



# Real-time changepoint detection in a nonlinear expectile model

Gabriela Ciuperca<sup>1</sup> · Matúš Maciak<sup>2</sup> · Michal Pešta<sup>2</sup> 

Received: 20 January 2022 / Accepted: 15 March 2023 / Published online: 30 March 2023  
© The Author(s) 2023

## Abstract

An online changepoint detection procedure based on conditional expectiles is introduced. The key contribution is threefold: nonlinearity of the underlying model improves the overall flexibility while a parametric form of the unknown regression function preserves a simple and straightforward interpretation; The conditional expectiles, well-known in econometrics for being the only coherent and elicitable risk measure, introduce additional robustness—especially with respect to asymmetric error distributions common in various types of data; The proposed statistical test is proved to be consistent and the distribution under the null hypothesis does not depend on the functional form of the underlying model nor the unknown parameters. Empirical properties of the proposed real-time changepoint detection test are investigated in a simulation study and a practical applicability is illustrated using the Covid-19 prevalence data from Prague.

**Keywords** Asymmetric least squares · Changepoint test · Conditional expectiles · Online detection · Coherent risk measure

## 1 Introduction

It is a common task—not only in statistics—to provide procedures for detecting and estimating changepoints in all kinds of mathematical and stochastic models. Such pro-

---

✉ Michal Pešta  
michal.pesta@mff.cuni.cz

Gabriela Ciuperca  
gabriela.ciuperca@univ-lyon1.fr

Matúš Maciak  
matus.maciak@mff.cuni.cz

<sup>1</sup> Institut Camille Jordan, Université Lyon 1, 43 blvd du 11 Novembre 1918, Lyon 69622, France

<sup>2</sup> Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Charles University, Sokolovská 49/83, Prague 18675, Czech Republic

cedures are also important from a practical point of view and they may be often crucial in many real life problems. For instance, detecting a changepoint in some data generating model may trigger some model retraining mechanisms or, more frequently, it may govern important decisions effecting specific subjects or even the whole population—such as different pandemic restrictions related to the recent Covid-19 infection spread. On the other hand, the estimation of changepoints may lead to some correction procedures, specific treatment implementations, additional target-specific decisions, or just a deeper understanding of the underlying data generating process.

Considering the basic stochastic principles of the changepoint detection and various estimation methods, two different approaches are usually adopted in practical implementations. If the whole data sample is available at the very beginning of the analysis, the detection algorithm is called an *offline procedure*. If the data arrive in time (usually in an observation-by-observation manner) and the changepoint detection algorithm runs concurrently as new observations appear, such algorithms are referred to as *online procedures*.

In this paper, we focus on the *online regime*, where the proposed changepoint detection algorithm will be applied for a nonlinear parametric regression model. In addition to this nonlinearity, the conditional expectile estimation of the unknown parameters is adopted—similarly as in Newey and Powell (1987) where, however, the authors investigated a simple linear model instead—to have a coherent risk measure while also accounting for possibly asymmetric random error distributions. The changepoint detection itself is performed in terms of a consistent statistical test which is based on an accumulating dataset used in each consecutive step of the proposed *online procedure*.

There is a vast literature available on both—the offline and online changepoint detection strategies considering different models and various technical assumptions. Bearing in mind just the *online procedures*, Nedényi (2018) proposed an online testing approach based on a CUSUM test statistic to detect changes in a parameter of a discrete-time stochastic process. Linear regression models with independent error terms are considered in Chu et al (1996) and Horváth et al (2004), where a standard least squares estimator is employed. Possible detection delays in a sequential changepoint test for a multiple linear regression model are discussed in Aue et al (2009). Linear regression models with dependent observations are investigated in Fremdt (2015) and the online changepoint detection procedures within autoregressive times series are studied, for instance, in Hušková et al (2007). Some generalizations for multivariate cases can be found in Aue et al (2009) or Hoga (2017) and their results are further generalized in Barassi et al (2020) where a semiparametric CUSUM test is proposed to perform the online changepoint detection for various correlation structures of nonlinear multivariate regression models with dynamically evolving volatilities. Nonlinear integer-valued times series are also discussed from this perspective in Lee and Lee (2019). A very nice overall review of the *online procedures* can be also found in Basseville and Nikiforov (1993).

The method presented in this paper advocates the idea of semi-parametric CUSUM approaches in a combination with some robustness with respect to the underlying error terms. Firstly, a nonlinear regression model is assumed to govern the data generating process. Although the underlying regression function is deterministic, it is allowed to be nonlinear with respect to a set of unknown parameters. This introduces a rela-

tively flexible class of possible functions. Second, despite the independent error terms assumed for the proposed online detection regime, there are no restrictive assumptions imposed on the underlying error distribution and, in particular, substantial robustness is achieved with the proposed expectile estimation that also allows for asymmetric and heavy-tailed error distributions. The conditional expectiles define the only coherent and elicitable risk measure (see, for instance, Bellini et al (2018) or Ziegel (2016)) which is particularly important in situations where some risk related assessment is needed. Moreover, despite many similarities with conditional quantiles, the conditional expectiles are well-known to be viable also in situations when the conditional quantiles fail (see Philipps (2022) for a more comprehensive comparison). Third, the proposed test statistic follows, under the null hypothesis of no change, a relatively simple distribution which does not depend on the underlying regression function nor the set of the unknown parameters. Finally, the whole procedure can be implemented in a straightforward way and all necessary calculations performed within the proposed *online regime* can be easily obtained. Thus, the presented real-time changepoint detection method has a great potential for a practical applicability which goes way beyond the Covid-19 example illustrated at the end.

The rest of the paper is structured as follows: The underlying data and the corresponding changepoint model are described in the next section. A real-time changepoint detection in terms of a formal statistical test is introduced in Sect. 3. The asymptotic properties of the proposed test are also detailed there. In Sect. 4, finite sample properties are investigated and the Covid-19 prevalence data from Prague, Czech Republic, are analysed using the proposed methodological framework. Section 5 concludes with some final remarks. All theoretical proofs and further technical details are postponed to the Appendix.

## 2 Asymmetric least squares with changepoint

Let us consider a set of *historical data* denoted as  $\{(Y_i, \mathbf{X}_i^\top)^\top; i = 1, \dots, m\}$  for some deterministic  $q$ -dimensional vector of explanatory variables  $\mathbf{X}_i = (X_{i1}, \dots, X_{iq})^\top$  and some integer  $m \in \mathbb{N}$ . The data are assumed to follow a general nonlinear parametric regression model

$$Y_i = f(\mathbf{X}_i, \boldsymbol{\beta}) + \varepsilon_i, \quad i = 1, \dots, m, \quad (1)$$

where  $f(\cdot, \boldsymbol{\beta})$  is an explicit function depending on some unknown vector parameter  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top \in \Gamma \subseteq \mathbb{R}^p$  with the true (unknown) value denoted as  $\boldsymbol{\beta}^0 \in \mathbb{R}^p$ . A different approach could consider  $\mathbf{X}_i$ 's as random vectors, however, we concentrate on the fixed design as we want to adopt a robust (i.e., distribution-free) approach with only minimal assumptions being imposed on the underlying data distribution. Nevertheless, with respect to the forthcoming theory, analogous results for the random design can be derived as well (all under some technical assumptions needed for the deterministic convergences to become convergences in probability).

After the *historical data* are observed, another  $T_m \in \mathbb{N}$  observations are measured instantly for both—the response variable  $Y_i$  and the explanatory vector  $\mathbf{X}_i \in \mathcal{Y} \subseteq \mathbb{R}^q$ , both for  $i = m + 1, \dots, m + T_m$ . The underlying model for these new observations—

online data—is assumed to take an analogous form

$$Y_i = f(\mathbf{X}_i, \beta_i) + \varepsilon_i, \quad i = m + 1, \dots, m + T_m, \tag{2}$$

where the underlying regression functional form remains the same and  $\beta_i \in \mathbb{R}^p$ . For the parameter vectors  $\{\beta_i\}_{i=m+1}^{m+T_m}$  in (2), it is either assumed that their true (unknown) values are all equal to  $\beta^0$  (thus, there is no changepoint present in the overall combined model (1) and (2)) or, instead, there is some specific index  $k_m^0 \in \{1, \dots, T_m - 1\}$  such that  $\beta_i = \beta^0$  for all  $i = m + 1, \dots, m + k_m^0$ , while  $\beta_i \neq \beta^0$  for  $i = m + k_m^0 + 1, \dots, m + T_m$ . In such case, there is a changepoint (located at  $k_m^0$ ) present in the model generating the *online data*  $\{(Y_i, \mathbf{X}_i^\top)^\top; i = m + 1, \dots, m + T_m\}$ .

The error terms  $\{\varepsilon_i\}_{1 \leq i \leq m+T_m}$  from the overall model (1) and (2) are assumed to be independent and, moreover, they all follow the same distribution. A generic random error term from the underlying distribution is denoted as  $\varepsilon$ . The idea is to use the *historical data* to estimate the unknown parameter vector  $\beta \in \mathbb{R}^p$ . Later, the *online data*—starting from the observation index  $i = m + 1$ —are measured in real-time while asking a question for each new observation  $i \geq m + 1$  whether the underlying model remains unchanged (i.e.,  $\beta_i = \beta^0$ ) or there is some change detected in terms of the unknown parameter vectors  $\beta_i \in \mathbb{R}^p$ . If there is no changepoint detected for the given  $i$  then all available observations are used in the next step to ask the same question regarding the new—most recent observation. The whole changepoint detection process stops at the first observation  $i \in \{m + 1, \dots, m + T_m\}$  for which there is a statistical evidence that  $\beta_i \neq \beta^0$ .

From a formal theoretical point view, at the first step, the *historical data*  $\{(Y_i, \mathbf{X}_i^\top)^\top : i = 1, \dots, m\}$  are used to obtain a conditional expectile estimator for the unknown parameter vector  $\beta \in \mathbb{R}^p$ . In particular, for a given expectile index  $\tau \in (0, 1)$  the expectile function is defined as

$$\rho_\tau(x) = \left| \tau - \mathbb{I}_{\{x < 0\}} \right| x^2, \quad \text{for } x \in \mathbb{R}, \tag{3}$$

and the corresponding expectile estimator of the unknown (true) parameter vector  $\beta^0 \in \mathbb{R}^p$  from the model in (1) is defined as

$$\widehat{\beta}_m \equiv \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^m \rho_\tau(Y_i - f(\mathbf{X}_i, \beta)), \tag{4}$$

where  $\widehat{\beta}_m = (\widehat{\beta}_{m1}, \dots, \widehat{\beta}_{mp})^\top \in \mathbb{R}^p$ . It is straightforward to verify that for  $\tau = 1/2$  the expectile estimate  $\widehat{\beta}_m$  defined by (4) reduces to a standard (nonlinear) least squares (LS) estimator of  $\beta^0 \in \mathbb{R}^p$ . In general, the  $\tau^{th}$  expectile of the given distribution can be interpreted as a hypothetical mean of some other distribution that would be obtained if the values above the expectile in the original distribution would occur  $\frac{\tau}{1-\tau}$  times more frequently. Thus, the choice of  $\tau \in (0, 1)$  can be also seen in terms of some “exploratory” approach that somehow “balances” the distribution towards the (zero) mean and it provides a useful information about the skewness and possible outly-

ing/extreme observations. Also note, that depending on the choice of the regression function  $f$ , the minimization problem in (4) may or may not be a convex problem. This restricts the choice of the algorithm used to obtain the final solution. For numerical issues and different techniques for fitting nonlinear models we refer to Chambers (1973). Computational aspects are further discussion in Sect. 4.

In the second step, the expectile estimator  $\widehat{\beta}_m$  obtained from the *historical data*  $\{(Y_i, \mathbf{X}_i^\top)^\top; i = 1, \dots, m\}$  is used to perform a real-time changepoint detection in the *online data*  $\{(Y_i, \mathbf{X}_i^\top)^\top; i = m + 1, \dots, m + T_m\}$  in terms of a formal statistical test of the null hypothesis

$$H_0 : \beta_i = \beta^0, \quad \text{for } i = m + 1, \dots, m + T_m; \tag{5}$$

against the alternative hypothesis

$$H_1 : \exists k_m^0 \in \{1, \dots, T_m - 1\} \\ \text{such that } \begin{cases} \beta_i = \beta^0, & i = m + 1, \dots, m + k_m^0; \\ \beta_i = \beta^1 & i = m + k_m^0 + 1, \dots, m + T_m, \end{cases} \tag{6}$$

where  $\beta^0 \neq \beta^1$ . The proposed test statistic, sensitive to the null hypothesis, is defined as

$$\mathcal{T}(m) = \sup_{1 \leq k \leq T_m} \frac{\|\mathbf{S}(m, k)\|_\infty}{z(m, k, \gamma)}, \tag{7}$$

for a standard supremum norm  $\|\cdot\|_\infty$ , a regularization function  $z(m, k, \gamma) \equiv m^{1/2}(1 + k/m)(k/(k + m))^\gamma$  for some  $\gamma \in [0, 1/2)$ , and

$$\mathbf{S}(m, k) \equiv \mathbf{J}_m^{-1/2}(\widehat{\beta}_m) \sum_{i=m+1}^{m+k} \nabla f(\mathbf{X}_i, \widehat{\beta}_m) g_\tau(\widehat{\varepsilon}_i), \tag{8}$$

where  $g_\tau(x) \equiv \rho'_\tau(x) = 2\tau x \mathbb{I}_{\{x \geq 0\}} + 2(1 - \tau)x \mathbb{I}_{\{x < 0\}}$  stands the first derivative of the expectile function  $\rho_\tau(x)$  and  $\widehat{\varepsilon}_i = Y_i - f(\mathbf{X}_i, \widehat{\beta}_m)$  are so-called expectile residuals for  $i = 1, \dots, m, m + 1, \dots, m + T_m$ . Similarly, by  $h_\tau(x) \equiv \rho''_\tau(x) = 2\tau \mathbb{I}_{\{x \geq 0\}} + 2(1 - \tau) \mathbb{I}_{\{x < 0\}}$  we denote the second derivative of  $\rho_\tau(x)$ . In addition,  $\nabla f(\mathbf{X}_i, \widehat{\beta}_m)$  stands for a  $p$ -dimensional vector of the first partial derivatives  $\frac{\partial}{\partial \beta} f(\mathbf{X}_i, \widehat{\beta}_m)$  evaluated at the expectile estimate  $\widehat{\beta}_m$ , and

$$\mathbf{J}_m(\widehat{\beta}_m) \equiv \frac{\text{Var}[g_\tau(\varepsilon)]}{m} \sum_{i=1}^m \nabla f(\mathbf{X}_i, \widehat{\beta}_m) \nabla^\top f(\mathbf{X}_i, \widehat{\beta}_m), \tag{9}$$

where  $\mathbf{J}_m^{-1/2}(\widehat{\beta}_m)$  in (8) denotes the inverse of the square root matrix (in a sense of the Cholesky factorization) of  $\mathbf{J}_m(\widehat{\beta}_m)$ . A formal decision with respect to the null hypothesis

esis in (5) is done by comparing the test statistic in (7) with the corresponding quantile of the limit distribution, which is a functional of a Wiener process (see Theorem 2). Details regarding the behaviour of the test statistic under the null hypothesis and the alternative hypothesis are derived in the next section.

**Remark 1** In practical applications, the theoretical quantity  $\text{Var}[g_\tau(\varepsilon)]$  in (9) is typically unknown. However, the corresponding finite sample counterpart  $S_{g_\tau}^2 = \frac{1}{(m-1)} \sum_{i=1}^m [g_\tau(\widehat{\varepsilon}_i)]^2$  can be used instead as a plug-in estimate, where  $\widehat{\tau} \in (0, 1)$  (implicitly) solves  $\frac{1}{m} \sum_{i=1}^m g_\tau(\widehat{\varepsilon}_i) = 0$  (i.e., the empirical version of the theoretical assumption  $\mathbb{E}[g_\tau(\varepsilon)] = 0$ ) and  $\{\widehat{\varepsilon}_i\}_{i=1}^m$  are the model-based residuals. The empirical estimates for the theoretical quantities  $\text{Var}[g_\tau(\varepsilon)]$ ,  $\mathbb{E}[g_\tau(\varepsilon)]$ , and  $\tau \in (0, 1)$  are all based on the historical data  $\{(Y_i, \mathbf{X}_i^\top)^\top; i = 1, \dots, m\}$ .

### 3 Theoretical results

Besides a  $p$ -dimensional vector  $\nabla f(\mathbf{x}, \boldsymbol{\beta}) = \partial f(\mathbf{x}, \boldsymbol{\beta})/\partial \boldsymbol{\beta}$ , for any  $\mathbf{x} \in \mathcal{Y}$  and  $\boldsymbol{\beta} \in \Gamma$  let, analogously,  $\nabla^2 f(\mathbf{x}, \boldsymbol{\beta}) \equiv \partial^2 f(\mathbf{x}, \boldsymbol{\beta})/\partial \boldsymbol{\beta}^2$  be a  $(p \times p)$ -dimensional matrix of the second partial derivatives. In addition, let  $\nabla_j^2 f(\mathbf{x}, \boldsymbol{\beta}) \equiv (\partial^2 f(\mathbf{x}, \boldsymbol{\beta})/(\partial \beta_i \partial \beta_j))_{1 \leq i \leq p}$ , which is again a  $p$ -vector for each  $j \in \{1, \dots, p\}$ . Finally,  $\mathbf{V}_m(\boldsymbol{\beta})$  stands for a  $(p \times p)$ -dimensional matrix being defined as  $\mathbf{V}_m(\boldsymbol{\beta}) \equiv m^{-1} \sum_{i=1}^m \nabla f(\mathbf{X}_i, \boldsymbol{\beta}) \nabla^\top f(\mathbf{X}_i, \boldsymbol{\beta})$  and for any two constants  $a, b \in \mathbb{R}$  let  $a \vee b = \max(a, b)$  and  $a \wedge b = \min(a, b)$ .

#### 3.1 Model assumptions

Considering the overall changepoint model in (1) and (2), the theoretical results formulated in this section rely on the set of assumptions stated below. For a better organization of the whole paper, the assumptions are split into five groups, (A)–(E).

ASSUMPTION (A):

- (A1) The parameter space  $\Gamma \subseteq \mathbb{R}^p$  is a compact set and the design space  $\mathcal{Y} \subseteq \mathbb{R}^q$  is assumed to be bounded;
- (A2) For each  $i \in \{1, \dots, m, m + 1, m + T_m\}$ , the partial derivatives  $\nabla f(\mathbf{X}_i, \boldsymbol{\beta})$  and  $\nabla^2 f(\mathbf{X}_i, \boldsymbol{\beta})$  all exist and, moreover,  $\nabla f(\mathbf{X}_i, \boldsymbol{\beta})$  is continuous on  $\mathcal{Y} \times \Gamma$ ;
- (A3) For  $q_m(\boldsymbol{\beta}) \equiv \text{Card}\{i \in \{1, \dots, m\}; f(\mathbf{X}_i, \boldsymbol{\beta}) \neq f(\mathbf{X}_i, \boldsymbol{\beta}^0)\}$  and every  $\boldsymbol{\beta} \in \Gamma$  such that  $\boldsymbol{\beta} \neq \boldsymbol{\beta}^0$  it holds, that  $0 < \lim_{m \rightarrow \infty} q_m/m \leq 1$ .

ASSUMPTION (B): The density function of the random error terms  $\{\varepsilon_i\}_{i=1}^{m+T_m}$  (the generic error term  $\varepsilon$  respectively) is continuous and strictly positive in zero.

ASSUMPTION (C): There exists a positive definite matrix  $\mathbf{V}(\boldsymbol{\beta}^0)$  such that  $\mathbf{V}_m(\boldsymbol{\beta}^0) = m^{-1} \sum_{i=1}^m \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0) \nabla^\top f(\mathbf{X}_i, \boldsymbol{\beta}^0) \rightarrow \mathbf{V}(\boldsymbol{\beta}^0)$  for  $m \rightarrow \infty$ .

ASSUMPTION (D): The model errors  $\{\varepsilon_i\}_{i=1}^{m+T_m}$  are independent and identically distributed (i.i.d.) with a continuous distribution, such that  $\mathbb{E}[\varepsilon_i^4] < \infty$  and  $\mathbb{E}[g_\tau(\varepsilon_i)] = 0$ .

Assumptions (A), (B), and (C) are common conditions needed to show a strong consistency of the conditional expectile estimate  $\widehat{\boldsymbol{\beta}}_m$  defined in (4). Analogous conditions are used, for instance, by Choi et al (2003). Similarly, Assumption (D) is quite

standard for the expectile models (e.g., Gu and Zou (2016), Kim and Lee (2016), or Ciuperca (2022)).

### 3.2 Asymptotic behaviour of the expectile estimator

In order to study the asymptotic behaviour of the expectile estimator  $\widehat{\beta}_m$  defined in (4) let us consider the  $p$ -square matrix

$$\Omega \equiv \mathbb{E}[h_\tau(\varepsilon)]\mathbf{V}(\beta^0).$$

In addition to Assumption (A2), it is also required to impose slightly stricter assumptions on the matrix of the second partial derivatives  $\nabla^2 f(\mathbf{x}, \beta)$ .

**ASSUMPTION (E):** The elements of  $\nabla^2 f(\mathbf{x}, \beta)$  are all bounded for any  $\mathbf{x} \in \Upsilon$  and for  $\beta$  from a neighborhood of  $\beta^0$  of radius of the order  $m^{-1/2}$ .

The assumption above is a common property which is—under Assumption (A1)—satisfied by any function  $f$  which is continuous on  $\Upsilon \times \Gamma$ . It is considered, for instance, for a sequential test in a nonlinear changepoint model in Ciuperca (2013) where an ordinary least squares (LS) estimation framework was used instead. For the expectile estimation framework proposed in this paper, the asymptotic behaviour of the estimator in (4) is formulated in the next proposition.

**Proposition 1** *Under Assumptions (A)–(E),*

$$\widehat{\beta}_m = \beta^0 + \Omega^{-1} \frac{1}{m} \sum_{i=1}^m \nabla f(\mathbf{X}_i, \beta^0) g_\tau(\varepsilon_i) + o_{\mathbb{P}}(m^{-1/2}), \quad asm \rightarrow \infty.$$

If the regression function  $f$  is linear in  $\beta \in \Gamma$ , then the asymptotic behaviour in the proposition reduces to a special case of Proposition 1 from Ciuperca (2022). Similarly, if the regression function  $f$  in (1) is nonlinear in  $\beta \in \Gamma$ , but the random error terms follow some normal distribution  $N(0, \sigma^2)$  with  $\sigma^2 < \infty$ , the asymptotic behaviour in Proposition 1 gives the results of Theorem 2.1 in Seber and Wild (2003).

### 3.3 Test statistic under $H_0$ and $H_1$

The asymptotic behaviour of the test statistic defined in (7) is investigated in this section under both—the null hypothesis in (5) and the alternative hypothesis in (6). Note that that the vectors of parameters  $\beta^0, \beta^1 \in \Gamma$ , where  $\beta^0 \neq \beta^1$ , are both unknown. Let  $\mathbf{J}_m(\beta) \equiv \text{Var}[g_\tau(\varepsilon)]\mathbf{V}_m(\beta)$ , for  $\beta \in \Gamma$ , be a  $p \times p$  matrix—a theoretical (deterministic) version of its empirical counterpart—the  $p \times p$  matrix  $\mathbf{J}_m(\widehat{\beta}_m)$  defined in (9). Considering the size  $m \in \mathbb{N}$  for the *historical data* and the size  $T_m \in \mathbb{N}$  for the *online data* there are two specific possibilities which should be considered separately.

- if  $\lim_{m \rightarrow \infty} T_m/m = \infty$  for either  $T_m = \infty$  or  $T_m < \infty$ , then such a scenario is called an *open-end procedure*;
- if  $\lim_{m \rightarrow \infty} T_m/m = T$  for  $T_m < \infty$  where  $T \in (0, \infty)$ , then such a scenario is called a *closed-end procedure*.

By a common convention, it is usually assumed that for the open-end procedures it holds that  $T = \infty$ .

**Theorem 2** *Let Assumptions (A)–(E) be satisfied. Then, under  $H_0$ ,*

$$T(m) \equiv \sup_{1 \leq k \leq T_m} \frac{\|\mathbf{S}(m, k)\|_\infty}{z(m, k, \gamma)} \xrightarrow{m \rightarrow \infty} \sup_{0 < t < L(T)} \frac{\|\mathbf{W}_p(t)\|_\infty}{t^\gamma},$$

where  $\{\mathbf{W}_p(t); t \in (0, \infty)\}$  is a  $p$ -dimensional Wiener process where  $L(T) = 1$  for the open-end procedure and  $L(T) = T/(1 + T)$  for the closed-end procedure.

The test statistic in Theorem 2 is based on the expectile estimator  $\widehat{\boldsymbol{\beta}}_m$  of the true parameter vector  $\boldsymbol{\beta}^0 \in \Gamma$  calculated from the *historical data*. However, the limit process is the same as for the expectile estimator in the linear model considered in Ciuperca (2022), or the quantile estimator proposed in Zhou et al (2015). On the other hand, the test statistic is different from that proposed by Ciuperca (2013) or Horváth et al (2004) where the authors rather considered the CUSUM type statistic based on the least squares residuals of the linear model or the nonlinear model respectively.

In addition, the asymptotic behaviour of the test statistic under the null hypothesis in Theorem 2 does not depend on the underlying form of the nonlinear regression function  $f$  nor the true value  $\boldsymbol{\beta}^0$  as was the case for the test statistic applied for the parametric nonlinear model proposed in Ciuperca (2013). Therefore, the test statistic in Theorem 2 generally less restrictive, it is easier to use, and more straightforward to apply also for the least squares estimation (i.e., when  $\tau = 1/2$ ).

For the behaviour of the test statistic under the alternative hypothesis, more caution is needed. The model in (1) changes after the *historical data* and this change must be identifiable. Consequently, some reasonable assumptions are needed for the difference between the true parameter vectors  $\boldsymbol{\beta}^0$  and  $\boldsymbol{\beta}^1$  and, also, the underlying regression function  $f$ . Specific details are formulated in the next theorem.

**Theorem 3** *Let Assumptions (A)–(E) be satisfied and let  $m^{1/2}\|\boldsymbol{\beta}^0 - \boldsymbol{\beta}^1\|_2 \rightarrow \infty$  as  $m \rightarrow \infty$ . If there exists  $C > 0$  such that*

$$\frac{1}{m^s} \left\| \sum_{i=m+k_m^0+1}^{m+\widetilde{k}_m} c_i \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0) [f(\mathbf{X}_i, \boldsymbol{\beta}^1) - f(\mathbf{X}_i, \boldsymbol{\beta}^0)] \right\|_\infty > C > 0$$

for some constants  $|c_i| \in [2(\tau \wedge (1 - \tau)), 2(\tau \vee (1 - \tau))]$ , then

$$\sup_{1 \leq k \leq T_m} \frac{\|\mathbf{S}(m, k)\|_\infty}{z(m, k, \gamma)} \xrightarrow{m \rightarrow \infty} \mathbb{P} \rightarrow \infty.$$

Considering the assertions of both theorems together, the statistical test based on the proposed test statistics in (7) is proved to be consistent. The decision rule can be defined directly by considering the corresponding quantiles of the limit process from Theorem 2.



**Example 1** For a simple linear function  $f(x, \beta) = \beta_0 + \beta_1 x$ , the unknown vector parameters  $\beta^0 = (\beta_1^0, \beta_2^0)^\top \in \Gamma$  and  $\beta^1 = (\beta_1^1, \beta_2^1)^\top \in \Gamma$ , and  $x \in \Upsilon \subseteq \mathbb{R}$  one just need that  $\Gamma \subseteq \mathbb{R}^2$  is a compact set,  $\Upsilon$  is bounded, and  $(\beta_1^1 - \beta_1^0)(\beta_2^1 - \beta_2^0) \neq 0$  for the assumptions in (A1)–(A3) to hold. Assumption (B) is typically valid for common (continuous) error distributions. Assumptions (C) and (E) are satisfied trivially by the linearity of  $f$ . Finally, Assumption (D) can not be usually verified in a straightforward way but a sample estimate for  $\tau \in (0, 1)$  can be used such that the empirical counterpart of the equation  $\mathbb{E}[g_\tau(\varepsilon_i)] = 0$  is satisfied.

**Example 2** For a nonlinear function  $f(x, \beta) = \exp\{-\beta_1 e^{-\beta_2 x}\}$  (the Gompertz curve for  $x \in \Upsilon$ ) with  $\beta^0 = (\beta_1^0, \beta_2^0)^\top \in \Gamma$  and  $\beta^1 = (\beta_1^1, \beta_2^1)^\top \in \Gamma$  for  $\Gamma = (0, \infty) \times (0, \infty)$  and some bounded  $\Upsilon \subset \mathbb{R}$  it is easy to see that  $\nabla f(x, \beta) = (-f(x, \beta)e^{-\beta_2 x}, f(x, \beta)\beta_1 x e^{-\beta_2 x})^\top$  is continuous on  $\Upsilon \times \Gamma$  and  $\nabla^2 f(x, \beta)$  exists. Thus, for (A1)–(A3) to hold, one just needs that  $(\beta_1^1 - \beta_1^0)(\beta_2^1 - \beta_2^0) \neq 0$ . Assumption (C) can be shown in a straightforward way and the remaining assumptions are analogous to Example 1.

On the basis of the results obtained above one can define a stopping time—i.e., the first observation for which the null hypothesis in (5) is rejected in favor of the alternative hypothesis—considering the significance level  $\alpha \in (0, 1)$ . The corresponding changepoint estimate is defined as

$$\widehat{k}_m \equiv \begin{cases} \inf \left\{ k \geq 1; \sup_{1 \leq k \leq T_m} \frac{\|S(m,k)\|_\infty}{z(m,k,\gamma)} > c_\alpha(\gamma) \right\}; \\ \infty, & \text{if } \sup_{1 \leq k \leq T_m} \frac{\|S(m,k)\|_\infty}{z(m,k,\gamma)} \leq c_\alpha(\gamma) \text{ for all } k = 1, \dots, T_m, \end{cases}$$

where  $c_\alpha(\gamma)$  is  $(1 - \alpha)$ -quantile of the distribution of  $\sup_{0 < t < L(T)} \|W_p(t)\|_\infty / t^\gamma$ . Note, that  $\widehat{k}_m$  is the corresponding index referring to the online data only (i.e.,  $\widehat{k}_m \in \{1, \dots, T_m\}$ ). Thus, from the overall point of view, the underlying model changes after  $m + \widehat{k}_m$  observations. It holds that  $\lim_{m \rightarrow \infty} \mathbb{P}[\widehat{k}_m < \infty \mid H_0 \text{ true}] = \alpha$  and, similarly,  $\lim_{m \rightarrow \infty} \mathbb{P}[\widehat{k}_m < \infty \mid H_1 \text{ true}] = 1$ . Hence, the proposed test is consistent.

## 4 Empirical study

Finite sample properties of the proposed real-time changepoint detection method based on the expectile estimator defined in (4) are closely investigated in this section. Firstly, the empirical level of the test is assessed under various settings and the empirical power of the test is investigated for various changepoint scenarios. In the second part, the proposed methodology is also applied to analyze the Covid-19 prevalence data from Prague, Czech Republic, in order to link some authorities’ decisions to the real-time pandemic situation.

### 4.1 Simulation experiment

The main concept of the simulation study is analogous to that presented in Choi et al (2003). However, instead of a simple exponential function used for the underlying

**Table 1** Simulation results under the null hypothesis (with the theoretical value of  $\tau = 0.5$  for the symmetric distributions and the empirical estimate  $\hat{\tau} = 0.0719$  in terms of Remark 1 for the asymmetric distribution)

Distribution	$m$	$\hat{\beta}_1$   Std.Er. $\beta_1 = 10.00$	$\hat{\beta}_2$   Std.Er. $\beta_2 = 5.00$	Scenario 1 $T_m = 10$	Scenario 2 $T_m = m/2$	Scenario 3 $T_m = m \log(m)$
$N(0, 1)$ $\tau = 0.5000$	20	10.52 (2.832)	5.32 (1.825)	7.74%	7.74%	7.52%
	50	10.32 (2.857)	5.20 (1.382)	4.92%	6.08%	5.64%
	200	10.24 (2.766)	5.03 (0.808)	5.08%	5.58%	6.54%
$N(1, 1)$ $\hat{\tau} = 0.0719$	20	10.18 (2.889)	5.61 (2.020)	5.76%	5.76%	6.28%
	50	10.27 (2.869)	5.39 (1.638)	4.08%	3.26%	4.46%
	200	10.30 (2.833)	5.08 (0.977)	4.40%	4.26%	4.88%
$L(0, 1)$ $\tau = 0.5000$	20	10.50 (2.842)	5.32 (1.797)	7.84%	7.84%	9.86%
	50	10.37 (2.850)	5.18 (1.374)	4.08%	5.12%	7.38%
	200	10.29 (2.769)	5.02 (0.799)	5.06%	4.90%	5.58%

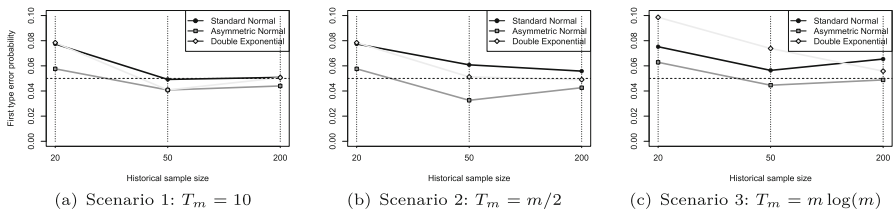
The parameter estimates are reported with the corresponding standard errors (in parentheses) over 5000 Monte Carlo simulations. Relative proportions of false rejections are given for three different scenarios for  $T_m$  reflecting the open-end and closed-end procedures. The nominal level of all the tests is always set to be  $\alpha = 0.05$

regression, a more complex Gompertz curve of the form

$$f(x, \boldsymbol{\beta}) = \exp\{-\beta_1 e^{-\beta_2 x}\}$$

is employed, where  $\boldsymbol{\beta}^0 = (\beta_1, \beta_2)^\top \equiv (10, 5)^\top$  and  $x \in (0, 1)$ . The reason is that the function used in Choi et al (2003) becomes very insensitive to any parameter change for large  $x_t = t$  (even for  $t \geq 10$ ). A simple *iterative grid search algorithm* is implemented to solve (4) and the changepoint test is performed in terms of Theorem 2. For the length of the historical period there are three different options considered ( $m \in \{20, 50, 200\}$ ). Analogously as in Choi et al (2003), three error distributions are used: a symmetric standard normal distribution (with  $\tau = 0.5$ ), asymmetric normal distribution with the mean and variance being equal to one ( $\hat{\tau} = 0.0719$ ), and, finally, a heavy-tailed (symmetric) Laplace distribution with the zero mean and unit variance (again,  $\tau = 0.5$  due to the symmetric property). In order to mimic both situations—the closed-end scenario and the open-end scenario—there are again three options considered for  $T_m \in \{10, m/2, m \log m\}$ . The empirical results under the null hypothesis (of no change in the model) are summarized in Table 1 and in Fig. 1. Different values for the regularization parameter  $\gamma \in [0, 1/2)$  were considered as well but no substantial differences were found, therefore, all reported results are for  $\gamma = 0.1$  only.

The empirical level of the test seems to properly keep the nominal level of  $\alpha = 0.05$  for all considered scenarios. The results are slightly conservative for the symmetric distributions (the normal distribution  $N(0, 1)$  and the double exponential distribution  $L(0, 1)$ ). On the other hand, a slightly underestimated nominal level is observed for the asymmetric error distribution (the normal distribution  $N(1, 1)$ ) but the actual differences are rather negligible. The corresponding expectile estimates of the unknown (true) parameters  $\beta_1 = 10$  and  $\beta_2 = 5$  seem both to be consistent for all considered scenarios and no inconsistencies are observed in Table 1.



**Fig. 1** The asymptotic behavior of the empirical Type I error probabilities for three different values of  $m \in \{20, 50, 200\}$ , three error distributions (standard normal, asymmetric normal with unit mean and variance, and double exponential with zero mean and unit variance) and three different choices of  $T_m$  in order to mimic the closed-end and open-end scenarios

On the other hand, the situation under the alternative hypothesis becomes slightly more comprehensive as there might be many different changepoint scenarios to possibly consider and take into account. For brevity purposes, there are only the results for one representative situation provided in this manuscript, but, any other situations were considered and compared with rather analogous results among all.

In particular, the following simulation scenarios under the alternative hypothesis were considered:

- A change occurs either in  $\beta_1$ , or in  $\beta_2$ , or in both elements of  $\beta = (\beta_1, \beta_2)^T$  simultaneously;
- A change occurs immediately after the historical data or the changepoint occurs after the first half of the online data;
- The magnitude of the change is relatively small compared to the true parameter values (20% change with respect to the true value) or the change is relatively large (the parameter(s) after the changepoint is(are) doubled);
- Finally, if the changepoint occurs in both elements of  $\beta = (\beta_1, \beta_2)^T$ , the corresponding effects of the changes may act against each other—thus, the resulting regression function after the change is very similar to the regression function before the change—or, alternatively, the effects of the changes aim at the same direction—thus, the regression function after the change is quite different from the underlying regression function before the change and there is also more power in the data to reveal such change.

All these situations have, of course, an important impact on the simulation results and, in particular, the performance of the proposed test in terms of its empirical power. For illustration purpose, one particular scheme (with the changepoint in  $\beta_2$  only and the change magnitude being equal to the true value of  $\beta_2$ ) is reported in Table 2. It is obvious from the table, that the performance of the proposed test (in terms of the empirical power) mostly depends on the true changepoint location and the length of the *online data* but in all considered situations the proposed test seems to be consistent.

Note that for the situations where the changepoint occurs in the first half of the *online data* (the rows denoted as  $k_m^{(1)}$  in Table 2), there are some false rejections (roughly 5%) of the observed rejections in the first half of the online data before the actual change appears. Such false rejection are not considered in Table 2 and only the rejections after the first half of the *online data* are reported. This is also reflected by

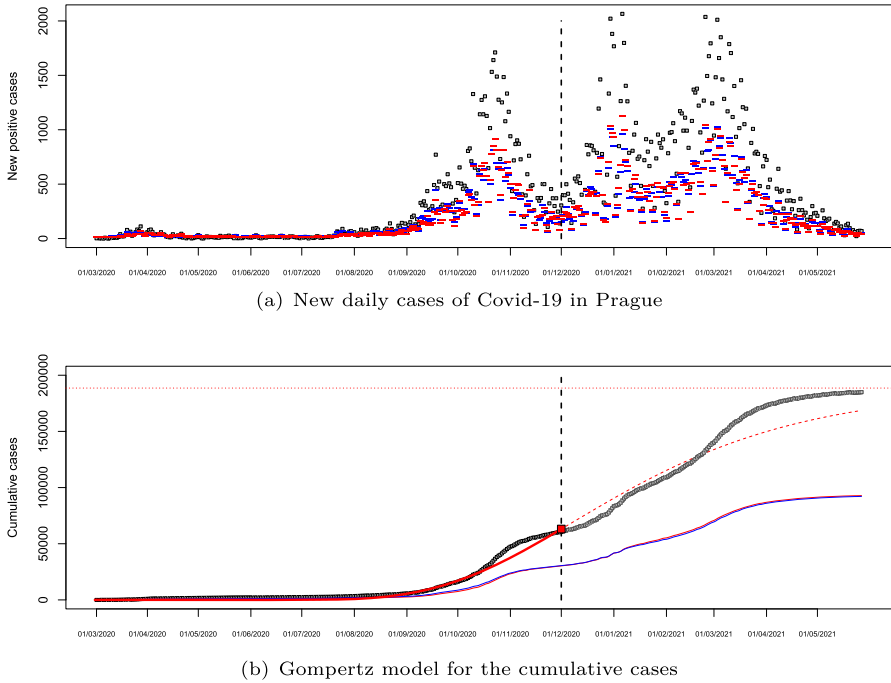
**Table 2** Empirical powers of the proposed real-time changepoint test based on 5000 Monte Carlo simulations given for various simulation settings

Dist.	$m$	$k_m^0$	Scenario 1 $T_m = 10$	Scenario 2 $T_m = m/2$	Scenario 3 $T_m = m \log(m)$
$N(0, 1)$ $\tau = 0.5000$	20	$k_m^{(1)}$	2.79% [0.77   0.78]	2.79% [0.77   0.78]	2.44% [0.72   0.71]
		$k_m^{(2)}$	12.80% [0.31   0.22]	12.80% [0.31   0.22]	20.40% [0.19   0.12]
	50	$k_m^{(1)}$	4.10% [0.73   0.76]	4.10% [0.76   0.78]	5.20% [0.93   0.93]
		$k_m^{(2)}$	13.30% [0.33   0.11]	14.60% [0.44   0.46]	35.30% [0.32   0.25]
	200	$k_m^{(1)}$	6.52% [0.87   0.89]	8.99% [0.80   0.82]	21.76% [0.83   0.85]
		$k_m^{(2)}$	16.70% [0.46   0.44]	47.60% [0.46   0.49]	96.40% [0.15   0.10]
$N(1, 1)$ $\hat{\tau} = 0.0719$	20	$k_m^{(1)}$	2.51% [0.71   0.67]	2.51% [0.71   0.67]	2.42% [0.62   0.55]
		$k_m^{(2)}$	11.20% [0.37   0.39]	11.20% [0.37   0.39]	23.30% [0.29   0.22]
	50	$k_m^{(1)}$	4.20% [0.83   0.83]	4.30% [0.83   0.83]	5.00% [0.85   0.89]
		$k_m^{(2)}$	12.20% [0.43   0.46]	13.30% [0.49   0.33]	32.30% [0.48   0.58]
	200	$k_m^{(1)}$	5.52% [0.82   0.84]	7.20% [0.55   0.55]	21.71% [0.83   0.85]
		$k_m^{(2)}$	14.21% [0.42   0.44]	38.00% [0.48   0.49]	93.80% [0.18   0.15]
$L(0, 1)$ $\tau = 0.5000$	20	$k_m^{(1)}$	2.64% [0.77   0.78]	2.64% [0.77   0.78]	2.29% [0.74   0.78]
		$k_m^{(2)}$	11.50% [0.36   0.22]	11.50% [0.36   0.22]	21.20% [0.24   0.16]
	50	$k_m^{(1)}$	3.95% [0.75   0.78]	4.10% [0.78   0.79]	4.99% [0.81   0.86]
		$k_m^{(2)}$	12.90% [0.43   0.44]	13.20% [0.48   0.50]	37.70% [0.30   0.21]
	200	$k_m^{(1)}$	6.70% [0.75   0.67]	6.93% [0.80   0.79]	24.04% [0.81   0.83]
		$k_m^{(2)}$	15.60% [0.40   0.47]	37.90% [0.46   0.47]	96.40% [0.18   0.12]

The change in  $\beta_2$  occurs either in the first half of the online data for  $k_m^{(1)} = \lfloor T_m/2 \rfloor + 1$ , or the change occurs at the very beginning of the online data, thus  $k_m^{(2)} = 1$ . The empirical powers for different  $T_m$ , different error distributions, and three sizes of the historical data are given in terms of the relative proportions (using  $\tau = 0.5$  for the symmetric distributions and  $\hat{\tau} = 0.0719$  for the asymmetric distribution). Two changepoint indicators are also used in the brackets: the average changepoint location index and the median changepoint location index. Values close to zero stand for an early changepoint detection (zero standing for the detection at the first available observation) and values close to one mean late changepoint discoveries (one standing for the detection at the last available observation)

the fact that the average and median *changepoint location indicators* in the brackets are always greater than 0.50—which stands for the half of the *online data* sequence.

The average changepoint location indicator of, let us say, 0.25 indicates that the changepoint was estimated (when averaged over all simulations) after the first quarter of the *online data*. If the median location indicator (the second value in the brackets) is higher than the average, then the majority of the changepoint recoveries occurred after the first quarter, but there were also some relatively rear although very early recoveries (including also the very first online observation). On the other hand, for the median location indicator being smaller than the average indicator, the majority of



**Fig. 2** Covid-19 positive cases in Prague, Czech Republic. The overall daily increments in the upper panel and the cumulative counts in the lower panel are—just for better illustration—provided also separately for males (blue) and females (red). The vertical lines represent the date when the strict pandemic restrictions being in effect before Christmas 2020 were relaxed. The Gompertz population model in (10) is fitted on the *historical data*—the data before the restrictions release in December 1, 2020. The projection of the model is provided for the future in dashed red. The estimated saturation of  $\hat{K} = 189\,616$  is visualized in dotted red

the changepoint recoveries occurred before the first quarter, but there were also some very late recoveries (including the very last observations).

### 4.2 Covid-19 prevalence

Relatively recently, the world society was very much effected by the Covid-19 pandemic, therefore, we tried to apply the proposed estimation and changepoint detection method for a nonlinear parametric population risk model—a three parameter Gompertz curve—to model the cumulative counts of the Covid-19 positive cases in Prague, the capital of the Czech Republic, over the period from the first positive case appearance (March 1, 2020) until the end of May 2021. The data, provided for academic purposes by the Institute of Health Information and Statistics of the Czech Republic are assumed to follow a typical nonlinear (growth) model in (1), where

$$f(X_i, \beta) = K \exp \left\{ -\beta_1 e^{-\beta_2 x_i} \right\} \tag{10}$$

**Table 3** Parameter estimates for the underlying Gompertz model in (10) for three different data scenarios: firstly, historical data until restrictions release are considered; Second, the proposed online testing procedure is applied until the null hypothesis is rejected and the model is retrained. Finally, the all available data are used to estimate the overall model

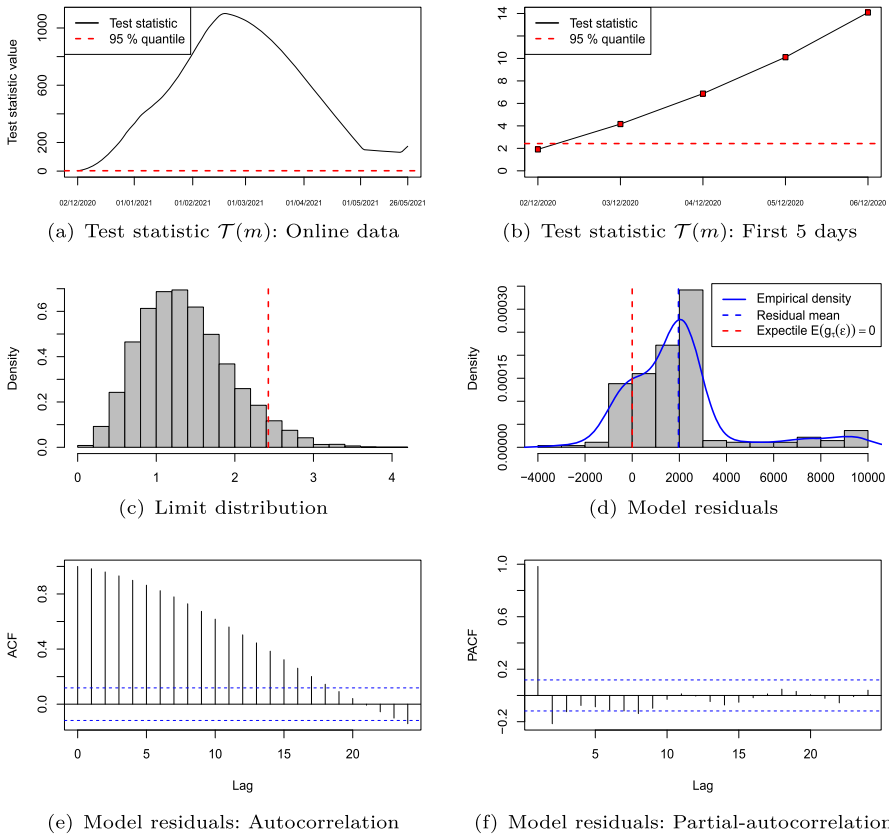
Estimation method and data	$\beta_1$	$\beta_2$	$K$	Objective function
<i>Historical data (until 01/12/2020)</i>				
Symmetric least squares ( $\tau = 0.50$ )	36.04	0.0129	187 811	$3.00 \times 10^6$
Expectile method ( $\tau = 0.11$ )	37.97	0.0129	188 576	$0.78 \times 10^6$
<i>Re-estimation after the change detection (03/12/2020)</i>				
Symmetric least squares ( $\tau = 0.50$ )	35.38	0.0130	180 174	$3.27 \times 10^6$
Expectile method ( $\tau = 0.19$ )	36.04	0.0129	179 718	$1.92 \times 10^6$
<i>All available data (until 26/05/2021)</i>				
Symmetric least squares ( $\tau = 0.50$ )	20.17	0.0096	256 970	$10.9 \times 10^6$
Expectile method ( $\tau = 0.26$ )	20.18	0.0095	255 032	$8.53 \times 10^6$

for the unknown parameter vector  $\boldsymbol{\beta} = (\beta_1, \beta_2, K)^\top \in \mathbb{R}_+^3$ . The univariate explanatory variables  $X_i \equiv x_i$  stand for the current day and the dependent random variables  $Y_i$  in (1) reflect the cumulative Covid-19 positive cases at the given day. A similar population growth model—a five parameter logistic curve—was recently applied in Chen et al (2020) to predict the overall number of positive Covid-19 cases in the US. The resulting model, however, turned out to heavily underestimate the true number of positive cases, which could be also caused by the underlying distributional symmetry assumption.

In our approach, instead of trying to predict the overall positive cases, we pursue a slightly different goal: Firstly, the data are split into two parts—the *historical data* from the very first Covid-19 positive case in Prague until December 1, 2020 (when a rather populist and quite much criticized government decision waved off some of the strict pandemic restrictions before Christmas) and the *online data*—arriving after December 1, 2020. Second, the proposed changepoint test is adopted to test whether the model before the government decision and the model after the government decision is the same, or not. Finally, the model can be also used to get some predictions of the overall Covid-19 positive cases over the overall follow-up period.

The data—daily positive cases—are visualized in Fig. 2a. The corresponding cumulative counts are given in the panel below—Fig. 2b. The Gompertz model from (10) is fitted on the *historical data* thus, the period from March 1, 2020 until December 1, 2020. The estimated parameters are provided in Table 3. The estimated number of the overall Covid-19 positive cases is  $\hat{K} = 188\,576$ , while the true number of all positive cases reported until May 26, 2021, is 184 959.

The proposed changepoint detection test based on (7) is performed to verify the stability of the model trained on the *historical data*, for  $m = 275$ , while new *online data* are arriving in a step-by-step manner (for  $T_m = 176$ ). The values for the test statistic in (7) at each step of the online testing regime are plotted in Fig. 3a. The null hypothesis of no changepoint in the vector parameter  $\boldsymbol{\beta} = (\beta_1, \beta_2, K)^\top$  is rejected relatively



**Fig. 3** The test statistic profile for the *online data* in panel (a) and the first five days only for a more detailed insight in panel (b); The limit distribution from Theorem 2 with the corresponding 95% sample quantile  $c_{0.95}(\gamma) = 2.4260$ , for  $\gamma = 0.1$ , in panel (c); The model residuals from (10) with the corresponding density estimate, the empirical mean, and the empirical expectile for  $\hat{\tau} = 0.11$ , such that the empirical counterpart of  $E[g_{\tau}(\epsilon)]$  equals to zero—all in panel (d); Finally, the residual autocorrelation and partial-autocorrelation plots in panels (e) and (f) respectively

fast—just two days after the government reduced—the restrictions—the corresponding test statistic is  $T(m) = 4.1618$  for  $m = 275$  and the corresponding 95% quantile of the limit distribution from Theorem 2 is  $c_{0.95}(\gamma) = 2.4260$  for  $\gamma = 0.1$ . This may suggest that the actual change in the model occurred already before the online data—which can be also seen in Fig. 2—either from the first peak and the consecutive drop-off in panel (a) or some evident underestimation at the end of the historical data in panel (b). The estimated parameters for the retrained model after the changepoint detection are, for comparison, also reported in Table 3. Alternatively, one could also consider another set of the historical data (and maybe slightly more representative)—from the very first case until the first culmination (i.e., the beginning of November 2020, thus  $m = 245$ ) and to test whether the model changes significantly after the peak as the daily Covid-19 cases start to decrease. The estimated parameters are very similar ( $\hat{\beta}_1 = 88.15$ ,  $\hat{\beta}_2 = 0.0166$ , and  $\hat{K} = 197264$ ) but it takes 8 days for

the proposed test statistic to detect a significant change in the model. Nevertheless, despite some obvious correlation among the model-based residuals (Fig. 3d and e) the estimated model seems to be relatively stable and the proposed changepoint detection test performs very well.

## 5 Conclusions

In this paper, we proposed the *online procedure* for testing stability of a nonlinear parametric regression model while taking into account the conditional expectile estimation framework. There are three main pivots behind the proposed methodology: Firstly, the nonlinear parametric form of the unknown regression function improves the overall flexibility of the model while the dependence on the unknown parameters still preserves a relatively simple and straightforward interpretation of the overall regression function estimate. Second, the expectile estimation method allows for some additional robustness especially with respect to asymmetric distributions. The estimation algorithm depends on the “asymmetry index”  $\tau \in (0, 1)$ , which is usually unknown, but it can be either anticipated from the data generating mechanism or some plug-in estimate can be used instead. Third, the online regime for the changepoint detection makes the proposed method instantly applicable, which may turn out to be convenient in situations when real-time decisions and model adaptations are required. Finally, given the underlying regression function, the whole minimization problem formulated in (4) does not have to be convex—therefore, we proposed a widely applicable general iterative grid search algorithm which can be effectively used in practical applications.

The proposed methodological framework enriches the class of *online procedures* for changepoint detections. To our best knowledge, the specific model setup considered in this paper has not been studied in the literature yet. The empirical performance is illustrated through an extensive simulation study. A practical applicability of the whole methodological framework is illustrated on a real data example concerning some of the most recent challenges related to online decision making—especially essential decisions related to the Covid-19 pandemics made by local and global authorities.

**Acknowledgements** The research of Matúš Maciak and Michal Pešta was supported by the Czech Science Foundation project GAČR No. 21-10768 S.

**Funding** Open access publishing supported by the National Technical Library in Prague.

## Declarations

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.



### Appendix A Proofs and technical lemma

**Proof of Proposition 1** The expectile estimator  $\widehat{\beta}_m$  from (4) is the solution of the system of equations

$$\frac{\partial}{\partial \beta} \sum_{i=1}^m \rho_{\tau}(Y_i - f(\mathbf{X}_i, \beta)) = \mathbf{0}_p,$$

where  $\mathbf{0}_p$  stands for a zero vector of dimension  $p$ . Then, with a probability equal to 1, it holds that

$$\sum_{i=1}^m \nabla f(\mathbf{X}_i, \widehat{\beta}_m) g_{\tau}(\varepsilon_i + f(\mathbf{X}_i, \beta^0) - f(\mathbf{X}_i, \widehat{\beta}_m)) = \mathbf{0}_p. \tag{A1}$$

Since under Assumptions (A), (B), (C) we have  $\widehat{\beta}_m \xrightarrow[m \rightarrow \infty]{a.s.} \beta^0$  (see Theorem 2.2 of Choi et al (2003)) and since for  $t \rightarrow 0$  we have  $g_{\tau}(\varepsilon - t) = g_{\tau}(\varepsilon) - h_{\tau}(\varepsilon)t + o_{\mathbb{P}}(t)$ , then using Assumptions (A1) and (A2), we obtain

$$g_{\tau}(\varepsilon_i + f(\mathbf{X}_i, \beta^0) - f(\mathbf{X}_i, \widehat{\beta}_m)) = g_{\tau}(\varepsilon_i) - h_{\tau}(\varepsilon_i)(f(\mathbf{X}_i, \widehat{\beta}_m) - f(\mathbf{X}_i, \beta^0)) + o_{\mathbb{P}}(f(\mathbf{X}_i, \widehat{\beta}_m) - f(\mathbf{X}_i, \beta^0)).$$

Substituting into the relation in (A1) we obtain

$$\mathbf{0}_p = \sum_{i=1}^m \nabla f(\mathbf{X}_i, \widehat{\beta}_m) \left[ g_{\tau}(\varepsilon_i) - h_{\tau}(\varepsilon_i)(f(\mathbf{X}_i, \widehat{\beta}_m) - f(\mathbf{X}_i, \beta^0)) + o_{\mathbb{P}}(f(\mathbf{X}_i, \widehat{\beta}_m) - f(\mathbf{X}_i, \beta^0)) \right]$$

and, since  $\widehat{\beta}_m \xrightarrow[m \rightarrow \infty]{a.s.} \beta^0$ , using also Assumption (A2) and the Taylor expansion, we have, for any  $j = 1, \dots, p$ , that

$$0 = \sum_{i=1}^m \left[ \frac{\partial f}{\partial \beta_j}(\mathbf{X}_i, \beta^0) + \nabla_j^2 f^{\top}(\mathbf{X}_i, \beta^0 + \theta_{ji}(\widehat{\beta}_m - \beta^0))(\widehat{\beta}_m - \beta^0) \right] \times \left[ g_{\tau}(\varepsilon_i) - h_{\tau}(\varepsilon_i)(f(\mathbf{X}_i, \widehat{\beta}_m) - f(\mathbf{X}_i, \beta^0)) + o_{\mathbb{P}}(f(\mathbf{X}_i, \widehat{\beta}_m) - f(\mathbf{X}_i, \beta^0)) \right],$$

with  $\nabla_j^2 f$  the column  $p$ -vector  $(\partial^2 f / (\partial \beta_j \partial \beta_k))_{1 \leq k \leq p}$  of derivatives of order 2 and  $\theta_{ji}$  a random variable such that  $0 \leq \theta_{ji} \leq 1$  with probability 1. Hence, using Assumptions (A1), (A2), (C), and (E), together with the Cauchy–Schwarz inequality and the Taylor

expansion, we obtain

$$\mathbf{0}_p = \frac{1}{m} \sum_{i=1}^m \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0) g_\tau(\varepsilon_i) - \frac{1}{m} \sum_{i=1}^m h_\tau(\varepsilon_i) \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0) \nabla f^\top(\mathbf{X}_i, \boldsymbol{\beta}^0) (\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}^0) + o_{\mathbb{P}}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}^0). \tag{A2}$$

By Assumption (C) we can define a  $p$ -square invertible matrix

$$\boldsymbol{\Omega}_m \equiv m^{-1} \sum_{i=1}^m h_\tau(\varepsilon_i) \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0) \nabla f^\top(\mathbf{X}_i, \boldsymbol{\beta}^0)$$

and using the relation in (A2) we obtain

$$\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}^0 = \boldsymbol{\Omega}_m^{-1} \frac{1}{m} \sum_{i=1}^m \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0) g_\tau(\varepsilon_i) (1 + o_{\mathbb{P}}(1)), \tag{A3}$$

or, again, using Assumptions (C) and (D) to get

$$\widehat{\boldsymbol{\beta}}_m = \boldsymbol{\beta}^0 + \boldsymbol{\Omega}^{-1} \frac{1}{m} \sum_{i=1}^m \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0) g_\tau(\varepsilon_i) (1 + o_{\mathbb{P}}(1)),$$

which proves the given proposition. □

In order to show the asymptotic behavior of the test statistic under the null hypothesis in (5) and the alternative hypothesis in (6) let us define a stochastic process

$$\mathbf{r}_{m,k}(\mathbf{u}) \equiv \sum_{i=m+1}^{m+k} \mathbf{R}_i(\mathbf{u}), \tag{A4}$$

for  $k = 1, \dots, T_m$ , where  $\mathbf{u} \in \mathbb{R}^p$ , such that  $\|\mathbf{u}\|_2 \leq C$  for some constant  $C < \infty$ . In addition, let

$$\mathbf{R}_i(\mathbf{u}) \equiv \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u}) g_\tau(Y_i - f(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u})) - \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0) g_\tau(\varepsilon_i),$$

for  $i = m + 1, \dots, m + T_m$ , where the convergence rate of  $\widehat{\boldsymbol{\beta}}_m$  derived above is used. The following lemma is crucial for the proofs of the main theorems.

**Lemma 4** *Let Assumptions (A1), (A2), (D), and (E) be satisfied and let the null hypothesis in (5) hold. Then, for any constants  $C_1, C_2 > 0$  and all  $k \in \mathbb{N}$  large enough, there exists a constant  $C_3 > 0$  such that*

$$\mathbb{P} \left[ \sup_{\|\mathbf{u}\|_2 \leq C_1} \|\mathbf{r}_{m,k}(\mathbf{u}) - \mathbb{E}[\mathbf{r}_{m,k}(\mathbf{u})]\|_1 \geq p C_2 C_3 \sqrt{\frac{k}{m}} \sqrt{\log k} \right] \leq 4k^{-C_2^2}$$

for  $m \in \mathbb{N}$  sufficiently large.

**Proof of Lemma 4** For any observation  $i \in \{m + 1, \dots, m + T_m\}$ , any vector  $\mathbf{u} = (u_1, \dots, u_p) \in \mathbb{R}^p$  such that  $\|\mathbf{u}\|_2 \leq C_1$ , we can express  $\mathbf{R}_i(\mathbf{u})$  as

$$\begin{aligned} \mathbf{R}_i(\mathbf{u}) &= \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u}) \left[ g_\tau(Y_i - f(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u})) - g_\tau(\varepsilon_i) \right] \\ &\quad + g_\tau(\varepsilon_i) \left[ \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u}) - \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0) \right] \equiv \mathbf{B}_i^{(1)}(\mathbf{u}) + \mathbf{B}_i^{(2)}(\mathbf{u}). \end{aligned} \tag{A5}$$

Both terms in (A5) will be studied separately. Let us start with  $\mathbf{B}_i^{(1)}(\mathbf{u})$ . The function  $g_\tau$  can be expressed as  $g_\tau(x) = 2\tau x + 2x(1 - 2\tau)\mathbb{I}_{\{x < 0\}}$ . Using the notation  $v_i \equiv f(\mathbf{X}_i, \boldsymbol{\beta}^0) - f(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u})$  it holds with probability 1 that

$$\begin{aligned} g_\tau(Y_i - f(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u})) - g_\tau(\varepsilon_i) &= 2\tau v_i + 2(1 - 2\tau) \left[ (\varepsilon_i + v_i)\mathbb{I}_{\{\varepsilon_i < -v_i\}} - \varepsilon_i\mathbb{I}_{\{\varepsilon_i < 0\}} \right]. \end{aligned} \tag{A6}$$

Let us consider a random variable  $W_i \equiv \varepsilon_i [\mathbb{I}_{\{\varepsilon_i < -v_i\}} - \mathbb{I}_{\{\varepsilon_i < 0\}}]$ . It holds that

$$|W_i| = |\varepsilon_i| \mathbb{I}_{\{\min(0, -v_i) \leq \varepsilon_i < \max(0, -v_i)\}},$$

which also implies

$$\mathbb{P}[|W_i| < |v_i|] = 1. \tag{A7}$$

The relation in (A6) can be written as

$$g_\tau(\varepsilon_i + v_i) - g_\tau(\varepsilon_i) = 2\tau v_i + 2(1 - 2\tau) [W_i + v_i \mathbb{I}_{\{\varepsilon_i < -v_i\}}], \tag{A8}$$

which holds with probability one. On the other hand, by Assumptions (A1) and (A2) we have that  $\|\nabla f(\mathbf{x}, \boldsymbol{\beta})\|_2$  is bounded for all  $\mathbf{x} \in \mathcal{X}$  and all  $\boldsymbol{\beta} \in \mathcal{V}_m(\boldsymbol{\beta}^0, \mathbf{u})$ , with  $\mathcal{V}_m(\boldsymbol{\beta}^0, \mathbf{u}) = \{\boldsymbol{\beta}; \|\boldsymbol{\beta} - \boldsymbol{\beta}^0\|_2 \leq m^{-1/2}\|\mathbf{u}\|_2\}$ . For the right-hand side of (A8), using the fact that  $\|\nabla f(\mathbf{x}, \boldsymbol{\beta})\|_2$  is bounded, together with relation (A7), by applying the first order Taylor expansion to  $f(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u})$ , we have that there exists  $C > 0$  such that we can write for the left-hand side of (A8)

$$\mathbb{P}[-Cm^{-1/2} < g_\tau(\varepsilon_i + v_i) - g_\tau(\varepsilon_i) < Cm^{-1/2}] = 1. \tag{A9}$$

Using the relations in (A6), (A8), and (A9), together with the fact that  $\|\nabla f(\mathbf{x}, \boldsymbol{\beta})\|_2$  is bounded for all  $\mathbf{x} \in \mathcal{X}$  and  $\boldsymbol{\beta} \in \mathcal{V}_m(\boldsymbol{\beta}^0, \mathbf{u})$  and applying the Hoeffding inequality, we obtain that

$$\mathbb{P} \left[ \left| \sum_{i=m+1}^{m+k} \frac{\partial f}{\partial \beta_j}(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u}) G_\tau(\varepsilon_i, v_i) \right| \geq t \right] \leq 2 \exp \left( -\frac{2t^2}{4C^2k/m} \right),$$

for all  $t \in \mathbb{R}$  and  $j = 1, \dots, p$ , where, for brevity, use used the notation

$$G_\tau(\varepsilon_i, v_i) = [g_\tau(\varepsilon_i - v_i) - g_\tau(\varepsilon_i)] - \mathbb{E}[g_\tau(\varepsilon_i - v_i) - g_\tau(\varepsilon_i)].$$

Next, similarly as in the proof of Lemma 1 in Ciuperca (2022), under Assumptions (A1) and (A2), using the last relation above, we have that for all constants  $C_1, C_4 > 0$ , there exists a constant  $\tilde{C} > 0$  such that

$$\begin{aligned} & \mathbb{P} \left[ \left\| \sum_{i=m+1}^{m+k} (\mathbf{B}_i^{(1)}(\mathbf{u}) - \mathbb{E}[\mathbf{B}_i^{(1)}(\mathbf{u})]) \right\|_1 \geq \sqrt{2}C_4\tilde{C}p\sqrt{\frac{k}{m}}\sqrt{\log k} \right] \\ & \leq \mathbb{P} \left[ \max_{1 \leq j \leq p} \left| \sum_{i=m+1}^{m+k} \frac{\partial f}{\partial \beta_j}(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u})G_\tau(\varepsilon_i, v_i) \right| \geq \sqrt{2}C_4\tilde{C}\sqrt{\frac{k}{m}}\sqrt{\log k} \right] \\ & \leq 2k^{-C_4^2}. \end{aligned} \tag{A10}$$

Next, we proceed by studying the random vector  $\mathbf{B}_i^{(2)}(\mathbf{u})$  from the relation in (A5). Let us denote its  $j$ -th elemen, for  $j = 1, \dots, p$ , as

$$\begin{aligned} B_{ij}^{(2)}(\mathbf{u}) &= \left[ \frac{\partial f}{\partial \beta_j}(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u}) - \frac{\partial f}{\partial \beta_j}(\mathbf{X}_i, \boldsymbol{\beta}^0) \right] g_\tau(\varepsilon_i) \\ &= m^{-1/2}\mathbf{u}\nabla_j^2 f(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\theta_{ji}\mathbf{u})g_\tau(\varepsilon_i), \end{aligned}$$

for some constants  $\theta_{ji} \in [0, 1]$ . Under Assumption (D), it holds that  $\text{Var}[g_\tau(\varepsilon_i)] < \infty$  and also  $\mathbb{E}[g_\tau^2(\varepsilon_i)] < C < \infty$ . Using these two relations, we obtain that

$$\text{Var}[B_{ij}^{(2)}(\mathbf{u})] \leq \mathbb{E}[(B_{ij}^{(2)}(\mathbf{u}))^2] \leq Cm^{-1}\|\mathbf{u}\|_2^2\|\nabla_j^2 f(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\theta_{ji}\mathbf{u})\|_2^2,$$

for each  $j \in \{1, \dots, p\}$ , where the Cauchy–Schwarz inequality was applied in the last step. Moreover, using Assumption (E) we have that for all  $C_1 > 0$  there exists a constant  $C_5 > 0$  such that

$$\max_{m+1 \leq i \leq m+k} \left( \max_{\|\mathbf{u}\|_2 \leq C_1} \|\nabla_j^2 f(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u})\|_2 \right) \leq C_5.$$

Therefore, we also obtain that

$$\sum_{i=m+1}^{m+k} \text{Var}[B_{ij}^{(2)}(\mathbf{u})] = O(k/m). \tag{A11}$$

For  $\mathbf{B}_i^{(2)}(\mathbf{u})$ , taking into account the relation in (A11) and the fact that  $B_{ij}^{(2)}(\mathbf{u})$  is uniformly bounded by  $Cm^{-1/2}$ , we can use Lemma 4.1 of Ciuperca (2017) for

$\delta_k = k/m$ . Then, for  $k \in \mathbb{N}$  being sufficiently large, for any constant  $C_6 > 0$  and  $\|\mathbf{u}\|_2 < C_1$ , we have

$$\mathbb{P} \left[ \left| \sum_{i=m+1}^{m+k} [B_{ij}^{(2)}(\mathbf{u}) - \mathbb{E}[B_{ij}^{(2)}(\mathbf{u})]] \right| \geq (1 + C_6) \sqrt{\frac{k}{m}} \sqrt{\log k} \right] \leq 2k^{-C_6},$$

which further implies, similarly as for (A10), that

$$\mathbb{P} \left[ \left\| \sum_{i=m+1}^{m+k} [\mathbf{B}_i^{(2)}(\mathbf{u}) - \mathbb{E}[\mathbf{B}_i^{(2)}(\mathbf{u})]] \right\|_1 \geq p(1 + C_6) \sqrt{\frac{k}{m}} \sqrt{\log k} \right] \leq 2k^{-C_6}. \tag{A12}$$

Taking now  $C_6 = C_4^2$  and  $\tilde{C} = 2^{-1/2}(1 + C_4^2)/C_4$  we get

$$\mathbb{P} \left[ \left\| \sum_{i=m+1}^{m+k} [\mathbf{B}_i^{(2)}(\mathbf{u}) - \mathbb{E}[\mathbf{B}_i^{(2)}(\mathbf{u})]] \right\|_1 \geq \sqrt{2} p \tilde{C} C_4 \sqrt{\frac{k}{m}} \sqrt{\log k} \right] \leq 2k^{-C_4^2}. \tag{A13}$$

Moreover, for any constant  $c > 0$  and any two random vectors  $\mathbf{V}_1$  and  $\mathbf{V}_2$  of the same size, it holds that

$$\begin{aligned} \mathbb{P}[\|\mathbf{V}_1 + \mathbf{V}_2\|_1 \geq c] &\leq \mathbb{P}[\|\mathbf{V}_1\|_1 + \|\mathbf{V}_2\|_1 \geq c] \leq \mathbb{P} \left[ \left( \|\mathbf{V}_1\|_1 \geq \frac{c}{2} \right) \cup \left( \|\mathbf{V}_2\|_1 \geq \frac{c}{2} \right) \right] \\ &\leq \mathbb{P} \left[ \|\mathbf{V}_1\|_1 \geq \frac{c}{2} \right] + \mathbb{P} \left[ \|\mathbf{V}_2\|_1 \geq \frac{c}{2} \right]. \end{aligned} \tag{A14}$$

For the first inequality in (A14) we used the fact that for any constant  $c > 0$  the event  $c \leq \|\mathbf{V}_1 + \mathbf{V}_2\|_1 \leq \|\mathbf{V}_1\|_1 + \|\mathbf{V}_2\|_1$  implies, with probability one, also the random event

$$(\|\mathbf{V}_1 + \mathbf{V}_2\|_1 \geq c) \subseteq (\|\mathbf{V}_1\|_1 + \|\mathbf{V}_2\|_1 \geq c) \subseteq \left( \max(\|\mathbf{V}_1\|_1, \|\mathbf{V}_2\|_1) \geq c/2 \right),$$

where we considered  $\mathbf{V}_1 = \mathbf{B}_i^{(1)}(\mathbf{u}) - \mathbb{E}[\mathbf{B}_i^{(1)}(\mathbf{u})]$  and  $\mathbf{V}_2 = \mathbf{B}_i^{(2)}(\mathbf{u}) - \mathbb{E}[\mathbf{B}_i^{(2)}(\mathbf{u})]$ . Taking now the relations in (A10), (A13) and (A14), and the constants  $C_4 = C_2$ ,  $C_3 = 2^{-3/2} \min(\tilde{C}, \tilde{C})$ , and  $c = pC_2C_3(k/m)^{1/2}(\log k)^{1/2}$ , we obtain that

$$\mathbb{P} \left[ \sum_{i=m+1}^{m+k} \|\mathbf{R}_i(\mathbf{u}) - \mathbb{E}[\mathbf{R}_i(\mathbf{u})]\|_1 \geq pC_2C_3 \sqrt{\frac{k}{m}} \sqrt{\log k} \right] \leq 4k^{-C_2^2},$$

for all  $C_2 > 0$ , which completes the proof of the lemma. □

**Proof of Theorem 2** If  $k$  is fixed then  $z(m, k, \gamma) \xrightarrow{m \rightarrow \infty} \infty$ . With respect to Assumptions (A1), (A2), (B), and (D), we have that

$$\left\| \mathbf{J}_m^{-1/2}(\widehat{\boldsymbol{\beta}}_m) \sum_{i=m+1}^{m+k} \nabla f(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}_m) g_\tau(\widehat{\varepsilon}_i) \right\|_\infty \xrightarrow{m \rightarrow \infty} 0.$$

Therefore, in this case, we also obtain that

$$\frac{\|\mathbf{S}(m, k)\|_\infty}{z(m, k, \gamma)} \xrightarrow{m \rightarrow \infty} 0.$$

If  $k \rightarrow \infty$  (i.e.,  $k \equiv k_m \rightarrow \infty$  for  $m \rightarrow \infty$ ), then by Lemma 4 we have for  $\mathbf{u} \in \mathbb{R}^p$ , with  $\|\mathbf{u}\|_1 \leq C_1$ :  $\mathbf{r}_{m,k}(\mathbf{u}) = \mathbb{E}[\mathbf{r}_{m,k}(\mathbf{u})] + O_{\mathbb{P}}(m^{-1/2}k^{1/2}(\log k)^{1/2})$ .

On the other hand, for  $\mathbb{E}[\mathbf{r}_{m,k}(\mathbf{u})]$  we have, using Assumptions (A1), (A2), (B), (D), and (E), that

$$\begin{aligned} & \mathbb{E}[\mathbf{r}_{m,k}(\mathbf{u})] \\ &= \sum_{i=m+1}^{m+k} \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u}) \mathbb{E}\left[ g_\tau(\varepsilon_i) - h_\tau(\varepsilon_i) (f(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u}) - f(\mathbf{X}_i, \boldsymbol{\beta}^0)) \right. \\ & \quad \left. + o_{\mathbb{P}}(f(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u}) - f(\mathbf{X}_i, \boldsymbol{\beta}^0)) \right] \\ &= - \sum_{i=m+1}^{m+k} \left[ \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u}) m^{-1/2} \nabla f^\top(\mathbf{X}_i, \boldsymbol{\beta}^0) \mathbf{u} \mathbb{E}[h_\tau(\varepsilon)] + o_{\mathbb{P}}(m^{-1/2}) \right] \\ &= -\mathbb{E}[h_\tau(\varepsilon)] m^{-1/2} \sum_{i=m+1}^{m+k} \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0) \nabla f^\top(\mathbf{X}_i, \boldsymbol{\beta}^0) \mathbf{u} + o(km^{-1/2}). \end{aligned}$$

Thus, we also have that

$$\begin{aligned} \mathbf{r}_{m,k}(\mathbf{u}) &= -m^{-1/2} \mathbb{E}[h_\tau(\varepsilon)] \sum_{i=m+1}^{m+k} \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0) \nabla f^\top(\mathbf{X}_i, \boldsymbol{\beta}^0) \mathbf{u} \\ & \quad + O_{\mathbb{P}}(m^{-1/2}k^{1/2}(\log k)^{1/2}). \end{aligned} \tag{A15}$$

Using the relation in (A3) and taking  $\mathbf{u} = m^{1/2}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}^0)$  in the relation in (A15), we obtain

$$\begin{aligned} \mathbf{r}_{m,k}(m^{1/2}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}^0)) &= -m^{-1} \mathbf{V}_m^{-1}(\boldsymbol{\beta}^0) \left( \sum_{i=m+1}^{m+k} \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0) \nabla f^\top(\mathbf{X}_i, \boldsymbol{\beta}^0) \right) \\ & \quad \times \left( \sum_{i=1}^m \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0) g_\tau(\varepsilon_i) + o_{\mathbb{P}}(m^{1/2}) \right) \\ & \quad + O_{\mathbb{P}}(m^{-1/2}k^{1/2}(\log k)^{1/2}) \end{aligned}$$

and, using Assumption (C), also

$$\begin{aligned} \mathbf{r}_{m,k}(m^{1/2}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}^0)) &= -km^{-1} \left( \sum_{i=1}^m \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0) g_\tau(\varepsilon_i) + o_{\mathbb{P}}(m^{1/2}) \right) \\ &\quad + O_{\mathbb{P}}(m^{-1/2}k^{1/2}(\log k)^{1/2}). \end{aligned}$$

On the other hand, by the definition of the random process  $\mathbf{r}_{m,k}$ , we get, with probability one, that

$$\mathbf{r}_{m,k}(m^{1/2}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}^0)) = \sum_{i=m+1}^{m+k} [\nabla f(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}_m) g_\tau(\widehat{\varepsilon}_i) - \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0) g_\tau(\varepsilon_i)].$$

The last two relations imply

$$\begin{aligned} \sum_{i=m+1}^{m+k} \nabla f(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}_m) g_\tau(\widehat{\varepsilon}_i) &= \sum_{i=m+1}^{m+k} \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0) g_\tau(\varepsilon_i) - km^{-1} \sum_{i=1}^m \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0) g_\tau(\varepsilon_i) \\ &\quad + o_{\mathbb{P}}(km^{-1/2}) + O_{\mathbb{P}}(m^{-1/2}k^{1/2}(\log k)^{1/2}). \end{aligned} \tag{A16}$$

The rest of the proof follows the same lines as the proof of Theorem 1 in Ciuperca (2022) using the Komlós–Major–Tusnády (KMT) approximation for independent random vectors and Theorem 2.1 of Horváth et al (2004). Let us only sketch the main idea of the end of the proof. By the KMT approximation for independent random variables not identically distributed (see Götze and Zaitsev (2009)) for each component of the random vectors  $(\mathbf{J}_m^{-1/2}(\boldsymbol{\beta}^0) \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0) g_\tau(\varepsilon_i))_{1 \leq i \leq m}$  and of  $(\mathbf{J}_m^{-1/2}(\boldsymbol{\beta}^0) \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0) g_\tau(\varepsilon_i))_{m+1 \leq i \leq m+T_m}$ , we have that for all  $\nu > 3$  and  $m \rightarrow \infty$ , there exists two Wiener processes  $\{\mathbf{W}_{1,m}(t), t \in [0, \infty)\}$  and  $\{\mathbf{W}_{2,m}(t), t \in [0, \infty)\}$  of the dimension  $p$  such that for the two terms of the right hand side of the relation in (A16) it holds

$$\sup_{1 \leq k < \infty} k^{-1/\nu} \left\| \mathbf{J}_m^{-1/2}(\boldsymbol{\beta}^0) \sum_{i=m+1}^{m+k} \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0) g_\tau(\varepsilon_i) - \mathbf{W}_{1,m}(k) \right\|_\infty = o_{\mathbb{P}}(1) \tag{A17}$$

and

$$\left\| \mathbf{J}_m^{-1/2}(\boldsymbol{\beta}^0) \sum_{i=1}^m \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0) g_\tau(\varepsilon_i) - \mathbf{W}_{2,m}(m) \right\|_\infty = o_{\mathbb{P}}(m^{1/\nu}). \tag{A18}$$

By the proof of Theorem 2.1 of Horváth et al (2004), for any  $T > 0$ , we have

$$\begin{aligned} \sup_{1 \leq k < \infty} \frac{\left\| \mathbf{W}_{1,m}(k) - \frac{k}{m} \mathbf{W}_{2,m}(m) \right\|_\infty}{z(m, k, \gamma)} \\ \stackrel{\mathcal{L}}{\underset{m \rightarrow \infty}{\leq}} \max_{1 \leq k \leq mT} \sqrt{m} \frac{\left\| W_1(k/m) - (k/m) W_2(1) \right\|_\infty}{z(m, k, \gamma)} \end{aligned} \tag{A19}$$

and also

$$\max_{1 \leq k \leq mT} \sqrt{m} \frac{\|W_1(k/m) - (k/m)W_2(1)\|_\infty}{z(m, k, \gamma)} \xrightarrow[m \rightarrow \infty]{\mathcal{L}} \sup_{0 < t < L(T)} \frac{\|\mathbf{W}_p(t)\|_\infty}{t^\gamma}. \tag{A20}$$

On the other hand, by Assumptions (C) and (D), together with Proposition 1, we have that  $\mathbf{J}_m^{-1/2}(\widehat{\boldsymbol{\beta}}_m)$  converges to  $\mathbf{J}_m^{-1/2}(\boldsymbol{\beta}^0)$  with the convergence rate of the order  $m^{-1/2}$ . Thus,  $\mathbf{J}_m^{-1/2}(\widehat{\boldsymbol{\beta}}_m) = \mathbf{J}_m^{-1/2}(\boldsymbol{\beta}^0) + O_{\mathbb{P}}(m^{-1/2})$ . The proof of the theorem now follows by combining this last relation together with the relations in (A16), (A17), (A18), and (A19).  $\square$

**Proof of Theorem 3** It will be shown that there exists an observation with an index  $\tilde{k}_m \in \{m + 1, \dots, m + T_m\}$  for which  $\|\mathbf{S}(m, \tilde{k}_m)\|_\infty / z(m, \tilde{k}_m, \gamma) \xrightarrow[m \rightarrow \infty]{\mathbb{P}} \infty$ .

We consider the open-end procedure case. Let us consider  $\tilde{k}_m = k_m^0 + m^s$ , with  $s > 1$ . Since the function  $(x + 1)(x/(1 + x))^{-\gamma}$  is increasing in  $x > 0$  for  $\gamma \in [0, 1/2)$ , we have, as in the proof of Theorem 1, that there exists  $C > 0$  such that,

$$\frac{\|\mathbf{J}_m^{-1/2}(\widehat{\boldsymbol{\beta}}_m) \sum_{i=m+1}^{m+k_m^0} \nabla f(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}_m) g_\tau(\varepsilon_i + f(\mathbf{X}_i, \boldsymbol{\beta}^0) - f(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}_m))\|_\infty}{z(m, \tilde{k}_m, \gamma)} < C < \infty, \tag{A21}$$

with probability converging to 1, when  $m \rightarrow \infty$ .

It remains to study  $\sum_{i=m+k_m^0+1}^{m+\tilde{k}_m} \nabla f(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}_m) g_\tau(\varepsilon_i + f(\mathbf{X}_i, \boldsymbol{\beta}^1) - f(\mathbf{X}_i, \widehat{\boldsymbol{\beta}}_m))$  or more precisely, taking into account the convergence rate of  $\widehat{\boldsymbol{\beta}}_m$ , we are going to study

$$\sum_{i=m+k_m^0+1}^{m+\tilde{k}_m} \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u}) g_\tau(\varepsilon_i + f(\mathbf{X}_i, \boldsymbol{\beta}^1) - f(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u}))$$

with  $\mathbf{u} \in \mathbb{R}^p$ ,  $\|\mathbf{u}\|_2 < C$ . Consider the following sum

$$\begin{aligned} \sum_{i=m+k_m^0+1}^{m+\tilde{k}_m} \mathbf{R}_i(\mathbf{u}) = & - \sum_{i=m+k_m^0+1}^{m+\tilde{k}_m} \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0) g_\tau(\varepsilon_i) \\ & + \sum_{i=m+k_m^0+1}^{m+\tilde{k}_m} \left[ \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u}) g_\tau(\varepsilon_i \right. \\ & \left. + f(\mathbf{X}_i, \boldsymbol{\beta}^1) - f(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u}) \right]. \end{aligned} \tag{A22}$$



Then, taking into account Assumption (D), the expectation  $\mathbb{E} \left[ \sum_{i=m+k_m^0+1}^{m+\tilde{k}_m} \mathbf{R}_i(\mathbf{u}) \right]$  on the left hand-side of the equation in (A22) equals to

$$\sum_{i=m+k_m^0+1}^{m+\tilde{k}_m} \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u}) \mathbb{E} [g_\tau(\varepsilon_i + f(\mathbf{X}_i + \boldsymbol{\beta}^1) - f(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u}))].$$

Therefore, it remains to study  $\mathbb{E} [g_\tau(\varepsilon_i + f(\mathbf{X}_i + \boldsymbol{\beta}^1) - f(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u}))] = \mathbb{E} [g_\tau(\varepsilon_i + f(\mathbf{X}_i + \boldsymbol{\beta}^1) - f(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u})) - g_\tau(\varepsilon_i)]$ , where Assumption (D) was applied. By Lemma 2 of Gu and Zou (2016), we have that there exists a constant  $c_i$ , with  $|c_i| \in [2\underline{c}, 2\bar{c}]$ , such that  $g_\tau(\varepsilon_i + f(\mathbf{X}_i + \boldsymbol{\beta}^1) - f(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u})) - g_\tau(\varepsilon_i) = c_i [f(\mathbf{X}_i, \boldsymbol{\beta}^1) - f(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u})]$ , with probability one. Thus, we obtain

$$\begin{aligned} & \frac{1}{\tilde{k}_m - k_m^0} \left\| \sum_{i=m+k_m^0+1}^{m+\tilde{k}_m} \mathbb{E}[\mathbf{R}_i(\mathbf{u})] \right\|_\infty \\ &= \frac{1}{\tilde{k}_m - k_m^0} \left\| \sum_{i=m+k_m^0+1}^{m+\tilde{k}_m} \frac{c_i \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u})}{[f(\mathbf{X}_i, \boldsymbol{\beta}^1) - f(\mathbf{X}_i, \boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u})]^{-1}} \right\|_\infty. \end{aligned}$$

Taking into account this last relation, taking into account the fact that for all  $c_i$  such that  $|c_i| \in [2\underline{c}, 2\bar{c}]$  there exists some constant  $C > 0$  such that

$$\frac{1}{m^s} \left\| \sum_{i=m+k_m^0+1}^{m+\tilde{k}_m} c_i \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0) [f(\mathbf{X}_i, \boldsymbol{\beta}^1) - f(\mathbf{X}_i, \boldsymbol{\beta}^0)] \right\|_\infty > C > 0,$$

using the Kolmogorov strong Law of Large Numbers for independent not identically distributed random variables  $R_{ij}(\mathbf{u})$ , we finally obtain that

$$\left\| \sum_{i=m+k_m^0+1}^{m+\tilde{k}_m} \mathbf{R}_i(\mathbf{u}) \right\|_\infty = O_{\mathbb{P}}((\tilde{k}_m - k_m^0) \|\boldsymbol{\beta}^1 - \boldsymbol{\beta}^0\|_2). \tag{A23}$$

On the other hand, by the Central Limit Theorem, we also have

$$\sum_{i=m+k_m^0+1}^{m+\tilde{k}_m} \nabla f(\mathbf{X}_i, \boldsymbol{\beta}^0) g_\tau(\varepsilon_i) = O_{\mathbb{P}}(\tilde{k}_m - k_m^0)^{1/2} = O_{\mathbb{P}}(m^{s/2}). \tag{A24}$$

Therefore, the relations in (A22), (A23), and (A24) together imply

$$\frac{\|\mathbf{J}_m^{-1/2}(\boldsymbol{\beta}^0 + \frac{\mathbf{u}}{\sqrt{m}}) \sum_{i=m+k_m^0+1}^{m+\tilde{k}_m} \mathcal{N}_i\|_\infty}{z(m, \tilde{k}_m, \gamma)} = \frac{O_{\mathbb{P}}(m^{s/2}) + O_{\mathbb{P}}(m^s \|\boldsymbol{\beta}^1 - \boldsymbol{\beta}^0\|_2)}{z(m, \tilde{k}_m, \gamma)},$$

where, again for brevity,  $\mathcal{N}_i$  is defined as

$$\mathcal{N}_i = \nabla f\left(\mathbf{X}_i, \boldsymbol{\beta}^0 + \frac{\mathbf{u}}{\sqrt{m}}\right) g_\tau\left(\varepsilon_i + f(\mathbf{X}_i + \boldsymbol{\beta}^1) - f\left(\mathbf{X}_i, \boldsymbol{\beta}^0 + \frac{\mathbf{u}}{\sqrt{m}}\right)\right).$$

Moreover, it also hold that

$$\frac{O_{\mathbb{P}}(m^{s/2}) + O_{\mathbb{P}}(m^s \|\boldsymbol{\beta}^1 - \boldsymbol{\beta}^0\|_2)}{z(m, \tilde{k}_m, \gamma)} \xrightarrow{m \rightarrow \infty} \infty. \quad (\text{A25})$$

Since  $\|\mathbf{J}_m^{-1/2}(\boldsymbol{\beta}^0 + m^{-1/2}\mathbf{u}) - \mathbf{J}_m^{-1/2}(\boldsymbol{\beta}^0)\|_2 \xrightarrow{m \rightarrow \infty} 0$ , where  $\|\cdot\|_2$  is the spectral norm, the theorem follows by relations in (A21) and (A25).  $\square$

## References

- Aue A, Horváth L, Reimherr M (2009) Delay times of sequential procedures for multiple time series regression models. *J Economet* 149(2):174–190
- Barassi M, Horváth L, Zhao Y (2020) Change-point detection in the conditional correlation structure of multivariate volatility models. *J Bus Econ Stat* 38(2):340–349
- Basseville M, Nikiforov I (1993) Detection of abrupt changes: theory and applications. Prentice-Hall, Upper Saddle River
- Bellini F, Bignozzi V, Puccetti G (2018) Conditional expectiles, time consistency and mixture convexity properties. *Insur Math Econ* 82(2018):117–123
- Chambers J (1973) Fitting nonlinear models: numerical techniques. *Biometrika* 60(1):1–13
- Chen D, Chen X, Chen J (2020) Reconstructing and forecasting the covid-19 epidemic in the united states using a 5-parameter logistic growth model. *Global Health Res Policy* 32(1):1–7
- Choi S, Kim H, Lee Y (2003) Nonlinear asymmetric least squares estimators. *J Korean Statistical Soc* 32(1):47–64
- Chu CSJ, Stinchcombe M, White H (1996) Monitoring structural change. *Econometrica* 64(1):1045–1065
- Ciuperca G (2013) Two tests for sequential detection of a change-point in a nonlinear model. *J Stat Plan Infer* 143(10):1621–1834
- Ciuperca G (2017) Real time change-point detection in a nonlinear quantile model. *Seq Anal* 36(1):1–23
- Ciuperca G (2022) Real-time detection of a change-point in a linear expectile model. *Statistical Papers* (to appear)
- Fremdt S (2015) Page's sequential procedure for change-point detection in time series regression. *Statistics* 49(1):128–155
- Götze F, Zaitsev A (2009) Bounds for the rate of strong approximation in the multidimensional invariance principle. *Theory Probab Appl* 53(1):59–80
- Gu Y, Zou H (2016) High-dimensional generalizations of asymmetric least squares regression and their applications. *Ann Stat* 44(6):2661–2694
- Hoga Y (2017) Monitoring multivariate time series. *J Multivar Anal* 155(1):105–121
- Horváth L, Hušková M, Kokoszka P et al (2004) Monitoring changes in linear models. *J Stat Plan Infer* 126(1):225–251
- Hušková M, Prášková Z, Steinebach J (2007) On the detection of changes in autoregressive time series. *J Stat Plan Infer* 137(4):1243–1259
- Kim M, Lee S (2016) Nonlinear expectile regression with application to value-at-risk and expected shortfall estimation. *Comput Stat Data Anal* 94(1):1–19
- Lee Y, Lee S (2019) CUSUM test for general nonlinear integer-valued GARCH models: comparison study. *Ann Inst Stat Math* 71(5):1033–1057
- Nedényi F (2018) An online change detection test for parametric discrete-time stochastic processes. *Seq Anal* 37(2):246–267
- Newey W, Powell J (1987) Asymmetric least squares estimation and testing. *Econometrica* 55(4):819–847

- Philipps C (2022) Interpreting expectiles. SSRN, pp 1–39. <https://doi.org/10.2139/ssrn.3881402>
- Seber G, Wild C (2003) Nonlinear regression. Wiley, New York
- Zhou M, Wang H, Tang Y (2015) Sequential change point detection in linear quantile regression models. *Statist Probab Lett* 100(1):98–103
- Ziegel J (2016) Coherence and elicibility. *Math Financ* 26(4):901–918

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.