Taylor & Francis
Taylor & Francis Group

Sequential Analysis
Design Methods and Applications

ISSN: 0747-4946 (Print) 1532-4176 (Online) Journal homepage: http://www.tandfonline.com/loi/lsqa20

An online change detection test for parametric discrete-time stochastic processes

Fanni K. Nedényi

To cite this article: Fanni K. Nedényi (2018) An online change detection test for parametric discrete-time stochastic processes, Sequential Analysis, 37:2, 246-267, DOI: 10.1080/07474946.2018.1466540

To link to this article: https://doi.org/10.1080/07474946.2018.1466540

SEQUENTIAL ANALYSIS
2018, VOL. 37, NO. 2, 246–267
https://doi.org/10.1080/07474946.2018.1466540

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

An online change detection test for parametric discrete-time stochastic processes

Fanni K. Nedényi

MTA-SZTE Analysis and Stochastics Research Group, Bolyai Intitute, University of Szeged, Szeged, Hungary

ABSTRACT
Detecting a change as fast as possible in an observed stochastic process is an important task. In this article, an online procedure is presented to detect changes in the parameter of general discrete-time parametric stochastic processes. As examples, regression models, autoregressive processes, and Galton–Watson processes are investigated. The test is called cumulative sum (CUSUM) type because it is based on the cumulated sums of the estimates of certain martingale difference sequences belonging to the process. In case of a single change alternative hypothesis, the procedure is examined in terms of consistency. Due to the online manner, the time of change can also be estimated.

ARTICLE HISTORY
Received 17 July 2017
Revised 14 January 2018
Accepted 13 April 2018

KEYWORDS
Change-point detection; online procedure; parametric process; rejection time

SUBJECT CLASSIFICATIONS
60F05; 60J80; 62F03

1. Introduction

In the literature on statistics, offline and online procedures have both been introduced to detect changes in stochastic systems. We call a procedure offline if the whole sample is given at the time of the testing and online if the testing is performed in a sequential manner, taking observations one by one. The aim of this article is to perform online change-point detection on the parameter of a certain vector-valued parametric process $X_1, X_2, \ldots$

The online procedure is considered the following way. Throughout the article, we assume that the so-called noncontamination assumption holds for some positive integer m, meaning that the parameter is unchanged until time m. This assumption is regular in the context of online procedures and allows us to estimate the default value of the parameter in question. For the sake of generality we fix a constant $T > 0$ and define the test based on the observations $X_1, \ldots, X_m, X_{m+1}, \ldots, X_{m+\lfloor Tm \rfloor}$. If $T = \infty$, then the test is called open-ended; otherwise, it is called closed-ended. The goal is to test the null hypothesis that there is no change in the parameter on the entire given time horizon. In the online case, test statistics of the form $\tau_{m,k} = \tau_{m,k}(X_1, \ldots, X_{m+k})$, $k = 1, 2, \ldots,$

CONTACT F. K. Nedényi nfanni@math.u-szeged.hu Bolyai Institute, University of Szeged, Aradi vértanúk tere 1, H-6720 Szeged, Hungary.
Recommended by Marie Huskova

are considered, and a rejection is made if $\sup_{1\le k\le \lfloor Tm\rfloor} \tau_{m,k} > x_\alpha$, where x_α is the critical value corresponding to the significance level $\alpha \in (0,1)$. The value κ is called a rejection time if $\tau_{m,\kappa} > x_\alpha$. The theoretical background of the procedure is that under the null hypothesis and certain regularity conditions $\sup_{1\le k\le \lfloor Tm\rfloor} \tau_{m,k} \to_D \tau_T$, $m\to\infty$, for some random variable τ_T that depends on the model and the constant T. Then an approximation of the critical value x_α can be derived from the distribution of τ_T by solving $P(\tau_T > x_\alpha) = \alpha$ for x_α. Indeed, if x_α is a continuity point of the distribution function of the limit variable τ_T, then

$$P\Big(\sup_{1\le k\le \lfloor Tm\rfloor} \tau_{m,k} > x_\alpha\Big) \to \alpha, \qquad m\to\infty,$$

meaning that x_α is an asymptotically correct critical value corresponding to the significance level α.

Online change-point detection has been an investigated area in the last decades. The above-discussed noncontamination assumption was first introduced in Chu et al. (1996). In Chu et al. (1996) and Horváth et al. (2004), a statistical methodology was developed that supplies a limit theorem establishing an online procedure. The statistics in these papers are special cases of ours, having the form $\tau_{m,k} = ||S_{m,k}||$, where $S_{m,k}$ is defined in (2.2). In Horváth et al. (2004, 2007) and Aue et al. (2006), this general methodology is applied to linear regression models in an open-ended manner. Under a single change alternative hypothesis, their tests are shown to be consistent and they investigate the distribution of the rejection times as well. In Kirch and Tadjuidje Kamgaing (2011), open-ended and closed-ended procedures are given to test for a change in special functional autoregressive models. Our aim is to generalize these results to discrete-time stochastic processes satisfying certain general regularity conditions. Our article and the above-mentioned references contain statistics based on the cumulative sums (CUSUMs) of suitable estimators of certain martingale difference sequences of the process. Such statistics are called CUSUM-type. Note that another CUSUM-type statistic is also frequently applied in online change-point detection that is based on the cumulated sums of likelihood quotients.

The main results of the article are presented in Section 2, with the proofs given in Section 3. Subsection 2.3 contains a discussion of some examples of processes that fit into our model.

2. Main results

2.1. Model and test statistics

In our model, the observations are $\mathbb{R}^q \times \mathbb{R}^r$-valued random pairs $(\mathbb{X}_n, \mathbb{Y}_n)$, $n = 1, 2, \ldots$, with some positive integers q and r. Let $\mathcal{F}_{n-1}$ stand for the σ-algebra generated by the random vectors $\{\mathbb{X}_k, \mathbb{Y}_{k-1} : k \le n\}$. Throughout the article we will assume that

$$E[\mathbb{Y}_n \,|\, \mathcal{F}_{n-1}] = E[\mathbb{Y}_n \,|\, \mathbb{X}_n] = f(\mathbb{X}_n, \boldsymbol{\theta}_{\mathrm{n}}), \qquad n = 1, 2, \ldots, \tag{2.1}$$

where $f : \mathbb{R}^q \times \Theta \to \mathbb{R}^r$ is a known measurable function with components $f_1, \ldots, f_r$, Θ is a measurable subset of a finite dimensional Euclidean space, and $\boldsymbol{\theta}_{\mathrm{n}} \in \Theta$ is a parameter of the joint distribution of $\mathbb{X}_n$ and $\mathbb{Y}_n$. Note that here and throughout the article, the equations concerning the conditional expectations are understood in an almost sure sense.

For any fixed, known positive integer m, by the noncontamination assumption it is *a priori* known that $\boldsymbol{\theta}_{\mathrm{n}} = \boldsymbol{\theta}_0$ for $n = 1, \ldots, m$ with a fixed but unknown $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$. The aim of online change detection is to test whether $\boldsymbol{\theta}_{\mathrm{m+1}} = \cdots = \boldsymbol{\theta}_{\mathrm{m+\lfloor Tm \rfloor}} = \boldsymbol{\theta}_0$ with a given $T \in (0, \infty]$. For this goal, we will test the null hypothesis

$$\mathcal{H}_0: \qquad E[\mathbb{Y}_n \mid \mathbb{X}_n] = f(\mathbb{X}_n, \boldsymbol{\theta}_0), \qquad n = m+1, \ldots, m + \lfloor Tm \rfloor.$$

Note that this null hypothesis is weaker than the equality of the parameters. It is easy to see that without further assumptions, the dynamics of the underlying model could be unchanged with different parameters; for example, if the function f does not depend on all of the components of its second argument. However, in case of many applications the two are equivalent; see, for example, the one discussed in Subsection 2.3.2.

We would like to obtain asymptotical results, namely, when m, the size of the training sample, and therefore the number of observations goes to infinity. One could define a triangular array with rows $(\mathbb{X}_n, \mathbb{Y}_n), n = 1, \ldots, m + \lfloor Tm \rfloor$, where $m = 1, 2, \ldots$ Then for every $m = 1, 2, \ldots$, the m th row is the input for the corresponding testing, where the first m pairs serve as the training sample, and we test the above-introduced $\mathcal{H}_0$ corresponding to the given m. Therefore, for the asymptotical results we assume that every row satisfies the noncontamination assumption and the related null hypothesis. Then the variables $\mathbb{U}_n := \mathbb{Y}_n - f(\mathbb{X}_n, \boldsymbol{\theta}_0)$, $n = 1, 2, \ldots$, form a martingale difference sequence with respect to the filtration $\mathcal{F}_0, \mathcal{F}_1, \ldots$ For a given positive integer m, we consider an estimator $\hat{\boldsymbol{\theta}}_m$ of the true parameter $\boldsymbol{\theta}_0$ based on the training sample $(\mathbb{X}_1, \mathbb{Y}_1), \ldots, (\mathbb{X}_m, \mathbb{Y}_m)$, and we define an estimator of the martingale difference sequence by $\hat{\mathbb{U}}_{m,n} := \mathbb{Y}_n - f(\mathbb{X}_n, \hat{\boldsymbol{\theta}}_m)$, $n = 1, 2, \ldots$, which variables our testing method is based on.

We summarize our regularity conditions and some additional notations in the following assumption. Throughout the article, the vector norm is the Euclidean norm, and $\mathbb{1}_A$ is the indicator of the event A. The notations $\mathbb{Z}_+$, $\mathbb{Z}_{++}$ and $\mathcal{B}(\mathbb{R}^q)$ stand for the set of nonnegative integers, positive integers, and the Borel σ-algebra of the space $\mathbb{R}^q$, respectively.

Assumption 2.1.

i. *The process $\mathbb{X}_n$, $n \in \mathbb{Z}_{++}$, is strictly stationary and ergodic or it is an aperiodic positive Harris recurrent Markov chain. The notation $\tilde{\mathbb{X}}_0$ stands for an arbitrary random vector whose distribution is the same as the unique stationary distribution of this process.*
ii. *Suppose that $E[\mathbb{Y}_n \mid \mathbb{X}_n] = f(\mathbb{X}_n, \boldsymbol{\theta}_0)$ for every $n \in \mathbb{Z}_{++}$.*
iii. *There exists an open neighborhood $\Theta_0 \subseteq \Theta$ of $\boldsymbol{\theta}_0$ such that the functions $f_i(\mathbf{x}, \boldsymbol{\theta})$, $\mathrm{i} = 1, \ldots, \mathrm{r}$, are continuously differentiable with respect to the variable $\boldsymbol{\theta}$ at every point $(\mathbf{x}, \boldsymbol{\theta}) \in \mathbb{R}^{\mathrm{q}} \times \Theta_0$. Let $\nabla_{\boldsymbol{\theta}} f_i(\mathbf{x}, \boldsymbol{\theta})$ stand for the vector of partial derivatives.*
iv. *There exists a real number $a > 0$ and a measurable function $h: \mathbb{R}^q \to [0, \infty)$ such that*

$$\|\nabla_{\boldsymbol{\theta}} f_i(\mathbf{x}, \boldsymbol{\theta}) - \nabla_{\boldsymbol{\theta}} \mathrm{f_i}(\mathbf{x}, \boldsymbol{\theta}_0)\| \leq \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^{\mathrm{a}} \mathrm{h}(\mathbf{x}), \ \mathbf{x} \in \mathbb{R}^{\mathrm{q}}, \ \boldsymbol{\theta} \in \Theta_0,$$

for $i = 1, \ldots, r$.
v. *The expectations $Eh(\tilde{\mathbb{X}}_0)$ and $E\nabla_{\boldsymbol{\theta}} f_i(\tilde{\mathbb{X}}_0, \boldsymbol{\theta}_0)$, $i = 1, \ldots, r$, are finite.*

vi. *We have an estimator* $\hat{\boldsymbol{\theta}}_m$ *of* $\boldsymbol{\theta}_0$ *based on the training sample* $(\mathbb{X}_1, \mathbb{Y}_1), \ldots, (\mathbb{X}_m, \mathbb{Y}_m)$ *such that* $m^{1/2}(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0) = O_P(1)$.

vii. *There exists an* $\varepsilon > 0$ *such that* $\sup_{n\geq 1} E||\mathbb{U}_n||^{2+\varepsilon}$ *is finite. Note that if this holds for any* $\varepsilon > 0$, *then the constant* $v_0 := \sup_{n\geq 1} E||\mathbb{U}_n||^2$ *is finite as well.*

viii. *There exists a nonsingular matrix* $\mathbf{C}_0 \in \mathbb{R}^{r\times r}$ *such that one of the following convergences holds as* $m \to \infty$:

$$\frac{1}{m}\sum_{n=1}^{m} \mathbb{U}_n\mathbb{U}_n^\top \xrightarrow{P} \mathbf{C}_0, \qquad \frac{1}{m}\sum_{n=1}^{m} E\,[\mathbb{U}_n\mathbb{U}_n^\top \,|\, \mathcal{F}_{n-1}] \xrightarrow{P} \mathbf{C}_0.$$

ix. *The matrix* $\mathbf{C}_0$ *has a weakly consistent positive semidefinite estimator* $\hat{\mathbf{C}}_m \in \mathbb{R}^{r\times r}$ *based on the sample* $(\mathbb{X}_1, \mathbb{Y}_1), \ldots, (\mathbb{X}_m, \mathbb{Y}_m)$.

We note that the estimators $\hat{\boldsymbol{\theta}}_m$ and $\hat{\mathbf{C}}_m$ do not need to be well defined with probability 1 for every m; it is enough if they exist with asymptotic probability 1 as $m \to \infty$. The following statements on $\hat{\mathbf{C}}_m$ hold in the same sense, with asymptotic probability 1 as $m \to \infty$. Based on Assumption 2.1, the matrices $\mathbf{C}_0$ and $\hat{\mathbf{C}}_m$ are positive semidefinite, which implies that they have unique square roots $\mathbf{C}_0^{1/2}$ and $\hat{\mathbf{C}}_m^{1/2}$ among positive semidefinite matrices. Also, assumption (viii) ensures that the estimator $\hat{\mathbf{C}}_m$ is nonsingular with asymptotic probability 1, meaning that $\hat{\mathbf{C}}_m^{1/2}$ is invertible in the same sense.

In Subsection 2.3 we show examples of the considered model along with some remarks on how to check the introduced assumptions.

Similar to Horváth et al. (2004, 2007), Aue et al. (2006), and Kirch and Tadjuidje Kamgaing (2011), we consider the weight function

$$g_\gamma(m,k) = m^{1/2}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^\gamma, \qquad m,k \in \mathbb{Z}_{++},$$

where $\gamma \in [0, 1/2)$ is an arbitrary tuning parameter, and introduce the random vectors

$$\mathbb{S}_{m,k} := \hat{\mathbf{C}}_m^{-1/2}\frac{\sum_{n=m+1}^{m+k}\hat{\mathbb{U}}_{m,n} - \frac{k}{m}\sum_{n=1}^{m}\hat{\mathbb{U}}_{m,n}}{g_\gamma(m,k)}, \qquad m,k \in \mathbb{Z}_{++}. \tag{2.2}$$

Our main result is stated in the following theorem, where $\mathcal{W}(t) = [W_1(t), \ldots, W_r(t)]^\top$, $t \geq 0$, is an r-dimensional standard Wiener process. Here and throughout the article we use the convention $0/0 := 0$, and for $T = \infty$ let $T/(T+1) := 1$.

Theorem 2.1. *Suppose that the sequence* $(\mathbb{X}_n, \mathbb{Y}_n)$, $n = 1, 2, \ldots$, *satisfies (2.1) and the noncontamination assumption. If Assumption 2.1 holds, implying that* $\mathcal{H}_0$ *is true for every* $m \in \mathbb{Z}_{++}$, *then for any continuous function* $\psi : \mathbb{R}^r \to \mathbb{R}$ *and for any* $T \in (0, \infty]$ *we have the convergence*

$$\sup_{1\leq k\leq \lfloor Tm\rfloor} \psi(\mathbb{S}_{m,k}) \xrightarrow{\mathcal{D}} \sup_{0\leq t\leq T/(T+1)} \psi(\mathcal{W}(t)/t^\gamma), \qquad m \to \infty.$$

Let us note that by the law of the iterated logarithm, the process $\mathcal{W}(t)/t^\gamma$ is sample continuous on the interval $[0,1]$. This implies that the limit in Theorem 2.1 is a finite random variable. As a result, the null hypothesis $\mathcal{H}_0$ can be tested as described in Section 1 by using the statistics $\tau_{m,k} = \psi(\mathbb{S}_{m,k})$. In the next theorem, we present three examples for such statistics, which can be obtained by using the scaling property of the Wiener process with the norm-like functions

$$\psi_1(\mathbf{y}) = ||\mathbf{y}||, \qquad \psi_2(\mathbf{y}) = \max_{1 \le i \le r} |y_i|, \qquad \psi_3(\mathbf{y}) = |\mathbf{c}^\top \mathbf{y}|, \tag{2.3}$$

where $\mathbf{y} = [y_1, ..., y_r]^\top, \mathbf{c} \in \mathbb{R}^r$. The variables $S_{m,k,1}, ..., S_{m,k,r}$ stand for the components of the random vector $\mathbb{S}_{m,k}$.

Theorem 2.2. *Suppose that the conditions of Theorem 2.1 hold. Then for arbitrary constants $T \in (0, \infty]$ and $\mathbf{c} \in \mathbb{R}^r$ we have that*

$$\sup_{1 \le k \le \lfloor Tm \rfloor} ||\mathbb{S}_{m,k}|| \xrightarrow{\mathcal{D}} \left(\frac{T}{1+T}\right)^{1/2-\gamma} \sup_{0 \le t \le 1} \frac{||\mathcal{W}(t)||}{t^\gamma},$$

$$\sup_{1 \le k \le \lfloor Tm \rfloor} \max_{1 \le i \le r} |S_{m,k,i}| \xrightarrow{\mathcal{D}} \left(\frac{T}{1+T}\right)^{1/2-\gamma} \max_{1 \le i \le r} \sup_{0 \le t \le 1} \frac{|W_i(t)|}{t^\gamma},$$

$$\sup_{1 \le k \le \lfloor Tm \rfloor} |\mathbf{c}^\top \mathbb{S}_{m,k}| \xrightarrow{\mathcal{D}} \left(\frac{T}{1+T}\right)^{1/2-\gamma} ||\mathbf{c}|| \sup_{0 \le t \le 1} \frac{|W_1(t)|}{t^\gamma},$$

as $m \to \infty$.

We omit the proof of this simple theorem. The main advantage of the three tests based on the functions in (2.3) is that the critical values corresponding to the closed-ended case can be easily calculated from the critical value x_α of the open-ended test in the form $(T/(1+T))^{1/2-\gamma} x_\alpha$. Also note that the limit variables are continuous, which implies that there exist asymptotically correct critical values for any significance level $\alpha \in (0, 1)$. The test based on the function ψ_1 is the classical one introduced by Chu et al. (1996) and investigated by several authors in the last two decades. Horváth et al. (2004) published a table of the critical values in the case $r=1$ based on computer simulation. However, the quantiles of the limit variable $\sup_{0 \le t \le 1} ||\mathcal{W}(t)||/t^\gamma$ are not available for every positive integer r. This fact motivates the second test based on the function ψ_2, having critical values that can be determined by using only the quantiles of the one-dimensional case. Indeed, let x_β be the critical value of the one-dimensional limit process corresponding to the significance level $\beta = 1 - (1-\alpha)^{1/r}$. Then,

$$P\left(\max_{i=1,\dots,r} \sup_{0 \le t \le 1} \frac{|W_i(t)|}{t^\gamma} \le x_\beta\right) = P\left(\sup_{0 \le t \le 1} \frac{|W_1(t)|}{t^\gamma} \le x_\beta\right)^r = (1-\beta)^r = 1 - \alpha,$$

meaning that x_β is the critical value corresponding to the r-dimensional limit process and significance level α. We note that in several applications the components of the statistics $\mathbb{S}_{m,k}$ have different sensitivities for the model change, and a suitable linear combination of them can improve the power of the method. This is the concept of the test corresponding to the function ψ_3.

2.2. Results under the alternative hypothesis

In this subsection, we investigate the test statistics under the alternative hypothesis that there is a single change in the dynamics of the system. To ensure that the noncontamination assumption holds, we consider a sequence of nonnegative integers k_m^*, $m \in \mathbb{Z}_{++}$, and assume that for any m the change happens at the time point $m + k_m^*$. For simplicity, we investigate only the open-ended case, and we assume that the dynamics before and after the change do not depend on the values m and k_m^*. The goal is to show the consistency of the test under some suitable conditions of the model and to investigate the time of rejection as a function of m.

To formalize the model, consider a sequence of $\mathbb{R}^q \times \mathbb{R}^r$-valued observations $(\mathbb{X}_n, \mathbb{Y}_n)$, $n \in \mathbb{Z}_{++}$, satisfying Assumption 2.1, and additionally $\mathbb{R}^q \times \mathbb{R}^r$-valued random pairs $(\mathbb{X}_{m,m+k_m^*+n}, \mathbb{Y}_{m,m+k_m^*+n})$, $m, n \in \mathbb{Z}_{++}$. For a given m we will perform the test based on the sample $(\mathbb{X}_{m,1}, \mathbb{Y}_{m,1}), (\mathbb{X}_{m,2}, \mathbb{Y}_{m,2}), \dots$, where $(\mathbb{X}_{m,n}, \mathbb{Y}_{m,n}) := (\mathbb{X}_n, \mathbb{Y}_n)$ for $n \le m + k_m^*$. As a consequence of this construction, for every m the dynamics of the system does not change before the $(m + k_m^*)$ th step, and some additional regularity conditions summarized in the next assumption will ensure that after this time point the system follows another dynamics starting from the initial value $(\mathbb{X}_{m,m+k_m^*}, \mathbb{Y}_{m,m+k_m^*})$. To perform the test, we introduce the random vectors

$$\mathbb{U}_{m,n} := \mathbb{Y}_{m,n} - E[\mathbb{Y}_{m,n} \mid \mathbb{X}_{m,n}], \qquad \hat{\mathbb{U}}_{m,n} := \mathbb{Y}_{m,n} - f(\mathbb{X}_{m,n}, \hat{\boldsymbol{\theta}}_m) \qquad m, n \in \mathbb{Z}_{++},$$

and we define $\mathbb{S}_{m,k}$ by formula (2.2).

Assumption 2.2.

i. *The processes* $\{\mathbb{X}_{m,m+k_m^*+n}, n \in \mathbb{Z}_{++}\}$, $m \in \mathbb{Z}_{++}$, *are strictly stationary with the same finite dimensional distributions, or they are positive Harris recurrent Markov chains with the same transition probability kernel. Let* $\tilde{\mathbb{X}}_A$ *be an arbitrary* $\mathbb{R}^q$*-valued random vector whose distribution is the same as the unique stationary distribution of the processes.*

ii. *We have* $E[\mathbb{Y}_{m,n} \mid \mathbb{X}_{m,n}] = f(\mathbb{X}_{m,n}, \boldsymbol{\theta}_{\mathrm{A}})$ *for every integer* $m \ge 1$ *and* $n \ge m + k_m^* + 1$ *with some* $\boldsymbol{\theta}_{\mathrm{A}} \in \Theta_0$ *and with the function* f *introduced in Assumption 2.1.*

iii. *The expectations* $Eh(\tilde{\mathbb{X}}_A)$, $Ef(\tilde{\mathbb{X}}_A, \boldsymbol{\theta}_0)$, $Ef(\tilde{\mathbb{X}}_A, \boldsymbol{\theta}_{\mathrm{A}})$, *and* $E\nabla_{\boldsymbol{\theta}} f_i(\tilde{\mathbb{X}}_A, \boldsymbol{\theta}_0)$, $i = 1, \dots, r$, *are finite, where h is the function defined in (iv) of Assumption 2.1.*

iv. *There exists a positive integer* m_A *such that*

$$v_A := \sup_{m \ge m_A} \sup_{n \ge m + k_m^* + 1} E||\mathbb{U}_{m,n}||^2 < \infty.$$

In this subsection, we work under the alternative hypothesis

$$\mathcal{H}_A : \qquad \Delta := Ef(\tilde{\mathbb{X}}_A, \boldsymbol{\theta}_{\mathrm{A}}) - Ef(\tilde{\mathbb{X}}_A, \boldsymbol{\theta}_0) \neq 0.$$

We will test whether the dynamics of the process $(\mathbb{X}_{m,n}, \mathbb{Y}_{m,n})$, $n \in \mathbb{Z}_{++}$, are unchanged over time under this single change alternative hypothesis by using the test statistics $\tau_{m,k} := \psi(\mathbb{S}_{m,k})$ introduced in Section 1, where $\psi : \mathbb{R}^r \to \mathbb{R}$ is an arbitrary continuous function. With a given critical value, x_α corresponding to a significance level α the time of the first rejection after the $(m + \ell)$ th step is defined by $\kappa_{m,\ell} := \min\{k > \ell : \tau_{m,k} > x_\alpha\}$. In particular, for every m, the variables $\kappa_{m,0}$ and κ_{m,k_m^*} stand for the first time of rejection after the last element of the training sample and after the time of the actual model change, respectively. The following result is motivated by the similar theorems of Horváth et al. (2004) and Aue et al. (2006) stated for their linear regression models.

Theorem 2.3. *Assume that Assumptions 2.1 and 2.2 and the alternative hypothesis $\mathcal{H}_A$ are satisfied, and $\lim_{||\mathbf{x}||\to\infty}\psi(\mathbf{x})=\infty$.*

i. *For any sequence k_m^* of nonnegative integers we have $\kappa_{m,k_m^*}-k_m^*=o_P(m+k_m^*)$ as $m\to\infty$. It is a direct consequence that the related test is consistent.*
ii. *If $k_m^*=\lfloor cm^b\rfloor$ for every m with some constants $b,c\geq 0$, then $\kappa_{m,k_m^*}-k_m^*=O_P(m^\beta)$, where*

$$\beta=\begin{cases}(1-2\gamma)/(2-2\gamma), & 0\leq b\leq(1-2\gamma)/(2-2\gamma),\\ 1/2-\gamma(1-b), & (1-2\gamma)/(2-2\gamma)<b\leq 1,\\ b-1/2, & 1<b.\end{cases}$$

Let us note that the functions ψ_1 and ψ_2 defined by (2.3) satisfy the conditions of the theorem, which means that the results of statements (i) and (ii) are valid for the related tests. Although the limit $\lim_{||\mathbf{x}||\to\infty}\psi_3(\mathbf{x})$ does not exist, we show in Remark 3.1 after the proof of the latter theorem that with some minor changes in the calculations one can obtain the same rates for the function ψ_3 under the additional assumption that $\mathbf{c}^\top\mathbf{C}_0^{-1/2}\Delta\neq 0$.

In Theorem 2.3, we examined the first time of rejection after the model change. However, in the applications we may meet false alarms, when the test detects the change of the model too early, before the actual time of the change, $m+k_m^*$. Using our notations, the false alarm is the event $\{\kappa_{m,0}\leq k_m^*\}$. In our last result, we examine the asymptotic probability of this event.

Theorem 2.4. *Assume that Assumption 2.1 is satisfied and consider any of the three testing methods of Theorem 2.2. If $k_m^*=\lfloor cm^b\rfloor$ for every m with some constants $b\geq 0$ and $c>0$, then*

$$P(\kappa_{m,0}\leq k_m^*)\to\begin{cases}0, & b<1,\\ \alpha^*, & b=1,\\ \alpha, & b>1,\end{cases}$$

where $\alpha^*\in(0,\alpha)$.

2.3. Some general remarks and examples

Let us present some ideas how to check the conditions of Assumption 2.1 in applications. In most cases, condition (i) has to be verified based on *a priori* information on the model. Positive Harris recurrence is already proved for many discrete-time Markov chains, which can be shown along with (v) by using the Foster–Lyapunov criteria (14.3) in chapter 14 of Meyn and Tweedie (2009). In the simple case when the process $\mathbb{X}_n$, $n\in\mathbb{Z}_{++}$, has countable state space, (i) of Assumption 2.1 holds if the process has exactly one positive recurrent class and it is aperiodic and reached within finitely many steps starting from any initial distribution with probability 1.

Assumptions (iii) and (iv) are analytical conditions, which must be checked by standard calculations. We note that these conditions are satisfied with $a=1$ and $h(\mathbf{x})=\max_{i=1,\dots,r}\sup_{\boldsymbol{\theta}\in\Theta}||\nabla^2_{\boldsymbol{\theta}}f_i(\mathbf{x},\boldsymbol{\theta})||$ if the function f is twice continuously differentiable with respect to $\boldsymbol{\theta}$ on $\mathbb{R}^q\times\Theta_0$. In many applications, we find models where the function is linear in the form $f(\mathbf{x},\mathbf{A}) = \mathbf{A}\mathbf{x}$, $\mathbf{x} \in \mathbb{R}^q$, with coefficient and parameter $\mathbf{A} \in \mathbb{R}^{r\times q}$. Although this model is not parameterized by vectors, is has a natural reparameterization by using $\boldsymbol{\theta} = \boldsymbol{\theta}(\mathbf{A}) \in \mathbb{R}^{\mathrm{rq}}$ defined as the the vector of the columns of $\mathbf{A}$. The partial derivatives of the function $\mathbf{Ax}$ are linear and do not depend on $\mathbf{A}$, which implies that (iv) holds with $h = 0$. As a consequence of these, in this linear case (v) is satisfied if the variable $\tilde{\mathbb{X}}_0$ has finite mean.

Note that (viii) of Assumption 2.1 is required because we would like to use the martingale central limit theorem. By theorem 3.33 in chapter VIII of Jacod and Shiryaev (2003), under (vii) of Assumption 2.1 the conditions of (viii) of Assumption 2.1 are equivalent. In many applications, the martingale differences $\mathbb{U}_n$, $n \in \mathbb{Z}_{++}$, are independent and identically distributed (i.i.d.), then (viii) of Assumption 2.1 is satisfied with $\mathbf{C}_0 := E(\mathbb{U}_1\mathbb{U}_1^\top)$ by the law of large numbers.

For certain models, the matrix $\mathbf{C}_0$ is singular. The matrix $\mathbf{C}_0$ is the limit of covariance matrices. Therefore, the singularity of this matrix indicates that asymptotically the components of $\mathbb{U}_n$ are linearly dependent, meaning that some components can be expressed as the linear combinations of others. In such cases, it can help to remove the corresponding components of the process $\mathbb{Y}_n$, $n \in \mathbb{Z}_{++}$. Then, the matrix $\mathbf{C}_0$ related to this modified process possibly becomes non singular.

The method to estimate the parameter $\boldsymbol{\theta}$ depends on the concrete model. Possible estimations are the least squares, conditional least squares (CLS), weighted conditional least squares (WCLS), maximum likelihood, or Yule-Walker. Note that if we apply the CLS estimation for $\boldsymbol{\theta}$, and for every $1 \le i \le r$ the function $\nabla_{\boldsymbol{\theta}} f_i(\mathbf{x},\boldsymbol{\theta})$ has a constant, non-zero component, then the statistic $\mathbb{S}_{m,k}$ reduces to

$$\mathbb{S}_{m,k} = \hat{\mathbf{C}}_m^{-1/2}\frac{\sum_{n=m+1}^{m+k}\hat{\mathbb{U}}_{m,n}}{g_\gamma(m,k)}, \qquad m,k \in \mathbb{Z}_{++}.$$

In some cases, $\mathbf{C}_0 = \mathbf{C}_0(\boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$. Then, $\hat{\mathbf{C}}_m := \mathbf{C}_0(\hat{\boldsymbol{\theta}}_m)$ is a weakly consistent estimator of $\mathbf{C}_0$.

2.3.1. Regression and autoregressive models

Consider the model $\xi_n = \phi(\boldsymbol{\zeta}_{\mathrm{n}},\boldsymbol{\theta}) + \eta_{\mathrm{n}}$, $\mathrm{n} \in \mathbb{Z}_{++}$, where $\phi : \mathbb{R}^q \times \boldsymbol{\Theta} \to \mathbb{R}$ and $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \ldots$ is a sequence of $\mathbb{R}^q$-valued input variables. Furthermore, $\eta_1, \eta_2, \ldots$ are error terms with mean 0 and variance σ^2, independent of the previous sequence. In this model, we can test the change of the parameter $\boldsymbol{\theta}$ by using Theorem 2.1 with the setup $\mathbb{X}_n = \boldsymbol{\zeta}_{\mathrm{n}}$, $\mathbb{Y}_n = \xi_n$, $f(\mathbf{x},\boldsymbol{\theta}) = \phi(\mathbf{x},\boldsymbol{\theta})$, and $\mathbb{U}_n = \eta_n = \xi_n - \phi(\boldsymbol{\zeta}_{\mathrm{n}},\boldsymbol{\theta})$. Also, we can test the change of both $\boldsymbol{\theta}$ and σ with $\mathbb{X}_n = \boldsymbol{\zeta}_{\mathrm{n}}$, $\mathbb{Y}_n = [\xi_n, \eta_n^2]^\top$,

$$f(\mathbf{x},\boldsymbol{\theta},\sigma) = \begin{bmatrix}\phi(\mathbf{x},\boldsymbol{\theta})\\ \sigma^2\end{bmatrix}, \qquad \mathbb{U}_{\mathrm{n}} = \begin{bmatrix}\eta_{\mathrm{n}}\\ \eta_{\mathrm{n}}^2 - \sigma^2\end{bmatrix} = \begin{bmatrix}\xi_{\mathrm{n}} - \phi(\boldsymbol{\zeta}_{\mathrm{n}},\boldsymbol{\theta})\\ [\xi_{\mathrm{n}} - \phi(\boldsymbol{\zeta}_{\mathrm{n}},\boldsymbol{\theta})]^2 - \sigma^2\end{bmatrix}.$$

Although in the applications the exact values of the error terms are not available, the test can be performed without this information. Because $\mathbb{U}_n$ can be represented as a function of the parameters and the known pair $(\boldsymbol{\zeta}_{\mathrm{n}}, \xi_n)$, the variables $\hat{\mathbb{U}}_{m,n}$ can be written up by using some estimators $\hat{\boldsymbol{\theta}}_m$ and $\hat{\sigma}_m$ based on the real observations $(\boldsymbol{\zeta}_1, \xi_1), \dots, (\boldsymbol{\zeta}_m, \xi_m)$.

If $\boldsymbol{\zeta}_n = [\xi_{n-1}, \dots, \xi_{n-q}]^\top$ for every $n \in \mathbb{Z}_{++}$ with some $q \in \mathbb{Z}_{++}$ and initial vector $[\xi_0, \dots, \xi_{1-q}]$, then ξ_n, $n \in \mathbb{Z}_{++}$, is an autoregressive process that behaves similar to the regression model in terms of the above-described method.

One can consider, for example, the least squares, conditional least squares, or Yule-Walker method to obtain applicable estimators.

2.3.2. Homogeneity of independent observations

Consider independent random variables $\xi_0, \xi_1, \dots$ coming from a parametric family parameterized by $\boldsymbol{\theta}$. We can test the change of this parameter with the setup $\mathbb{X}_n = \xi_{n-1}$, $\mathbb{Y}_n = [\phi_1(\xi_n), \dots, \phi_r(\xi_n)]^\top$,

$$f(\mathbf{x}, \boldsymbol{\theta}) = \mathrm{f}(\boldsymbol{\theta}) = \begin{bmatrix} \mathrm{E}_{\boldsymbol{\theta}}\phi_1(\xi_1) \\ \vdots \\ \mathrm{E}_{\boldsymbol{\theta}}\phi_\mathrm{r}(\xi_1) \end{bmatrix}, \qquad \mathbb{U}_\mathrm{n} = \begin{bmatrix} \phi_1(\xi_\mathrm{n}) - \mathrm{E}_{\boldsymbol{\theta}}\phi_1(\xi_1) \\ \vdots \\ \phi_\mathrm{r}(\xi_\mathrm{n}) - \mathrm{E}_{\boldsymbol{\theta}}\phi_\mathrm{r}(\xi_1) \end{bmatrix},$$

where $\phi_1, \dots, \phi_r : \mathbb{R} \to \mathbb{R}$ are arbitrary such that $f(\boldsymbol{\theta})$ exists. Choose functions $\phi_1, \dots, \phi_r$ that characterize the parameter $\boldsymbol{\theta}$ by a resulting bijective $f(\boldsymbol{\theta})$ function. Then, a change of $f(\boldsymbol{\theta})$ is equivalent to a change in the parameter $\boldsymbol{\theta}$ itself.

Now assume that $\xi_0, \xi_1, \dots$ are independent but not necessarily from a parametric family. Again, consider the same setup for $\mathbb{X}_n$, $\mathbb{Y}_n$, and some functions $\phi_1, \dots, \phi_r : \mathbb{R} \to \mathbb{R}$. Then we can test for a change in the parameter

$$f(\mathbf{x}, \boldsymbol{\theta}) := \boldsymbol{\theta} := \begin{bmatrix} \mathrm{E}\phi_1(\xi_1) \\ \vdots \\ \mathrm{E}\phi_\mathrm{r}(\xi_1) \end{bmatrix}.$$

For example, one can test for a change in the first r moments of the variables by choosing the functions $\phi_1(x) = x, \dots, \phi_r(x) = x^r$.

2.3.3. Multitype Galton–Watson processes

Consider a positive integer p and a random or deterministic, $\mathbb{Z}_+^p$-valued vector $\boldsymbol{\xi}_0$. The $\mathbb{Z}_+^p$-valued process $\boldsymbol{\xi}_\mathrm{n} = [\xi_{n,1}, \dots, \xi_{n,p}]^\top$, $n \in \mathbb{Z}_+$, is a multitype Galton–Watson process if it can be represented in the form

$$\boldsymbol{\xi}_\mathrm{n} = \sum_{k=1}^{\xi_{n-1,1}} \boldsymbol{\zeta}_1(n,k) + \cdots + \sum_{k=1}^{\xi_{n-1,p}} \boldsymbol{\zeta}_\mathrm{p}(n,k) + \boldsymbol{\eta}(\mathrm{n}), \qquad \mathrm{n} \in \mathbb{Z}_{++},$$

where

$$\boldsymbol{\xi}_0,\ \boldsymbol{\zeta}_\mathrm{i}(n,k),\ \boldsymbol{\eta}(\mathrm{n}), \qquad \mathrm{k}, \mathrm{n} \in \mathbb{Z}_{++}, \qquad \mathrm{i} = 1, \dots, \mathrm{p},$$

are $\mathbb{Z}_+^p$-valued random vectors being independent of each other, and the offspring variables $\boldsymbol{\zeta}_{\mathrm{i}}(n,k)$, $k \in \mathbb{Z}_{++}$, are identically distributed for every i and n.

Our goal is to test whether the distributions of the offsprings and the innovations are unchanged over time. For this goal, we consider two tests. With the first one, we test whether the means of the distributions are unchanged. With the second one, we test whether both the means and variances are unchanged. Under the null hypothesis, we refer to the offspring and innovation distributions by $\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_{\mathrm{p}}, \boldsymbol{\eta}$, because their distributions do not depend on the parameters n and k. Also, we introduce the matrix

$$\mathbf{M} := [E\boldsymbol{\zeta}_1, \ldots, E\boldsymbol{\zeta}_{\mathrm{p}}, E\boldsymbol{\eta}] \in \mathbb{R}^{\mathrm{p}\times(\mathrm{p}+1)}$$

and we define the first test by setting

$$\mathbb{X}_n := \begin{bmatrix} \boldsymbol{\xi}_{\mathrm{n}-1} \\ 1 \end{bmatrix} = [\xi_{n-1,1}, \ldots, \xi_{n-1,p}, 1]^\top, \qquad \mathbb{Y}_n := \boldsymbol{\xi}_{\mathrm{n}}, \qquad n \in \mathbb{Z}_{++},$$

resulting in $f(\mathbf{x}, \mathbf{M}) = \mathbf{M}\mathbf{x}$ and $\mathbb{U}_n = \boldsymbol{\xi}_{\mathrm{n}} - \mathbf{M}[\boldsymbol{\xi}_{\mathrm{n}-1}^\top, 1]^\top$.

For the second test, under the null hypothesis we consider the matrix

$$\mathbf{V} := [D^2\boldsymbol{\zeta}_1, \ldots, D^2\boldsymbol{\zeta}_{\mathrm{p}}, D^2\boldsymbol{\eta}] \in \mathbb{R}^{\mathrm{p}\times(\mathrm{p}+1)},$$

where the variance of a vector is understood componentwise. Then, by the results of Nedényi (2015), one can test the change of $(\mathbf{M}, \mathbf{V})$ by the setup

$$\mathbb{X}_n = \begin{bmatrix} \boldsymbol{\xi}_{\mathrm{n}-1} \\ 1 \end{bmatrix}, \qquad \mathbb{Y}_n = \begin{bmatrix} \boldsymbol{\xi}_{\mathrm{n}} \\ (\boldsymbol{\xi}_{\mathrm{n}} - \mathbf{M}\mathbb{X}_n)^2 \end{bmatrix}, \qquad f(\mathbf{x}, \mathbf{M}, \mathbf{V}) = \begin{bmatrix} \mathbf{M} \\ \mathbf{V} \end{bmatrix} \mathbf{x}.$$

Then, $\mathbb{U}_n = [(\boldsymbol{\xi}_{\mathrm{n}} - \mathbf{M}\mathbb{X}_n)^\top, ((\boldsymbol{\xi}_{\mathrm{n}} - \mathbf{M}\mathbb{X}_n)^2 - \mathbf{V}\mathbb{X}_n)^\top]^\top$. We suggest applying the CLS and WCLS methods to achieve the necessary parameter estimators in both cases. The estimators are detailed in Nedényi (2015).

3. Proofs

Lemma 3.1. *Consider a measurable set $S \subseteq \mathbb{R}^q$ and an array of S-valued random vectors with rows $\{\mathbb{M}_{m,0}, \mathbb{M}_{m,1}, \ldots\}$, $m \in \mathbb{Z}_{++}$, that satisfies any of the following assumptions:*

i. *The rows of the array are strictly stationary ergodic processes with the same finite dimensional distributions.*
ii. *The rows are positive Harris recurrent Markov chains with the same probability transition kernel. Furthermore, the process of the initial values $\{\mathbb{M}_{m,0} : m \in \mathbb{Z}_{++}\}$ is strictly stationary or it is an aperiodic positive Harris recurrent Markov chain.*

In both cases, let π denote the unique stationary distribution of the rows. Consider a measurable function $\phi : S \to \mathbb{R}^r$ such that $\int_S \|\phi(\mathbf{x})\| \pi(d\mathbf{x}) < \infty$, and introduce

$$\mathbb{A}_{m,k} := \frac{1}{k}\sum_{n=1}^{k} \phi(\mathbb{M}_{m,n}) - \int_S \phi(\mathbf{x})\pi(d\mathbf{x}), \qquad m, k \in \mathbb{Z}_{++}.$$

Then, for any real sequence a_m tending to infinity, we have $\sup_{k\geq a_m} \|\mathbb{A}_{m,k}\| = o_P(1)$ and $\sup_{k\geq 1} \|\mathbb{A}_{m,k}\| = O_P(1)$ as $m \to \infty$.

Proof. If the array satisfies condition (i), then for any m we have

$$\frac{1}{k}\sum_{n=1}^{k}\phi(\mathbb{M}_{m,n}) \overset{\mathcal{D}}{=} \frac{1}{k}\sum_{n=1}^{k}\phi(\mathbb{M}_{1,n}) \to \int_S \phi(\mathbf{x})\pi(d\mathbf{x}), \qquad k\to\infty,$$

where the convergence holds with probability 1, proving both statements. In the remaining of the proof we show that the statements are true under assumption (ii) as well.

Let π' stand for the unique stationary distribution of the process $\mathbb{M}_{m,0}$, $m\in\mathbb{Z}_{++}$, and let p_m denote the distribution of the random vector $\mathbb{M}_{m,0}$. If the initial values form an aperiodic positive Harris recurrent Markov chain, then by theorem 13.0.1 of Meyn and Tweedie (2009) the transition probabilities of the chain converge to the stationary distribution in the total variation metric. From this we obtain that

$$\sup_{B\in\mathcal{B}(S)}|p_m(B)-\pi'(B)| \le \int_S \sup_{B\in\mathcal{B}(S)}|P(\mathbb{M}_{m,0}\in B\,|\,\mathbb{M}_{1,0}=\mathbf{x})-\pi'(B)|\,p_1(d\mathbf{x}) \to 0, \tag{3.1}$$

as $m\to\infty$. Note that the convergence in (3.1) is obvious if the process $\mathbb{M}_{m,0}$, $m\in\mathbb{Z}_{++}$, is strictly stationary. Also, theorem 17.0.1 of Meyn and Tweedie (2009) implies the "law of large numbers" $\mathbb{A}_{1,k}\to 0$, $k\to\infty$, in case of any distribution p_1, where the convergence is understood in an almost sure sense. Hence, we have $\sup_{k\ge a_m}\mathbb{A}_{1,k}\to_P 0$ as $m\to\infty$ on the event $\{\mathbb{M}_{1,0}=\mathbf{x}\}$ in case of an arbitrary $\mathbf{x}\in S$. This implies the convergence

$$\rho_m(\mathbf{x},\delta) := P(\sup_{k\ge a_m}||\mathbb{A}_{1,k}||>\delta\,|\,\mathbb{M}_{1,0}=\mathbf{x}) \to 0, \qquad m\to\infty,$$

for any fixed value $\delta>0$. Note that by the Markov property

$$P(\sup_{k\ge a_m}||\mathbb{A}_{1,k}||>\delta\,|\,\mathbb{M}_{1,0}=\mathbf{x}) = P(\sup_{k\ge a_m}||\mathbb{A}_{m,k}||>\delta\,|\,\mathbb{M}_{m,0}=\mathbf{x}), \quad m\in\mathbb{Z}_{++},$$

for every $\mathbf{x}\in S$. By using this consequence of the Markov property and the dominated convergence it follows that

$$\begin{aligned}
P\Big(\sup_{k\ge a_m}||\mathbb{A}_{m,k}||>\delta\Big) &= \int_S \rho_m(\mathbf{x},\delta)p_m(d\mathbf{x})\\
&\le \left|\int_S \rho_m(\mathbf{x},\delta)(p_m-\pi')(d\mathbf{x})\right| + \int_S \rho_m(\mathbf{x},\delta)\pi'(d\mathbf{x})\\
&\le \sup_{\mathbf{x}\in S}\rho_m(\mathbf{x},\delta)\sup_{B\in\mathcal{B}(S)}|p_m(B)-\pi'(B)| + \int_S \rho_m(\mathbf{x},\delta)\pi'(d\mathbf{x}) \to 0,
\end{aligned}$$

as $m\to\infty$.

For the second statement, let us recall that $\mathbb{A}_{1,k}\to 0$, $k\to\infty$, almost surely, which implies that the sequence $\mathbb{A}_{1,k}$, $k\in\mathbb{Z}_{++}$, is bounded stochastically. From this we get the convergence

$$\rho(\mathbf{x},c) := P(\sup_{k\ge 1}||\mathbb{A}_{1,k}||>c\,|\,\mathbb{M}_{1,0}=\mathbf{x}) \to 0, \qquad c\to\infty,$$

for any $\mathbf{x}\in S$. Because $\rho(\mathbf{x},c)$ is a measurable function of the variable $\mathbf{x}$ in case of any fixed $c>0$, the sets

$$S(c) = \{\mathbf{x}\in S : \rho(\mathbf{x},c)\le\varepsilon/3\}, \qquad c>0,$$

form an increasing system of measurable subsets of S with limit set $\cup_{c>0}S(c) = S$ for every $\varepsilon > 0$. This implies that there exists $c_0 > 0$ such that $\pi'(S(c_0)) \geq 1 - \varepsilon/3$ and $\sup_{\mathbf{x}\in S(c_0)} \rho(\mathbf{x}, c_0) \leq \varepsilon/3$. By using the Markov property, we obtain the inequalities

$$\begin{aligned} P\Big(\sup_{k\geq 1}||\mathbb{A}_{m,k}|| > c_0\Big) &= \int_S \rho(\mathbf{x}, c_0)p_m(d\mathbf{x}) \\ &\leq \left|\int_S \rho(\mathbf{x}, c_0)(p_m - \pi')(d\mathbf{x})\right| + \int_{S(c_0)} \rho(\mathbf{x}, c_0)\pi'(d\mathbf{x}) + \int_{S\setminus S(c_0)} \rho(\mathbf{x}, c_0)\pi'(d\mathbf{x}) \\ &\leq \sup_{\mathbf{x}\in S}\rho(\mathbf{x}, c_0) \sup_{B\in\mathcal{B}(S)} |p_m(B) - \pi'(B)| + \varepsilon/3 + \varepsilon/3. \end{aligned}$$

Because the first term converges to 0 by (3.1), it follows that $P(\sup_{k\geq 1} ||\mathbb{A}_{m,k}|| > c_0) \leq \varepsilon$ if m is large enough, completing the proof of the second statement. □

For every positive integer m, consider the processes

$$\hat{\mathcal{X}}_m(t) := \frac{\sum_{n=m+1}^{m+\lfloor tm \rfloor} \hat{\mathbb{U}}_{m,n} - \frac{\lfloor tm \rfloor}{m}\sum_{n=1}^{m} \hat{\mathbb{U}}_{m,n}}{g_\gamma(m, \lfloor tm \rfloor)}, \qquad \mathcal{X}(t) := \mathbf{C}_0^{1/2}\frac{\mathcal{W}(\frac{t}{1+t})}{(\frac{t}{1+t})^\gamma}, \qquad t \geq 0,$$

and let $\mathcal{X}_m$ be the theoretical counterpart of $\hat{\mathcal{X}}_m$, which is obtained by replacing the vectors $\hat{\mathbb{U}}_{m,n}$ by $\mathbb{U}_n$, respectively. The processes $\mathcal{X}_m$ and $\hat{\mathcal{X}}_m$ are random elements of the Skorokhod space $\mathcal{D}^r[0, \infty)$ of $\mathbb{R}^r$-valued càdlàg functions defined on $[0, \infty)$. (For the topology of $\mathcal{D}^r[0, \infty)$, see chapter VI of Jacod and Shiryaev [2003] or see section 16 of Billingsley [1999] for the case $r = 1$.) Additionally, the law of the iterated logarithm implies that $\mathcal{X}$ is a random element of the space $\mathcal{C}^r[0, \infty) \subseteq \mathcal{D}^r[0, \infty)$ of continuous functions.

The theoretical base of our main results is the fact that the process $\hat{\mathcal{X}}_m$ converges in distribution to $\mathcal{X}$ in $\mathcal{D}^r[0, \infty)$ if Assumption 2.1 is satisfied. This convergence is a direct consequence of Lemmas 3.2 and 3.3 stated below. We note that under some additional regularity conditions one can also construct copies $\mathcal{X}^{(1)}, \mathcal{X}^{(2)}, \ldots$ of the process $\mathcal{X}$ such that $\sup_{t\geq 0} ||\hat{\mathcal{X}}_m(t) - \mathcal{X}^{(m)}(t)|| \to_P 0$ as $m \to \infty$. This stronger tool was used by Horváth et al. (2004), Aue et al. (2006), and Kirch and Tadjuidje Kamgaing (2011) to prove results similar to those of our Theorems 2.1 and 2.3. □

Lemma 3.2. *If (i)–(vi) of Assumption 2.1 hold, then* $\sup_{t\geq 0}||\hat{\mathcal{X}}_m(t) - \mathcal{X}_m(t)||\to_P 0$ *as* $m \to \infty$.

Proof. Consider Θ_0, an open sphere with center $\boldsymbol{\theta}_0$. Because $\hat{\boldsymbol{\theta}}_m$ is a weakly consistent estimator of $\boldsymbol{\theta}_0$ by (vi) of Assumption 2.1, we have $P(\hat{\boldsymbol{\theta}}_m \in \Theta_0) \to 1$ as $m \to \infty$. Our goal is to prove a stochastic convergence, which means that we can condition on the event $\{\hat{\boldsymbol{\theta}}_m \in \Theta_0\}$ for every m. We will often use the inequalities

$$g_\gamma(m, k) = m^{1/2}\left(1 + \frac{k}{m}\right)\left(\frac{k}{m+k}\right)^\gamma \geq \begin{cases} c_\gamma m^{1/2-\gamma}k^\gamma, & k \leq m, \\ c_\gamma m^{-1/2}k, & k > m, \end{cases}$$

where c_γ is a suitable positive constant not depending on m and k.

Because the lemma follows from the stochastic convergence of the suprema of the norms of the components of the process $\hat{\mathcal{X}}_m(t) - \mathcal{X}(t)$, $t \geq 0$, it is enough to prove the statement for $r = 1$. Because $\hat{\mathcal{X}}_m$ and $\mathcal{X}_m$ are step functions defined on the same partition, we must show that

$$\sup_{k\geq 1}\frac{\left|\left(\sum_{n=m+1}^{m+k}\hat{\mathbb{U}}_{m,n}-\frac{k}{m}\sum_{n=1}^{m}\hat{\mathbb{U}}_{m,n}\right)-\left(\sum_{n=m+1}^{m+k}\mathbb{U}_n-\frac{k}{m}\sum_{n=1}^{m}\mathbb{U}_n\right)\right|}{g_\gamma(m,k)}=o_P(1)\quad(3.2)$$

as $m\to\infty$. From (iii) of Assumption 2.1, it follows that for each m and n there exists a parameter $\boldsymbol{\theta}_{\mathrm{m,n}}\in\Theta$ such that $||\boldsymbol{\theta}_{\mathrm{m,n}}-\boldsymbol{\theta}_0||\leq||\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}_0||$ and

$$\begin{aligned}\hat{\mathbb{U}}_{m,n}-\mathbb{U}_n&=f(\mathbb{X}_n,\boldsymbol{\theta}_0)-f(\mathbb{X}_n,\hat{\boldsymbol{\theta}}_m)=(\boldsymbol{\theta}_0-\hat{\boldsymbol{\theta}}_m)^\top\nabla_{\boldsymbol{\theta}}f(\mathbb{X}_n,\boldsymbol{\theta}_{\mathrm{m,n}})\\&=(\boldsymbol{\theta}_0-\hat{\boldsymbol{\theta}}_m)^\top[\mathbb{D}_{m,n}+\phi(\mathbb{X}_n)+E\nabla_{\boldsymbol{\theta}}f(\tilde{\mathbb{X}}_0,\boldsymbol{\theta}_0)],\end{aligned}$$

where

$$\mathbb{D}_{m,n}=\nabla_{\boldsymbol{\theta}}f(\mathbb{X}_n,\boldsymbol{\theta}_{\mathrm{m,n}})-\nabla_{\boldsymbol{\theta}}f(\mathbb{X}_n,\boldsymbol{\theta}_0),\qquad\phi(\mathbf{x})=\nabla_{\boldsymbol{\theta}}f(\mathbf{x},\boldsymbol{\theta}_0)-E\nabla_{\boldsymbol{\theta}}f(\tilde{\mathbb{X}}_0,\boldsymbol{\theta}_0),\qquad\mathbf{x}\in S.$$

Because $\hat{\boldsymbol{\theta}}_m\in\Theta_0$, we also have $\boldsymbol{\theta}_{\mathrm{m,n}}\in\Theta_0$, and (iv) of Assumption 2.1 implies the inequality $||\mathbb{D}_{m,n}||\leq||\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}_0||^a h(\mathbb{X}_n)$. By (i) of Assumption 2.1, we can apply Lemma 3.1 to the array of random vectors $\{\mathbb{X}_m,\mathbb{X}_{m+1},\dots\}$, $m\in\mathbb{Z}_{++}$, and we get that

$$\begin{aligned}&\sup_{k\geq 1}\frac{\sum_{n=m+1}^{m+k}||\mathbb{D}_{m,n}||}{g_\gamma(m,k)}\leq||\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}_0||^a\sup_{1\leq k\leq m}\left(\frac{k}{m}\right)^{1-\gamma}\frac{\sum_{n=m+1}^{m+k}h(\mathbb{X}_n)}{c_\gamma m^{-1/2}k}\\&+||\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}_0||^a\sup_{k>m}\frac{\sum_{n=m+1}^{m+k}h(\mathbb{X}_n)}{c_\gamma m^{-1/2}k}\leq\frac{2m^{1/2}}{c_\gamma}||\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}_0||^a\sup_{k\geq 1}\frac{\sum_{n=m+1}^{m+k}h(\mathbb{X}_n)}{k}=o_P(m^{1/2}),\end{aligned}$$

as $m\to\infty$. Similarly, from ergodicity it follows that

$$\begin{aligned}&\sup_{k\geq 1}\frac{\frac{k}{m}\sum_{n=1}^{m}||\mathbb{D}_{m,n}||}{g_\gamma(m,k)}\leq||\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}_0||^a\sup_{1\leq k\leq m}\left(\frac{k}{m}\right)^{1-\gamma}\frac{\sum_{n=1}^{m}h(\mathbb{X}_n)}{c_\gamma m^{1/2}}\\&+||\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}_0||^a\sup_{k>m}\frac{\sum_{n=1}^{m}h(\mathbb{X}_n)}{c_\gamma m^{1/2}}\leq\frac{2m^{1/2}}{c_\gamma}||\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}_0||^a\frac{\sum_{n=1}^{m}h(\mathbb{X}_n)}{m}=o_P(m^{1/2}),\end{aligned}$$

as $m\to\infty$. Using (v) of Assumption 2.1 and the same steps as in the last formula, one can also show that

$$\sup_{k\geq 1}\frac{\frac{k}{m}||\sum_{n=1}^{m}\phi(\mathbb{X}_n)||}{g_\gamma(m,k)}\leq\frac{2m^{1/2}}{c_\gamma}\frac{||\sum_{n=1}^{m}\phi(\mathbb{X}_n)||}{m}=o_P(m^{1/2}),\qquad m\to\infty.$$

Finally, from Lemma 3.1 with $a_m = m^{1/2}$, it follows that

$$\sup_{k\ge 1} \frac{||\sum_{n=m+1}^{m+k} \phi(\mathbb{X}_n)||}{g_\gamma(m,k)} \le \sup_{1\le k\le m^{1/2}} \left(\frac{k}{m}\right)^{1-\gamma} \frac{\sum_{n=m+1}^{m+k} |\phi(\mathbb{X}_n)|}{c_\gamma m^{-1/2} k}$$
$$+ \sup_{m^{1/2}<k\le m} \left(\frac{k}{m}\right)^{1-\gamma} \frac{\sum_{n=m+1}^{m+k} |\phi(\mathbb{X}_n)|}{c_\gamma m^{-1/2} k} + \sup_{k>m} \frac{\sum_{n=m+1}^{m+k} |\phi(\mathbb{X}_n)|}{c_\gamma m^{-1/2} k}$$
$$\le \frac{m^{\gamma/2}}{c_\gamma} \sup_{1\le k\le m^{1/2}} \frac{\sum_{n=m+1}^{m+k} |\phi(\mathbb{X}_n)|}{k} + \frac{2m^{1/2}}{c_\gamma} \sup_{k>m^{1/2}} \frac{\sum_{n=m+1}^{m+k} |\phi(\mathbb{X}_n)|}{k} = o_P(m^{1/2}).$$

By summarizing the last four formulae, we obtain the approximations

$$\sup_{k\ge 1} \frac{|\sum_{n=m+1}^{m+k} (\hat{\mathbb{U}}_{m,n} - \mathbb{U}_n) - k(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}_m)^\top E\nabla_{\boldsymbol{\theta}} f(\tilde{\mathbb{X}}_0, \boldsymbol{\theta}_0)|}{g_\gamma(m,k)} = ||\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0|| o_P(m^{1/2}) = o_P(1),$$

and

$$\sup_{k\ge 1} \frac{|\frac{k}{m}\sum_{n=1}^{m} (\hat{\mathbb{U}}_{m,n} - \mathbb{U}_n) - k(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}_m)^\top E\nabla_{\boldsymbol{\theta}} f(\tilde{\mathbb{X}}_0, \boldsymbol{\theta}_0)|}{g_\gamma(m,k)} = ||\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0|| o_P(m^{1/2}) = o_P(1), \tag{3.3}$$

as $m \to \infty$. From these (3.2) follows, and the proof is complete. □

Lemma 3.3. *If (ii), (vii), and (viii) of Assumption 2.1 hold, then $\mathcal{X}_m \to_{\mathcal{D}} \mathcal{X}$ as $m \to \infty$ in the space $\mathcal{D}^r[0,\infty)$.*

Proof. Our goal is to apply the multivariate martingale central limit theorem (theorem 3.33 in chapter VIII of Jacod and Shiryaev [2003]) to the martingale difference sequences $\{\mathbb{U}_1/m^{1/2}, \mathbb{U}_2/m^{1/2}, \ldots\}, m \in \mathbb{Z}_{++}$. Note that for any values $t, \delta > 0$ we have the convergence

$$\frac{1}{m}\sum_{n=1}^{\lfloor mt\rfloor} E[||\mathbb{U}_n||^2 \ \mathbb{1}_{\{||\mathbb{U}_n|| > \delta m^{1/2}\}} \,|\, \mathcal{F}_{n-1}] \le \frac{1}{\delta^\varepsilon m^{1+\varepsilon/2}} \sum_{n=1}^{\lfloor mt\rfloor} E[||\mathbb{U}_n||^{2+\varepsilon} \,|\, \mathcal{F}_{n-1}] \xrightarrow{P} 0,$$

as $m \to \infty$, because by (vii) of Assumption 2.1 the variable on the right side converges to zero in an L_1 sense. This means that the conditional Lindeberg condition is satisfied, and one can show similarly that (viii) of Assumption 2.1 implies that at least one of conditions $[\gamma_6{}'-D]$ and $[\hat{\gamma}_6{}'-D]$ to the same theorem holds as well. As a result, the martingale central limit theorem can be applied, and it implies the weak convergence of

$$\mathcal{U}_m(t) := m^{-1/2} \sum_{n=1}^{\lfloor mt\rfloor} \mathbb{U}_n, \qquad t \ge 0,$$

to $\mathbf{C}_0^{1/2}\mathcal{W}(t)$, $t \ge 0$, in $\mathcal{D}^r[0,\infty)$ as $m \to \infty$. (Let us recall that $\mathcal{W}$ is an r-dimensional standard Wiener process.) Introduce the processes

$$\mathcal{Y}_m(t) := \frac{1}{m^{1/2}} \left(\sum_{n=m+1}^{m+\lfloor mt \rfloor} \mathbb{U}_n - \frac{\lfloor mt \rfloor}{m} \sum_{n=1}^{m} \mathbb{U}_n \right), \qquad \mathcal{Y}(t) := \mathbf{C}_0^{1/2}(t+1)\mathcal{W}\left(\frac{t}{t+1}\right),$$

defined for $t \geq 0$. From the convergence of $\mathcal{U}_m$, we obtain that

$$\mathcal{Y}_m = \left[\mathcal{U}_m(t+1) - \frac{\lfloor m(t+1) \rfloor}{m}\mathcal{U}_m(1)\right]_{t\geq 0} \xrightarrow{\mathcal{D}} \left[\mathbf{C}_0^{1/2}\mathcal{W}(t+1) - (t+1)\mathbf{C}_0^{1/2}\mathcal{W}(1)\right]_{t\geq 0},$$

as $m \to \infty$. Because the limit is a Gaussian process with the same mean and covariance function as $\mathcal{Y}$, we get that $\mathcal{Y}_m \to_{\mathcal{D}} \mathcal{Y}$ holds in $\mathcal{D}^r[0, \infty)$.

For every positive integer ν, introduce the function

$$\Phi_\nu : \mathcal{D}^r[0,\infty) \times \mathcal{D}[1/\nu, \infty) \to \mathcal{D}^r[0,\infty), \qquad \Phi_\nu(y, w)(t) = y(t)w(t) \ \mathbb{1}_{\{t \geq 1/\nu\}}.$$

By the results in chapter VI of Jacod and Shiryaev (2003), the Borel σ-algebra generated by the Skorokhod topology on the space $\mathcal{D}^r[0,\infty)$ is identical to the σ-algebra generated by the finite dimensional projections, and the convergence to a continuous function in the Skorokhod sense is equivalent to the local uniform convergence. These facts imply that the function Φ_ν is measurable, and it is continuous at the elements of the set $\mathcal{C}^r[0,\infty) \times \mathcal{C}[1/\nu, \infty)$. For the shorter notations, introduce the processes $\mathcal{X}_{m,\nu}(t) := \mathcal{X}_m(t)\mathbb{1}_{\{t\geq 1/\nu\}}$ and $\mathcal{X}_{0,\nu}(t) := \mathcal{X}(t)\mathbb{1}_{\{t\geq 1/\nu\}}$, along with the functions

$$w(t) := \left[(1+t)\left(\frac{t}{1+t}\right)^{\gamma}\right]^{-1}, \qquad w_m(t) := \frac{m^{1/2}}{g_\gamma(m, \lfloor mt \rfloor)} = w\left(\frac{\lfloor mt \rfloor}{m}\right), \qquad t \geq 1/\nu.$$

Because $\mathcal{Y}_m \to_{\mathcal{D}} \mathcal{Y}$ and w_m converges to w uniformly on the interval $[1/\nu, \infty)$, we get that $(\mathcal{Y}_m, w_m) \to_{\mathcal{D}} (\mathcal{Y}, w)$, and using the continuous mapping theorem we get the convergence

$$\mathcal{X}_{m,\nu} = \Phi_\nu(\mathcal{Y}_m, w_m) \xrightarrow{\mathcal{D}} \Phi_\nu(\mathcal{Y}, w) = \mathcal{X}_{0,\nu}, \qquad m \to \infty.$$

Let us recall that by the law of the iterated logarithm we have $\lim_{t\to 0} ||\mathcal{X}(t)|| = 0$ almost surely. This implies that the process $\mathcal{X}_{0,\nu}$ converges to $\mathcal{X}$ in the supremum distance with probability 1 as $\nu \to \infty$, resulting in convergence of the distributions as well.

To finish the proof of the statement, we only need to show that the processes $\mathcal{X}_{m,\nu}$ are uniformly close to $\mathcal{X}_m$. Let $U_{n,1}, \ldots, U_{n,r}$ stand for the components of the random vector $\mathbb{U}_n$ and note that $U_{1,j}, U_{2,j}, \ldots$ is a martingale difference sequence for every j. Theorem 1 of Chow (1960) states that for a non increasing sequence of positive numbers, $c_1, c_2, \ldots$, a submartingale sequence of random variables, $Z_1, Z_2, \ldots$, and $\varepsilon > 0$, it holds for every $\ell \in \mathbb{Z}_{++}$ that

$$\begin{aligned} \varepsilon P(\max_{1\leq k\leq \ell} c_k Z_k \geq \varepsilon) &\leq \sum_{k=1}^{\ell-1}(c_k - c_{k+1})E(Z_k^+) + c_\ell E(Z_\ell^+) \\ &= c_1 E(Z_1^+) + \sum_{k=2}^{\ell-1} c_k[E(Z_k^+) - E(Z_{k-1}^+)], \end{aligned}$$

where $Z^+ := \max(Z, 0)$ for any random variable Z. For a fixed $m \in \mathbb{Z}_{++}$ and

$j \in \{1, \ldots, r\}$, identify the sequences as $c_k := 1/g_\gamma^2(m,k)$ and $Z_k := \left(\sum_{n=m+1}^{m+k} U_{n,j}\right)^2$, $k \in \mathbb{Z}_{++}$. Because $U_{1,j}, U_{2,j}, \ldots$ is a martingale difference sequence, the sequence Z_k, $k \in \mathbb{Z}_{++}$ is a submartingale. Note that

$$\left\{ \max_{1 \le k \le \lfloor m/\nu \rfloor} \frac{\|\sum_{n=m+1}^{m+k} \mathbb{U}_n\|}{g_\gamma(m,k)} \ge \varepsilon \right\} \subseteq \bigcup_{j=1}^{r} \left\{ \max_{1 \le k \le \lfloor m/\nu \rfloor} \frac{\left(\sum_{n=m+1}^{m+k} U_{n,j}\right)^2}{g_\gamma(m,k)^2} \ge \frac{\varepsilon^2}{r} \right\}. \tag{3.4}$$

Then applying Chow's inequality, we get that

$$\begin{aligned} &P\left(\max_{1 \le k \le \lfloor m/\nu \rfloor} \frac{\|\sum_{n=m+1}^{m+k} \mathbb{U}_n\|}{g_\gamma(m,k)} \ge \varepsilon \right) \\ &\le \sum_{j=1}^{r} P\left(\max_{1 \le k \le \lfloor m/\nu \rfloor} \frac{\left(w(k/m)\sum_{n=m+1}^{m+k} U_{n,j}\right)^2}{m} \ge \frac{\varepsilon^2}{r} \right) \\ &\le \sum_{j=1}^{r} \frac{r}{\varepsilon^2} \sum_{k=1}^{\lfloor m/\nu \rfloor} \frac{w^2(k/m) E U_{m+k,j}^2}{m} \le \frac{r^2 v_0}{\varepsilon^2} \int_0^{1/\nu} \frac{1}{t^{2\gamma}} dt = \frac{r^2 v_0}{\varepsilon^2 (1-2\gamma)\nu^{1-2\gamma}} \to 0 \end{aligned}$$

as $\nu \to \infty$. Also, the convergence of the process $\mathcal{U}_m$ implies that the variables $||\mathcal{U}_m(1)||$ are stochastically bounded, which results in the convergence

$$\max_{1 \le k \le \lfloor m/\nu \rfloor} \frac{\frac{k}{m}\|\sum_{n=1}^{m} \mathbb{U}_n\|}{g_\gamma(m,k)} = ||\mathcal{U}_m(1)|| \max_{1 \le k \le \lfloor m/\nu \rfloor} \frac{k}{m} w(\frac{k}{m}) \le ||\mathcal{U}_m(1)|| \frac{1}{\nu^{1-\gamma}} \xrightarrow{P} 0,$$

uniformly in m as $\nu \to \infty$. From these we get that

$$\sup_{0 \le t \le 1/\nu} \|\mathcal{X}_m(t) - \mathcal{X}_{m,\nu}(t)\| = \max_{1 \le k \le \lfloor m/\nu \rfloor} ||\mathcal{X}_m(k/m)|| \xrightarrow{P} 0, \qquad \nu \to \infty,$$

uniformly in m. Note that $\mathcal{X}_{0,\nu} \to \mathcal{X}$ almost surely as $\nu \to \infty$. Then, theorem 3.2 of Billingsley (1999) implies that the process $\mathcal{X}_m$ converges in distribution to $\mathcal{X}$ as $m \to \infty$ in the space $\mathcal{D}^r[0,\infty)$. □

Proof of Theorem 2.1. By the properties of the Skorokhod topology, Lemmas 3.2 and 3.3 imply the convergence $\hat{\mathcal{X}}_m \to_{\mathcal{D}} \mathcal{X}$ in the space $\mathcal{D}^r[0,\infty)$ as $m \to \infty$. Because $\hat{\mathbf{C}}_m^{-1/2}$ is a weakly consistent estimator of $\mathbf{C}_0^{-1/2}$, we also get that $\hat{\mathbf{C}}_m^{-1/2} \hat{\mathcal{X}}_m \to_{\mathcal{D}} \mathbf{C}_0^{-1/2} \mathcal{X}$ as $m \to \infty$.

Consider the function $\Psi_T : \mathcal{D}^r[0,\infty) \to \mathbb{R}$ defined as $\Psi_T(y) := \sup_{0 \le t \le T} \psi(y(t))$. It can be shown that Ψ_T is measurable for any $T \in (0,\infty]$, and by proposition 2.4 of Jacod and Shiryaev (2003) it is continuous at the elements of the set $\mathcal{C}^r[0,\infty)$ if T is finite. Because $\mathbf{C}_0^{-1/2}\mathcal{X}$ is a sample continuous process, it follows from the continuous mapping theorem (see theorem 2.7 of Billingsley [1999]) that

$$\sup_{1 \le k \le \lfloor Tm \rfloor} \psi(\mathbb{S}_{m,k}) = \Psi_T(\hat{\mathbf{C}}_m^{-1/2} \hat{\mathcal{X}}_m) \xrightarrow{\mathcal{D}} \Psi_T(\mathbf{C}_0^{-1/2}\mathcal{X}) = \sup_{0 \le t \le T/(1+T)} \psi(\mathcal{W}(t)/t^\gamma), \tag{3.5}$$

for any finite T as $m \to \infty$. Unfortunately, this argument does not work for $T = \infty$, because in case of an arbitrary continuous ψ the function Ψ_∞ is not continuous on

$\mathcal{C}^r[0,\infty)$. In the remainder of the proof, we show that the statement is true for $T=\infty$ by using a different method.

Because the random vectors $\mathbb{U}_1, \mathbb{U}_2, \ldots$ have bounded second moments, the martingale law of large numbers (see, e.g., theorem 3 in section VII.9 in Feller [1971]) implies the almost sure convergence

$$\mathcal{X}_m\left(\frac{k}{m}\right) = m^{1/2}\left(1+\frac{m}{k}\right)^{\gamma}\left[\frac{1}{m+k}\sum_{n=1}^{m+k}\mathbb{U}_n - \frac{1}{m}\sum_{n=1}^{m}\mathbb{U}_n\right] \to -\frac{1}{m^{1/2}}\sum_{n=1}^{m}\mathbb{U}_n, \qquad (3.6)$$

$k\to\infty$. In the next step, we show that this convergence is uniform in m. Let $\mathcal{X}_m^*$ denote the process $\mathcal{X}_m$ with fixed parameter $\gamma=0$. From (3.6), it follows for any $T\in(0,\infty)$ and $k\geq Tm$ that

$$\mathcal{X}_m^*\left(\frac{k}{m}\right) - \mathcal{X}_m^*(T) = \frac{m^{1/2}}{m+k}\sum_{n=m+\lfloor Tm\rfloor+1}^{m+k}\mathbb{U}_n - \frac{m^{1/2}(k-\lfloor Tm\rfloor)}{(m+k)(m+\lfloor Tm\rfloor)}\sum_{n=1}^{m+\lfloor Tm\rfloor}\mathbb{U}_n.$$

By using again the Hájek–Rényi type inequality (3.4), we get that

$$P\left(\sup_{k\geq Tm}\frac{\|\sum_{n=m+\lfloor Tm\rfloor+1}^{m+k}\mathbb{U}_n\|}{m^{-1/2}(m+k)}\geq\varepsilon\right) \leq \sum_{j=1}^{r}P\left(\sup_{k\geq Tm}\frac{\left(\sum_{n=m+\lfloor Tm\rfloor+1}^{m+k}U_{n,j}\right)^2}{m^{-1}(m+k)^2}\geq\frac{\varepsilon^2}{r}\right)$$
$$\leq\sum_{j=1}^{p}\frac{r}{\varepsilon^2}\sum_{k=\lfloor Tm\rfloor+1}^{\infty}\frac{EU_{m+k,j}^2}{m(1+k/m)^2}\leq\frac{rv_0}{\varepsilon^2}\int_{T-1}^{\infty}\frac{1}{(1+t)^2}dt=\frac{rv_0}{\varepsilon^2 T}\to 0, \qquad T\to\infty.$$

Also, the tightness of the variables $\mathcal{U}_m(1)$, $m\in\mathbb{Z}_{++}$, implies that

$$\sup_{k\geq Tm}\frac{m^{1/2}(k-\lfloor Tm\rfloor)}{(m+k)(m+\lfloor Tm\rfloor)}\|\sum_{n=1}^{m+\lfloor Tm\rfloor}\mathbb{U}_n\|$$
$$=\sup_{k\geq Tm}\left(\frac{m}{m+\lfloor Tm\rfloor}\right)^{1/2}\frac{(k-\lfloor Tm\rfloor)}{m+k}\frac{\|\sum_{n=1}^{m+\lfloor Tm\rfloor}\mathbb{U}_n\|}{\sqrt{m+\lfloor Tm\rfloor}}\leq\frac{\|\mathcal{U}_{m+\lfloor Tm\rfloor}(1)\|}{T^{1/2}}\xrightarrow{P}0$$

holds uniformly in m as $T\to\infty$. As a result, we get the convergence

$$\sup_{t\geq T}\|\mathcal{X}_m^*(t)-\mathcal{X}_m^*(T)\| = \sup_{k\geq Tm}\|\mathcal{X}_m^*(k/m)-\mathcal{X}_m^*(T)\|\xrightarrow{P}0, \qquad T\to\infty,$$

uniformly in m. Because for any fixed $T\geq 0$ the variables $\mathcal{X}_m^*(T)$, $m\in\mathbb{Z}_{++}$, are tight, it also follows that $\sup_{t\geq T}\|\mathcal{X}_m^*(t)\|=O_P(1)$. We already proved that the statement is true for any finite T. Using this result with function $\psi(\mathbf{x})=\|\mathbf{x}\|$, $\mathbf{x}\in\mathbb{R}^r$, we get that $\sup_{0\leq t\leq T}\|\mathcal{X}_m^*(t)\|=O_P(1)$, resulting in the rate $\sup_{t\geq 0}\|\mathcal{X}_m^*(t)\|=O_P(1)$.

Let $\gamma\in[0,1/2)$ be an arbitrary value and note that $\mathcal{X}_m(t)=(1+m/\lfloor tm\rfloor)^{\gamma}\mathcal{X}_m^*(t)$, where the function $(1+m/\lfloor tm\rfloor)^{\gamma}$, $t\geq T$, is decreasing and it has finite limit at infinity. Then, for any $T>1$, by using the triangular inequality, we get the convergence

$$\begin{aligned}\sup_{t\geq T}\|\mathcal{X}_m(t)-\mathcal{X}_m(T)\| &\leq \left(1+\frac{m}{\lfloor Tm\rfloor}\right)^{\gamma}\sup_{t\geq T}\|\mathcal{X}^*_m(t)-\mathcal{X}^*_m(T)\| \\ &+\sup_{t\geq T}\left[\left(1+\frac{m}{\lfloor Tm\rfloor}\right)^{\gamma}-\left(1+\frac{m}{\lfloor tm\rfloor}\right)^{\gamma}\right]\sup_{t\geq T}||\mathcal{X}^*_m(t)|| \\ &\leq 2^{\gamma}\sup_{t\geq T}\|\mathcal{X}^*_m(t)-\mathcal{X}^*_m(T)\|+\left(1+\frac{1}{T-1}\right)^{\gamma}\sup_{t\geq 0}||\mathcal{X}^*_m(t)||\xrightarrow{P}0,\end{aligned} \tag{3.7}$$

uniformly in m as $T\to\infty$. From this one can prove that $\sup_{t\geq 0}||\mathcal{X}_m(t)||=O_P(1)$ similar to how we obtained the related rate for the process $\mathcal{X}^*_m$.

Consider arbitrary values $\varepsilon,\delta,\delta'>0$. By the uniform stochastic boundedness, there exists a constant K such that $P(\sup_{t\geq 0}||\mathcal{X}_m||\leq K)\geq 1-\varepsilon$ holds for $m\in\mathbb{Z}_{++}$. By using this bound, Lemma 3.2, the uniform convergence in (3.7), and the weak consistency of the estimator $\hat{\mathbf{C}}_m$ imply that there exist positive values $T_0,m_0\geq 0$ depending only on ε, δ', and K, such that

$$P\left(\sup_{t\geq 0}||\hat{\mathcal{X}}_m(t)||\leq 2K,\qquad \sup_{t\geq T}||\hat{\mathbf{C}}_m^{-1/2}\hat{\mathcal{X}}_m(t)-\hat{\mathbf{C}}_m^{-1/2}\hat{\mathcal{X}}_m(T)\|\leq\delta'\right)\geq 1-2\varepsilon \tag{3.8}$$

holds for every $T\geq T_0$ and $m\geq m_0$. Because the function ψ is continuous, it is uniformly continuous on the r-dimensional closed sphere having radius $2K$ and having center at the origin. This means that the δ' can be chosen such that $||\psi(\mathbf{x})-\psi(\mathbf{y})||\leq\delta$ for every element $\mathbf{x}$ and $\mathbf{y}$ of the sphere satisfying $||\mathbf{x}-\mathbf{y}||\leq\delta'$. By using this property along with (3.8), we get that

$$\begin{aligned}&p_{T,m}(\delta):=P(\boldsymbol{\Psi}_{\infty}(\hat{\mathbf{C}}_m^{-1/2}\hat{\mathcal{X}}_m)-\boldsymbol{\Psi}_T(\hat{\mathbf{C}}_m^{-1/2}\hat{\mathcal{X}}_m)>\delta)\\ &\leq P(\sup_{t\geq T}\psi(\hat{\mathbf{C}}_m^{-1/2}\hat{\mathcal{X}}_m(t))-\psi(\hat{\mathbf{C}}_m^{-1/2}\hat{\mathcal{X}}_m(T))>\delta)\leq 2\varepsilon\end{aligned}$$

for every $T\geq T_0$ and $m\geq m_0$. Because ε is an arbitrary positive value, if follows that

$$\lim_{m\to\infty}\limsup_{T\to\infty}p_{T,m}(\delta)=0$$

holds for every $\delta>0$. Note that $\boldsymbol{\Psi}_T(\mathcal{X})\to\boldsymbol{\Psi}_{\infty}(\mathcal{X})$ almost surely as $T\to\infty$. Then, the convergence in (3.5) proved for any finite T and theorem 3.2 of Billingsley (1999) implies that the result in (3.5) is true for $T=\infty$ as well. This argument completes the proof of the theorem. □

Proof of Theorem 2.3. Consider an arbitrary integer-valued sequence $k_m\geq k^*_m+1$, $m\in\mathbb{Z}_{++}$. Let us note that $\mathbb{X}_{m,n}=\mathbb{X}_n$ and $\mathbb{U}_{m,n}=\mathbb{U}_n$ hold for any positive integers m and $n\leq m+k^*_m$, and we have $\hat{\mathbb{U}}_{m,n}-\mathbb{U}_{m,n}=f(\mathbb{X}_{m,n},\boldsymbol{\theta}_{\mathrm{A}})-f(\mathbb{X}_{m,n},\hat{\boldsymbol{\theta}}_m)$ for every m and $n>m+k^*_m$. Then it follows that

$$\sum_{n=m+1}^{m+k_m} \hat{\mathbb{U}}_{m,n} - \frac{k_m}{m} \sum_{n=1}^{m} \hat{\mathbb{U}}_{m,n} = \left[\sum_{n=m+1}^{m+k_m^*} \hat{\mathbb{U}}_{m,n} - \frac{k_m^*}{m} \sum_{n=1}^{m} \hat{\mathbb{U}}_{m,n} \right] + \sum_{n=m+k_m^*+1}^{m+k_m} \mathbb{U}_{m,n}$$
$$+ \sum_{n=m+k_m^*+1}^{m+k_m} [f(\mathbb{X}_{m,n}, \boldsymbol{\theta}_{\mathrm{A}}) - f(\mathbb{X}_{m,n}, \boldsymbol{\theta}_0)] + \sum_{n=m+k_m^*+1}^{m+k_m} [f(\mathbb{X}_{m,n}, \boldsymbol{\theta}_0) - f(\mathbb{X}_{m,n}, \hat{\boldsymbol{\theta}}_m)]$$
$$- \frac{k_m - k_m^*}{m} \sum_{n=1}^{m} \mathbb{U}_n - \frac{k_m - k_m^*}{m} \sum_{n=1}^{m} [\hat{\mathbb{U}}_{m,n} - \mathbb{U}_n].$$

First, consider the case $r = 1$. Because $g_\gamma(m,k)$ is an increasing function of k, Theorem 2.2 implies that

$$\frac{\left|\sum_{n=m+1}^{m+k_m^*} \hat{\mathbb{U}}_{m,n} - \frac{k_m^*}{m} \sum_{n=1}^{m} \hat{\mathbb{U}}_{m,n}\right|}{g_\gamma(m,k_m)} \le |\hat{\mathcal{X}}_m(k_m^*/m)| \le \sup_{t \ge 0} |\hat{\mathcal{X}}_m(t)| = O_P(1).$$

Let us note that $\sup_{k \ge 1} k/g_\gamma(m,k) = O(m^{1/2})$. Using this rate and the weak convergence of the process $\mathcal{U}_m$, which was shown in the proof of Lemma 3.3, we obtain that

$$\frac{\frac{k_m - k_m^*}{m} \left|\sum_{n=1}^{m} \mathbb{U}_n\right|}{g_\gamma(m,k_m)} \le \sup_{k \ge 1} \frac{k|\mathcal{U}_m(1)|}{m^{1/2} g_\gamma(m,k)} = O_P(1).$$

Also, from equation (3.3), it follows that

$$\frac{\frac{k_m - k_m^*}{m} \left|\sum_{n=1}^{m} (\hat{\mathbb{U}}_{m,n} - \mathbb{U}_n)\right|}{g_\gamma(m,k_m)} \le \sup_{k \ge 1} \frac{k O_P(m^{-1/2})}{g_\gamma(m,k)} + o_P(1) = O_P(1), \qquad m \to \infty.$$

Because the random variables $\mathbb{U}_{m,1}, \mathbb{U}_{m,2}, \ldots$ form a martingale difference sequence, they are pairwise uncorrelated. Then, for any $m \ge m_A$, by using (iv) of Assumption 2.2 we get that

$$\operatorname{Var}\left(\frac{\sum_{n=m+k_m^*+1}^{m+k_m} \mathbb{U}_{m,n}}{g_\gamma(m,k_m)} \right) \le \frac{(k_m - k_m^*)\nu_A}{g_\gamma^2(m,k_m)} \le \left(\frac{k_m}{m + k_m} \right)^{1-2\gamma} \nu_A \le \nu_A.$$

From this, the Chebyshev inequality implies that the variable on the left side is of rate $O_P(1)$ as $m \to \infty$.

Consider the constant Δ defined by the alternative hypothesis $\mathcal{H}_A$ and assume that $k_m - k_m^* \to \infty$ as $m \to \infty$. By the assumptions, we can apply Lemma 3.1 to the array of variables $\{\mathbb{X}_{m,m+k_m^*+\ell}, \ell \in \mathbb{Z}_+\}$, $m \in \mathbb{Z}_{++}$, with the function $\phi(\mathbf{x}) = f(\mathbf{x}, \boldsymbol{\theta}_{\mathrm{A}}) - f(\mathbf{x}, \boldsymbol{\theta}_0)$, and we obtain the equation

$$\sum_{n=m+k_m^*+1}^{m+k_m} [f(\mathbb{X}_{m,n}, \boldsymbol{\theta}_{\mathrm{A}}) - f(\mathbb{X}_{m,n}, \boldsymbol{\theta}_0)] = (k_m - k_m^*)[\Delta + o_P(1)], \qquad m \to \infty.$$

Similar arguments result in

$$\sum_{n=m+k_m^*+1}^{m+k_m} [f(\mathbb{X}_{m,n},\boldsymbol{\theta}_0) - f(\mathbb{X}_{m,n},\hat{\boldsymbol{\theta}}_m)] = o_P(k_m - k_m^*), \qquad m \to \infty.$$

By summarizing the results of the current proof, the weak consistency of the estimator $\hat{\mathbf{C}}_m$ implies that

$$\mathbb{S}_{m,k_m} = \hat{\mathbf{C}}_m^{-1/2} \frac{\sum_{n=m+1}^{m+k_m} \hat{\mathbb{U}}_{m,n} - \frac{k_m}{m}\sum_{n=1}^{m} \hat{\mathbb{U}}_{m,n}}{g_\gamma(m,k_m)} = \frac{k_m - k_m^*}{g_\gamma(m,k_m)} [\mathbf{C}_0^{-1/2}\Delta + o_P(1)] + O_P(1) \tag{3.9}$$

as $m \to \infty$ in the case $r = 1$. From this it follows that (3.9) holds for an arbitrary r as well, because in the general case the equation is understood componentwise.

(i). Consider the sequence $k_m = k_m^* + \lfloor \varepsilon(m + k_m^*) \rfloor$, $m \in \mathbb{Z}_{++}$, with an arbitrary $\varepsilon > 0$. If m is large enough, we obtain the inequality

$$\frac{k_m - k_m^*}{g_\gamma(m,k_m)} \geq \frac{\sqrt{m}\lfloor \varepsilon(m+k_m^*)\rfloor}{m + k_m^* + \lfloor \varepsilon(m+k_m^*)\rfloor} \geq \frac{\sqrt{m}\lfloor \varepsilon(m+k_m^*)\rfloor}{(1+\varepsilon)(m+k_m^*)},$$

and the right side converges to infinity as $m \to \infty$. Because $\mathbf{C}_0$ is nonsingular and $\Delta \neq 0$ by the alternative hypothesis, we have $\mathbf{C}_0^{-1/2}\Delta \neq 0$. This means that $||\mathbb{S}_{m,k_m}|| \to_P \infty$, implying the convergence $\psi(\mathbb{S}_{m,k_m}) \to_P \infty$. Let x_α stand for the critical value of the test corresponding to an arbitrary significance level $\alpha \in (0,1)$. Then we have the convergence

$$P(\kappa_{m,k_m^*} - k_m^* \leq \varepsilon(m + k_m^*)) \geq P(\psi(\mathbb{S}_{m,k_m}) > x_\alpha) \to 1, \qquad m \to \infty,$$

proving the first statement.

(ii) To prove the second statement, consider the values $k_m = k_m^* + \lfloor Cm^\beta \rfloor$, $m \in \mathbb{Z}_{++}$, with an arbitrary $C > 0$ and with the β defined by the theorem. The conditions on the function ψ imply that there exists a real value $K > 0$ such that $\psi(\mathbf{x}) > x_\alpha$ if $||\mathbf{x}|| > K$. By standard calculations, one can verify that

$$\lim_{m\to\infty} \frac{k_m - k_m^*}{g_\gamma(m,k_m)} = H(C) := \begin{cases} C\left(C + a\mathbb{1}_{\left\{b = \frac{1-2\gamma}{2-2\gamma}\right\}}\right)^{-\gamma}, & 0 \leq b \leq \dfrac{1-2\gamma}{2-2\gamma}, \\ Ca^{-\gamma}(1 + a\mathbb{1}_{\{b=1\}})^{\gamma-1}, & \dfrac{1-2\gamma}{2-2\gamma} < b \leq 1, \\ Ca^{-1}, & 1 < b. \end{cases}$$

From this and from equation (3.9) it follows that for any fixed $C > 0$, if m is large enough, then

$$||\mathbb{S}_{m,k_m}|| \geq \frac{H(C)}{2}[||\mathbf{C}_0^{-1/2}\Delta|| + o_P(1)] + O_P(1), \tag{3.10}$$

where the terms $o_P(1)$ and $O_P(1)$ are the same as in (3.9) and do not depend on C. Fix an arbitrary real number $\delta > 0$. Because $\lim_{C\to\infty} H(C) = \infty$, the right side of (3.10) converges to infinity as $C \to \infty$ with probability 1. This implies that the value C can be chosen in such a way that the right side of (3.10) is greater than K with a probability at least $1 - \delta$. Using this C, we obtain the inequalities

$$P(\kappa_{m,k_m^*} - k_m^* \leq Cm^\beta) \geq P(\psi(\mathbb{S}_{m,k_m}) > x_\alpha) \geq P(||\mathbb{S}_{m,k_m}|| > K) \geq 1 - \delta$$

for every large enough m. Because δ is an arbitrary positive number, the probability on the left side converges to 1 as $m \to \infty$, proving the second statement of the theorem. □

Remark 3.1. Because the functions ψ_1 and ψ_2 of (2.3) satisfy the assumptions of Theorem 2.3, the results are valid for the related test statistics. Unfortunately, we cannot apply the theorem for the test statistics corresponding to the third convergence, because the limit $\lim_{\|\mathbf{x}\|\to\infty} \psi_3(\mathbf{x})$ does not exist. However, with some modifications in the proof, one can show that the results of Theorem 2.3 are valid for ψ_3, if $\mathbf{c}^\top \mathbf{C}_0^{-1/2}\Delta \neq 0$. For this goal, note that the base idea of the proof is formula (3.9), which ensures that the vector $\mathbb{S}_{m,k_m}$ is "large" in some sense. From this equation, we get that

$$\psi_3(\mathbb{S}_{m,k_m}) = |\mathbf{c}^\top \mathbb{S}_{m,k_m}| = \frac{k_m - k_m^*}{g_\gamma(m,k_m)}\left[|\mathbf{c}^\top \mathbf{C}_0^{-1/2}\Delta| + o_P(1)\right] + O_P(1), \tag{3.11}$$

implying that $\psi_3(\mathbb{S}_{m,k_m})$ is "large" as well, if $\mathbf{c}^\top \mathbf{C}_0^{-1/2}\Delta \neq 0$. Then the results of Theorem 2.3 can be obtained for the function ψ_3 by using (3.11) in parts (i) and (ii) of the proof.

Proof of Theorem 2.4. Let us note that in the open-ended case all of the three convergences in Theorem 2.2 can be written in the form $\sup_{k\geq 1} \psi(\mathbb{S}_{m,k}) \to_{\mathcal{D}} Z$, $m \to \infty$, where $\psi : \mathbb{R}^r \to \mathbb{R}$ is one of the functions in (2.3), and Z is a non-negative-valued absolute continuous random variable with unbounded support. Let F_Z stand for the distribution function of Z and let x_α be the critical value of the open-ended test corresponding to the significance level α.

If $b < 1$, then consider an arbitrary value $\varepsilon > 0$. Because $k_m^* < \varepsilon m$ if m is large enough, we get that

$$P(\kappa_{m,0} \leq k_m^*) \leq P\left(\sup_{1\leq k\leq \lfloor \varepsilon m\rfloor} \psi(\mathbb{S}_{m,k}) > x_\alpha\right) \to 1 - F_Z\left(\left(\frac{1+\varepsilon}{\varepsilon}\right)^{1/2-\gamma} x_\alpha\right),$$

as $m \to \infty$. Because the limit can be arbitrarily small by choosing a sufficiently small ε, the left side converges to 0 as $m \to \infty$.

If $b = 1$, then the identity $F_Z(x_\alpha) = 1 - \alpha$ implies the convergence

$$P\big(\kappa_{m,0} \leq k_m^*\big) = P\left(\sup_{1\leq k\leq \lfloor cm\rfloor} \psi\big(\mathbb{S}_{m,k}\big) > x_\alpha\right) \to 1 - F_Z\left(\left(\frac{1+c}{c}\right)^{1/2-\gamma} x_\alpha\right) \in (0,\alpha),$$

as $m \to \infty$.

If $b > 1$, then consider an arbitrary $T > 0$ and note that for every large enough m we have the inequality and the convergence

$$\begin{aligned} 1 - F_Z\left(\left(\frac{1+T}{T}\right)^{1/2-\gamma} x_\alpha\right) &\leftarrow P\left(\sup_{1\leq k\leq \lfloor Tm\rfloor} \psi\big(\mathbb{S}_{m,k}\big) > x_\alpha\right) \leq P(\kappa_{m,0} \leq k_m^*) \\ &\leq P\left(\sup_{k\geq 1} \psi\big(\mathbb{S}_{m,k}\big) > x_\alpha\right) \to 1 - F_Z(x_\alpha) = \alpha, \end{aligned}$$

as $m \to \infty$. Because by increasing T the left side can be arbitrarily close to α, the probability in question goes to α as $m \to \infty$. This argument completes the proof of the theorem. □

Acknowledgments

The author is thankful to her supervisor Gábor Szűcs for introducing her to the problem investigated in this article and for his help during the research. The work of the Editor and the Referee is greatly appreciated.

Funding

This research was supported by the ÚNKP-ÚNKP-16-3 New National Excellence Program of the Ministry of Human Capacities.

References

Aue, A., Horváth, L., Hušková, M., and Kokoszka, P. (2006). Change-Point Monitoring in Linear Models, *Econometrics Journal* 9**:** 373–403.

Billingsley, P. (1999). *Convergence of Probability Measures*, New York: Wiley.

Chow, Y. S. (1960). A Martingale Inequality and the Law of Large Numbers, *Proceedings of American Mathematical Society* 11**:** 107–111.

Chu, C.-S. J., Stinchcombe, M., and White, H. (1996). Monitoring Structural Change, *Econometrica* 64**:** 1045–1065.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, vol. 2, New York: Wiley.

Horváth, L., Hušková, M., Kokoszka, P., and Steinebach, J. (2004). Monitoring Changes in Linear Models, *Journal of Statistical Planning and Inference* 126**:** 225–251.

Horváth, L., Kokoszka, P., and Steinebach, J. (2007). On Sequential Detection of Parameter Changes in Linear Regression, *Statistics & Probability Letters* 77**:** 885–895.

Jacod, J. and Shiryaev, A. N. (2003). *Limit Theorems for Stochastic Processes*, Berlin: Springer.

Kirch, C. and Tadjuidje Kamgaing, J. (2011). *An Online Approach to Detecting Changes in Nonlinear Autoregressive Models*, https://kluedo.ub.uni-kl.de/frontdoor/index/index/docId/2772

Meyn, S. and Tweedie, R. L. (2009). *Markov Chains and Stochastic Stability*, Cambridge: Cambridge University Press.

Nedényi, F. (2015). Conditional Least Squares Estimators for Multitype Galton–Watson Processes, *Acta Scientiarum Mathematicarum (Szeged)* 81**:** 325–348.