Detection of breaks in weak location time series models with quasi-Fisher scores

Christian Francq*, Lorenzo Trapani† and Jean-Michel Zakoïan‡

March 4, 2024

Abstract

Based on Godambe's theory of estimating functions, we propose a class of cumulative sum, CUSUM, statistics to detect breaks in the dynamics of time series under weak assumptions. First, we assume a parametric form for the conditional mean, but make no specific assumption about the data-generating process (DGP) or even about the other conditional moments. The CUSUM statistics we consider depend on a sequence of weights that influence their asymptotic accuracy. Data-driven procedures are proposed for the optimal choice of the sequence of weights, in the sense of Godambe. We also propose modified versions of the tests that allow to detect breaks in the dynamics even when the conditional mean is misspecified. Our results are illustrated using Monte Carlo experiments and real financial data.

MSC2020 subject classifications: Primary 62M10, 91B84; secondary 60G10, 62F12.

JEL Classification: C13, C14, C22 and C58

Keywords: Change-points, CUSUM, Estimating functions, Quasi-likelihood estimator (QLE).

*University of Lille and CREST-ENSAE. Email: christian.francq@univ-lille3.fr

†University of Leicester. Email: lt285@leicester.ac.uk

‡Corresponding author Jean-Michel Zakoïan: CREST-ENSAE, 5 Avenue Henri Le Chatelier, 91120 Palaiseau, France. Email: zakoian@ensae.fr

1 Introduction

The estimating function (EF) approach was originally developed by Godambe (1960) for fully parametric estimation and by Durbin (1960) for estimating a simple autoregressive structure. Since the sixties, its extension to more general stochastic processes, has led to a number of articles, including Godambe (1985), Godambe and Heyde (1987), and more recently Jacod and Sørensen (2018).

The EF approach is particularly attractive for time series where the dynamics is not fully specified, but the conditional mean is assumed to be a given function of past observations and a finite-dimensional parameter. Unlike "strong" models, which are generally determined by a sequence of innovations (often assumed to be independent and identically distributed (iid)), models characterized only by the first conditional moment (so-called "weak" location models) are not naturally amenable to (quasi) likelihood-based inference. For example, the consistency of the Gaussian quasi-maximum likelihood (QMLE) estimator requires, among other regularity assumptions, the correct specification of the first two conditional moments. In contrast, the quasi-likelihood estimator (QLE), which is obtained by solving estimating equations derived from the first conditional moment, can be consistent and asymptotically normal without the correct specification of the second conditional moment (see Francq and Zakoïan, 2023). In this approach, only the conditional mean needs to be correctly specified.

In many applications, however, structural breaks (or change-points) can affect the conditional mean. The literature on structural breaks has a long history, dating back to Page (1955). Originally developed in the independent case, structural break analysis has long been extended to time series. For a thorough account of structural breaks in the time series literature, see Aue and Horváth (2013) and Horváth and Rice (2023). This topic is important because the analysis and prediction of a time series can be invalidated by the presence of undetected change-points.

This paper is concerned with the detection of structural breaks in the (parametric) conditional mean, relying on the EF approach to estimate the parameter. Testing structural breaks in the conditional or unconditional mean has sparked a variety of methods (see Csörgö and Horváth (1997) and Horváth and Parzen (1994) for pioneer works, and Horváth and Rice (2023) for a recent account). Among them, the cumulative sum (CUSUM) procedure, which is based on partial sums of the observed process, is arguably the most popular. To test for the constancy of the conditional mean, we consider a CUSUM process based on partial sums of the variables involved in the estimating equations.

This paper deals with the detection of structural breaks in the conditional (parametric) mean, using the EF approach to estimate the parameter. Testing for structural breaks in the unconditional or conditional mean has given rise to a variety of methods (see Csörgö and Horváth (1997) and Horváth and Parzen (1994) for pioneering work, and Horv ath and Rice (2023) for a recent review). Of these, the cumulative sum procedure (CUSUM), which is based on the partial sums of the Fisher score, is undoubtedly the most popular. To test the constancy of the conditional mean, we consider a CUSUM procedure based on the quasi-Fisher partial sums involved in the EF approach. The procedure depends on the choice of a sequence of weights, leading to a potentially infinite number of consistent tests. In this paper we show that the best test is, in a sense, related to Godambe's optimal QLE. Data-driven procedures are proposed for this optimal choice of weights. We not only consider tests for breaks, but also study the detection of time breaks. Heteroskedasticity and Autocorrelation Consistent (HAC) versions of the tests are proposed to detect breaks in the dynamics even when the conditional mean is misspecified.

The paper is organized as follows. Section 2 presents the EF approach we follow. The main assumptions and the existing estimation results we need are recalled. Section 3 defines our new family of change-point tests and establishes the asymptotic behavior of these tests under the null of no break. Section 4 shows that under local break alternatives the test with the optimal local asymptotic power is obtained by using the optimal QLE in the Godambe sense. Section 5 discusses the case where the conditional mean may not be correctly specified. The exact break time cannot be consistently estimated, but Section 6 shows that the average change-point location can be consistently estimated. Numerical illustrations are provided in Section 7.

2 Model and estimating function approach

Consider a real time series $(y_t)_{t\in\mathbb{Z}}$ and the sigma-field $\mathcal{F}_t$ generated by $\{y_u : u \leq t\}$. Assume the existence of a measurable function $m_t(\boldsymbol{\theta}) = m(\boldsymbol{\theta}; y_{t-1}, y_{t-2}, \dots)$, depending on some parameter $\boldsymbol{\theta} \in \Theta$ where Θ is a compact subset of $\mathbb{R}^d$, such that for some parameter value $\boldsymbol{\theta}_0$ and all $t \in \mathbb{Z}$,

$$m_t := m_t(\boldsymbol{\theta}_0) = E_{t-1}(y_t), \qquad \text{where } E_t(\cdot) = E(\cdot \mid \mathcal{F}_t). \tag{1}$$

It is assumed that $m_t(\cdot)$ is a well defined measurable function of $\{y_u : u \leq t\}$ and that it is almost surely continuous over Θ. It is always possible to write the model as

$$y_t = m_t + \epsilon_t, \tag{2}$$

where $m_t \in \mathcal{F}_{t-1}$ and (ϵ_t) is such that $E_{t-1}(\epsilon_t) \equiv 0$, or equivalently,

$$E_{t-1}\epsilon_t(\boldsymbol{\theta}_0) = 0, \quad \epsilon_t(\boldsymbol{\theta}) = y_t - m_t(\boldsymbol{\theta}). \tag{3}$$

We refer to Model (2) as a *weak location model*, by contrast with models in which strong assumptions are imposed on the error term, such as iidness.

Denote by $\kappa_{2t}(\boldsymbol{\theta}) = \kappa_2(\boldsymbol{\theta}; y_{t-1}, y_{t-2}, \dots)$ an *assumed* proxy of $\sigma_t^2(\boldsymbol{\theta}) := E_{t-1}\{y_t - m_t(\boldsymbol{\theta})\}^2$. Note that $m_t(\boldsymbol{\theta})$ and $\kappa_{2t}(\boldsymbol{\theta})$ are not statistics when they depend on the non-observed values $\{y_u : u \leq 0\}$. Moreover, $\kappa_{2t}(\boldsymbol{\theta}) = \kappa_{2t}(\boldsymbol{\theta}, \boldsymbol{\gamma})$ may depend on some unknown nuisance parameter $\boldsymbol{\gamma}$ (as well as $\sigma_t^2(\boldsymbol{\theta})$). Let $\mathcal{I}_t$ be the sigma-field generated by $\{y_u : 1 \leq u \leq t\}$, the information available at time t. Note that $\mathcal{I}_t \subset \mathcal{F}_t$. Let $\widetilde{m}_t(\boldsymbol{\theta})$ be a $\mathcal{I}_t$-measurable approximation[1] of $m_t(\boldsymbol{\theta})$. Similarly, $\widetilde{\kappa}_{2t}$ stands for a $\mathcal{I}_t$-measurable approximation of κ_{2t}, which may be constant or may depend on $\boldsymbol{\theta}$ (see examples below). When $\kappa_{2t}(\boldsymbol{\theta}) = \kappa_{2t}(\boldsymbol{\theta}, \boldsymbol{\gamma})$ depends on a nuisance parameter, and $\widehat{\boldsymbol{\gamma}}_n$ is an estimator of $\boldsymbol{\gamma}$, then $\widetilde{\kappa}_{2t} = \widetilde{\kappa}_{2t}(\boldsymbol{\theta}, \widehat{\boldsymbol{\gamma}}_n)$ denotes a $\mathcal{I}_n$-measurable approximation of $\kappa_{2t}(\boldsymbol{\theta})$.[2]

The parameter $\boldsymbol{\theta}_0$ can be estimated by solving

$$\sum_{t=1}^{n} \frac{\partial \widetilde{m}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\widetilde{\epsilon}_t(\boldsymbol{\theta})}{\widetilde{\kappa}_{2t}} = 0, \quad \widetilde{\epsilon}_t(\boldsymbol{\theta}) = y_t - \widetilde{m}_t(\boldsymbol{\theta}). \tag{4}$$

Any measurable solution of the estimating equation (4) is called QLE. Equation (4) is called an estimating equation, and the left-hand side is called an estimating function (see *e.g.* Bera *et al.* (2006) or Heyde (2008) for general references on this literature). This estimating function corresponds to Fisher's score for particular conditional distributions of y_t (see Remark 2 below) and in general it can be interpreted as a quasi-score, hence the title of the paper. All the QLEs that we consider hereafter only differ by the weighting sequence $(\widetilde{\kappa}_{2t})$.

Remark 1 (Examples of QLEs). *When $\widetilde{\kappa}_{2t}$ is a non zero constant, the solution of (4) is the Least Squares (LS) estimator. When $y_t \geq 0$ and $\widetilde{\kappa}_{2t}$ is proportional to $\widetilde{m}_t(\boldsymbol{\theta}) > 0$ (respectively $\widetilde{m}_t^2(\boldsymbol{\theta}) > 0$), the solution of (4) is the Poisson (respectively exponential) QMLE, obtained by minimizing*

$$\sum_{t=1}^{n} \widetilde{m}_t(\boldsymbol{\theta}) - y_t \log \widetilde{m}_t(\boldsymbol{\theta}) \quad \textit{(respectively } \textstyle\sum_{t=1}^{n} y_t/\widetilde{m}_t(\boldsymbol{\theta}) + \log \widetilde{m}_t(\boldsymbol{\theta})\textit{)}.$$

[1]See **A2** below for a precise assumption.

[2]maybe we should write $\widetilde{\kappa}_{2,n,t}$ instead of $\widetilde{\kappa}_{2t}$ but we found this notation too cumbersome.

Remark 2 (Case where the QLE is the Maximum Likelihood Estimator (MLE))**.** *Assume that the distribution of y_t given $\mathcal{F}_{t-1}$ belongs to the one-parameter exponential family. This means that, with respect to a σ-finite measure, the conditional distribution admits a density of the form*

$$g_{m_t}(y) = k(y)\exp\{\eta(m_t)y - a(m_t)\}, \tag{5}$$

for some positive function k and twice differentiable functions $\eta(\cdot)$ and $a(\cdot)$. It is known that $\eta'(m_t) = a'(m_t)/m_t = 1/\sigma_t^2$. It follows that

$$\frac{\partial \log g_{m_t(\boldsymbol{\theta})}(y_t)}{\partial \boldsymbol{\theta}} = \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\epsilon_t(\boldsymbol{\theta})}{\sigma_t^2(\boldsymbol{\theta})}.$$

The QLE is thus the MLE (only approximately when $m_t \neq \widetilde{m}_t$).

An estimating function is said to be unbiased if its expectation is zero at the true value. Godambe (1985) introduced the concept of 'optimal' unbiased estimating functions, similar to the BLUE property for unbiased estimators. He showed (see also Chandra and Taniguchi, 2001) that among the unbiased estimating functions of the form $\sum_{t=1}^n \boldsymbol{a}_{t-1}(\boldsymbol{\theta})\widetilde{\epsilon}_t(\boldsymbol{\theta})$, with $\boldsymbol{a}_t(\boldsymbol{\theta})$ any d-dimensional $\mathcal{I}_t$-measurable vector, optimal ones are of the form (4) when $\widetilde{\kappa}_{2t}$ is proportional to $\sigma_t^2(\boldsymbol{\theta})$ (which is however generally unknown).

In order to discuss consistency and asymptotic normality of the QLEs, we introduce the following set of assumptions.

A1 The process $(y_t)_{t\in\mathbb{Z}}$ is strictly stationary and ergodic.

A2 There exists $\rho \in [0,1)$ such that, almost surely $\sup_{\boldsymbol{\theta}\in\Theta} |m_t(\boldsymbol{\theta}) - \widetilde{m}_t(\boldsymbol{\theta})| \leq K_t\rho^t$, where K_t is a generic $\mathcal{F}_{t-1}$-measurable random variable such that $\sup_t EK_t^r < \infty$ for some $r > 0$.

A3 Let $\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}) = \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\epsilon_t(\boldsymbol{\theta})}{\kappa_{2t}(\boldsymbol{\theta})}$. If $E\{\boldsymbol{\Upsilon}_t(\boldsymbol{\theta})\} = 0$ for some $\boldsymbol{\theta} \in \Theta$, then $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. The parameter $\boldsymbol{\theta}_0$ belongs to the interior of the compact set Θ.

A4 The function $\boldsymbol{\theta} \mapsto m_t(\boldsymbol{\theta})$ is twice continuously differentiable, and

$$\sup_{\boldsymbol{\theta}\in\Theta} \left\| \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial \widetilde{m}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| \leq K_t\rho^t, \qquad a.s.$$

where K_t is as in **A2**, $\|\cdot\|$ denotes any norm on $\mathbb{R}^d$. Moreover, for some $s > 0$, $E|y_t|^s < \infty$ and $E\sup_{\boldsymbol{\theta}\in\Theta}\left\{|m_t(\boldsymbol{\theta})|^s + \left\|\frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\|^s\right\} < \infty$.

A5 If $\boldsymbol{\lambda}^\top \frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = 0$ a.s. then $\boldsymbol{\lambda} = \mathbf{0}_d$.

A6 There exists a constant $\underline{\kappa} > 0$ such that $\inf_{\boldsymbol{\theta}\in\Theta} \kappa_{2t}(\boldsymbol{\theta}) \geq \underline{\kappa}$ a.s.

A7 For all $\boldsymbol{\theta} \in \Theta$ the sequence $\{\kappa_{2t}(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ is stationary, ergodic and $\mathcal{F}_{t-1}$-measurable, the function $\boldsymbol{\theta} \mapsto \kappa_{2t}(\boldsymbol{\theta})$ is continuously differentiable, there exist $\rho \in [0,1)$ and K_t as in **A2** such that, almost surely,

$$\sup_{\boldsymbol{\theta}\in\Theta} |\kappa_{2t}(\boldsymbol{\theta}) - \widetilde{\kappa}_{2t}(\boldsymbol{\theta})| \leq K_t \rho^t$$

for n large enough.[3]

A8 We have

$$E \sup_{\boldsymbol{\theta}\in\Theta} \|\boldsymbol{\Upsilon}_t(\boldsymbol{\theta})\|^2 < \infty \quad \text{and} \quad E \sup_{\boldsymbol{\theta}\in\Theta} \left\| \frac{\partial \boldsymbol{\Upsilon}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \right\| < \infty.$$

Remark 3 (On the assumptions)**.** *Assumptions* **A1**-**A8** *(under slightly different forms) have been commented on in Francq and Zakoïan (2023). They can be made explicit for particular models, such as ARMA or GARCH. As an illustration, assume the INGARCH model obtained when* y_t *given* $\mathcal{F}_{t-1}$ *follows a Poisson distribution with intensity parameter* $m_t = c_0 + a_0 y_{t-1} + b_0 m_{t-1}$, *with obvious notation. In this case, it is known that* **A1** *holds true when* $a_0 + b_0 < 1$, *and that the strictly stationarity solution of the INGARCH model even admits moments at any order. It is easy to see that, when* $|b| < 1$ *for all* $\boldsymbol{\theta} \in \Theta$, **A2** *holds with* $K_t = \sup_{\boldsymbol{\theta}\in\Theta} \{|a||y_0| + |b||m_0(\boldsymbol{\theta})|\}$[4]. *Similarly* **A7** *holds, but with another expression of* $K_t = K$. *Assumption* **A5** *holds if* $a_0 \neq 0$ *and* $\inf_{\boldsymbol{\theta}\in\Theta} c > 0$. *Since* y_t *admits moments of any order,* **A8** *is always satisfied. Is is also clear that all the other assumptions hold true for many weighting sequences* $(\widetilde{\kappa}_{2t})$.

Denote by $X_n \xrightarrow{\mathcal{L}} X$, or simply $X_n \xrightarrow{\mathcal{L}} P_X$, when a sequence of random vectors X_n convergences in distribution to a random vector X of distribution P_X. For a sequence of random functions $\{X_n(u), u \in [0,1]\}$ tending weakly to $\{X(u), u \in [0,1]\}$, we denote $X_n(\cdot) \Longrightarrow X(\cdot)$. Under a slightly reformulated version of the previous assumptions, the following result can be found in Francq and Zakoïan (2023).

Theorem 1 (Francq and Zakoïan, (2023))**.** *Under Assumptions* **A1**-**A8**, *a QLE* $\widehat{\boldsymbol{\theta}}$ *of* $\boldsymbol{\theta}_0$, *such that*

$$\sum_{t=1}^{n} \widetilde{\boldsymbol{\Upsilon}}_t(\widehat{\boldsymbol{\theta}}) = 0, \qquad \widetilde{\boldsymbol{\Upsilon}}_t(\boldsymbol{\theta}) = \frac{\partial \widetilde{m}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\widetilde{\epsilon}_t(\boldsymbol{\theta})}{\widetilde{\kappa}_{2t}(\boldsymbol{\theta})},$$

exists[5] *for* n *large enough, and as* $n \to \infty$ *we have* $\widehat{\boldsymbol{\theta}} \to \boldsymbol{\theta}_0$ *a.s. and*

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) = -\boldsymbol{J}^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0) + o_P(1) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma} := \boldsymbol{J}^{-1}\boldsymbol{I}\boldsymbol{J}^{-1}\right)$$

[3]Recall that $\widetilde{\kappa}_{2t}$ may depend on $\widehat{\gamma}_n$.

[4]here $K_t = K$ but this variable can be time-varying for other models.

[5]uniqueness has been shown under the extra contraction assumption **A10** of Francq and Zakoïan (2023).

with

$$\boldsymbol{J} = E\left(\frac{-1}{\kappa_{2t}(\boldsymbol{\theta}_0)}\frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^\top}\right), \qquad \boldsymbol{I} = E\left(\frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\kappa_{2t}^2(\boldsymbol{\theta}_0)}\frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^\top}\right).$$

If $\kappa_{2t}(\boldsymbol{\theta}_0) \propto \sigma_t^2(\boldsymbol{\theta}_0)$ *then the asymptotic variance of the QLE, which is equal to*

$$\boldsymbol{\Sigma}_{op} = \left\{E\frac{1}{\sigma_t^2(\boldsymbol{\theta}_0)}\frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^\top}\right\}^{-1},$$

is optimal in the sense that $\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_{op}$ *is semi positive definite.*

Remark 4 (Selection of $\widetilde{\kappa}_{2t}$ by QLIK). *In view of Remark 1, natural candidates for the weighting sequence are* $\widetilde{\kappa}_{2t} \propto 1$, $\widetilde{\kappa}_{2t} \propto \widetilde{m}_t(\boldsymbol{\theta})$ *(for positive data) or* $\widetilde{\kappa}_{2t} \propto \widetilde{m}_t^2(\boldsymbol{\theta})$*, among an infinite number of other possibilities. Thus assume we want to select the weights over a finite set of potential weighting sequences, say* $\left\{\widetilde{\kappa}_{2t}^{(i)}(\boldsymbol{\theta})\right\}_t$ *for* $i \in \{1, \dots, I\}$*. Since the optimal weighting sequence is the conditional variance–up to any non zero multiplicative constant–and since the conditional variance is solution to the quasi-likelihood (QLIK) loss function, Francq and Zakoïan (2023) proposed a data-driven selection of the weights by minimizing over i the empirical QLIK loss function*

$$QLIK_n\left(\widetilde{\kappa}_{2\cdot}^{(i)}(\widehat{\boldsymbol{\theta}})\right) = \frac{1}{n}\sum_{t=1}^{n}\left\{\frac{\widetilde{\epsilon}_t^2(\widehat{\boldsymbol{\theta}})}{\widehat{c}_n^{(i)}\widetilde{\kappa}_{2t}^{(i)}(\widehat{\boldsymbol{\theta}})} + \log\left(\widehat{c}_n^{(i)}\widetilde{\kappa}_{2t}^{(i)}(\widehat{\boldsymbol{\theta}})\right)\right\}, \qquad \widehat{c}_n^{(i)} = \frac{1}{n}\sum_{t=1}^{n}\frac{\widetilde{\epsilon}_t^2(\widehat{\boldsymbol{\theta}})}{\widetilde{\kappa}_{2t}^{(i)}(\widehat{\boldsymbol{\theta}})},$$

where $\widehat{\boldsymbol{\theta}}$ *is a first step estimator of* $\boldsymbol{\theta}_0$.

Remark 5 (Estimation of $\widetilde{\kappa}_{2t}$ by GARCH-X). *It is also natural to estimate the conditional variance by fitting a GARCH-type model on* $\widetilde{\epsilon}_t = y_t - \widetilde{m}_t(\widehat{\boldsymbol{\theta}})$*, for* $t = 1, \dots, n$*, where* $\widehat{\boldsymbol{\theta}}$ *is a first step consistent estimator of* $\boldsymbol{\theta}_0$*. This leads to the simple GARCH(1,1) estimator*

$$\widetilde{\kappa}_{2t} = \widehat{\omega} + \widehat{\alpha}\widetilde{\epsilon}_{t-1}^2 + \widehat{\beta}\widetilde{\kappa}_{2,t-1} \tag{6}$$

or to extended GARCH-X estimators like

$$\widetilde{\kappa}_{2t} = \widehat{\omega} + \widehat{\alpha}\widetilde{\epsilon}_{t-1}^2 + \widehat{\beta}\widetilde{\kappa}_{2,t-1} + \widehat{\pi}_1|\widetilde{m}_t(\widehat{\boldsymbol{\theta}})| \tag{7}$$

or

$$\widetilde{\kappa}_{2t} = \widehat{\omega} + \widehat{\alpha}\widetilde{\epsilon}_{t-1}^2 + \widehat{\beta}\widetilde{\kappa}_{2,t-1} + \widehat{\pi}_1|\widetilde{m}_t(\widehat{\boldsymbol{\theta}})| + \widehat{\pi}_2\widetilde{m}_t^2(\widehat{\boldsymbol{\theta}}). \tag{8}$$

For instance, (8) allows weights proportional to the conditional mean or its square, and thus can target the Poisson and exponential QMLEs (see Remark 1).

3 Change-point tests

We are interested in testing the assumption that the parameter $\boldsymbol{\theta}_0$ does not change over time. In other words, assuming the observations $y_1, \ldots, y_n$ satisfy $m_t(\boldsymbol{\theta}_t) = E_{t-1}(y_t)$ for $t = 1, \ldots, n$, where $\boldsymbol{\theta}_t \in \Theta$, we consider testing

$$\mathbf{H_0}: \quad \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \cdots = \boldsymbol{\theta}_n$$

against the alternative of at least one unknown breakpoint.

Inspired by basic CUSUM statistics used in changepoint problems, we consider the process, defined for $u \in [0,1]$ by

$$\widetilde{\boldsymbol{T}}_n(u) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nu]} \widetilde{\boldsymbol{\Upsilon}}_t(\widehat{\boldsymbol{\theta}}).$$

By convention $\widetilde{\boldsymbol{T}}_n(0) = 0$ and, by definition of the QLE, we also have $\widetilde{\boldsymbol{T}}_n(1) = 0$. A natural statistic for testing $\mathbf{H_0}$ is

$$\widetilde{S}_n = \sup_{u \in (0,1)} \widetilde{S}_n(u) = \max_{k \in \{1,\ldots,n-1\}} \widetilde{S}_n(k/n), \qquad \widetilde{S}_n(u) = \widetilde{\boldsymbol{T}}_n^\top(u) \boldsymbol{I}_n^{-1} \widetilde{\boldsymbol{T}}_n(u)$$

where $\boldsymbol{I}_n$ denotes a non singular consistent estimator of $\boldsymbol{I}$. Note that **A5** entails that $\boldsymbol{I}$ is not singular. Horváth and Parzen (1994) introduced a CUSUM statistic similar to $\widetilde{S}_n$, but based on Fisher's score, for testing iidness against "abrupt change". Berkes, Horváth and Kokoszka (2004) adapted the approach to test parameter constancy in GARCH model, using a QMLE-based score. Negri and Nishiyama (2017) considered more general models with applications to ergodic and non-ergodic diffusion processes. Kutoyants (2016) considered CUSUM statistics based on Fishers's score and studied goodness-of-fit tests for diffusion processes and nonlinear time series models. Lee *et al.* (2003) considered a CUSUM based on $\widehat{\boldsymbol{\theta}}_k - \widehat{\boldsymbol{\theta}}_n$ for $k = 1, \ldots, n$, where $\widehat{\boldsymbol{\theta}}_k$ is an estimator based on the first k observations. Shao and Zhang (2010) proposed a self-normalized Kolmogorov-Smirnov test for a change point in the mean of a time series. Aue and Horváth (2013) showed how CUSUM statistics can be used to detect breaks in the unconditional and conditional means and variances of time series. Note that there exist numerous CUSUM statistics and many other techniques to detect breaks. The literature on change point detection is actually so vast that it is not possible to provide an exhaustive review. For a recent overview of change point detection algorithms, see e.g. Truong et al. (2020).

Inspired by these references, we will show that, under the null of no break, $\widetilde{\boldsymbol{T}}_n(u)$ converges weakly to a Gaussian process $\boldsymbol{T}(u) = (T_1(u), \ldots, T_d(u))^\top$ with covariance structure $E\boldsymbol{T}(u) = 0$ and $\mathrm{Cov}(\boldsymbol{T}(u), \boldsymbol{T}(v)) = \boldsymbol{I}\{\min(u,v) - uv\}$. Thus, each component of the vector $\boldsymbol{I}^{-1/2}\boldsymbol{T}(u)$ is a standard Brownian bridge $\{B(u), u \in [0,1]\}$, with $B(u) = W(u) - uW(1)$ where $\{W(u), u \in [0,1]\}$ denotes a standard Brownian motion.

Theorem 2. *Under Assumptions* **A1-A9**, *including* $\mathbf{H_0}$, *we have*

$$\widetilde{S}_n \xrightarrow{\mathcal{L}} S := \sup_{u\in(0,1)} \boldsymbol{T}^\top(u)\boldsymbol{I}^{-1}\boldsymbol{T}(u) = \sup_{u\in(0,1)} \sum_{j=1}^{d}\{B_j(u)\}^2,$$

where $B(u) = (B_1(u), \ldots, B_d(u))^\top$ *is a* d*-dimensional standard Brownian bridge.*

Remark 6. The distribution of S (for $d \leq 10$) is tabulated in Lee, Ha, Na and Na (2003).

Remark 7. The estimation of $\boldsymbol{I}$ cannot be achieved by plug-in, using the formula displayed in Theorem 1. Indeed, the conditional variance function $\sigma_t^2(\cdot)$ is generally unknown. However, noting that $\boldsymbol{I} = E\left\{\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0)\boldsymbol{\Upsilon}_t^\top(\boldsymbol{\theta}_0)\right\}$, a consistent estimator is

$$\boldsymbol{I}_n = \frac{1}{n}\sum_{t=1}^{n} \widetilde{\boldsymbol{\Upsilon}}_t(\widehat{\boldsymbol{\theta}}_n)\widetilde{\boldsymbol{\Upsilon}}_t^\top(\widehat{\boldsymbol{\theta}}_n). \tag{9}$$

Remark 8. Nyblom (1989) proposed a general theory for testing the constancy of the parameters involved in the conditional distribution of a time series model. Applied to our semiparametric framework, the Nyblom test[6] replaces the supremum with a mean. More specifically, the test rejects the parameter constancy for large values of

$$\widetilde{S}_n^N := \frac{1}{n}\sum_{k=1}^{n} \widetilde{S}_n(k/n)$$

which, by the continuous mapping theorem, has the asymptotic distribution $\int_0^1 \sum_{j=1}^d \{B_j(u)\}^2 du$ under the assumptions of Theorem 2. The Nyblom test, which enjoys some optimality properties under the alternative that the parameter process follows a martingale, is widely used in econometrics (for an example see Hansen, Lunde and Voev (2014), where the test is used to assess the constancy of a correlation).

Remark 9. The CUSUM test also has optimality properties, but for different alternatives than the Nyblom test. At the very beginning of their book, Horváth and Rice (2023) give an example where the (normalized) CUSUM test coincides with a likelihood ratio test and thus enjoys its general good asymptotic properties. Inspired by this example, consider a sequence of independent and Gaussian vectors $Y_1, \ldots, Y_k, \ldots, Y_n$ such that $Y_t \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ for $t \leq k$ and $Y_t \sim \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$ for $t > k$, where $\boldsymbol{\Sigma}$ is a known non singular variance matrix. Let the null

[6] We are grateful to P.R. Hansen for pointing out that we can use this test in our framework.

$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ and the alternative $H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$. The unknown parameters of interest are k and $\boldsymbol{\theta} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$. Note that, at $\boldsymbol{\theta}_0 = (\boldsymbol{\mu}_0, \boldsymbol{\mu}_0) \in H_0$, the likelihood $L_n(Y_1, \dots, Y_n; \boldsymbol{\theta}_0, k)$ does not depend on k. With obvious notations, the standard likelihood ratio leads to reject H_0 for large values of $LR = \sup_{1 \le k \le n} LR(k)$, where

$$LR(k) = \log \frac{L_n(Y_1, \dots, Y_n; \widehat{\boldsymbol{\theta}}, k)}{L_n(Y_1, \dots, Y_n, \widehat{\boldsymbol{\theta}}_0, k)} = \frac{nk}{2(n-k)} (\widehat{\mu}_1 - \widehat{\mu}_0)^\top \boldsymbol{\Sigma}^{-1} (\widehat{\mu}_1 - \widehat{\mu}_0),$$

up to unimportant additive constants, noting that $\widehat{\mu}_2 - \widehat{\mu}_1 = \frac{n}{n-k}(\widehat{\mu}_0 - \widehat{\mu}_1)$ and $\widehat{\mu}_2 - \widehat{\mu}_0 = \frac{k}{n-k}(\widehat{\mu}_0 - \widehat{\mu}_1)$. This likelihood ratio is directly related to the weighted CUSUM by

$$2LR = \sup_{u \in (0,1)} \frac{\widetilde{S}_n(u)}{u(1-u)} + o_P(1).$$

Proof of Theorem 2. Let

$$\boldsymbol{T}_n(u) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nu]} \boldsymbol{\Upsilon}_t(\widehat{\boldsymbol{\theta}}).$$

A Taylor expansion around $\boldsymbol{\theta}_0$ yields

$$\boldsymbol{T}_n(u) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nu]} \boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0) + u \left(\frac{1}{nu} \sum_{t=1}^{[nu]} \frac{\partial}{\partial \boldsymbol{\theta}^\top} \boldsymbol{\Upsilon}_t(\widehat{\boldsymbol{\theta}}^*) \right) \sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$$

where $\widehat{\boldsymbol{\theta}}^*$ is between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$.

By Theorem 1, it follows that

$$\begin{aligned} \boldsymbol{T}_n(u) =& \frac{1}{\sqrt{n}} \left(\sum_{t=1}^{[nu]} \boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0) - u \sum_{t=1}^{n} \boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0) \right) + \frac{1}{n} \sum_{t=1}^{[nu]} \left(\frac{\partial}{\partial \boldsymbol{\theta}^\top} \boldsymbol{\Upsilon}_t(\widehat{\boldsymbol{\theta}}^*) - \boldsymbol{J} \right) \sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + u o_P(1) \\ :=& \boldsymbol{T}_n^0(u) + \boldsymbol{R}_n(u) \sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + o_P(1), \end{aligned} \tag{10}$$

where the reminder term is independent of u. It follows from the functional central limit theorem for stationary, ergodic martingale differences (see e.g. Theorem 18.3 of Billingsley, 1986) that

$$\boldsymbol{T}_n^0(\cdot) \Longrightarrow \boldsymbol{T}(\cdot).$$

It remains to show that

$$\sup_{u \in (0,1)} \|\boldsymbol{R}_n(u)\| = o_P(1), \qquad \sup_{u \in (0,1)} \|\boldsymbol{T}_n(u) - \widetilde{\boldsymbol{T}}_n(u)\| = o_P(1). \tag{11}$$

Let $V_k(\boldsymbol{\theta}_0)$ be the ball of center $\boldsymbol{\theta}_0$ and radius $1/k$. The strong consistency of $\widehat{\boldsymbol{\theta}}$ entails that

$$\|\boldsymbol{R}_n(u)\| \le \left\| \frac{1}{n} \sum_{t=1}^{[nu]} \left(\frac{\partial}{\partial \boldsymbol{\theta}^\top} \boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0) - \boldsymbol{J} \right) \right\| + \sup_{\boldsymbol{\theta} \in V_k(\boldsymbol{\theta}_0)} \left\| \frac{1}{n} \sum_{t=1}^{[nu]} \left(\frac{\partial}{\partial \boldsymbol{\theta}^\top} \boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0) - \frac{\partial}{\partial \boldsymbol{\theta}^\top} \boldsymbol{\Upsilon}_t(\boldsymbol{\theta}) \right) \right\|.$$

We have,

$$\sup_{\boldsymbol{\theta}\in V_k(\boldsymbol{\theta}_0)}\sup_{u\in(0,1)}\left\|\frac{1}{n}\sum_{t=1}^{[nu]}\left(\frac{\partial}{\partial\boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0)-\frac{\partial}{\partial\boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t(\boldsymbol{\theta})\right)\right\|\leq\frac{1}{n}\sum_{t=1}^{n}\sup_{\boldsymbol{\theta}\in V_k(\boldsymbol{\theta}_0)}\left\|\frac{\partial}{\partial\boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0)-\frac{\partial}{\partial\boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t(\boldsymbol{\theta})\right\|$$

which tends, as $n\to\infty$, to

$$E\sup_{\boldsymbol{\theta}\in V_k(\boldsymbol{\theta}_0)}\left\|\frac{\partial}{\partial\boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0)-\frac{\partial}{\partial\boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t(\boldsymbol{\theta})\right\|.$$

by the ergodic theorem. By Fatou's lemma, and using the continuity assumptions in **A2** and **A7**, the latter expectation can be made arbitrarily small by choosing k sufficiently large. Let (u_n) be a deterministic sequence converging to infinity slower than n (i.e. $n/u_n\to\infty$). Let $\boldsymbol{Y}_t=\frac{\partial}{\partial\boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0)-\boldsymbol{J}$. We have

$$\sup_{u\in(0,1)}\left\|\frac{1}{n}\sum_{t=1}^{[nu]}\left(\frac{\partial}{\partial\boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0)-\boldsymbol{J}\right)\right\|\leq\sup_{1\leq k\leq u_n}\left\|\frac{1}{n}\sum_{t=1}^{k}\boldsymbol{Y}_t\right\|+\sup_{u_n\leq k\leq n}\left\|\frac{1}{n}\sum_{t=1}^{k}\boldsymbol{Y}_t\right\|.$$

We have $\frac{1}{k}\sum_{t=1}^{k}\boldsymbol{Y}_t\to 0$ a.s. as $k\to\infty$, hence the last term in the previous inequality converges to 0 a.s. Moreover, by Markov inequality, for any $\iota>0$,

$$P\left(\sup_{1\leq k\leq u_n}\left\|\frac{1}{n}\sum_{t=1}^{k}\boldsymbol{Y}_t\right\|>\iota\right)\leq P\left(\frac{1}{n}\sum_{t=1}^{u_n}\|\boldsymbol{Y}_t\|>\iota\right)\leq\frac{u_n}{n\iota}E\|\boldsymbol{Y}_1\|\to 0,$$

as $n\to\infty$, from which we deduce that the first convergence in (11) holds.

Now, we have

$$\begin{aligned}\sup_{u\in(0,1)}\|\boldsymbol{T}_n(u)-\widetilde{\boldsymbol{T}}_n(u)\|&\leq\sup_{u\in(0,1)}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{[nu]}\frac{\partial\widetilde{m}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\widetilde{\epsilon}_t(\boldsymbol{\theta})}{\widetilde{\kappa}_{2t}(\boldsymbol{\theta})}-\frac{\partial m_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\epsilon_t(\boldsymbol{\theta})}{\kappa_{2t}(\boldsymbol{\theta})}\right\|\\&\leq\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sup_{\boldsymbol{\theta}\in\Theta}\left(\left\|\frac{\partial m_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}-\frac{\partial\widetilde{m}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|\frac{|\epsilon_t(\boldsymbol{\theta})|}{\kappa_{2t}(\boldsymbol{\theta})}\right)\\&+\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sup_{\boldsymbol{\theta}\in\Theta}\left(|m_t(\boldsymbol{\theta})-\widetilde{m}_t(\boldsymbol{\theta})|\left\|\frac{\partial\widetilde{m}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|\frac{1}{\kappa_{2t}(\boldsymbol{\theta})}\right)\\&+\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sup_{\boldsymbol{\theta}\in\Theta}\left(\left|\frac{1}{\kappa_{2t}(\boldsymbol{\theta})}-\frac{1}{\widetilde{\kappa}_{2t}(\boldsymbol{\theta})}\right|\left\|\frac{\partial\widetilde{m}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\||\widetilde{\epsilon}_t(\boldsymbol{\theta})|\right).\end{aligned}$$

By Assumptions **A6-A7**, the first term in the right-hand side is bounded by

$$\frac{1}{\underline{\kappa}\sqrt{n}}\sum_{t=1}^{\infty}\sup_{\boldsymbol{\theta}\in\Theta}|\epsilon_t(\boldsymbol{\theta})|K_t\rho^t=O\left(\frac{1}{\sqrt{n}}\right),\quad a.s.$$

because the summands have finite s-th order moment by **A4**. The same upper bound holds for the other two terms of the right-hand side of the previous inequality. Hence the second convergence in (11) is established. □

We can now construct a test for $\mathbf{H_0}$. At the significance level $\alpha \in (0,1)$, an asymptotic critical region is given by

$$\max_{1\le k\le n} \widetilde{S}_n(k/n) > S_{1-\alpha} \tag{12}$$

where $S_{1-\alpha}$ is the $(1-\alpha)$ quantile of the law of S.

Example 1 (Conditional mean of the weak AR(1) process). Suppose that

$$y_t = \theta_0 y_{t-1} + \epsilon_t, \quad t \in \mathbb{Z}, \quad |\theta_0| < 1,$$

where (ϵ_t) is strictly stationary, ergodic and satisfies $E_{t-1}(\epsilon_t) = 0$. Then Assumption **A1** is satisfied and it is clear that, with $\Theta = [-1,1]$, **A2** and **A4** are also satisfied. Moreover, $E\boldsymbol{\Upsilon}_t(\theta) = E\left(\frac{y_{t-1}^2}{\kappa_{2t}(\theta)}\right)(\theta-\theta_0)$, showing that **A3** holds true. Under the assumption $\sigma_t^2(\theta_0) > 0$ in the first part of **A5**, the second part holds true. A sufficient condition for Assumption **A8** to hold is

$$E\sup_\theta \left\{\frac{\sigma_t^2(\theta_0)}{\kappa_{2t}^2(\theta)}y_{t-1}^2 + \frac{y_{t-1}^4}{\kappa_{2t}^2(\theta)} + \left|\frac{1}{\kappa_{2t}^2(\theta)}\frac{\partial \kappa_{2t}}{\partial\theta}\right|(|y_{t-1}| + y_{t-1}^2)\right\} < \infty.$$

It should be noted that if, for instance, κ_{2t} is of the form $a + by_{t-1}^2$ with $a, b > 0$, and similarly for $\sigma_t^2(\theta_0)$, the latter conditions may only require $Ey_t^2 < \infty$. Finally, $\boldsymbol{I} = E\left(\frac{\sigma_t^2(\boldsymbol{\theta}_0)}{\kappa_{2t}^2(\boldsymbol{\theta}_0)}y_{t-1}^2\right)$.

4 Power comparisons

The estimator defined in (4), with $\kappa_{2t}(\boldsymbol{\theta}_0)$ proportional to $\sigma_t^2(\boldsymbol{\theta}_0)$ under **A7**, is optimal in the Godambe sense within the class of EF estimators solving

$$\sum_{t=1}^n \boldsymbol{a}_{t-1}(\boldsymbol{\theta})\widetilde{\epsilon}_t(\boldsymbol{\theta}) = 0, \tag{13}$$

where $\boldsymbol{a}_{t-1}(\boldsymbol{\theta})$ is a $d \times 1$ vector belonging to $\mathcal{F}_{t-1}$ (see Chandra and Taniguchi, 2001). In this section we show that Godambe's optimal QLEs lead to optimal tests, in the sense that they optimize some local asymptotic power (LAP).

We will consider a sequence of "local breaks" occurring at a proportion $u_0 \in (0,1)$ of the observations. The simplest example of such a local break is obtained by assuming that $y_1, \dots, y_n$ are independent and Gaussian with variance σ^2, and that $y_t = y_{t,n}$ has mean $\theta_0 + \delta_1/\sqrt{[nu_0]}$ when $t \le [nu_0]$ and $\theta_0 + \delta_2/\sqrt{n-[nu_0]}$ when $t > [nu_0]$. Note that we then have

$$\frac{1}{\sqrt{[nu_0]}}\sum_{t=1}^{[nu_0]}(y_t-\theta_0) \sim \mathcal{N}\left(\delta_1, \sigma^2\right), \quad \frac{1}{\sqrt{n-[nu_0]}}\sum_{t=[nu_0]+1}^{n}(y_t-\theta_0) \sim \mathcal{N}\left(\delta_2, \sigma^2\right),$$

and thus

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(y_t-\theta_0)\sim\mathcal{N}\left(\delta_3,\sigma^2\right),\quad \delta_3=\sqrt{u_0}\delta_1+\sqrt{1-u_0}\delta_2.$$

Note that, in this simple example, $\overline{y}=n^{-1}\sum_{t=1}^{n}y_t$ is the Q(M)LE of θ_0 (under the null $\delta_1=\delta_2=0$ of no local break), $\widetilde{T}_n(u)=T_n(u)=n^{-1/2}\sum_{t=1}^{[nu]}(y_t-\overline{y})$ is the usual CUSUM process, and

$$\widetilde{S}_n=\sup_{u\in(0,1)}\frac{1}{n\widehat{\sigma}_y^2}\left\{\sum_{t=1}^{[nu]}(y_t-\overline{y})\right\}^2,\quad \widehat{\sigma}_y^2=\frac{1}{n}\sum_{t=1}^{n}(y_t-\overline{y})^2,$$

is nothing else than the Kolmogorov test statistic. Note also that (10)–(11) hold with $R_n(u)=u-[nu]/n$ and $o_P(1)=0$. The asymptotic distribution of the Kolmogorov test statistic under such local breaks can be obtained as a corollary of the next result.

Let us now return to the general situation. Suppose the conditional distribution of y_t changes at a single point, which is located at a fixed proportion $u_0\in(0,1)$ of the observations. Let $\widehat{\boldsymbol{\theta}}_{(1)}$ be the QLE computed on $y_1,\dots,y_{[u_0n]}$ and $\widehat{\boldsymbol{\theta}}_{(2)}$ the QLE computed on $y_{[u_0n]+1},\dots,y_n$. Recall that $\widehat{\boldsymbol{\theta}}$ is the QLE computed on all the observations $y_1,\dots,y_n$. Let the local alternatives $H_1=H_1(\boldsymbol{\delta}_1,\boldsymbol{\delta}_2)$ such that, for $\boldsymbol{\Upsilon}_t=\boldsymbol{\Upsilon}_{t,n}(\boldsymbol{\theta}_0)$, as $n\to\infty$

$$\sqrt{nu_0}\left(\widehat{\boldsymbol{\theta}}_{(1)}-\boldsymbol{\theta}_0\right)=-\boldsymbol{J}^{-1}\frac{1}{\sqrt{nu_0}}\sum_{t=1}^{[nu_0]}\boldsymbol{\Upsilon}_t+o_P(1)\xrightarrow{\mathcal{L}}\mathcal{N}\left(\boldsymbol{\delta}_1,\boldsymbol{J}^{-1}\boldsymbol{I}\boldsymbol{J}^{-1}\right),\tag{14}$$

$$\sqrt{n(1-u_0)}\left(\widehat{\boldsymbol{\theta}}_{(2)}-\boldsymbol{\theta}_0\right)=-\boldsymbol{J}^{-1}\frac{1}{\sqrt{n(1-u_0)}}\sum_{t=[nu_0]+1}^{n}\boldsymbol{\Upsilon}_t+o_P(1)\xrightarrow{\mathcal{L}}\mathcal{N}\left(\boldsymbol{\delta}_2,\boldsymbol{J}^{-1}\boldsymbol{I}\boldsymbol{J}^{-1}\right).\tag{15}$$

Under mild regularity conditions (for example under mixing conditions), $\widehat{\boldsymbol{\theta}}_{(1)}$ and $\widehat{\boldsymbol{\theta}}_{(2)}$ are asymptotically independent, and we then have

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)=-\boldsymbol{J}^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{\Upsilon}_t+o_P(1)\xrightarrow{\mathcal{L}}\mathcal{N}\left(\boldsymbol{\delta}_3,\boldsymbol{J}^{-1}\boldsymbol{I}\boldsymbol{J}^{-1}\right)\tag{16}$$

with $\boldsymbol{\delta}_3=\sqrt{u_0}\boldsymbol{\delta}_1+\sqrt{1-u_0}\boldsymbol{\delta}_2$.

Theorem 3. *Under $H_1(\boldsymbol{\delta}_1,\boldsymbol{\delta}_2)$ and regularity conditions ensuring $\boldsymbol{I}_n\to\boldsymbol{I}$ almost surely, $\boldsymbol{I}$ non singular, (10)–(11) and (14)–(16), for all $u\in(0,1)$, $\widetilde{S}_n(u)/u(1-u)$ converges in distribution to a noncentral chi-square distribution with d degrees of freedom. When $\sqrt{1-u_0}\boldsymbol{\delta}_1\neq\sqrt{u_0}\boldsymbol{\delta}_2$, the noncentrality parameter is not equal to 0 and the best LAP is obtained for the optimal QLE.*

Proof. Since $\boldsymbol{\Upsilon}_t=\boldsymbol{\Upsilon}_{t,n}$ is such that

$$\frac{1}{\sqrt{nu_0}}\sum_{t=1}^{[nu_0]}\boldsymbol{\Upsilon}_t\xrightarrow{\mathcal{L}}\mathcal{N}\left(\boldsymbol{J}\boldsymbol{\delta}_1,\boldsymbol{I}\right),\quad \frac{1}{\sqrt{n(1-u_0)}}\sum_{t=[nu_0]+1}^{n}\boldsymbol{\Upsilon}_t\xrightarrow{\mathcal{L}}\mathcal{N}\left(\boldsymbol{J}\boldsymbol{\delta}_2,\boldsymbol{I}\right),$$

for $u \le u_0$ we have

$$\frac{1}{\sqrt{nu}} \sum_{t=1}^{[nu]} \left(\boldsymbol{\Upsilon}_t - \frac{1}{\sqrt{nu_0}} \boldsymbol{J} \boldsymbol{\delta}_1 \right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \boldsymbol{I}\right),$$

and thus

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[nu]} \boldsymbol{\Upsilon}_t \xrightarrow{\mathcal{L}} \mathcal{N}\left(\frac{u}{\sqrt{u_0}} \boldsymbol{J} \boldsymbol{\delta}_1, u\boldsymbol{I} \right), \quad \frac{1}{\sqrt{n}} \sum_{t=[nu]+1}^{[nu_0]} \boldsymbol{\Upsilon}_t \xrightarrow{\mathcal{L}} \mathcal{N}\left(\frac{u_0 - u}{\sqrt{u_0}} \boldsymbol{J} \boldsymbol{\delta}_1, (u_0 - u)\boldsymbol{I} \right),$$
$$\frac{1}{\sqrt{n}} \sum_{t=[nu_0]+1}^{n} \boldsymbol{\Upsilon}_t \xrightarrow{\mathcal{L}} \mathcal{N}\left(\sqrt{1-u_0} \boldsymbol{J} \boldsymbol{\delta}_2, (1-u_0)\boldsymbol{I} \right),$$

and for $u \ge u_0$ we have

$$\frac{1}{\sqrt{n}} \sum_{t=[nu_0]+1}^{[nu]} \boldsymbol{\Upsilon}_t \xrightarrow{\mathcal{L}} \mathcal{N}\left(\frac{u-u_0}{\sqrt{1-u_0}} \boldsymbol{J} \boldsymbol{\delta}_2, (u-u_0)\boldsymbol{I} \right), \quad \frac{1}{\sqrt{n}} \sum_{t=[nu]+1}^{n} \boldsymbol{\Upsilon}_t \xrightarrow{\mathcal{L}} \mathcal{N}\left(\frac{1-u}{\sqrt{1-u_0}} \boldsymbol{J} \boldsymbol{\delta}_2, (1-u)\boldsymbol{I} \right).$$

We therefore have

$$\begin{aligned}
\boldsymbol{T}_n^0(u) &= \frac{1}{\sqrt{n}} \left(\sum_{t=1}^{[nu]} \boldsymbol{\Upsilon}_t - u \sum_{t=1}^{n} \boldsymbol{\Upsilon}_t \right) \\
&= \frac{1}{\sqrt{n}} \sum_{t=1}^{[nu]} (1-u) \boldsymbol{\Upsilon}_t - u \frac{1}{\sqrt{n}} \sum_{t=[nu]+1}^{[nu_0]} \boldsymbol{\Upsilon}_t - u \frac{1}{\sqrt{n}} \sum_{t=[nu_0]+1}^{n} \boldsymbol{\Upsilon}_t \\
&\xrightarrow{\mathcal{L}} \boldsymbol{T}_{u_0}(u) \sim \mathcal{N}\left\{ \boldsymbol{J} \boldsymbol{\delta}_{u_0}(u), u(1-u)\boldsymbol{I} \right\}, \quad \boldsymbol{\delta}_{u_0}(u) = \frac{u(1-u_0)}{\sqrt{u_0}} \boldsymbol{\delta}_1 - u\sqrt{1-u_0} \boldsymbol{\delta}_2,
\end{aligned}$$

when $u \le u_0$, and

$$\begin{aligned}
\boldsymbol{T}_n^0(u) &= \frac{1}{\sqrt{n}} \sum_{t=1}^{[nu_0]} (1-u) \boldsymbol{\Upsilon}_t + (1-u) \frac{1}{\sqrt{n}} \sum_{t=[nu_0]+1}^{[nu]} \boldsymbol{\Upsilon}_t - u \frac{1}{\sqrt{n}} \sum_{t=[nu]+1}^{n} \boldsymbol{\Upsilon}_t \\
&\xrightarrow{\mathcal{L}} \boldsymbol{T}_{u_0}(u) \sim \mathcal{N}\left\{ \boldsymbol{J} \boldsymbol{\delta}_{u_0}(u), u(1-u)\boldsymbol{I} \right\}, \quad \boldsymbol{\delta}_{u_0}(u) = \sqrt{u_0}(1-u) \boldsymbol{\delta}_1 - \frac{u_0(1-u)}{\sqrt{1-u_0}} \boldsymbol{\delta}_2,
\end{aligned}$$

when $u \ge u_0$. It follows that, for all $u \in (0,1)$, $\boldsymbol{T}_{u_0}^\top(u) \boldsymbol{I}^{-1} \boldsymbol{T}_{u_0}(u)/u(1-u)$ follows a chi-square distribution with d degrees of freedom and noncentrality parameter

$$\frac{1}{u(1-u)} \boldsymbol{\delta}_{u_0}^\top(u) \boldsymbol{J} \boldsymbol{I}^{-1} \boldsymbol{J} \boldsymbol{\delta}_{u_0}(u),$$

which is maximal for the optimal QLE. We conclude by noting that the noncentral chi-squared distribution satisfies the stochastic-equal-mean order property: the larger the mean (*i.e.* the noncentrality parameter) is, the larger is the cdf, at any point. Note that the noncentrality parameter is maximal at u_0. □

As an illustration of the computation made in the proof of Theorem 3, let us compare the LAPs of the CUSUM, NYBLOM (see Remark 8) and Weighted CUSUM (W-CUSUM) in the

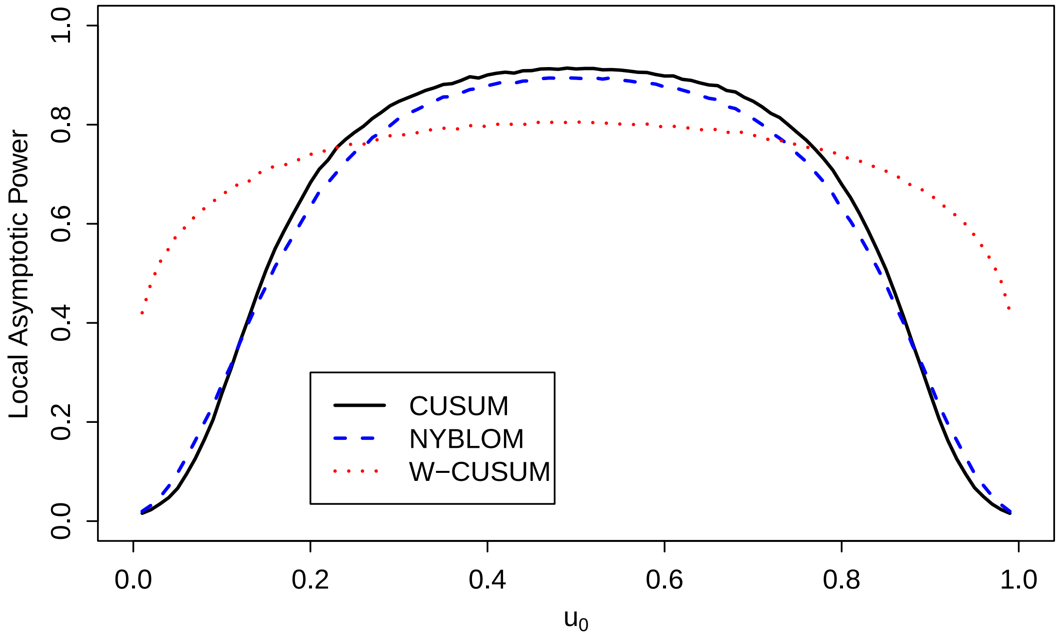

Figure 1: Powers of the CUSUM, Nyblom, and Weighted CUSUM tests as a function of the break date u_0.

simple case where we want to test for the existence of a local break in the mean of a sequence of independent Gaussian variables. The 3 tests reject for large values of $\widetilde{S}_n$ $\widetilde{S}_n^N$ and $\widetilde{S}_n^W$ defined by

$$\widetilde{S}_n = \max_{1\le k<n} \widetilde{S}_n\left(\frac{k}{n}\right), \quad \widetilde{S}_n^N = \frac{1}{n}\sum_{k=1}^{n} \widetilde{S}_n\left(\frac{k}{n}\right), \quad \widetilde{S}_n^W = \max_{1\le k<n} \frac{n^2}{k(n-k)}\widetilde{S}_n\left(\frac{k}{n}\right)$$

with $\widetilde{S}_n(k/n) = \left\{\sum_{t=1}^{k}(y_t-\overline{y})\right\}^2/(n\widehat{\sigma}_y^2)$. The critical values of the tests as well as the LAPs are evaluated by using $50,000$ independent replications of the test statistics with $n = 1,000$. Figure 1 shows the LAPs for the nominal leavel $\alpha = 1\%$ and the alternatives $H_1(\delta_1,\delta_2)$ with $\delta_1 = -\delta_2 = 3$ for a grid of values of $u_0 \in \{0.01, 0.02, \dots, 0.99\}$. As expected, the weighted CUSUM is more powerful than the unweighted version when the local break u_0 comes early or late. Note that the CUSUM and Nyblom tests have similar power, often with a slight advantage for the CUSUM.

5 Change-point tests with misspecified conditional mean

In Sections 2-4, we assumed that $m_t(\boldsymbol{\theta}_0)$ is truly the conditional mean of y_t given $\mathcal{F}_{t-1}$. In this section, we propose to relax this assumption. The intuition is that, even if the conditional mean is not correctly specified, its estimated value should not vary too much when the DGP is stable.

Replace **A3** and **A5** by:

A3* Let $\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}) = \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{y_t - m_t(\boldsymbol{\theta})}{\kappa_{2t}(\boldsymbol{\theta})}$ where $m_t()$ is $\mathcal{F}_{t-1}$-measurable. If $E\{\boldsymbol{\Upsilon}_t(\boldsymbol{\theta})\} = 0$ for some $\boldsymbol{\theta} \in \Theta$, then $\boldsymbol{\theta} = \boldsymbol{\theta}_0^*$, where the so-called pseudo-true parameter $\boldsymbol{\theta}_0^*$ belongs to the interior of the compact set Θ.

Under **A8**, let $\boldsymbol{J}^* = E\frac{\partial}{\partial \boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0^*)$ and assume:

A5* $\boldsymbol{J}^*$ is non singular.

Let $\mu_t = E(y_t \mid \mathcal{F}_{t-1})$. Note that we may have $\mu_t \neq m_t(\boldsymbol{\theta}_0^*)$, and more generally $\mu_t \neq m_t(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \Theta$.

Example 2 (Approximating the conditional mean by an AR)**.** *Assume, perhaps wrongly, that* $m_t(\boldsymbol{\theta}) = a + by_{t-1}$ *with* $\boldsymbol{\theta} = (a, b)^\top$. *We then have*

$$\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}) = \begin{pmatrix} 1 \\ y_{t-1} \end{pmatrix} \frac{1}{\kappa_{2t}}(y_t - a - by_{t-1})$$

Then **A3*** *and* ***A5**** *are satisfied with*

$$\boldsymbol{\theta}_0^* = \boldsymbol{A}^{-1}\boldsymbol{b}, \quad \boldsymbol{b} = \begin{pmatrix} E\frac{y_t}{\kappa_{2t}} \\ E\frac{y_t y_{t-1}}{\kappa_{2t}} \end{pmatrix}, \quad \boldsymbol{A} = \begin{pmatrix} E\frac{1}{\kappa_{2t}} & E\frac{y_{t-1}}{\kappa_{2t}} \\ E\frac{y_{t-1}}{\kappa_{2t}} & E\frac{y_{t-1}^2}{\kappa_{2t}} \end{pmatrix}$$

when $\boldsymbol{b}$ *and* $\boldsymbol{A}$ *exist and* $\boldsymbol{A} = -\boldsymbol{J}^*$ *is invertible (which is for instance the case when* κ_{2t} *is constant and* $Var(y_t) > 0$*).*

Let $\boldsymbol{\Upsilon}_t^* = \boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0^*)$. We now need conditions ensuring the CLT

$$\frac{1}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{\Upsilon}_t^* \xrightarrow{\mathcal{L}} \mathcal{N}(0, \boldsymbol{I}^*) \tag{17}$$

for some long-run variance $\boldsymbol{I}^*$. Let $\{\alpha(h)\}_{h\geq 0}$ be the $\alpha-$mixing (strong mixing) coefficients of the process $(\boldsymbol{\Upsilon}_t^*)_{t\in\mathbb{Z}}$, defined by $\alpha(h) = \sup_{A\in\sigma(\boldsymbol{\Upsilon}_u^*, u\leq t), B\in\sigma(\boldsymbol{\Upsilon}_u^*, u\geq t+h)} |P(A\cap B) - P(A)P(B)|$. We reinforce **A1** by the following assumption.

A1* We have $\|\boldsymbol{\Upsilon}_1^*\|_{2+\nu} < \infty$ and $\sum_{h=1}^\infty \{\alpha(h)\}^{\nu/(2+\nu)} < \infty$ for some $\nu > 0$.

Note that, by Davydov's inequality, **A1*** entails the existence of the matrix $\boldsymbol{I}^*$.

Example 3 (Continuation of Example 2)**.** *Assumption A1** *is satisfied if, for instance,* $\kappa_{2t} \equiv \kappa > 0$, $\|y_1\|_{4+2\nu} < \infty$ *and* $\sum_{h=1}^\infty \{\alpha_y(h)\}^{\nu/(2+\nu)} < \infty$ *for some* $\nu > 0$, *where* $\{\alpha_y(h)\}$ *denotes the sequence of the* $\alpha-$*mixing coefficients of* (y_t). *A non constant weighting sequence* (κ_{2t}) *can reduce the moment requirement. In particular, if* κ_{2t} *is the volatility of an ARCH(1), or more generally* $\kappa_{2t} > c_1 + c_2 y_{t-1}^2$ *with positive constants* c_1 *and* c_2, *then only* $\|y_1\|_{2+\nu} < \infty$ *is required.*

Theorem 4. *Under Assumptions* **A1**, **A1***, **A2**, **A3***, **A4**, **A6**-**A8**, *there exists a QLE* $\widehat{\boldsymbol{\theta}}$ *satisfying*

$$\sum_{t=1}^{n} \widetilde{\boldsymbol{\Upsilon}}_t(\widehat{\boldsymbol{\theta}}) = 0, \quad \widetilde{\boldsymbol{\Upsilon}}_t(\boldsymbol{\theta}) = \frac{1}{\widetilde{\kappa}_{2t}(\boldsymbol{\theta})} \frac{\partial \widetilde{m}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \widetilde{\epsilon}_t(\boldsymbol{\theta}),$$

for n *large enough. Moreover* $\widehat{\boldsymbol{\theta}} \to \boldsymbol{\theta}_0^*$ *a.s. and , under* **A5***,

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0^*\right) = -\boldsymbol{J}^{*-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0^*) + o_P(1) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \boldsymbol{\Sigma}^* := \boldsymbol{J}^{*-1} \boldsymbol{I}^* \boldsymbol{J}^{*-1}\right) \quad as\ n \to \infty.$$

Proof. By standard arguments, it can be shown that **A2**, **A4** and **A7** entail

$$\sup_{\boldsymbol{\theta} \in \Theta} \left\| \sum_{t=1}^{n} \widetilde{\boldsymbol{\Upsilon}}_t(\boldsymbol{\theta}) - \boldsymbol{\Upsilon}_t(\boldsymbol{\theta}) \right\| \leq \sum_{t=1}^{\infty} K_t \rho^t < \infty \quad \text{a.s.} \tag{18}$$

This entails that the initial values that are generally used to compute recursively $\widetilde{m}_t(\boldsymbol{\theta})$ and $\widetilde{\kappa}_{2t}(\boldsymbol{\theta})$ have no consequence on the asymptotic behavior of the QLEs. In particular (18) and the ergodic theorem entail that, for any neighborhood $V(\boldsymbol{\theta}_1)$ of $\boldsymbol{\theta}_1 \in \Theta$,

$$\begin{aligned} \lim_{n\to\infty} \inf_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_1) \cap \Theta} \left\| n^{-1} \sum_{t=1}^{n} \widetilde{\boldsymbol{\Upsilon}}_t(\boldsymbol{\theta}) \right\| &\geq \lim_{n\to\infty} \inf_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_1) \cap \Theta} \left\| n^{-1} \sum_{t=1}^{n} \boldsymbol{\Upsilon}_t(\boldsymbol{\theta}) \right\| \\ &\geq \| E \boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_1) \| - E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_1) \cap \Theta} \| \boldsymbol{\Upsilon}_t(\boldsymbol{\theta}) - \boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_1) \| . \end{aligned}$$

If $V_m(\boldsymbol{\theta})$ denotes the ball of center $\boldsymbol{\theta}$ and radius $1/m$, by the dominated convergence theorem

$$E \sup_{\boldsymbol{\theta} \in V_m(\boldsymbol{\theta}_1) \cap \Theta} \| \boldsymbol{\Upsilon}_t(\boldsymbol{\theta}) - \boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_1) \|$$

is arbitrarily small when m is large enough. By **A3***, we also have $\| E \boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_1) \| > 0$ when $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_0^*$. We thus have shown that for any $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_0^*$, there exists a neighborhood $V(\boldsymbol{\theta}_1)$ of $\boldsymbol{\theta}_1$ such that

$$\liminf_{n\to\infty} \inf_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_1) \cap \Theta} \left\| n^{-1} \sum_{t=1}^{n} \widetilde{\boldsymbol{\Upsilon}}_t(\boldsymbol{\theta}) \right\| > 0, \quad \text{a.s.}$$

and that for any neighbourhood $V(\boldsymbol{\theta}_0^*)$ of $\boldsymbol{\theta}_0^*$

$$\limsup_{n\to\infty} \inf_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0^*) \cap \Theta} \left\| n^{-1} \sum_{t=1}^{n} \widetilde{\boldsymbol{\Upsilon}}_t(\boldsymbol{\theta}) \right\| = 0, \quad \text{a.s.}$$

By compactness of Θ, the existence and consistency of the QLE then follow.

A first order Taylor expansion, (18), and the consistency of $\widehat{\boldsymbol{\theta}}$ imply

$$\mathbf{0}_d = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \widetilde{\boldsymbol{\Upsilon}}_t(\widehat{\boldsymbol{\theta}}) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{\Upsilon}_t(\widehat{\boldsymbol{\theta}}) + o_P(1) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0^*) + \boldsymbol{J}_n^* \sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0^*) + o_P(1),$$

where the row i of $\boldsymbol{J}_n^*$ is of the form $n^{-1}\sum_{t=1}^n \frac{\partial}{\partial \boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_{it}(\boldsymbol{\theta}^*)$, and $\boldsymbol{\theta}^*$ is such that $\|\boldsymbol{\theta}^* - \boldsymbol{\theta}_0^*\| \leq \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0^*\|$. Using the consistency of $\widehat{\boldsymbol{\theta}}$, for n large enough we have

$$\|\boldsymbol{J}_n^* - \boldsymbol{J}^*\| \leq \frac{1}{n}\sum_{t=1}^n \sup_{\boldsymbol{\theta}\in V_m(\boldsymbol{\theta}_0^*)\cap\Theta} \left\| \frac{\partial}{\partial \boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}) - \frac{\partial}{\partial \boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0^*)\right\| + \left\| \boldsymbol{J}^* - \frac{1}{n}\sum_{t=1}^n \frac{\partial}{\partial \boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0^*)\right\|$$

for all m. The ergodic theorem entails that the a.s. limit as $n \to \infty$ of the right-hand side is arbitrarily small when m is large. The Bahadur representation, that is the expression of $\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0^*\right)$, follows. The last result follows from (17), which comes from **A1*** and the Central Limit Theorem of Herrndorf (1984). □

Standard estimators of a long-run variance of the form $\boldsymbol{I}^*$ are the Heteroskedasticity and Autocorrelation Consistent (HAC) estimators (see Newey and West (1987) and Andrews (1991)) and spectral density estimators (see den Haan and Levin, (1997)). Denote by $\boldsymbol{I}_n^*$ a consistent estimator of $\boldsymbol{I}^*$, and consider the process

$$\widetilde{S}_n^* = \sup_{u\in(0,1)} \widetilde{S}_n^*(u), \qquad \widetilde{S}_n^*(u) = \widetilde{\boldsymbol{T}}_n^\top(u)\boldsymbol{I}_n^{*-1}\widetilde{\boldsymbol{T}}_n(u).$$

Theorem 5. *Under Assumptions of Theorem 4, in particular the non-existence of a break, and if* $\boldsymbol{I}^*$ *is invertible we have* $\widetilde{S}_n^* \xrightarrow{\mathcal{L}} S$.

Proof. By the functional CLT of Herrndorf (1984) and the Cramer-Wold device, **A1*** entails

$$\frac{(\boldsymbol{I}^*)^{1/2}}{\sqrt{n}}\sum_{t=1}^{[n\cdot]} \boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0^*) \Longrightarrow \boldsymbol{W}(\cdot),$$

where $\boldsymbol{W}(\cdot)$ denotes a standard d-dimensional Brownian motion. The proof is therefore like that of Theorem 2. □

Remark 10. Note that $\boldsymbol{\theta}_0^*$ in **A3*** may vary with κ_{2t}, and thus the "optimal" test statistics is not necessarily obtained by choosing κ_{2t} proportional to the conditional variance, as is the case when $m_t(\cdot)$ corresponds to the well-specified conditional mean (see Theorems 1 and 3).

6 Change-point estimation

One of the main goals of change-point analysis is to estimate the location of breaks under the alternative. Results on this issue go back to Hinkley (1970) in the case of iid random variables. To cite just one recent reference for a general class of (strong) time series models, Ling (2016) derived asymptotic results on estimated change-points.

Assume that, for $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ belonging to Θ and for $u_0 \in (0, 1]$,

$$y_t = y_{t,n} = \begin{cases} m_t(\boldsymbol{\theta}_1) & \text{if } t \leq [nu_0] \\ m_t(\boldsymbol{\theta}_2) & \text{if } t > [nu_0] \end{cases} + \epsilon_t, \tag{19}$$

where (ϵ_t) is such that $E_{t-1}(\epsilon_t) \equiv 0$. Note that $u_0 = 1$ corresponds to the null hypothesis of no change-point. As in the previous sections, $\widehat{\boldsymbol{\theta}}$ denotes a QLE such that $\sum_{t=1}^n \widetilde{\boldsymbol{\Upsilon}}_t(\widehat{\boldsymbol{\theta}}) = 0$.

We will introduce two stationary processes, $(y_t^{(1)})_{t\in\mathbb{Z}}$ and $(y_t^{(2)})_{t\in\mathbb{Z}}$, which will be used to approximate the observed process before and after the break, respectively. For all $\boldsymbol{\theta} \in \Theta$, let $m_t^{(i)}(\boldsymbol{\theta}) = m(\boldsymbol{\theta}; y_{t-1}^{(i)}, y_{t-2}^{(i)}, \dots)$ and $\kappa_{2t}^{(i)}(\boldsymbol{\theta}) = \kappa_2(\boldsymbol{\theta}; y_{t-1}^{(i)}, y_{t-2}^{(i)}, \dots)$ be stationary approximations of the conditional mean and weight sequence before and after the break.

B1 For $i = 1, 2$, the process $y_t^{(i)} = m_t^{(i)}(\boldsymbol{\theta}_i) + \epsilon_t$, for $t \in \mathbb{Z}$, is strictly stationary and ergodic.

For all $\boldsymbol{\theta} \in \Theta$, let $\boldsymbol{\Upsilon}_t^{(i)}(\boldsymbol{\theta}) = \frac{\partial m_t^{(i)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\epsilon_t^{(i)}(\boldsymbol{\theta})}{\kappa_{2t}^{(i)}(\boldsymbol{\theta})}$ where $\epsilon_t^{(i)}(\boldsymbol{\theta}) = y_t^{(i)} - m_t^{(i)}(\boldsymbol{\theta})$. The pseudo-true parameter value is introduced as follows.

B2 For all $\boldsymbol{\theta}$ in Θ the variables $\boldsymbol{\Upsilon}_t^{(i)}(\boldsymbol{\theta})$ have finite variances, and there is a unique solution $\boldsymbol{\theta}_0^\star = \boldsymbol{\theta}_0^\star(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, belonging to the interior of Θ, to the equation

$$u_0 E\left\{\boldsymbol{\Upsilon}_t^{(1)}(\boldsymbol{\theta})\right\} + (1 - u_0) E\left\{\boldsymbol{\Upsilon}_t^{(2)}(\boldsymbol{\theta})\right\} = 0.$$

We make the following technical assumptions.

B3 For $i = 1, 2$, the function $\kappa_{2t}^{(i)}(\cdot)$ is continuously differentiable and $m_t^{(i)}(\cdot)$ is twice continuously differentiable. Moreover, there exists $\rho \in (0, 1)$ such that, almost surely, for $1 \leq t \leq [nu_0]$,

$$\sup_{\boldsymbol{\theta}\in\Theta} \left\{ |m_t^{(1)}(\boldsymbol{\theta}) - \widetilde{m}_t(\boldsymbol{\theta})| + \left\| \frac{\partial m_t^{(1)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial \widetilde{m}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| + |\kappa_{2t}^{(1)}(\boldsymbol{\theta}) - \widetilde{\kappa}_{2t}(\boldsymbol{\theta})| \right\} \leq K_t^{(1)} \rho^t,$$

and for $t > [nu_0]$,

$$\sup_{\boldsymbol{\theta}\in\Theta} \left\{ |m_t^{(2)}(\boldsymbol{\theta}) - \widetilde{m}_t(\boldsymbol{\theta})| + \left\| \frac{\partial m_t^{(2)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial \widetilde{m}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| + |\kappa_{2t}^{(2)}(\boldsymbol{\theta}) - \widetilde{\kappa}_{2t}(\boldsymbol{\theta})| \right\} \leq K_t^{(2)} \rho^{t-[nu_0]},$$

where $K_t^{(1)}$ is a measurable function of $\{y_u^{(1)} : u < t\}$ and $K_t^{(2)}$ is a measurable function of $\{y_u^{(1)}, y_u^{(2)} : u < t\}$, with $\sup_t E\{K_t^{(i)}\}^r < \infty$ for $i = 1, 2$ and some $r > 0$.

B4 For $i = 1, 2$ $E|y_t^{(i)}|^s < \infty$ and $E \sup_{\boldsymbol{\theta}\in\Theta} \left\{ |m_t^{(i)}(\boldsymbol{\theta})|^s + \left\| \frac{\partial m_t^{(i)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|^s + \left|\kappa_{2t}^{(i)}(\boldsymbol{\theta})\right|^s \right\} < \infty$, for some $s > 0$. Moreover, $\inf_{\boldsymbol{\theta}\in\Theta} \left|\kappa_{2t}^{(i)}(\boldsymbol{\theta})\right| \geq \underline{\kappa}$ a.s. for some constant $\underline{\kappa} > 0$.

B5 For $i = 1, 2$

$$E \sup_{\boldsymbol{\theta}\in\Theta} \left\| \boldsymbol{\Upsilon}_t^{(i)}(\boldsymbol{\theta}) \right\|^2 < \infty \quad \text{and} \quad E \sup_{\boldsymbol{\theta}\in\Theta} \left\| \frac{\partial \boldsymbol{\Upsilon}_t^{(i)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \right\| < \infty.$$

Example 4 (Illustration of Assumptions B1-B5). *Consider the simple case where* $y_t^{(i)} = \theta_i y_{t-1}^{(i)} + \epsilon_t$ *with* (ϵ_t) *a strong white noise. Note that we change the notation by unbolding the scalars. If* $|\theta_i| < 1$ *for* $i = 1, 2$*, then* **B1** *holds. Take* $\Theta \subset (-1, 1)$ *and assume for example that* $\widetilde{\kappa}_{2t} = \kappa_{2t} \propto 1$*,* $\widetilde{m}_t(\theta) = m_t(\theta) = \theta y_{t-1}$ *for* $t \geq 2$*, and* $\widetilde{m}_1(\theta) = 0$*, then* **B3** *holds. To show this result we note that* $y_t = y_t^{(1)}$ *for* $t \leq [nu_0]$*, that* $y_{[nu_0]+1}^{(i)} = \sum_{j=0}^{\infty} \theta_i^j \epsilon_{[nu_0]+1-j}$ *and* $|y_{[nu_0]+1+k}^{(2)} - y_{[nu_0]+1+k}| = |\theta_2|^k |y_{[nu_0]+1}^{(2)} - y_{[nu_0]+1}^{(1)}|$ *for* $k \geq 0$*. So we can choose* $K_t^{(1)} = |y_0|$ *and* $K_t^{(2)} = \sum_{j=0}^{\infty} |\theta_1^j - \theta_2^j| |\epsilon_{[nu_0]+1-j}|$ *in* **B3***. Assumption* **B4** *is always satisfied and if* $E\epsilon_t^4 < \infty$ *then* **B5** *also holds. Finally note that* **B2** *holds with*

$$E\left\{\Upsilon_t^{(1)}(\theta)\right\} = (\theta_i - \theta)\frac{E\epsilon_1^2}{1-\theta_i^2}, \qquad \theta_0^\star = \frac{\frac{u_0\theta_1}{1-\theta_1^2} + \frac{(1-u_0)\theta_2}{1-\theta_2^2}}{\frac{u_0}{1-\theta_1^2} + \frac{1-u_0}{1-\theta_2^2}}.$$

Let the change-point estimator

$$\widetilde{k} = \underset{k\in\{1,\ldots,n-1\}}{\arg\max}\ \widetilde{S}_n(k/n), \qquad \widetilde{S}_n(u) = \widetilde{\boldsymbol{T}}_n^\top(u)\boldsymbol{I}_n^{-1}\widetilde{\boldsymbol{T}}_n(u).$$

The consistency of the change-point estimator is established in the following result.

Theorem 6. *Under Assumptions* **B1**-**B5***, when* $u_0 \in (0,1)$ *and* $E\left\{\boldsymbol{\Upsilon}_t^{(1)}(\boldsymbol{\theta}_0^\star)\right\} \neq E\left\{\boldsymbol{\Upsilon}_t^{(2)}(\boldsymbol{\theta}_0^\star)\right\}$ *we have*

$$\frac{\widetilde{k}}{n} \to u_0, \quad \textit{in probability as } n \to \infty.$$

Proof. Let

$$\boldsymbol{T}_n(u) = \begin{cases} \frac{1}{\sqrt{n}}\sum_{t=1}^{[nu]} \boldsymbol{\Upsilon}_t^{(1)}(\widehat{\boldsymbol{\theta}}) & \text{if } u \leq u_0, \\ \frac{1}{\sqrt{n}}\sum_{t=1}^{[nu_0]} \boldsymbol{\Upsilon}_t^{(1)}(\widehat{\boldsymbol{\theta}}) + \frac{1}{\sqrt{n}}\sum_{t=[nu_0]+1}^{[nu]} \boldsymbol{\Upsilon}_t^{(2)}(\widehat{\boldsymbol{\theta}}) & \text{if } u > u_0. \end{cases}$$

i) We start by showing that

$$\sup_{u\in(0,1)} \|\boldsymbol{T}_n(u) - \widetilde{\boldsymbol{T}}_n(u)\| = o_P(1). \tag{20}$$

We first consider the supremum over $u \in (u_0, 1)$, which is the most complicated term. We have

$$\begin{aligned}
\sup_{u\in(u_0,1)} \|\boldsymbol{T}_n(u) - \widetilde{\boldsymbol{T}}_n(u)\| &\leq \sup_{\boldsymbol{\theta}\in\Theta} \left\| \frac{1}{\sqrt{n}} \sum_{t=1}^{[nu_0]} \frac{\partial \widetilde{m}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\widetilde{\epsilon}_t(\boldsymbol{\theta})}{\widetilde{\kappa}_{2t}(\boldsymbol{\theta})} - \frac{\partial m_t^{(1)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\epsilon_t^{(1)}(\boldsymbol{\theta})}{\kappa_{2t}^{(1)}(\boldsymbol{\theta})} \right\| \\
&+ \sup_{u\in(u_0,1)} \sup_{\boldsymbol{\theta}\in\Theta} \left\| \frac{1}{\sqrt{n}} \sum_{t=1+[nu_0]}^{[nu]} \frac{\partial \widetilde{m}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\widetilde{\epsilon}_t(\boldsymbol{\theta})}{\widetilde{\kappa}_{2t}(\boldsymbol{\theta})} - \frac{\partial m_t^{(2)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\epsilon_t^{(2)}(\boldsymbol{\theta})}{\kappa_{2t}^{(2)}(\boldsymbol{\theta})} \right\| \\
&\leq \frac{1}{\sqrt{n}} \sum_{t=1}^{[nu_0]} \sup_{\boldsymbol{\theta}\in\Theta} \left(\left\| \frac{\partial m_t^{(1)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial \widetilde{m}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| \frac{|\epsilon_t^{(1)}(\boldsymbol{\theta})|}{\kappa_{2t}^{(1)}(\boldsymbol{\theta})} \right) \\
&+ \frac{1}{\sqrt{n}} \sum_{t=1}^{[nu_0]} \sup_{\boldsymbol{\theta}\in\Theta} \left(|m_t^{(1)}(\boldsymbol{\theta}) - \widetilde{m}_t(\boldsymbol{\theta})| \left\| \frac{\partial \widetilde{m}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| \frac{1}{\kappa_{2t}^{(1)}(\boldsymbol{\theta})} \right) \\
&+ \frac{1}{\sqrt{n}} \sum_{t=1}^{[nu_0]} \sup_{\boldsymbol{\theta}\in\Theta} \left(\left| \frac{1}{\kappa_{2t}^{(1)}(\boldsymbol{\theta})} - \frac{1}{\widetilde{\kappa}_{2t}(\boldsymbol{\theta})} \right| \left\| \frac{\partial \widetilde{m}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| |\widetilde{\epsilon}_t(\boldsymbol{\theta})| \right) \\
&+ \frac{1}{\sqrt{n}} \sum_{t=1+[nu_0]}^{n} \sup_{\boldsymbol{\theta}\in\Theta} \left(\left\| \frac{\partial m_t^{(2)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial \widetilde{m}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| \frac{|\epsilon_t^{(2)}(\boldsymbol{\theta})|}{\kappa_{2t}^{(2)}(\boldsymbol{\theta})} \right) \\
&+ \frac{1}{\sqrt{n}} \sum_{t=1+[nu_0]}^{n} \sup_{\boldsymbol{\theta}\in\Theta} \left(|m_t^{(2)}(\boldsymbol{\theta}) - \widetilde{m}_t(\boldsymbol{\theta})| \left\| \frac{\partial \widetilde{m}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| \frac{1}{\kappa_{2t}^{(2)}(\boldsymbol{\theta})} \right) \\
&+ \frac{1}{\sqrt{n}} \sum_{t=1+[nu_0]}^{n} \sup_{\boldsymbol{\theta}\in\Theta} \left(\left| \frac{1}{\kappa_{2t}^{(2)}(\boldsymbol{\theta})} - \frac{1}{\widetilde{\kappa}_{2t}(\boldsymbol{\theta})} \right| \left\| \frac{\partial \widetilde{m}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| |\widetilde{\epsilon}_t(\boldsymbol{\theta})| \right).
\end{aligned}$$

By Assumptions **B3-B4**, the first term in the right-hand side is bounded by

$$\frac{1}{\underline{\kappa}\sqrt{n}} \sum_{t=1}^{\infty} \sup_{\boldsymbol{\theta}\in\Theta} |\epsilon_t^{(1)}(\boldsymbol{\theta})| K_t^{(1)} \rho^t = O\left(\frac{1}{\sqrt{n}}\right), \quad a.s.$$

using the existence of a bound for a small-order moment for $\sup_{\boldsymbol{\theta}\in\Theta} |\epsilon_t^{(1)}(\boldsymbol{\theta})| K_t^{(1)}$. The other terms can be handled in the same way. We similarly show that $\sup_{u\in(0,u_0)} \|\boldsymbol{T}_n(u) - \widetilde{\boldsymbol{T}}_n(u)\| = o_P(1)$. Thus (20) is established.

ii) Now we prove that

$$\widehat{\boldsymbol{\theta}} \to \boldsymbol{\theta}_0^{\star} \quad \text{a.s. as } n \to \infty. \tag{21}$$

We note that $\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\Theta} \left\| n^{-1} \sum_{t=1}^{n} \widetilde{\boldsymbol{\Upsilon}}_t(\boldsymbol{\theta}) \right\|$. Let $\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}) = \boldsymbol{\Upsilon}_t^{(1)}(\boldsymbol{\theta})$ if $t \leq [nu]$, and $\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}) = \boldsymbol{\Upsilon}_t^{(2)}(\boldsymbol{\theta})$ otherwise. For any neighborhood $V(\boldsymbol{\theta}_3)$ of $\boldsymbol{\theta}_3 \in \Theta$, using the fact that

$\sup_{\boldsymbol{\theta}\in\Theta}\left\|\sum_{t=1}^{\infty}\widetilde{\boldsymbol{\Upsilon}}_t(\boldsymbol{\theta})-\boldsymbol{\Upsilon}_t(\boldsymbol{\theta})\right\|<\infty$ a.s., we have

$$\begin{aligned}
&\lim_{n\to\infty}\inf_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_3)\cap\Theta}\left\|n^{-1}\sum_{t=1}^{n}\widetilde{\boldsymbol{\Upsilon}}_t(\boldsymbol{\theta})\right\| \\
&\geq \lim_{n\to\infty}\left\|n^{-1}\sum_{t=1}^{n}\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_3)\right\|-\lim_{n\to\infty}\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_3)\cap\Theta}\left\|n^{-1}\sum_{t=1}^{n}\boldsymbol{\Upsilon}_t(\boldsymbol{\theta})\right\| \\
&\geq \left\|u_0E\left\{\boldsymbol{\Upsilon}_t^{(1)}(\boldsymbol{\theta}_3)\right\}+(1-u_0)E\left\{\boldsymbol{\Upsilon}_t^{(2)}(\boldsymbol{\theta}_3)\right\}\right\|-\sup_{i=1,2}E\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_3)\cap\Theta}\left\|\boldsymbol{\Upsilon}_t^{(i)}(\boldsymbol{\theta})-\boldsymbol{\Upsilon}_t^{(i)}(\boldsymbol{\theta}_3)\right\|,
\end{aligned}$$

where the first term in the r.h.s. is positive for $\boldsymbol{\theta}_3\neq\boldsymbol{\theta}_0^*$, while the second term can be made arbitrarily small when the neighborhood shrinks to the singleton $\{\boldsymbol{\theta}_3\}$ by arguments already given. The consistency of $\widehat{\boldsymbol{\theta}}$ follows as in the proof of Theorem 4.

iii) Under **B5** let

$$\begin{aligned}
\boldsymbol{I}^\star &= u_0E\left\{\boldsymbol{\Upsilon}_t^{(1)}(\boldsymbol{\theta}_0^\star)\left(\boldsymbol{\Upsilon}_t^{(1)}(\boldsymbol{\theta}_0^\star)\right)^\top\right\}+(1-u_0)E\left\{\boldsymbol{\Upsilon}_t^{(2)}(\boldsymbol{\theta}_0^\star)\left(\boldsymbol{\Upsilon}_t^{(2)}(\boldsymbol{\theta}_0^\star)\right)^\top\right\}, \\
\boldsymbol{J}^\star &= u_0E\left\{\frac{\partial}{\partial\boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t^{(1)}(\boldsymbol{\theta}_0^\star)\right\}+(1-u_0)E\left\{\frac{\partial}{\partial\boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t^{(2)}(\boldsymbol{\theta}_0^\star)\right\}.
\end{aligned}$$

We will show that

$$\sup_{u\in(0,1)}\left|\frac{1}{n}\widetilde{S}_n(u)-L(u)\right|=o_P(1) \tag{22}$$

where

$$L(u)=\begin{cases}\{u(1-u_0)\}^2\boldsymbol{\Delta}'(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)\boldsymbol{I}^{\star-1}\boldsymbol{\Delta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) & \text{if } u\leq u_0\\ \{u_0(1-u)\}^2\boldsymbol{\Delta}'(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)\boldsymbol{I}^{\star-1}\boldsymbol{\Delta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) & \text{if } u> u_0\end{cases}$$

and, recalling that $\boldsymbol{\theta}_0^\star$ depends on $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$,

$$\boldsymbol{\Delta}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)=E\left\{\boldsymbol{\Upsilon}_t^{(1)}(\boldsymbol{\theta}_0^\star)\right\}-E\left\{\boldsymbol{\Upsilon}_t^{(2)}(\boldsymbol{\theta}_0^\star)\right\}.$$

By the arguments of the proof of Theorem 2 we have, with $\widehat{\boldsymbol{\theta}}_u^*$ between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0^\star$,

$$\frac{1}{\sqrt{n}}\boldsymbol{T}_n(u)=\frac{1}{n}\sum_{t=1}^{[nu]}\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0^\star)+u\left(\frac{1}{nu}\sum_{t=1}^{[nu]}\frac{\partial}{\partial\boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t(\widehat{\boldsymbol{\theta}}_u^*)\right)(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0^\star)$$

Given that $\widetilde{\boldsymbol{T}}_n(1)=0$ in view of (20) we have, for $\widehat{\boldsymbol{\theta}}^*$ between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0^\star$,

$$\boldsymbol{T}_n(1)=o_P(1)=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0^\star)+\boldsymbol{J}_n^\star\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0^\star),\quad \boldsymbol{J}_n^\star=\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t(\widehat{\boldsymbol{\theta}}^*).$$

Moreover, we have

$$\begin{aligned}
\boldsymbol{J}_n^\star =&\frac{1}{nu}\sum_{t=1}^{[nu]}\frac{\partial}{\partial\boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t(\widehat{\boldsymbol{\theta}}_u^\star)+\frac{1}{nu}\sum_{t=1}^{[nu]}\left\{\frac{\partial}{\partial\boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t(\widehat{\boldsymbol{\theta}}^\star)-\frac{\partial}{\partial\boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t(\widehat{\boldsymbol{\theta}}_u^\star)\right\}\\
&-\frac{1}{nu}\sum_{t=1}^{[nu]}\left(\frac{\partial}{\partial\boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t(\widehat{\boldsymbol{\theta}}^\star)-\boldsymbol{J}_n^\star\right)+\left(1-\frac{[nu]}{nu}\right)\boldsymbol{J}_n^\star.
\end{aligned}$$

Thus, by already given arguments

$$\begin{aligned}\frac{1}{\sqrt{n}}\boldsymbol{T}_n(u) =& \frac{1}{n}\left(\sum_{t=1}^{[nu]}\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0^\star) - u\sum_{t=1}^{n}\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0^\star)\right) + \frac{1}{n}\sum_{t=1}^{[nu]}\left(\frac{\partial}{\partial\boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t(\widehat{\boldsymbol{\theta}}_u^*) - \boldsymbol{J}_n^\star\right)(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0^\star) + uo_P(1)\\ :=& \frac{1}{\sqrt{n}}\boldsymbol{T}_n^0(u) + \boldsymbol{R}_n(u)(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0^\star) + o_P(1),\end{aligned} \tag{23}$$

where the reminder term is independent of u. We will show that, in contrast with the first convergence in (11) of the proof of Theorem 2, we have

$$\sup_{u\in(0,1)}\|\boldsymbol{R}_n(u)\| = O_P(1). \tag{24}$$

We have, using the consistency of $\widehat{\boldsymbol{\theta}}_u^*$ to $\boldsymbol{\theta}_0^\star$,

$$\begin{aligned}\|\boldsymbol{R}_n(u)\| \leq& \left\|\frac{1}{n}\sum_{t=1}^{[nu]}\left(\frac{\partial}{\partial\boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0^\star) - \boldsymbol{J}^\star\right)\right\| + \sup_{\boldsymbol{\theta}\in V_k(\boldsymbol{\theta}_0^\star)}\left\|\frac{1}{n}\sum_{t=1}^{[nu]}\left(\frac{\partial}{\partial\boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0^\star) - \frac{\partial}{\partial\boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t(\boldsymbol{\theta})\right)\right\|\\ &+ \|\boldsymbol{J}_n^\star - \boldsymbol{J}^\star\|.\end{aligned}$$

As in the proof of Theorem 2, it can be shown that the second term in the r.h.s. converges to 0 in probability, uniformly in $u \in (0,1)$. It can also be shown that the third term converges to 0 in probability. Moreover,

$$\begin{aligned}\sup_{u\in(0,u_0)}\left\|\frac{1}{n}\sum_{t=1}^{[nu]}\left(\frac{\partial}{\partial\boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0^\star) - \boldsymbol{J}^\star\right)\right\| \leq& \sup_{u\in(0,u_0)}\left\|\frac{1}{n}\sum_{t=1}^{[nu]}\left(\frac{\partial}{\partial\boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}_0^\star) - E\left\{\frac{\partial}{\partial\boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t^{(1)}(\boldsymbol{\theta}_0^\star)\right\}\right)\right\|\\ &+ \left\|E\left\{\frac{\partial}{\partial\boldsymbol{\theta}^\top}\boldsymbol{\Upsilon}_t^{(1)}(\boldsymbol{\theta}_0^\star)\right\} - \boldsymbol{J}^\star\right\|\end{aligned}$$

where the first term in the r.h.s. converges to 0 in probability by already given arguments. It can also be shown that a similar bound holds when the supremum of the l.h.s. term is taken over $(u_0,1)$. Thus (24) is established, from which it follows that the second term in the r.h.s. of (23) converges to 0 in probability as $n \to \infty$.

Now we have, for $u \leq u_0$,

$$\begin{aligned}\frac{1}{\sqrt{n}}\boldsymbol{T}_n^0(u) =& u(1-u)\frac{1}{[nu]}\sum_{t=1}^{[nu]}\boldsymbol{\Upsilon}_t^{(1)}(\boldsymbol{\theta}_0^\star) - u(u_0-u)\frac{1}{[nu_0]-[nu]}\sum_{t=[nu]+1}^{[nu_0]}\boldsymbol{\Upsilon}_t^{(1)}(\boldsymbol{\theta}_0^\star)\\ &- u(1-u_0)\frac{1}{n-[nu_0]}\sum_{t=[nu_0]+1}^{n}\boldsymbol{\Upsilon}_t^{(2)}(\boldsymbol{\theta}_0^\star) + o_P(1)\\ &\to u(1-u)E\left\{\boldsymbol{\Upsilon}_t^{(1)}(\boldsymbol{\theta}_0^\star)\right\} - u(u_0-u)E\left\{\boldsymbol{\Upsilon}_t^{(1)}(\boldsymbol{\theta}_0^\star)\right\}\\ &- u(1-u_0)E\left\{\boldsymbol{\Upsilon}_t^{(2)}(\boldsymbol{\theta}_0^\star)\right\}, \quad \text{in probability as } n\to\infty.\end{aligned}$$

Thus, for $u \leq u_0$,

$$\frac{1}{\sqrt{n}} \boldsymbol{T}_n^0(u) \to u(1-u_0)\boldsymbol{\Delta}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2), \quad \text{in probability as } n \to \infty,$$

and we similarly show that, for $u > u_0$,

$$\frac{1}{\sqrt{n}} \boldsymbol{T}_n^0(u) \to u_0(1-u)\boldsymbol{\Delta}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2), \quad \text{in probability as } n \to \infty.$$

Thus, (22) is not yet established but we have shown that $\frac{1}{\sqrt{n}} \boldsymbol{T}_n^0(u) \to \boldsymbol{T}(u)$ for all $u \in (0,1)$, where

$$\boldsymbol{T}(u) = \begin{cases} u(1-u_0)\boldsymbol{\Delta}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) & \text{if } u \leq u_0, \\ u_0(1-u)\boldsymbol{\Delta}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) & \text{if } u > u_0. \end{cases}$$

Now we have, letting $\boldsymbol{Y}_t^{(i)} = \boldsymbol{\Upsilon}_t^{(i)}(\boldsymbol{\theta}_0^\star) - E\boldsymbol{\Upsilon}_t^{(i)}(\boldsymbol{\theta}_0^\star)$ for $i = 1, 2$,

$$\sup_{u \in (0,u_0)} \left\| \frac{1}{\sqrt{n}} \boldsymbol{T}_n^0(u) - \boldsymbol{T}(u) \right\| \leq \sup_{u \in (0,u_0)} \left\| \frac{u}{[nu]} \sum_{t=1}^{[nu]} \boldsymbol{Y}_t^{(1)} \right\| + \left\| \frac{u_0 - u}{[nu_0] - [nu]} \sum_{t=[nu]+1}^{[nu_0]} \boldsymbol{Y}_t^{(1)} \right\| \\ + (1-u_0) \left\| \frac{1}{n - [nu_0]} \sum_{t=[nu_0]+1}^{n} \boldsymbol{Y}_t^{(2)} \right\| + o_p(1). \qquad (25)$$

The third term in the r.h.s. converges to 0 in probability as $n \to \infty$ in view of the stationarity and ergodicity of $\boldsymbol{Y}_t^{(2)}$. Moreover, by the arguments used in the proof of Theorem 2,

$$\sup_{u \in (0,u_0)} \left\| \frac{u}{[nu]} \sum_{t=1}^{[nu]} \boldsymbol{Y}_t^{(1)} \right\| \leq \sup_{1 \leq k \leq k_n} \left\| \frac{2}{n} \sum_{t=1}^{k} \boldsymbol{Y}_t^{(1)} \right\| + \sup_{k_n \leq k \leq n} \left\| \frac{2}{n} \sum_{t=1}^{k} \boldsymbol{Y}_t^{(1)} \right\|$$

where the last term in the r.h.s. converges to 0 a.s. and, by the Markov inequality, for any $\iota > 0$, and $k_n/n \to 0$,

$$P\left(\sup_{1 \leq k \leq u_n} \left\| \frac{1}{n} \sum_{t=1}^{k} \boldsymbol{Y}_t^{(1)} \right\| > \iota \right) \leq P\left(\frac{1}{n} \sum_{t=1}^{k_n} \left\| \boldsymbol{Y}_t^{(1)} \right\| > \iota \right) \leq \frac{k_n}{n\iota} E \left\| \boldsymbol{Y}_1^{(1)} \right\| \to 0,$$

as $n \to \infty$, from which we deduce that the first term in the r.h.s. of (25) converges in probability to 0. The second termcan be handled similarly. It follows that the l.h.s. of (25) converges in probability to 0. We similarly show that the same convergence holds when the supremum is taken over $(u_0, 1)$. We thus have shown that

$$\sup_{u \in (0,1)} \left\| \frac{1}{\sqrt{n}} \boldsymbol{T}_n^0(u) - \boldsymbol{T}(u) \right\| \to 0, \quad \text{in probability as } n \to \infty.$$

In view of equations (20), (23) and (24), we also have

$$\sup_{u \in (0,1)} \left\| \frac{1}{\sqrt{n}} \widetilde{\boldsymbol{T}}_n(u) - \boldsymbol{T}(u) \right\| \to 0, \quad \text{in probability as } n \to \infty. \qquad (26)$$

Noting that the matrix $\boldsymbol{I}_n$ converges in probability to $\boldsymbol{I}^\star$, the convergence in (22) is established. To conclude, it suffices to apply the argmax theorem (see Theorem 3.2.2 of Van der Vaart and Wellner, 1996). □

Example 5 (Unconditional mean of Gaussian variables). *Assume that $y_1, \dots, y_n$ are independent and Gaussian with variance σ^2, and mean equal to θ_1 when $t \le [nu_0]$ and to θ_2 when $t > [nu_0]$. We have $\widetilde{T}_n(u) = T_n(u) = n^{-1/2}\sum_{t=1}^{[nu]}(y_t - \overline{y})$ where $\overline{y} = n^{-1}\sum_{t=1}^n y_t$. Assumption* **B2** *is thus satisfied with $\theta_0^* = u_0\theta_1 + (1-u_0)\theta_2$. We also have $\boldsymbol{\Upsilon}_t^{(i)}(\theta) = y_t^{(i)} - \theta$ for $i = 1, 2$. Thus $\Delta(\theta_1, \theta_2) = \theta_2 - \theta_1$.*

7 Numerical illustrations

7.1 Monte Carlo experiments

Francq and Zakoian (2023) display Monte Carlo experiments for comparison of the finite-sample properties of a set of QLEs in different settings. They consider the case where the distribution of y_t conditional on $\mathcal{F}_{t-1}$ is a Gamma law of shape parameter $k_t = m_t^2/(k\sigma_t^2)$ and rate parameter $\theta_t = k\sigma_t^2/m_t$, in such a way that $E_{t-1}(y_t) = m_t$ and $\mathrm{Var}_{t-1}(y_t) = k\sigma_t^2$. We keep this framework, considering the ARMA(1,1)-type conditional mean $m_t = c_0 + a_0 y_{t-1} + b_0 m_{t-1}$ and 4 possibilities for the conditional variance: $\sigma_t^2 = 1$ in DGP A (as for a standard ARMA model), $\sigma_t^2 = m_t$ in DGP B (as for an INGARCH count time series model), $\sigma_t^2 = m_t^2$ in DGP C (as for an ACD duration model, or the square of a GARCH), $\sigma_t^2 = m_t^{3/2}$ in DGP D (which does not correspond to any standard model).

We considered 8 different QLEs solving (4): for the estimators A, B, C and D the weight sequence $\widetilde{\kappa}_{2t}$ is proportional to 1, $\widetilde{m}_t(\boldsymbol{\theta})$, $\widetilde{m}_t^2(\boldsymbol{\theta})$ and $\widetilde{m}_t^{3/2}(\boldsymbol{\theta})$, respectively, and for the 4 other estimators the weights are data driven. More precisely, the estimator Q is the one that minimizes the QLIK_n loss of Remark 4 over the 4 weighting sequences of the estimators A-D. The QLE called G estimates the weights by the GARCH(1,1) model (6), with the QLE A as first step estimator $\widehat{\boldsymbol{\theta}}$. The QLEs X1 and X2 estimate the weights by (7) and (8), respectively, also with the QLE A as first step estimator. The left part of Table 1 displays the empirical sizes of tests based on the 8 estimators for each of the 4 DGPs, when $n = 2,000$, $(c_0, a_0, b_0) = (0.01, 0.1, 0.89)$ and the nominal level $\alpha \in \{1\%, 5\%, 10\%\}$. The relative frequencies of rejection of the null $\mathbf{H_0}$ of no break, presented in Table 1 and in the other tables, are computed over $1,000$ independent replications of each DGP. For a test of level 1% (respectively 5%, respectively 10%), the empirical relative frequency of rejection over 1000 independent replications should vary between 0.4% and

1.7% (respectively 3.7% and 6.4%, respectively 8.2% and 11.9%) with a probability of about 95%. The empirical relative frequencies of rejection that fall outside these bounds are highlighted in color (red if the empirical frequency of rejection is too high and blue if it is too low). Looking at this table, it appears that the first-order errors are globally well controlled by the optimal tests and, more importantly, by the data-estimated optimal tests (except for a few slight under-rejections). However, non-optimal QLEs may have poor empirical sizes, which is the main motivation for using the proposed data-estimated optimal tests.

To compare the power of the different tests, we considered DGPs with a break at $t = 800$. For $t = 1, \dots, 800$ we took $(c_0, a_0, b_0) = (0.01, 0.1, 0.89)$ and for $t = 801, \dots, 2000$ we took $(c_0, a_0, b_0) = (0.15, 0.1, 0.75)$. The other parameters are unchanged, leading to DGPs A*-D*. Note that before and after the break the marginal mean $c_0/(1 - a_0 - b_0) = 1$ and the update parameter $a_0 = 0.1$ remain the same for all DGPs, only the persistence parameter b_0 changes. Despite the fact that the DGPs were chosen so that it is not possible for the eye to detect a change point on the trajectories, the right part of Table 1 shows that the tests are often able to detect the break. As expected, for the DGP $\mathrm{X} \in \{\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}\}$, the most powerful test is (or is close to) X among $\{\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}\}$. Interestingly, the data-selected estimators always perform very well, often as well as the optimal estimator. The poorer performing tests are highlighted in color. These underperforming tests are never the data-selected ones.

Figure 2 shows the empirical distributions of the change point estimates obtained with the 8 different tests. The simulated DGP is DGPA*, for simulation length $n = 8000$ and with the change point at $nu_0 = 3200$. As can be seen in the figure, the optimal and data-selected tests perform better than the other tests, not only in detecting the presence of a break, but also in estimating its position.

7.2 Application on exchange rates

As a simple real data illustration, consider the returns of the daily exchange rates of the US dollar (USD) and the Swiss franc (CHF) against the euro from 1999-01-04 to 2022-07-12 (corresponding to 6025 observations).

GARCH(1,1) models on the returns (i.e. ARMA(1,1) on the squared returns y_t) are estimated by QLEs and tests for breaks are performed using the test statistic $\widetilde{S}_n$, for which the optimal weights are estimated by the data-driven procedure (QLIK or based on the 3 GARCH models defined in the previous section).

Figure 3 shows that there is no evidence of breaks for USD, but strong evidence of breaks

α	A	B	C	D	Q	G	X1	X2	A	B	C	D	Q	G	X1	X2
				DGP A								DGP A*				
1%	1.0	2.7	4.9	4.7	0.9	0.6	0.7	0.8	79.8	32.6	20.2	26.0	71.5	77.2	76.8	77.2
5%	5.3	8.8	12.1	12.8	4.9	5.8	5.8	5.6	94.3	53.4	35.4	47.2	88.9	93.3	92.8	92.9
10%	10.1	14.6	18.3	18.7	9.5	11.0	10.2	10.1	97.0	65.7	46.3	60.4	93.1	96.7	96.6	96.6
				DGP B								DGP B*				
1%	0.9	0.7	1.7	1.1	0.7	0.4	0.7	0.7	59.9	80.5	30.0	69.6	79.6	77.2	82.1	82.3
5%	4.1	4.5	7.4	5.0	4.5	2.9	3.9	4.1	80.6	95.6	54.6	89.4	95.5	94.4	95.7	95.8
10%	9.8	8.4	13.0	10.8	8.4	8.4	8.3	8.8	89.7	98.4	68.5	96.0	98.4	98.3	98.7	98.5
				DGP C								DGP C*				
1%	5.8	1.1	0.7	0.7	0.7	0.6	0.8	0.8	49.7	70.9	81.8	83.2	81.8	81.7	83.5	88.8
5%	14.8	4.7	3.9	4.3	3.9	3.2	3.6	4.2	65.5	84.2	95.5	94.6	95.6	95.2	95.8	96.9
10%	22.9	9.9	9.3	8.8	9.3	7.3	8.0	8.7	74.0	91.2	97.9	98.0	97.9	97.6	97.7	98.5
				DGP D								DGP D*				
1%	1.7	1.1	1.7	0.9	0.9	0.7	1.0	1.1	52.3	79.7	67.1	86.7	82.6	81.1	81.6	84.5
5%	7.3	3.9	6.2	5.0	5.0	3.9	4.0	5.2	66.8	94.0	87.6	96.4	95.6	95.6	95.6	96.6
10%	13.6	8.3	10.6	9.7	9.8	8.8	8.7	9.5	76.1	97.2	94.9	98.0	97.7	98.1	98.1	98.1

Table 1: Empirical size (DGP A–D) and power (DGP A*–D*) of 8 QLE-based tests.

for CHF. The breakpoints are September 6, 2011 and January 15, 2015. In fact, the Swiss franc was pegged to the euro between these two dates.

References

[1] Aknouche, A. and Francq, C. (2022) Two-stage weighted least squares estimator of the conditional mean of observation-driven time series models. *Journal of Econometrics*, in press.

[2] Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

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

[4] Bera, A.K., Bilias, Y. and Simlai, P. (2006) Estimating functions and equations: An essay on historical developments with applications to econometrics. *Palgrave Handbook of Econometrics* 1, 427–476.

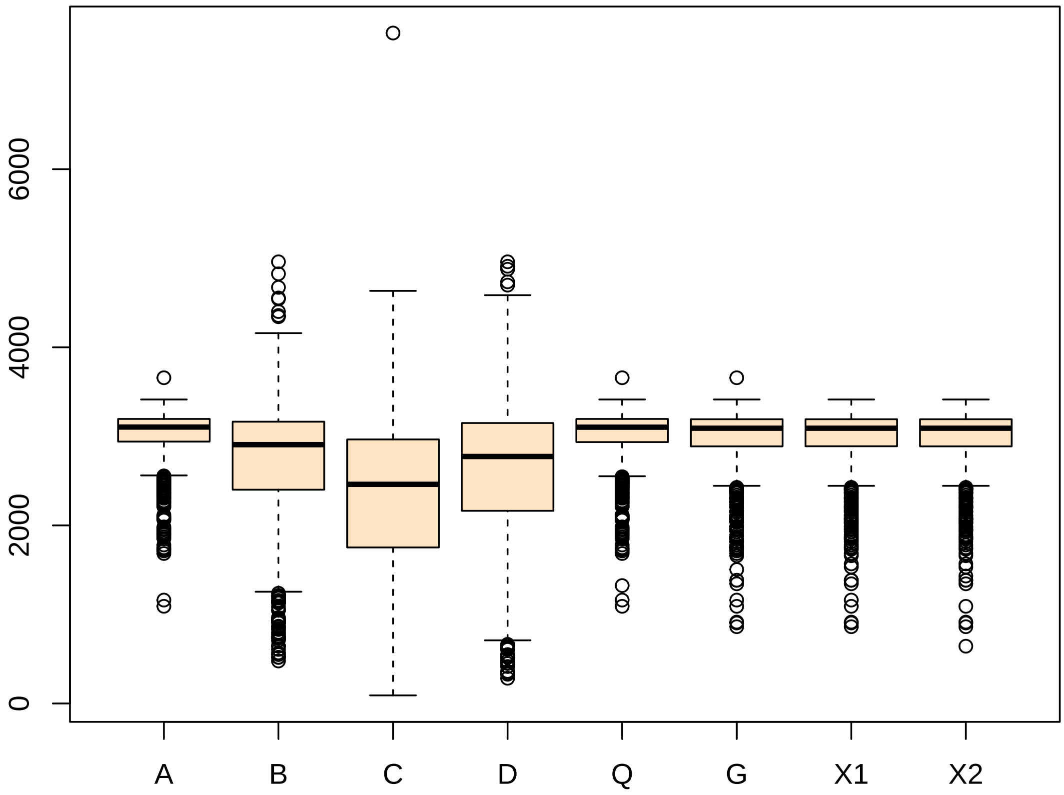

Figure 2: Distributions of the change point estimes

[5] Berkes, I., Horváth, L. and Kokoszka, P. (2004) Testing for parameter constancy in GARCH(p, q) models. *Statistics & Probability Letters* 70, 263–273.

[6] Chandra, A.S. and Taniguchi, M. (2001) Estimating functions for non-linear time series models. *Annals of the Institute of Statistical Mathematics* 53, 125–141.

[7] Csörgö, M. and Horváth, L. (1997) *Limit theorems in change-point analysis.* Chichester: Wiley.

[8] den Haan, W.J. and Levin, A. (1997) A Practitioner's Guide to Robust Covariance Matrix Estimation. *In Handbook of Statistics* 15, Rao, C.R. and G.S. Maddala (eds), 291–341.

[9] Durbin, J. (1960) Estimation of parameters in time-series regression models. *Journal of the Royal Statistical Society Series B* 22, 139–153.

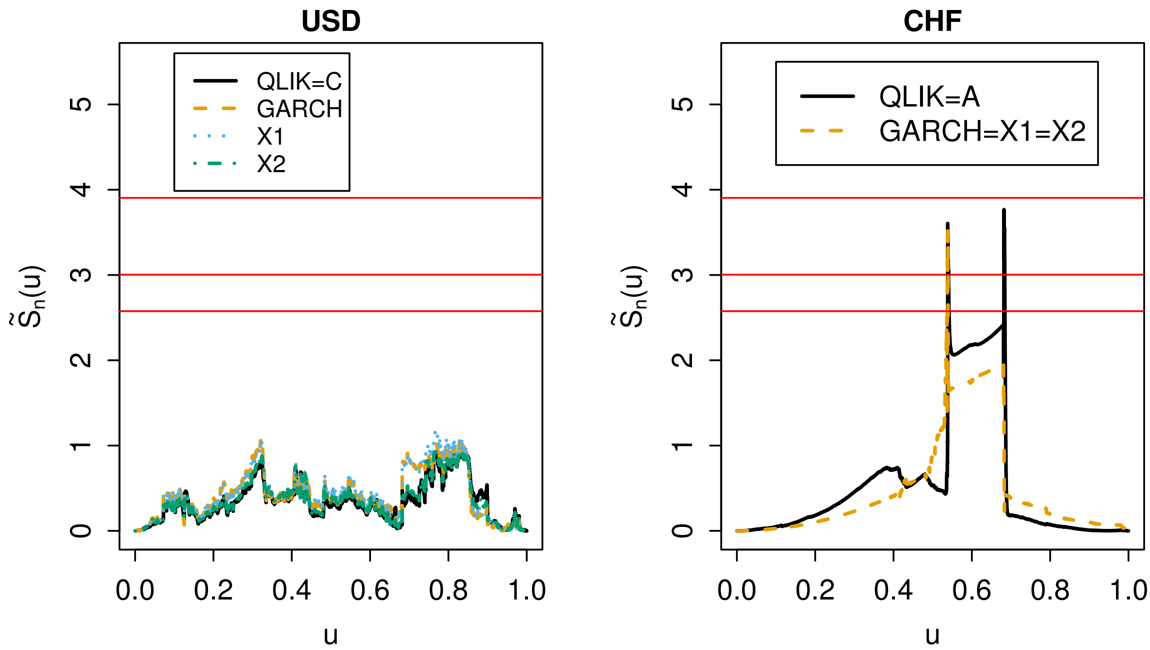

Figure 3: Trajectories of the CUSUM statistics $\widetilde{S}_n(u)$ for different QLEs and 2 exchange rates

[10] Francq, C. and Zakoïan, J-M. (2019) GARCH models: structure, statistical inference and financial applications. Chichester: John Wiley, second edition.

[11] Francq, C. and Zakoïan, J-M. (2023) Optimal estimating function for weak location-scale dynamic models. *Journal of Time Series Analysis* 44, 533-555.

[12] Godambe, V.P. (1960) An optimum property of regular maximum likelihood estimation. *Annals of Mathematical Statistics* 31, 1208–1212.

[13] Godambe, V. P. (1985) The foundations of finite sample estimation in stochastic processes. *Biometrika* 72, 419–428.

[14] Godambe, V.P. and Heyde, C.C. (1987) Quasi-likelihood and optimal estimation. *International Statistical Review* 55, 231–244.

[15] Górecki, T., Horváth, L. and Kokoszka, P. (2018) Change point detection in heteroscedastic time series. *Econometrics and statistics* 7, 63–88.

[16] Hansen, P.R., Lunde, A., and Voev, V. (2014) Realized beta GARCH: A multivariate GARCH model with realized measures of volatility. *Journal of Applied Econometrics* 29, 774–799.

[17] Heyde, C.C. (2008) *Quasi-likelihood and its application: a general approach to optimal parameter estimation.* Springer Science & Business Media.

[18] Herrndorf, N. (1984) A functional central limit theorem for weakly dependent sequences of random variables. *The Annals of Probability* 12, 141–153.

[19] Hinkley, D.V. (1970) Inference about the change-point in a sequence of random variables. *Biometrika* 57, 1–17.

[20] Horváth, L., and Parzen, E. (1994) Limit theorems for Fisher-score change processes. *Lecture Notes-Monograph Series* 157–169.

[21] Horváth, L., and G. Rice (2023) *Change point detection in time series.* To appear.

[22] Jacod, J. and Sørensen, M. (2018) A review of asymptotic theory of estimating functions. *Statistical Inference for Stochastic Processes* 21, 415–434.

[23] Kanai, H., Ogata, H. and Taniguchi, M. (2010) Estimating function approach for CHARN models. *Metron* 68, 1–21.

[24] Kutoyants, Y. A. (2016) On score-functions and goodness-of-fit tests for stochastic processes. *Mathematical Methods of Statistics* 25, 99–120.

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

[26] Li, D.X. and Turtle, H.J. (2000) Semiparametric ARCH models: an estimating function approach. *Journal of Business & Economic Statistics* 18, 174–186.

[27] Negri, I. and Nishiyama, Y. (2017) Z-process method for change point problems with applications to discretely observed diffusion processes. *Statistical Methods & Applications* 26, 231–250.

[28] Newey, W.K. and West, K.D. (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.

[29] Nyblom, J. (1989) Testing for the constancy of parameters over time. *Journal of the American Statistical Association* 84, 223–230.

[30] Page, E.S. (1955) A test for change in parameter occurring at an unknown point. *Biometrika* 41, 100–105.

[31] Pötscher, B.M. and Prucha, I. (1997) *Dynamic nonlinear econometric models: Asymptotic theory.* Springer Science & Business Media.

[32] Shao, X. and Zhang, X. (2010) Testing for change points in time series. *Journal of the American Statistical Association* 105, 1228–1240.

[33] Shorack, G.R., and Wellner, J.A. (2009) *Empirical processes with applications to statistics.* New York: Wiley.

[34] Truong, C., Oudre, L. and Vayatis, N. (2020) Selective review of offline change point detection methods. *Signal Processing* 167, 107299.

[35] van der Vaart, A. W. and Wellner, J. A. (1996) *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer-Verlag, New York.

[36] Zheng, T., Xiao, H. and Chen, R. (2022) Generalized autoregressive moving average models with GARCH errors. *Journal of Time Series Analysis* 43, 125–146.

[37] Zhu, K. (2016) Bootstrapping the portmanteau tests in weak autoregressive moving average models. *Journal of the Royal Statistical Society: Series B* 78, 463–485.

[38] Zhu, K. and Li, W.K. (2015) A bootstrapped spectral test for adequacy in weak ARMA models. *Journal of Econometrics* 187, 113–130.