partial sums of residuals in autoregressive and moving average models. *J. Time Ser. Anal.*, 14(3):247–260, 1993. ISSN 0143-9782.

Jushan Bai. Weak convergence of the sequential empirical processes of residuals in ARMA models. *Ann. Statist.*, 22(4):2051–2061, 1994. ISSN 0090-5364.

Patrick Billingsley. *Convergence of probability measures.* John Wiley & Sons, New York, 1968.

Yuan Shih Chow and Henry Teicher. *Probability theory: independence, interchangeability, martingales.* Springer Texts in Statistics. Springer-Verlag, New York, third edition, 1997. ISBN 0-387-98228-0.

Miklós Csörgő and Lajos Horváth. *Limit theorems in change-point analysis.* John Wiley & Sons, Chichester, 1997.

Richard A. Davis, Da Wei Huang, and Yi-Ching Yao. Testing for a change in the parameter values and order of an autoregressive model. *Ann. Statist.*, 23(1):282–304, 1995. ISSN 0090-5364.

John H. J. Einmahl and Ian W. McKeague. Empirical likelihood based hypothesis testing. *Bernoulli*, 9(2):267–290, 2003. ISSN 1350-7265.

T. W. Epps. Testing that a stationary time series is Gaussian. *Ann. Statist.*, 15(4): 1683–1698, 1987. ISSN 0090-5364.

T. W. Epps. Testing that a Gaussian process is stationary. *Ann. Statist.*, 16(4):1667–1683, 1988. ISSN 0090-5364.

I. Fazekas and O. Klesov. A general approach to the strong laws of large numbers. *Teor. Veroyatnost. i Primenen.*, 45(3):568–583, 2000. ISSN 0040-361X.

R. Fried and M. Imhoff. On the online detection of monotonic trends in time series. *Biom. J.*, 46(1):90–102, 2004. ISSN 0323-3847.

Edit Gombay and Daniel Serban. Monitoring parameter change in AR(p) time series models. *J. Multivariate Anal.*, 100(4):715–725, 2009. ISSN 0047-259X.

Z. Hlávka, M. Hušková, C. Kirch, and S. G. Meintanis. Bootstrap procedures for monitoring autoregressive models. In preparation, 2010.

Yongmiao Hong. Hypothesis testing in time series via the empirical characteristic function: a generalized spectral density approach. *J. Amer. Statist. Assoc.*, 94(448):1201–1220, 1999. ISSN 0162-1459.

Lajos Horváth. Change in autoregressive processes. *Stochastic Process. Appl.*, 44(2): 221–242, 1993. ISSN 0304-4149.

Lajos Horváth, Marie Hušková, Piotr Kokoszka, and Josef Steinebach. Monitoring changes in linear models. *J. Statist. Plann. Inference*, 126(1):225–251, 2004. ISSN 0378-3758.

Marie Hušková and Simos G. Meintanis. Change point analysis based on empirical characteristic functions. *Metrika*, 63(2):145–168, 2006a. ISSN 0026-1335.

Marie Hušková and Simos G. Meintanis. Change-point analysis based on empirical characteristic functions of ranks. *Sequential Anal.*, 25(4):421–436, 2006b. ISSN 0747-4946.

Marie Hušková, Zuzana Prášková, and Josef Steinebach. On the detection of changes in autoregressive time series. I. Asymptotics. *J. Statist. Plann. Inference*, 137(4): 1243–1259, 2007. ISSN 0378-3758.

Marie Hušková, Claudia Kirch, Zuzana Prášková, and Josef Steinebach. On the detection of changes in autoregressive time series. II. Resampling procedures. *J. Statist. Plann. Inference*, 138(6):1697–1721, 2008. ISSN 0378-3758.

M. Hušková and C. Kirch. Bootstrapping sequential change-point tests for linear regression. Submitted, 2010.

Claudia Kirch. Bootstrapping sequential change-point tests. *Sequential Anal.*, 27(3): 330–349, 2008. ISSN 0747-4946.

Sangyeol Lee, Youngmi Lee, and Okyoung Na. Monitoring distributional changes in autoregressive models. *Comm. Statist. Theory Methods*, 38(16-17):2969–2982, 2009. ISSN 0361-0926.

R. Ranga Rao. Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statist.*, 33:659–680, 1962. ISSN 0003-4851.

Robert J. Serfling. *Approximation theorems of mathematical statistics.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York, 1980. ISBN 0-471-02403-1.

Galen R. Shorack and Jon A. Wellner. *Empirical processes with applications to statistics.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York, 1986. ISBN 0-471-86725-X.

Kesar Singh. On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.*, 9(6): 1187–1195, 1981. ISSN 0090-5364.

Yi-Ching Yao. On the asymptotic behavior of a class of nonparametric tests for a change-point problem. *Statist. Probab. Lett.*, 9(2):173–177, 1990. ISSN 0167-7152.