Sequential monitoring for changes from stationarity to mild non-stationarity

Article

Accepted Version

Horváth, L., Liu, Z., Rice, G. and Wang, S. ORCID: https://orcid.org/0000-0003-2113-5521 (2020) Sequential monitoring for changes from stationarity to mild non-stationarity. Journal of Econometrics, 215 (1). pp. 209-238. ISSN 0304-4076 doi: 10.1016/j.jeconom.2019.08.010 Available at https://centaur.reading.ac.uk/85667/

To link to this article DOI: http://dx.doi.org/10.1016/j.jeconom.2019.08.010

Publisher: Elsevier

Sequential monitoring for changes from stationarity to mild non-stationarity

Lajos Horváth[a], Zhenya Liu[b,c], Gregory Rice[d], Shixuan Wang[e]

[a]*Department of Mathematics, University of Utah, Salt Lake City, UT 841120090, USA*
[b]*China Financial Policy Research Center, School of Finance, Renmin University of China, Beijing, 100872, China*
[c]*CERGAM, Aix-Marseille University, 13090 Aix-en-Provence Cedex 02, France*
[d]*Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, N2L 3G1, Canada*
[e]*Department of Economics, University of Reading, Reading, RG6 6AA, United Kingdom*

Abstract

We develop and study sequential testing procedures á la Chu et al. (1996) for on-line detection of changes in a time series from stationarity to mild forms of non-stationarity. The proposed tests are based on sequential CUSUM and KPSS-type detector processes, and are shown to provide consistent detection under a wide range of change point models, including changes in the parameters of ARMA and GARCH series from values within the model's stationarity parameter region to values close (converging) to the stationarity boundary. Local asymptotic results are established giving precise descriptions of the time to detection under several of these models, which show that such procedures are powerful to detect a wide range of non-stationary characteristics, including changes in mean, volatility, and unit root behaviour. The proposed methods are investigated by means of a simulation study and in applications to monitoring for changes in trend and unit root behaviour in macroeconomic production series, and to detect changes in volatility of the S&P-500 stock market index.

Keywords: change point detection, stationarity testing, normal approximation, non-stationary ARMA time series, non-stationary GARCH time series
JEL: C12, C22, C58

---

*Corresponding author
Email address: zhenya_liu@hotmail.com (Zhenya Liu)

1. Introduction

The literature on modelling and performing both retrospective and sequential detection for change points in econometric and financial time series is extensive. Typically, change points are modelled in terms of parameters (mean, variance, etc.) assumed to describe the underlying data generating process that differ before and after the time of change, and often then the series before and after the change can be thought of as arising from two, perhaps trend, stationary processes with unique distributions. Retrospective as well as sequential methods in change point analysis have been surveyed recently in Aue and Horváth (2013) and Horváth and Rice (2014).

A somewhat less explored class of models in change point analysis that are of potential interest when studying some economic and financial time series are those in which the series is assumed to evolve according to a stationary model with the exception of some short stretches, or "bubbles", that exhibit non-stationary or near non-stationary behaviour. For instance, production time series are often modelled as having stationary residuals around a piecewise linear trend with some potential stretches of unit root or random walk behaviour; see e.g. Nelson and Plosser (1982), Murray and Nelson (2000), Zivot and Andrews (2002). Similar models are also considered in the context of formal bubble detection with asset price data, which are typically modelled as change points from a random walk, or stationary first differenced series, to an explosive autoregressive one process with root that is close to (converging to) unity; see Phillips and Yu (2011), Phillips et al. (2015a,b), and Hafner (2018). Other unit root and stationarity tests that differentiate between $I(0)$ and $I(1)$ series can also be used for the purpose of retrospective bubble detection; see Kim (2000), Busetti and Taylor (2004), and Michaelides et al. (2014). We refer the reader to Gürkaynak (2005), Homm and Breitung (2012), and Linton (2019) for reviews and comparative studies of several of these methods.

In both of the above application arenas, online monitoring of the time series for such non-stationary characteristics is of special interest. Knowledge of such non-stationary behaviour may be used to improve economic forecasts or to inform policy decisions in real time. Based on the framework laid out in the seminal work of Chu et al. (1996) and in the context of financial bubble detection, Homm and Breitung (2012) describe adapting several bubble detection statistics for use in sequential monitoring for bubbles.

The theory surrounding such tests to date is lacking though in a number of areas. For example, for such sequential procedures it has not been quantified how "strong" or "long" a non-stationary bubble must be in order for such sequential procedures to be consistent. Given that many financial bubble models employ parametric models with parameters approaching the stationarity boundary, see for example Phillips and Yu (2011) and Phillips and Magdalinos (2007), such an analysis should evidently consider and quantify the effect of time series models that are close to their stationarity boundary. Additionally, the types of non-stationarities such methods are consistent against has been only lightly studied. Special attention has been paid in the literature to changes from stationartiy to I(1) processes, but little appears to be known regarding how such procedures behave when faced with non-stationarities characterized by, for example, changes in volatility.

In this paper, we develop and study a simple sequential monitoring procedure for detection of non-stationarities in time series. The proposed procedure is based on sequentially comparing a detector process to an upper boundary function, where the detectors are constructed from sequential CUSUM and KPSS (Kwiatkowski et al. (1992)) type statistics. It is shown under the null assumption that the observed series is stationary and weakly dependent that the boundary function can be calibrated in order to asymptotically control the false alarm rate. Further, we show that if non-stationarities modelled as mildly explosive ARMA and GARCH processes begin to form in the series, then the detection procedures are asymptotically consistent, and further local asymptotic results are obtained in these cases giving a precise description on the asymptotic detection time depending on the magnitude and size of the bubble. Simulation studies and data applications to monitor for changes in the trend and random walk behaviour of GDP series as well as to changes in the volatility in the price-to-dividend ratio of the S&P 500 stock market index show that the proposed procedures work well in practice, and are capable of detecting such features even before the economic events that these changes are attributed to occur.

The rest of the paper is organized as follows: In Section 2 we formulate the detection problem in the framework of sequential change point hypothesis testing, and describe several models considered in the sequel. Section 3 contains the definition of the detectors and boundary functions that we consider, along with their asymptotic properties assuming the series does not contain bubbles. We provide asymptotic results on the detectors under several nonstationary models in Section 4. Section 5 contains the results of a simulation study of the proposed detection procedures. Sections 6 and 7 provide presentations of the data applications.

2. Problem formulation and modeling of non-stationary bubbles

Consider a financial or economic time series from which we have observed a "stable" historical sample of length M, $X_1, \dots, X_M$. For instance, X_i might represent the first differenced price-to-dividend ratio of an asset on day i, with the historical sample taken over a time period that is thought to be stationary. As additional values of the series $X_{M+1}, X_{M+2}, \ldots$ are obtained, we are interested in detecting as soon as possible the existence of a point k^* after which a non-stationary "bubble" begins to build up in the series after observation X_{M+k^*}. We will describe more precisely below what we take non-stationarity to mean in this context. We consider here closed-ended procedures in which we stop the detection procedure after observing T observations if no non-stationarities have been detected. In order to make this precise, stability of the historical sample is characterized as follows:

Assumption 2.1. *There exists a sequence of standard Wiener processes* $\{W_{M,1}(t), 0 \le t \le M\}$ *and constants* μ *and* $\sigma > 0$ *such that*

$$\sup_{0 \le t \le M} |S_{M,1}(t) - \mu t - \sigma W_{M,1}(t)| = o_P(M^{1/2}),$$

where $S_{M,1}(t) = \sum_{0 \le s \le \lfloor t \rfloor} X_s$.

We then consider a sequential hypothesis testing problem with the null hypothesis given by

H_0: There exists a sequence of standard Wiener processes $\{W_{M,2}(t), 0 \le t < \infty\}$, independent of $\{W_{M,1}(t), 0 \le t \le M\}$, such that

$$\sup_{1 \le t \le T} |S_{M,2}(t) - \mu t - \sigma W_{M,2}(t)| = o_P(T^{1/2}), \quad for \ \ all \ \ T > 0,$$

where $S_{M,2}(t) = \sum_{M+1 \le s \le \lfloor t \rfloor} X_s$ and $\sigma > 0$ is defined in Assumption 2.1.

H_0 and Assumption 2.1 roughly specify that the combined historical sample and incoming data stream are weakly dependent and generated by the same underlying stochastic process, at least in an asymptotic sense. The functional central limit theorem assumed in H_0 and Assumption 2.1 is satisfied for a wide range of stationary processes, and Billingsley (1968) remains a basic reference for such results. Hall and Heyde (1980) not only establishes the functional central limit theorem for martingales and mixingales using the Skorokhod embedding scheme, but also provides results on the rate of convergence. The monographs of Bradley (2007) and Dedecker et al. (2007) provide introductions to mixing processes and comprehensive surveys. The proofs of the functional central limit theorem and moment inequalities for random variables approximable with sequences of finite dependence is established, for example, in Aue et al. (2014).

We wish to test H_0 against the alternative that we vaguely describe as

H_A: There exists an integer k^* with $1 \le k^* \le T$ such that $X_{M+k^*}, ..., X_{M+k^*+B}$ form a "bubble", where B is the length of the bubble.

Under H_A, the observations $X_{M+k^*}, ..., X_{M+k^*+B}$ are stochastically different from the historical sample. We consider several specific models for H_A below, including change point models with explosive AR(1), mildly nonstationary ARMA and GARCH sequences.

Example 2.1. (Change point in the mean) The most often used model for non-stationarity is the change in the mean model. It is usually assumed that $EX_s = \mu, 1 \le s \le M + k^*$, where μ is an unknown constant and

$$EX_s = \chi_M(s - (M + k^*)), \quad M + k^* + 1 \le s \le M + k^* + B$$

with some non-constant function χ_M. After the end of the non-stationary segment, the process might return to a stationary state. In this paper, we are not interested in detecting the end of the non-stationary segment of length B.

The next examples are inspired by those of Phillips et al. (2015a,b), and Lee and Phillips (2016). Throughout these examples, we assume that $\{\epsilon_s,\ s \in \mathbb{Z}\}$ is an independent and identically distributed innovation sequence with $E\epsilon_s = 0$.

Example 2.2. (Potential change point in the mean with explosive AR(1) errors) Assume that

$$X_s = \begin{cases} \mu + \eta_s, & \text{if } 1 \le s \le k^* + M, \\ \chi_M(s - (M + k^*)) + \eta_s, & \text{if } M + k^* + 1 \le s \le M + k^* + B, \end{cases} \tag{2.1}$$

where

$$\eta_s = \rho\eta_{s-1} + \bar{\epsilon}_s, \ -\infty < s \le M + k^* \text{ with some } \ |\rho| < 1,$$

and

$$\eta_{s+k^*+M} = b_s, \quad 1 \le s \le B,$$

with

$$b_s = \rho_M b_{s-1} + \bar{\epsilon}_{s+k^*+M} = \rho_M b_{s-1} + \epsilon_s, \quad 1 \le s \le B, \tag{2.2}$$

$\epsilon_s = \bar{\epsilon}_{s+k^*+M}$ and $b_0 = \eta_{M+k^*}$. We note that (2.1) allows for changes in the mean as well as in the structure of the errors. Phillips et al. (2015a,b) consider the case

$$\rho_M = 1 - \frac{a}{M}, \quad a \ne 0,$$

as the model for the bubble. In our sequential setting, the error term is a stationary AR(1) process until time $M + k^*$ and then it changes to another AR(1) process with parameter close to the boundary.

Following Aue and Horváth (2006) and Phillips and Magdalinos (2007), one can also consider the mildly explosive case:

Example 2.3. (Potential change point in the mean with mildly explosive AR(1) errors) We use the model in Example 2.2, but take the regression parameter in (2.2) to be given by

$$\rho_M = 1 - \frac{a_M}{M}, \text{ where } a_M \to \infty \text{ and } a_M/M \to 0. \tag{2.3}$$

We note that Phillips (2015a,b) (cf. also Chapter 9 of Linton (2019)) considers the case when (2.3) holds and $a_M \to -\infty$, i.e. ρ_M converges to 1 from above. The models in Examples 2.2 and 2.3 can be easily generalized to ARMA(p, q) sequences.

Example 2.4. (Potential change point in the mean with explosive ARMA(p, q) errors) It is assumed that model (2.1) holds. But instead of assuming that $\eta_t, \quad t \le M + k^*$ is a strictly stationary AR(1) sequence, we only require that $\eta_t, \ \ t \le M + k^*$ is a strictly stationary sequence with zero mean satisfying Assumption 2.1. In (2.2), the AR(1) sequence is replaced with the ARMA(p, q) equation with $\eta_{s+M+k^*} = b_s, 1 \le s \le B$, namely

$$b_s = \sum_{\ell=1}^{p} \beta_\ell b_{s-\ell} + \epsilon_s + \sum_{\ell=1}^{q} \alpha_\ell \epsilon_{s-\ell}, \quad 1 \le s \le B, \tag{2.4}$$

where $b_s = \eta_{s+M+k^*}, -p \le s \le 0$. Now the proximity to the boundary case is measured by how close

$$\rho_M = \beta_1 + \cdots + \beta_p \tag{2.5}$$

is to one.

Since the pioneering work of Engle (1982), non–linear time series are frequently used to model stock prices and returns. The GARCH(1,1) models of Bollerslev (1986) (cf. also Engle and Bollerslev (1986)) and its extensions are widely used in applications, including in economics and finance. For a review on GARCH we refer to Francq and Zakoian (2010) and its applications in finance to Hull (2000). Berkes et al. (2005) investigated “nearly–integrated” GARCH sequences which we use as a possible model for mild non-stationarity in financial data.

Example 2.5. (Potential change point in the mean with mildly explosive GARCH errors) We replace the AR(1) and ARMA(p, q) equations of (2.2) and (2.4) with

$$b_t = \sigma_t \epsilon_t \text{ and } \sigma_t^2 = \omega + \alpha b_{t-1}^2 + \beta \sigma_{t-1}^2, \ 1 \le t \le B, \tag{2.6}$$

where $\sigma_0^2 = \eta_{M+k^*}^2$, $b_t = \eta_{t+k^*}, t = 0, 1, \ldots, B$ and $E\epsilon_t^2 = 1$. It is assumed tht $\alpha = \alpha_M, \beta = \beta_M$ and the closeness to the boundary is measured by $\phi_M = 1 - (\alpha + \beta)$. The mildly explosive case means that $\phi_M \to 0$, as $M \to \infty$. For some theoretical results on mildly explosive GARCH (1,1) processes, we refer to Berkes et al. (2005).

3. Sequential testing procedure and main asymptotic results

Our detection procedure for H_A is built on the basic sequential change point testing framework of Chu et al. (1996). We define a detector $V_M(k)$, computed from the observations $X_1, X_2, \ldots, X_{M+k}$, which is compared to a boundary function $g_M(k)$. Introduce the stopping time

$$\tau_M = \inf\{k : \; V_M(k) \geq g_M(k), \; 1 \leq k \leq T \}.$$

We use the convention that $\inf \emptyset = \infty$, which implies that we set the stopping time to infinity if the detector does not cross the boundary during the observation period of length T. If $\tau_M < \infty$, then the procedure is terminated and we say that we have detected a non-stationary bubble at time $M + \tau_M$. We aim then to choose the boundary such that

$$\lim_{M \to \infty} P\{\tau_M < \infty\} = q \quad \text{under} \;\; H_0, \tag{3.1}$$

and

$$\lim_{M \to \infty} P\{\tau_M < \infty\} = 1 \quad \text{under} \;\; H_A, \tag{3.2}$$

where q is a given tolerance level for falsely detecting a non-existent change, and is selected by the practitioner.

We consider two types of detectors. The first detector is based on sequential CUSUM statistics. Although CUSUM based procedures were developed to detect change points in the mean of time series, we establish below that they also have nontrivial power to detect mildly non-stationary ARMA and GARCH segments. Let

$$Z_M^{(1)}(k) = k \left| \frac{1}{M} S_{M,1}(M) - \frac{1}{k} S_{M,2}(k) \right|, \tag{3.3}$$

i.e. we compare the sample mean of the historical sample sequentially with the sample means of the incoming data stream. The second detector is based on the KPSS statistic of Kwiatkowski et al. (1992) using the modification of Giraitis et al. (2003). Let

$$Z_M^{(2)}(k) = \left| S_{M,2}(k) - \frac{k}{M} S_{M,1}(M) - \frac{1}{k} \left(\sum_{\ell=1}^{k} \left(S_{M,2}(\ell) - \frac{\ell}{M} S_{M,1}(M) \right) \right) \right|. \tag{3.4}$$

For each detector we use a boundary function of the following form

Assumption 3.1.

$$g_M(k) = c(1 + d_0 M^{-\tau}) M^{1/2} \left(1 + \frac{k}{M}\right) f\left(\frac{k}{k+M}\right),$$

where $f(\cdot)$ *is a continuous function on* $[0,1]$, $\min_{z \le u \le 1} f(u) > 0$ *for all* $z > 0$, $\limsup_{u \to 0} u^\gamma / f(u) < \infty$ *with some* $\gamma < 1/2$, $d_0 \ge 0$ *and* $\tau \ge 0$.

The processes $Z_M^{(1)}(k)$ and $Z_M^{(2)}(k)$ are not asymptotically pivotal, since their asymptotic distributions depend on σ of Assumption 2.1. We assume that we can estimate σ from the historical sample with an estimator $\hat{\sigma}_M$ satisfying

Assumption 3.2. $\hat{\sigma}_M \to \sigma$ *in probability.*

Now the detectors are defined using the normalized $Z_M^{(i)}$ processes

$$V_M^{(1)}(k) = \frac{Z_M^{(1)}(k)}{\hat{\sigma}_M} \quad \text{and} \quad V_M^{(2)}(k) = \frac{Z_M^{(2)}(k)}{\hat{\sigma}_M}.$$

In order to derive the asymptotic properties of the proposed detection procedure, we assume that the length of the training sample and termination time of the closed-ended procedure are asymptotically proportional.

Assumption 3.3. *The time to termination of the sequential procedure* $T = T(M)$, *and*

$$\lim_{M \to \infty} \frac{T}{T+M} = \theta.$$

In addition to these assumptions, we also require estimates for the moments of the partial sums of $X_1, X_2, \ldots$

Assumption 3.4. *There is a* $\nu > 2$ *and constant* C *such that for all* $M + 1 \le \ell \le k \le M + T$

$$E\left|\sum_{i=\ell}^{k} (X_i - EX_i)\right|^\nu \le C(k - \ell + 1)^{\nu/2}.$$

We note that Assumption 3.4 is satisfied by a large class of weakly dependent sequences. Below let $W(t), 0 \le t < \infty$ be a Wiener process (standard Brownian motion).

Theorem 3.1. *If Assumptions 2.1–3.4 hold, then*

$$\lim_{M \to \infty} P\{\tau_M < \infty\} = P\left\{\sup_{0 \le u \le \theta} |W(u)|/f(u) \le c\right\}, \tag{3.5}$$

if the detector is given by $V_M^{(1)}(k)$ *and*

$$\lim_{M \to \infty} P\{\tau_M < \infty\} = P\left\{\sup_{0 \le u \le \theta} \frac{1}{f(u)} \left| W(u) - \frac{(1-u)^2}{u} \int_0^u \frac{1}{(1-x)^3} W(x)dx \right| \le c\right\}, \tag{3.6}$$

if the detector is given by $V_M^{(2)}(k)$.

Remark 3.1. Let $f(u) = u^\gamma$ with some $0 \le \gamma < 1/2$. The scale transformation of the Wiener process gives

$$\sup_{0\le u\le\theta} |W(u)|/u^\gamma = \sup_{0\le t\le 1} |W(t\theta)|/(t\theta)^\gamma \stackrel{\mathcal{D}}{=} \theta^{1/2-\gamma} \sup_{0\le t\le 1} |W(u)|/u^\gamma.$$

The distribution of $\sup_{0\le t\le 1}|W(t)|$ is well known and its table can be found, for example, in Shorack and Wellner (1986). Horváth et al. (2004) provide selected critical values for $\sup_{0\le t\le 1}|W(t)|/t^\gamma$ for $\gamma = 0, .15, .25, \ldots, .45$ and $.49$.

Remark 3.2. The proof of (3.6) shows (see (8.11)) that

$$\sup_{0\le u\le\theta} \frac{1}{f(u)} \left| W(u) - \frac{(1-u)^2}{u} \int_0^u \frac{1}{(1-x)^3} W(x)dx \right|$$
$$\stackrel{\mathcal{D}}{=} \sup_{0\le t\le \theta/(1-\theta)} \frac{1}{(1+t)f(t/(1+t))} \left| W_1(t) - tW_2(1) - \frac{1}{t}\int_0^t (W_1(u) - uW_2(1))du \right|,$$

where $W_1(t), 0 \le t < \infty$ is a Wiener process, $W_2(1)$ is a standard normal random variable, $\{W_1(t), 0 \le t < \infty\}$ and $W_2(1)$ are independent.

Remark 3.3. We note that the limits in Theorems 3.1 do not depend on the choices of d_0 nor on τ defining the boundary function because $d_0 M^{-\tau}$ disappears in the limit and it is used to improve the finite sample properties. We use them as tuning parameters to improve finite sample performance. Simulations show that the detector crosses the boundary too often after the first few observations under the null hypothesis. Including $d_0 M^{-\tau}$ in the boundary function, we increase the value of the boundary close to 0. Simulations shows that $d_0 = \hat\rho_w/(1-\hat\rho_w)$ and $\tau = 1/2$ provide good results, where $\hat\rho_w$ is an estimated autoregressive coefficient by fitting an AR(1) model to the historical data using the least squares principle. For the choice $f(u) = u^\gamma$, the number of observations needed to detect the change is a decreasing function of γ so larger γ will give faster detection. The choice of $\gamma = 1/2$ which is related to the square root boundary in classical sequential analysis is not allowed. Clearly, according to the law of the iterated logarithm $\sup_{0<u\le\theta}|W(u)|/u^{1/2} = \infty$ with probability one. Hence the limits in (3.5) and (3.6) are 0 for $\gamma = 1/2$. The rates of convergence in (3.5) and (3.6) are slower for γ close to 1/2.

4. Asymptotics for the time to detection under the alternative

Given the well established literature on sequential detection of change points in the mean, we consider only the change in the structure of the errors towards non-stationarity in this section, i.e. we assume that $\chi_M(u) = \mu$ in Examples 2.2–2.5.

4.1. Mildly explosive AR(1) observations

We assume that (2.2) holds and

Assumption 4.1. *$\{\epsilon_t, 0 \le t < \infty\}$ are independent and identically distributed random variables, independent of $\{X_s, s < k^*\}$, with $E\epsilon_0 = 0$, $0 < E\epsilon_0^2 = \sigma_\epsilon^2 < \infty$ and $E|\epsilon_0|^\kappa < \infty$ with some $\kappa > 2$.*

Theorem 4.1. *Assume that the detector is defined by $V_M^{(1)}(k)$ or $V_M^{(2)}(k)$,*

$$k^* = O(1), \tag{4.1}$$

$$0 < \delta_f = \lim_{u\to 0} f(u)/u^\gamma < \infty \quad \textit{with some } \gamma < 1/2, \tag{4.2}$$

(2.2), (2.3), *and Assumptions 2.1 and 4.1 hold.*
(i) If $P(|b_0| \neq 0) = 1$,

$$B \to \infty, \tag{4.3}$$

and

$$\frac{a_M}{M^{1/2+\gamma}} \to 0 \textit{ as } M \to \infty, \tag{4.4}$$

then we have

$$P(\tau_M \le k^* + 2) \to 1 \textit{ as } M \to \infty. \tag{4.5}$$

(ii) If

$$BM^{(2\gamma-1)/(3-2\gamma)} \to \infty, \textit{ as } M \to \infty, \tag{4.6}$$

and

$$\frac{a_M}{M^{1/2+\gamma}} = O(1), \tag{4.7}$$

then we have

$$\tau_M = O_P(M^{(1-2\gamma)/(3-2\gamma)}). \tag{4.8}$$

(iii) Let

$$A_M = \Big(\frac{a_M}{M^{1/2+\gamma}}\Big)^{1/(1/2-\gamma)}. \tag{4.9}$$

If

$$P\{b_0 = 0\} = 1 \tag{4.10}$$

$$\liminf_{M\to\infty} B/A_M = \infty, \tag{4.11}$$

$$\frac{a_M}{M^{1/2+\gamma}} \to \infty, \tag{4.12}$$

and τ_M is defined by the detector $V_M^{(1)}(k)$, then we have that

$$\lim_{M\to\infty} P\{\tau_M > xA_M\} = P\left\{\sup_{0\le u\le 1} |W(u)|/u^\gamma < cx^{\gamma-1/2}\frac{\delta_f\sigma}{\sigma_\epsilon}\right\}. \tag{4.13}$$

If τ_M is defined by the detector $V_M^{(2)}(k)$, then we have that

$$\lim_{M\to\infty} P\{\tau_M > xA_M\} = P\left\{\sup_{0\le u\le 1} u^{-\gamma}\left|W(u) - \frac{1}{u}\int_0^u W(t)dt\right| < cx^{\gamma-1/2}\frac{\delta_f\sigma}{\sigma_\epsilon}\right\}. \tag{4.14}$$

Conditions (4.3), (4.6), and (4.11) describe how large the bubble must be in order for it to be consistently detected relative to strength of the alternative. The result shows that if a_M is small, i.e. the observations are generated from a process that is closer to a unit root process, we need fewer observations to find the change and the size of the bubble could be smaller. The rate of τ_M in (4.5) suggests that one obtains faster detection asymptotically by taking γ to be close to 1/2. This observation is further studied and confirmed via simulation in Section 5 below.

Since the AR(1) process has been extensively studied in the literature on financial bubble detection when

$$\rho_M = 1 + \frac{a_M}{M}, \quad a_M > 0, \tag{4.15}$$

we consider the behaviour of our testing procedure when ρ_M converges to 1 from above.

Theorem 4.2. *Assume that the detector is defined by $V_M^{(1)}(k)$ and $V_M^{(2)}(k)$, Assumptions 2.1 and 4.1,* (2.2), (4.1), *and* (4.2) *hold, but* (2.3) *is replaced by (4.15).*
(i) If $P(|b_0| \neq 0) = 1$, (4.3) *and* (4.4) *hold, then we have* (4.5).
(ii) If (4.6) *and* (4.7) *hold, then we have* (4.8).
(iii) Assume that (4.10), (4.11) *and* (4.12) *hold,*

(iii-a) If

$$\limsup_{M\to\infty} \frac{a_M^{\frac{3}{2}-\gamma}}{M} < \infty,$$

then (4.13) *and* (4.14) *hold.*

(iii-b) If

$$\limsup_{M\to\infty} \frac{a_M^{\frac{3}{2}-\gamma}}{M} = \infty, \tag{4.16}$$

then

$$\tau_M = O_P\left((M/a_M)\log M\right)$$

To understand the difference between Theorems 4.1 and 4.2, we note that in case of $b_0 = 0$, $var(b_s)$ is proportional M/a_M in case of (2.3), to s if $\rho = 1$ and to $(\exp(sa_M/M) - 1)M/a_M$ if (4.15) holds. This means that the partial sums of the b_s depend on the number of terms in the sum. In case of (4.16) a_M is large, so b_s will be a large random variable even for small s.

We also note that except degenerate cases, $P(b_0 = 0) = 0$ if we start with an stationary AR(1) process during the training sample.

4.2. Mildly explosive ARMA(p, q) observations

We assume that the conditions of Example 2.4 hold. The coefficients $\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_p$ are allowed to depend on M, but such that

$$0 < \liminf_{M\to\infty} |\beta_p| \le \limsup_{M\to\infty} |\beta_p| < \infty. \tag{4.17}$$

The roots of the characteristic polynomial $\varphi(x) = 1 - \beta_1 x - \beta_2 x^2 - \ldots - \beta_p x^p$ are denoted by $r_{1,M}, r_{2,M}, \ldots, r_{p,M}$ such that $|r_{1,M}| \le \cdots \le |r_{p,M}|$, where in this case $|\cdot|$ denotes the complex modulus. First we consider the case when the roots of $\varphi(x)$ are distinct and exactly one of them approaches the boundary of the unit circle. Let

$$r_{1,M} \text{ be real, and } r_{1,M} = 1 + a_M/M, \quad a_M \to \infty \text{ and } a_M/M \to 0, \tag{4.18}$$

$$\lim_{M\to\infty} r_{i,M} = r_i, \ \ |r_i| > 1. \tag{4.19}$$

Under (4.18) and (4.19), it follows that asymptotically $\rho_M = \beta_1 + \cdots + \beta_m \approx 1 - a^*_M/M$ with a_M/a^*_M converging to a positive constant and $a^*_M/M \to 0$ as $M \to \infty$, in analogy with the mildly explosive AR(1) case. In order to have a non–degenerate limit for the normalized partial sums of mildly explosive ARMA(p, q) variables, we must assume in addition that

$$\lim_{M\to\infty} |1 + \cdots + \alpha_q| = \Xi > 0. \tag{4.20}$$

In the statement below we let $\boldsymbol{\eta}_p = (b_p, b_{p-1}, \ldots, b_0)^\top$.

Theorem 4.3. *Assume that the detector is defined by* $V^{(1)}_M(k)$ *or* $V^{(2)}_M(k)$, (2.4), (4.1), (4.2), (4.17)–(4.20), *and Assumptions 2.1 and 4.1 hold.*
(i) If (4.3) *and* (4.4) *are satisfied, and* $P(\|\boldsymbol{\eta}_p\| = 0) = 0$, *then* (4.5) *holds.*
(ii) If (4.6) *and* (4.7) *are satisfied, then* (4.8) *holds.*

(iii) If (4.11), (4.12) *hold and* τ_M *is defined by the detector* $V_M^{(1)}(k)$*, then we have that*

$$\lim_{M\to\infty} P\{\tau_M > xA_M\} = P\left\{\sup_{0\le u\le 1} |W(u)|/u^\gamma < cx^{\gamma-1/2}\frac{\delta_f\sigma}{\Xi^{1/2}\sigma_\epsilon}\right\},$$

where A_M *is defined in* (4.9)*. If* τ_M *is defined using* $V_M^{(2)}(k)$*, then we have that*

$$\lim_{M\to\infty} P\{\tau_M > xA_M\} = P\left\{\sup_{0\le u\le 1} u^{-\gamma}\left|W(u) - \frac{1}{u}\int_0^u W(t)dt\right| < cx^{\gamma-1/2}\frac{\delta_f\sigma}{\Xi^{1/2}\sigma_\epsilon}\right\}.$$

4.3. Mildly explosive AR(2,1) processes with double root

In Section 4.2 the roots of the characteristic equation of the ARMA(p,q) sequence are assumed distinct. We also consider the case when the characteristic polynomial has a double root that is approaching the unit circle. For the sake of simplicity, we assume the bubble is generated from an ARMA(2,1) process to illustrate this case. We replace (2.4) with

$$b_s = 2\beta b_{s-1} - \beta^2 b_{s-2} + \epsilon_s, \quad 1 \le s \le B. \tag{4.21}$$

If $|\beta| < 1$, then b_s converges a.s. to the stationary solution when $s \to \infty$. To study the boundary case we assume that

$$\beta = 1 - \frac{a_M}{M}, \text{ with } a_M \to \infty \text{ and } a_M/M \to 0. \tag{4.22}$$

In this case we are able to obtain the following upper bound on τ_M.

Theorem 4.4. *If the detector is defined by* $V_M^{(1)}(k)$ *or* $V_M^{(2)}(k)$*, Assumptions 2.1, 4.1,* (4.1)–(4.3), (4.20) *and* (4.21) *are satisfied, then there exists a positive sequence* J_M *so that*

$$\tau_M = O_P(J_M),$$

for all J_M *satisfying*

$$B/J_M \to \infty, \tag{4.23}$$

$$J_M\frac{a_M}{M} \to 0 \tag{4.24}$$

and

$$J_M M^{(2\gamma-1)/(5-2\gamma)} \to \infty. \tag{4.25}$$

One may obtain similar results for general ARMA(p, q) sequences for which the characteristic polynomial has roots with multiplicity at least two approaching the boundary of the unit circle. When $\gamma = 0$ in the definition of $f(u)$, the exact rate of τ_M can be obtained.

Remark 4.1. *We assume that Assumptions 2.1, 4.1,* (4.1)–(4.3), (4.20) *and* (4.21) *are satisfied,*

$$B/M^{1/5} \to \infty, \text{ and } a_M M^{-4/5} \to 0. \tag{4.26}$$

Then for all $x > 0$,

$$\lim_{M\to\infty} P(\tau_M > xM^{1/5}) = P\left(\sup_{0\le u\le 1}\left|\int_0^u \int_0^z W(y)dydz\right| < cx^{-5/2}\frac{\delta_f \sigma}{\sigma_\epsilon}\right).$$

4.4. Mildly explosive GARCH (1,1) observations.

In this section we assume that the equations in (2.6) hold and

$$\begin{aligned}&\epsilon_t,\ \ t\in\mathbb{Z} \text{ are independent and identically distributed random variables,} \\ &\text{independent of } \sigma_0^2, E\epsilon_i^2 = 1, \text{ and } E\epsilon_i^4 < \infty,\end{aligned} \tag{4.27}$$

$$E\sigma_0^2 < \infty \text{ and } E\sigma_0^4 < \infty \tag{4.28}$$

and

$$\omega > 0, \alpha \ge 0, \beta \ge 0 \text{ and } \alpha + \beta < 1. \tag{4.29}$$

Assumptions (4.27)–(4.29) are standard in the GARCH literature (see Francq and Zakoian (2010)). We assume that

$$\begin{aligned}&\alpha = \alpha_M, \beta = \beta_M, \alpha + \beta = 1 - \phi_M, \phi_M = \frac{a_M}{M}, \\ &a_M \to \infty \text{ and } a_M/M \to 0, \text{ as } M \to \infty\end{aligned} \tag{4.30}$$

and

$$M^{1/2}\alpha_M/a_M^{1/2} \to 0. \tag{4.31}$$

Assumption (4.31) is taken from Berkes et al. (2005). We note that in case of $\alpha + \beta = 1$ a stationary GARCH(1,1) might exists but will not have a finite second moment. As the GARCH becomes integrated, the variance of the CUSUM starts to increase, and hence the process will cross the boundary simply due to the fluctuations increasing in magnitude.

Theorem 4.5. *We assume that the detector is defined by* $V_M^{(1)}(k)$ *or* $V_M^{(2)}(k)$*, Assumption 2.1,* (2.6) , (4.1), (4.27)–(4.31) *hold.*

(i) If $a_M = O(M^{1/(2(1-\gamma))})$ *and*

$$Ba_M/M \to \infty,$$

then we have

$$\tau_M = O\left(\frac{M}{a_M}\right).$$

(ii) If $M^{1/(2(1-\gamma))}a_M \to \infty$ *and*

$$B\left(\frac{M^{2\gamma}}{a_M}\right)^{1/(1-2\gamma)} \to \infty,$$

then we have that

$$\tau_M = O\left(\left(\frac{a_M}{M^{2\gamma}}\right)^{1/(1-2\gamma)}\right).$$

5. Simulation Study

In this section, we present the results of a simulation study that aimed to assess the finite sample properties of the proposed sequential testing procedure under both H_0 and H_A. We generated data under a number of models, which we further detail below, and for several values of M. In each case we took $f(u) = u^\gamma$, $0 \le \gamma < 1/2$ to define the boundary function, and the length of the termination period $T = M$, so that $\theta = 1/2$ in Assumption 3.3. The sequential procedure was conducted 5000 times with independently generated samples, and the percentage of simulations for which the detector crossed the boundary function are reported for several values of γ. For the sake of brevity, we present here the results when using the detector $V_M^{(1)}$.

With $\theta = 1/2$, the asymptotic critical values $c = c^{(1)}(\gamma, q)$ in case of the detector $V_M^{(1)}(k)$, and $c = c^{(2)}(\gamma, q)$ in case of the detector $V_M^{(2)}(k)$ are defined by

$$P\left\{\sup_{0\le u\le 1/2} |W(u)|/u^\gamma > c^{(1)}(\gamma, q)\right\} = q \tag{5.1}$$

and

$$P\left\{\sup_{0\le u\le 1/2} \frac{1}{u^\gamma}\left|W(u) - \frac{(1-u)^2}{u}\int_0^u \frac{1}{(1-x)^3}W(x)dx\right| > c^{(2)}(\gamma, q)\right\} = q, \tag{5.2}$$

respectively. These are obtained via Monte Carlo simulation. The resulting critical values for several values of γ are displayed in Table 5.1.

Table 5.1: The critical values $c^{(1)}(\gamma, q)$ of (5.1) and $c^{(2)}(\gamma, q)$ of (5.2).

	$c^{(1)}(\gamma, q)$			$c^{(2)}(\gamma, q)$		
γ/q	0.01	0.05	0.10	0.01	0.05	0.10
0	1.96	1.57	1.38	1.17	0.97	0.86
0.15	2.21	1.80	1.59	1.34	1.12	1.00
0.25	2.41	1.99	1.78	1.48	1.24	1.13
0.35	2.68	2.25	2.03	1.67	1.41	1.29
0.45	3.14	2.68	2.46	1.94	1.69	1.57
0.49	3.56	3.05	2.81	2.17	1.90	1.76

In order to estimate σ^2 as defined in Assumptions 2.1 and Assumption 3.2, we employ a kernel lag-window estimator of the form

$$\hat{\sigma}_M^2 = \frac{1}{M}\sum_{\ell=1}^{M}(X_\ell - \hat{X}_M)^2 + 2\sum_{\ell=1}^{M-1} K\left(\frac{\ell}{h}\right)\hat{\gamma}_M(\ell), \quad \text{where } \bar{X}_M = \frac{1}{M}S_{M1}(M) = \frac{1}{M}\sum_{\ell=1}^{M} X_\ell,$$

$\hat{\gamma}_M(\ell)$ is the sample autocovariance at lag ℓ of the series based on the historical sample, K is a kernel function and h is a bandwidth parameter satisfying $h = h(M)$, $h/M + 1/h \to 0$ as $M \to \infty$. Under mild additional regularity conditions, this estimator satisfies Assumption 3.2; see Taniguchi and Kakizawa (2000) and Liu and Wu (2010). We compare below the Bartlett kernel

$$K_B(t) = (1 - |t|)I\{|t| \leq 1\}$$

and the quadratic spectral kernel

$$K_Q(t) = \frac{25}{12\pi^2 x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right) I\{|t| \leq 1\}.$$

Following Andrews (1991), we fit an AR(1) model to the historical data $X_1, \ldots, X_M$ using the least squares principle to produce an estimated autoregressive coefficient $\hat{\rho}_w = \hat{\rho}_w(M)$. The endogenous bandwidths are then defined as

$$h_B = 1.1447\left(\frac{4\hat{\rho}_w^2}{(1-\hat{\rho}_w)^2(1+\hat{\rho}_w)^2}M\right)^{1/3} \quad \text{and} \quad h_Q = 1.3221\left(\frac{4\hat{\rho}_w^2}{(1-\hat{\rho}_w)^4}\right)^{1/5} \tag{5.3}$$

for the Bartlett and quadratic spectral kernel, respectively (see also Müller (2005)). The results in terms of false positive rates were improved when the boundary function is "tuned" to the level of serial dependence in the data, i.e. using

$$g_M(k) = c^{(1)}(\gamma, q)\hat{\sigma}_M \left(1 + \frac{\hat{d}_0}{M^{1/2}}\right)\left(1 + \frac{k}{M}\right)\left(\frac{k}{k+M}\right)^{\gamma} \tag{5.4}$$

with $\hat{d}_0 = \hat{\rho}_w/(1-\hat{\rho}_w)$. Using this tuning parameter the finite sample false positive rates were improved.

5.1. Simulations under H_0

Throughout the presentation below we assume that $\{\epsilon_i,\ i \in \mathbb{Z}\}$ is a standard normal innovation sequence. In terms of linear time series satisfying H_0, we simulated data from an AR(1) process

$$X_s = e_s,\ \ e_s = \rho e_{s-1} + \epsilon_s, 1 \le s \le M+T \ \text{ with } \ \rho = .3, .5, .8, -.3, -.5 \text{ and } -.8.$$

as well as an ARMA(1,1) process

$$X_s = e_s,\ \ e_s = .5e_{s-1} + \epsilon_s + .5\epsilon_{s-1}.$$

We report the false positive rates from 5000 independent simulations with $\rho = 0.8$ in case of AR(1) generated data in Table 5.2, and for the ARMA(1,1) process in Table 5.3; the results for the other linear time series considered tended were somewhat better in terms of their false positive rates.

We observe that in general K_Q gave slightly better results when compared to K_B. In general the asymptotic result in Theorem 3.1 is less predictive in finite samples when γ approaches 1/2. This is expected since Theorem 3.1 does not hold when $f(u) = u^{1/2}$. However for nearly all setting of γ the procedure was fairly well sized for these linear time series examples.

We also considered a nonlinear time series model satisfying H_0, the GARCH(1,1) model:

$$X_i = \sigma_i\epsilon_i, \quad \sigma_i^2 = .25 + .25X_i^2 + .5\sigma_{i-1}^2. \tag{5.5}$$

The results are given in Table 5.4, we observed similarly strong performance. Although these data were not serially correlated, the results were not strongly effected by estimating the long run variance with a kernel based estimator, nor by using the tuned boundary function in (5.4).

Table 5.2: The empirical sizes of $V_M^{(1)}(k)$ for the AR(1) model ($e_i = 0.8e_{i-1} + \epsilon_i$).

		M=100, T=100			**M=500, T=500**			**M=1000, T=1000**		
Kernel	$\gamma \backslash q$	**0.01**	**0.05**	**0.1**	**0.01**	**0.05**	**0.1**	**0.01**	**0.05**	**0.1**
K_B	**0**	0.032	0.082	0.118	0.014	0.050	0.096	0.013	0.049	0.090
	0.15	0.030	0.073	0.115	0.013	0.050	0.085	0.012	0.048	0.087
	0.25	0.035	0.075	0.113	0.012	0.049	0.090	0.011	0.042	0.092
	0.35	0.030	0.073	0.110	0.011	0.046	0.085	0.012	0.043	0.084
	0.45	0.025	0.053	0.077	0.014	0.038	0.060	0.009	0.038	0.066
	0.49	0.017	0.040	0.058	0.007	0.025	0.044	0.005	0.027	0.046
K_Q	**0**	0.022	0.058	0.096	0.008	0.039	0.078	0.009	0.035	0.079
	0.15	0.022	0.055	0.093	0.010	0.035	0.078	0.008	0.039	0.078
	0.25	0.024	0.050	0.091	0.012	0.039	0.072	0.008	0.037	0.076
	0.35	0.020	0.050	0.079	0.009	0.035	0.065	0.008	0.040	0.071
	0.45	0.018	0.039	0.061	0.006	0.026	0.052	0.007	0.027	0.052
	0.49	0.011	0.031	0.040	0.003	0.020	0.032	0.003	0.017	0.032

Table 5.3: The empirical sizes of $V_M^{(1)}(k)$ for the ARMA(1,1) model ($e_i = 0.5e_{i-1} + \epsilon_i + 0.5\epsilon_{i-1}$).

		M=100, T=100			**M=500, T=500**			**M=1000, T=1000**		
Kernel	$\gamma \backslash q$	**0.01**	**0.05**	**0.1**	**0.01**	**0.05**	**0.1**	**0.01**	**0.05**	**0.1**
K_B	**0**	0.024	0.065	0.110	0.012	0.048	0.094	0.012	0.048	0.093
	0.15	0.025	0.067	0.110	0.012	0.049	0.092	0.012	0.049	0.093
	0.25	0.025	0.066	0.107	0.012	0.048	0.091	0.013	0.048	0.092
	0.35	0.023	0.065	0.101	0.011	0.047	0.087	0.011	0.048	0.089
	0.45	0.020	0.052	0.079	0.011	0.039	0.070	0.010	0.040	0.074
	0.49	0.014	0.038	0.059	0.006	0.028	0.051	0.006	0.030	0.056
K_Q	**0**	0.016	0.049	0.087	0.009	0.038	0.077	0.008	0.040	0.081
	0.15	0.016	0.049	0.086	0.009	0.037	0.075	0.008	0.040	0.080
	0.25	0.016	0.048	0.082	0.009	0.036	0.073	0.008	0.038	0.077
	0.35	0.015	0.047	0.077	0.008	0.036	0.070	0.007	0.038	0.073
	0.45	0.013	0.036	0.059	0.007	0.029	0.055	0.006	0.031	0.060
	0.49	0.008	0.025	0.041	0.004	0.019	0.039	0.003	0.021	0.043

Table 5.4: Size of $V_M^{(1)}(k)$ for the GARCH(1,1) case ($e_i = \nu_i \varepsilon_i$ and $\nu_i^2 = 0.25 + 0.25e_{i-1}^2 + 0.5\nu_{i-1}^2$)

		M=100, T=100			**M=500, T=500**			**M=1000, T=1000**		
Kernel	$\gamma \backslash \alpha$	**0.01**	**0.05**	**0.1**	**0.01**	**0.05**	**0.1**	**0.01**	**0.05**	**0.1**
K_B	**0**	0.021	0.073	0.134	0.013	0.050	0.103	0.013	0.055	0.104
	0.15	0.024	0.077	0.143	0.013	0.051	0.106	0.014	0.055	0.106
	0.25	0.027	0.084	0.144	0.014	0.051	0.105	0.013	0.055	0.109
	0.35	0.031	0.091	0.148	0.014	0.058	0.106	0.013	0.063	0.112
	0.45	0.038	0.091	0.138	0.022	0.068	0.111	0.024	0.072	0.118
	0.49	0.029	0.074	0.113	0.023	0.061	0.101	0.025	0.067	0.111
K_Q	**0**	0.021	0.065	0.117	0.012	0.054	0.105	0.009	0.047	0.101
	0.15	0.023	0.070	0.123	0.013	0.055	0.108	0.009	0.049	0.104
	0.25	0.027	0.074	0.129	0.014	0.056	0.109	0.011	0.050	0.103
	0.35	0.027	0.086	0.131	0.015	0.060	0.116	0.011	0.058	0.105
	0.45	0.036	0.087	0.124	0.023	0.066	0.111	0.022	0.065	0.113
	0.49	0.029	0.073	0.108	0.020	0.062	0.097	0.024	0.067	0.108

Table 5.5: The percentages of the crossings $V_M^{(1)}(k)$ and $g_M(k)$ of (5.4) with $q = .05$ under the alternative (5.6) and the long run variance estimator uses K_B.

	γ	**M=50, T=50 k*=0, B=50**	**M=100, T=100 k*=0, B=100**	**M=500, T=500 k*=0, B=500**	**M=50, T=50 k*=25, B=25**	**M=100, T=100 k*=50, B=50**	**M=500, T=500 k*=250, B=250**
$a_M = 5$	**0**	61.13%	87.28%	99.99%	34.89%	64.13%	99.47%
	0.15	62.57%	88.63%	100.00%	33.87%	63.20%	99.44%
	0.25	62.91%	89.12%	99.99%	32.37%	61.79%	99.38%
	0.35	62.78%	89.37%	100.00%	30.22%	59.55%	99.27%
	0.45	58.51%	87.52%	100.00%	24.18%	53.19%	98.85%
	0.49	53.08%	84.44%	100.00%	19.02%	47.14%	98.31%
$a_M = 3$	**0**	72.66%	92.26%	99.98%	46.79%	74.12%	99.66%
	0.15	73.80%	93.08%	99.97%	45.53%	73.43%	99.65%
	0.25	73.93%	93.38%	99.97%	43.91%	72.30%	99.62%
	0.35	73.68%	93.43%	99.97%	41.46%	70.39%	99.56%
	0.45	69.79%	91.99%	99.97%	34.55%	65.00%	99.30%
	0.49	64.99%	89.73%	99.95%	28.32%	59.61%	98.91%
$a_M = 1$	**0**	83.69%	94.55%	99.44%	63.90%	84.91%	99.35%
	0.15	84.13%	94.90%	99.44%	62.64%	84.39%	99.33%
	0.25	84.07%	94.99%	99.43%	61.00%	83.52%	99.30%
	0.35	83.55%	94.87%	99.39%	58.32%	82.26%	99.21%
	0.45	80.77%	93.83%	99.26%	51.08%	78.41%	98.98%
	0.49	77.34%	92.37%	99.14%	44.05%	74.39%	98.68%

Table 5.6: The percentages of the crossings $V_M^{(1)}(k)$ and $g_M(k)$ of (5.4) with $q = .05$ under the alternative (5.6) and the long run variance estimator uses K_Q.

	γ	**M=50, T=50 k*=0, B=50**	**M=100, T=100 k*=0, B=100**	**M=500, T=500 k*=0, B=500**	**M=50, T=50 k*=25, B=25**	**M=100, T=100 k*=50, B=50**	**M=500, T=500 k*=250, B=250**
$a_M = 5$	**0**	57.60%	86.28%	100.00%	31.97%	62.03%	99.37%
	0.15	59.00%	87.76%	100.00%	30.84%	61.13%	99.35%
	0.25	59.21%	88.42%	100.00%	29.34%	59.67%	99.27%
	0.35	58.92%	88.58%	100.00%	27.09%	57.29%	99.17%
	0.45	54.52%	86.57%	100.00%	21.21%	50.73%	98.70%
	0.49	48.79%	83.43%	99.99%	16.20%	44.48%	98.10%
$a_M = 3$	**0**	69.85%	91.02%	99.98%	43.30%	72.74%	99.65%
	0.15	70.95%	91.92%	99.98%	41.94%	71.91%	99.63%
	0.25	71.16%	92.26%	99.97%	40.18%	70.66%	99.59%
	0.35	70.72%	92.35%	99.97%	37.60%	68.63%	99.52%
	0.45	66.54%	90.74%	99.97%	30.64%	62.99%	99.19%
	0.49	61.39%	88.46%	99.95%	24.66%	57.41%	98.86%
$a_M = 1$	**0**	82.12%	94.38%	99.40%	61.27%	84.04%	99.40%
	0.15	82.60%	94.75%	99.40%	59.82%	83.44%	99.38%
	0.25	82.55%	94.77%	99.37%	58.06%	82.56%	99.35%
	0.35	82.04%	94.61%	99.34%	55.23%	81.15%	99.30%
	0.45	79.03%	93.49%	99.23%	47.73%	77.08%	99.10%
	0.49	75.23%	91.99%	99.11%	40.61%	72.86%	98.76%

Table 5.7: The percentages of the crossings of $V_M^{(1)}(k)$ and $g_M(k)$ of (5.4) with $q = .05$ under the alternative (5.6) when $a_M = \log M$.

	γ	**M=50, T=50** **k*=0, B=50**	**M=100, T=100** **k*=0, B=100**	**M=500, T=500** **k*=0, B=500**	**M=50, T=50** **k*=25, B=25**	**M=100, T=100** **k*=50, B=50**	**M=500, T=500** **k*=250, B=250**
K_Q	**0**	63.94%	87.17%	99.99%	37.77%	64.17%	99.27%
	0.15	65.28%	88.55%	100.00%	36.47%	63.24%	99.26%
	0.25	65.49%	89.14%	100.00%	34.76%	61.78%	99.15%
	0.35	65.14%	89.31%	100.00%	32.15%	59.55%	99.02%
	0.45	60.92%	87.49%	100.00%	25.78%	53.08%	98.42%
	0.49	55.58%	84.46%	99.99%	20.24%	46.90%	97.71%
K_B	**0**	67.03%	88.68%	99.99%	40.73%	66.21%	99.32%
	0.15	68.41%	89.86%	100.00%	39.57%	65.30%	99.29%
	0.25	68.75%	90.35%	100.00%	37.94%	63.84%	99.22%
	0.35	68.44%	90.50%	100.00%	35.62%	61.66%	99.08%
	0.45	64.37%	88.77%	100.00%	29.19%	55.46%	98.57%
	0.49	59.23%	86.05%	100.00%	23.39%	49.37%	97.83%

Table 5.8: The percentages of the crossings $V_M^{(1)}(k)$ and $g_M(k)$ of (5.4) with $q = .05$ under the alternative (5.7) and the long run variance estimator uses K_B.

	γ	**M=50, T=50 k*=0, B=50**	**M=100, T=100 k*=0, B=100**	**M=500, T=500 k*=0, B=500**	**M=50, T=50 k*=25, B=25**	**M=100, T=100 k*=50, B=50**	**M=500, T=500 k*=250, B=250**
$\alpha = 0.45$	**0**	85.35%	93.86%	99.91%	62.95%	73.77%	97.74%
	0.15	85.44%	93.90%	99.90%	62.48%	73.24%	97.60%
	0.25	85.34%	93.80%	99.89%	61.87%	72.62%	97.43%
	0.35	85.12%	93.68%	99.89%	60.88%	71.60%	97.17%
	0.45	84.03%	92.98%	99.86%	58.03%	68.82%	96.44%
	0.49	82.61%	92.00%	99.84%	55.36%	66.17%	95.60%
$\alpha = 0.47$	**0**	88.35%	95.75%	99.94%	67.40%	78.53%	98.65%
	0.15	88.46%	95.77%	99.94%	67.00%	78.11%	98.58%
	0.25	88.40%	95.76%	99.94%	66.35%	77.49%	98.49%
	0.35	88.19%	95.66%	99.94%	65.38%	76.63%	98.34%
	0.45	87.24%	95.10%	99.92%	62.76%	74.10%	97.93%
	0.49	85.96%	94.42%	99.90%	60.15%	71.79%	97.44%
$\alpha = 0.49$	**0**	90.83%	96.97%	99.96%	71.85%	82.62%	99.38%
	0.15	90.88%	96.95%	99.97%	71.41%	82.21%	99.35%
	0.25	90.77%	96.90%	99.97%	70.90%	81.69%	99.31%
	0.35	90.60%	96.81%	99.97%	70.08%	80.90%	99.25%
	0.45	89.79%	96.43%	99.96%	67.71%	78.75%	99.02%
	0.49	88.81%	95.95%	99.95%	65.38%	76.71%	98.67%

Table 5.9: The percentages of the crossings $V_M^{(1)}(k)$ and $g_M(k)$ of (5.4) with $q = .05$ under the alternative (5.7) and the long run variance estimator uses K_Q.

	γ	**M=50, T=50 k*=0, B=50**	**M=100, T=100 k*=0, B=100**	**M=500, T=500 k*=0, B=500**	**M=50, T=50 k*=25, B=25**	**M=100, T=100 k*=50, B=50**	**M=500, T=500 k*=250, B=250**
$\alpha = 0.45$	**0**	84.76%	93.23%	99.88%	61.35%	72.53%	97.58%
	0.15	84.82%	93.26%	99.88%	60.89%	72.00%	97.42%
	0.25	84.77%	93.19%	99.87%	60.26%	71.31%	97.23%
	0.35	84.52%	93.04%	99.86%	59.31%	70.28%	96.95%
	0.45	83.28%	92.23%	99.84%	56.56%	67.55%	96.22%
	0.49	81.74%	91.12%	99.82%	53.95%	64.77%	95.34%
$\alpha = 0.47$	**0**	87.79%	95.30%	99.96%	65.81%	77.42%	98.71%
	0.15	87.80%	95.32%	99.96%	65.39%	76.94%	98.64%
	0.25	87.71%	95.21%	99.96%	64.78%	76.26%	98.51%
	0.35	87.48%	95.05%	99.95%	63.85%	75.27%	98.34%
	0.45	86.47%	94.43%	99.94%	61.25%	72.90%	97.90%
	0.49	85.28%	93.70%	99.93%	58.73%	70.31%	97.36%
$\alpha = 0.49$	**0**	90.22%	96.67%	99.98%	70.43%	81.38%	99.33%
	0.15	90.27%	96.70%	99.97%	69.89%	80.89%	99.28%
	0.25	90.21%	96.63%	99.97%	69.24%	80.39%	99.22%
	0.35	89.99%	96.50%	99.97%	68.36%	79.64%	99.14%
	0.45	89.12%	95.95%	99.97%	65.84%	77.38%	98.88%
	0.49	88.08%	95.37%	99.97%	63.47%	75.15%	98.60%

5.2. Simulations under H_A

We also conducted some simulations to study the power of the proposed sequential bubble detection scheme. We initially consider data generated as described in Example 2.3. We assume that the mean is constant in both the historical sample and testing period, so we can assume without loss of generality that it is 0. The data are then generated to follow a stationary AR(1) process with regression parameter 0.5. The innovations ϵ_s are again independent standard normal random variables. Hence

$$X_s = \begin{cases} .5X_{s-1} + \epsilon_s, & \text{if } 1 \leq s \leq M + k^* \\ \rho_M X_{s-1} + \epsilon_s, & \text{if } M + k^* + 1 \leq s \leq M + k^* + B, \end{cases} \tag{5.6}$$

where ρ_M satisfies (2.3). We used the boundary function g_M of (5.4). The results are reported in Tables 5.5–5.7. As it is expected, the power is increasing when M is increasing or a_M is decreasing (getting closer to the stationarity boundary). The sequential detection works well when k^* is small, i.e. the change happens immediately of the beginning of the sample.
Figures 5.1 and 5.2 show the densities of the first exceedance time k of the boundary function $g_M(k)$ by $V_M^{(1)}(k)$. In Figure 5.1, we observe that larger γ gives in general faster detection, while the opposite relationship holds in Figure 5.2. In general the size of the procedure is improved by taking smaller values of γ. As a compromise, we use $\gamma = 0.35$ in the empirical study.

Figure 5.1: Density estimates of the detection time using $V_M^{(1)}(k)$ and $g_M(k)$ of (5.4) with $q = 0.05$ under the alternative (5.6) when $k^* = 0, M = 100, a_M = 5$ and the long run variance estimator uses K_Q.

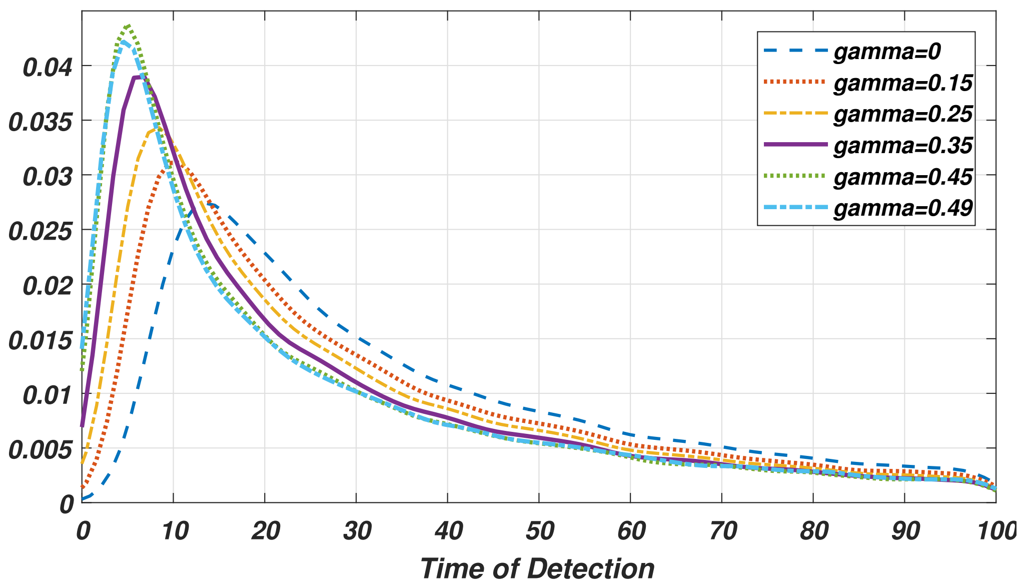
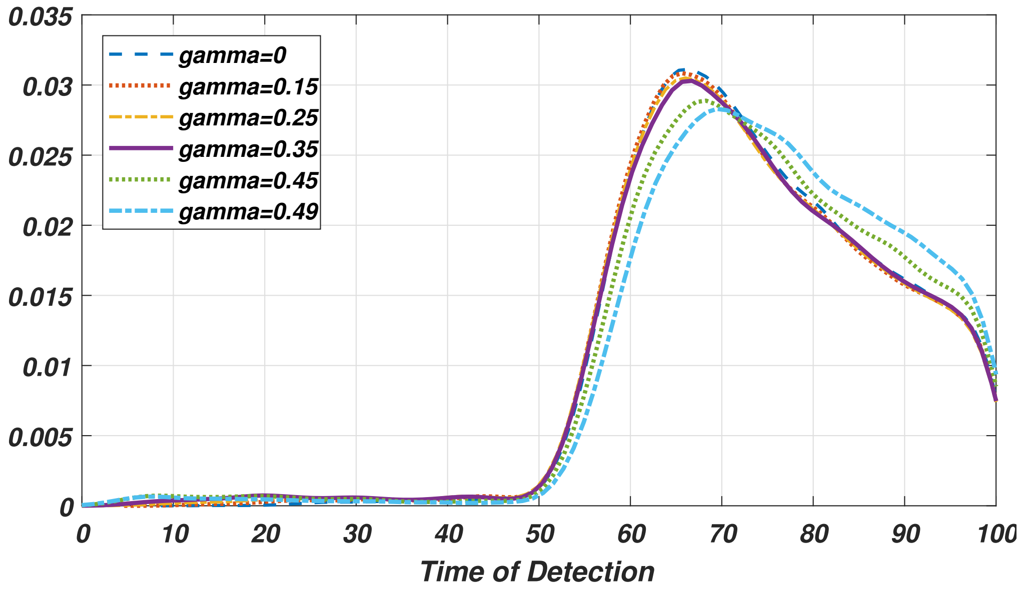


Figure 5.2: The densities of the times of crossings of $V_M^{(1)}(k)$ and $g_M(k)$ of (5.4) with $q = .05$ under the alternative (5.6) when $k^* = 50, M = 100, a_M = 5$ and the long run variance estimator uses K_Q.



We proceed to consider the data generated by GARCH(1,1) in Example 2.5.

$$\begin{cases} X_s = \sigma_i \epsilon_i, \quad \sigma_i^2 = .25 + .25 X_i^2 + .5\sigma_{i-1}^2, & \text{if } 1 \le s \le M + k^* \\ X_s = \sigma_i \epsilon_i, \quad \sigma_i^2 = .25 + \alpha X_i^2 + .5\sigma_{i-1}^2, & \text{if } M + k^* + 1 \le s \le M + k^* + B, \end{cases} \tag{5.7}$$

where $\alpha + 0.5$ is close to one. The results are reported in Tables 5.8–5.9. We observe a similar pattern as the result of the AR(1).

6. Application 1: monitoring for changes in real US GDP

Over the years, there has been intensive debate on whether the U.S. real GDP and other similar macro-economic production series follow a trend stationary process or a stochastic trend process. On the one hand, many empirical studies have found evidence in favor of a deterministic trend, especially when allowing for some structural breaks in the trend line (see Perron (1989); Diebold and Senhadji (1996); Cheung and Chinn (1997)). On the other hand, evidence has also been found to support the stochastic trend hypothesis as well (e.g. Nelson and Plosser (1982); Murray and Nelson (2000)). Thus, we take the position that the real GDP series can possibly change from a trend stationary process to alternative processes. We demonstrate how to use the proposed sequential procedure to monitor for non-stationarities in the real GDP series, which might be indicative of changes in the trend line or changes to a near unit root process.

We downloaded the quarterly US seasonally adjusted real GDP data from the Federal Reserve Bank of St. Louis[1], and considered monitoring for changes near two recent recession periods[2]. Two preprocessing steps were conducted before applying the tests. First, we consider the log of the real GDP, which is conventional. Second, in order to remove the deterministic trend, we estimate the trend line only within the training period and then extend the trend to the test period. After this, we are able to input the detrended series into our monitoring procedure and check whether the detector goes beyond the boundary at any point during the testing period. The monitoring procedure was carried out using the detector $V_M^{(1)}$ and the quadratic spectral kernel K_Q to estimate the long run variance parameter. The boundary function was determined by (5.4) with $q = 0.05$, $\gamma = 0.35$, and the critical value is calculated using the scale transformation in Remark 3.1.

Table 6.1 shows the specific information on the selected periods, along with the KPSS statistics and estimated autoregressive coefficients on the detrended series. The KPSS test suggests that the detrended series in the training periods are reasonably stationary, which meets with Assumption 2.1. Additionally, the estimated autoregressive parameters in both test periods are close to one. The results of the application to each period are detailed in the subsections below.

6.1. Period I: Q1 1985 to Q4 2001

The training period is set to be from Q1 1985 to Q4 1994, containing ten years worth of data (40 quarterly observations). We choose to monitor the detector from Q1 1995 until Q4 2001. The choice of the initial date in the test period is due to our attempt to find any early warning signal in the economy before the recession in 2001. The upper panel and middle panel in Figure 6.1 show the log series and detrended series of the U.S. real GDP, respectively. As can be observed, the detrended series were fluctuating around zero in the training period, but started to climb up after 1996 and reached a peak in 2000. The lower panel in Figure 6.1 presents the trajectory of the detector $V_{40}^{(1)}$ against the boundary function $g_{40}(k)$. The detector goes beyond the boundary in Q3 1999 for the first time. Given that the procedure used is powerful against changes in the

[1] `https://fred.stlouisfed.org/series/GDPC1`

[2] According the National Bureau of Economic Research, the recent two recessions are during *March 2001 to November 2001* and during *December 2007 and June 2009.*

Table 6.1: Description of detrended U.S. real GDP data in each period considered, including the time spans associated with each training and testing sets, the values of KPSS test statistics, and estimated AR(1) coefficients.

		Time Span	KPSS Stat	$\hat{\rho}$
Period *I*	**Training**	Q1 1985 – Q2 1994	0.166	0.901 (0.062)
	Test	Q1 1995 – Q4 2001	0.908***	0.944 (0.041)
Period *II*	**Training**	Q1 2002 – Q4 2006	0.095	0.632 (0.188)
	Test	Q1 2007 – Q4 2011	0.703**	0.946 (0.042)

Note: the setting of the KPSS test used does not include a linear trend component, and the bandwidth is set to $\lfloor 4(N/100)^{1/4} \rfloor$. The numbers in parentheses are the standard error of the estimation. KPSS critical values are 0.347 (10% level), 0.463 (5% level), 0.739 (1% level) in this setting. *** and ** indicate values significant to the 1% and 5% significance level of the asymptotic distribution.

mean as well as changes to stochastic trends, this result suggests that either the trend changed during the monitoring period, or the series begins to follow an approximate unit root process.

6.2. Period II: Q1 2002 to Q4 2011

The initial date in the training period is set to be Q1 2002, which is right after the end of Period *I*. The most recent recession due to subprime mortgage crisis started at the end of 2007. For the purpose of attempting to find possible early warning signs, we start to monitor for changes starting from Q1 2007. There are 20 quarterly observations in the training period between Q1 2002 and Q4 2006, and 20 observations in the test period between Q1 2007 and Q4 2011. The log U.S. real GDP with trend and detrended are shown in the upper panel and middle panel of Figure 6.2. The detrended series is fluctuating around zero, but it started to continuously decline from Q1 2007. The lower panel of Figure 6.2 plots the trajectory of the detector $V_{20}^{(1)}$ against the boundary function $g_{20}(k)$. The detector goes beyond the boundary in Q4 2007 for the first time.

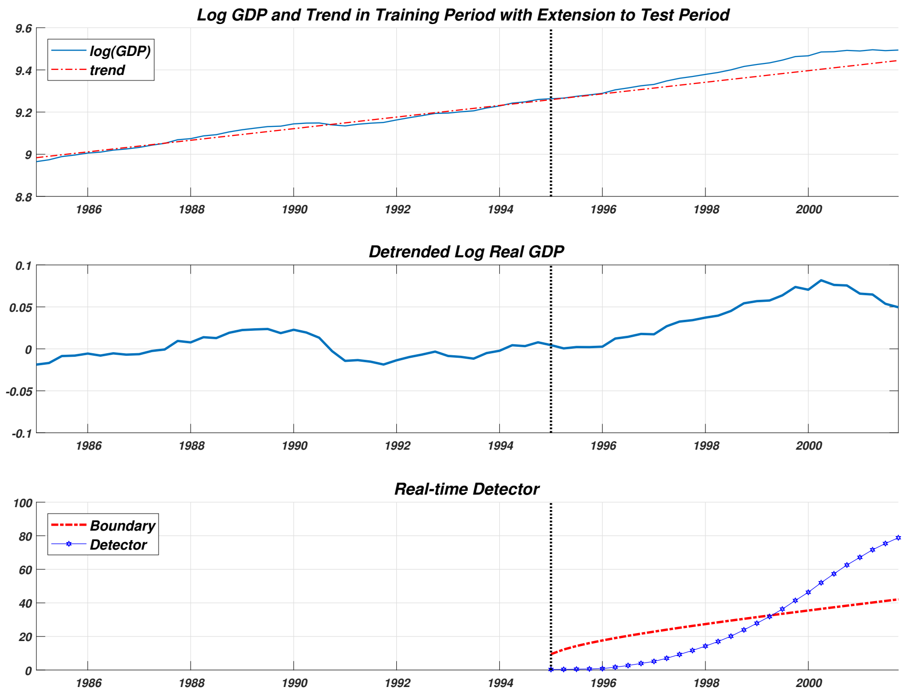


Figure 6.1: Upper panel: log of U.S. real GDP during Q1 1985 to Q4 2001 and the trend in the training period with extension to the test period. Middle Panel: detrended log of U.S. real GDP. Lower Panel: real-time Detector $V_{40}^{(1)}$ versus the boundary function $g_{40}(k)$ in the test period. The vertical dash line indicates the division between the training and test samples.

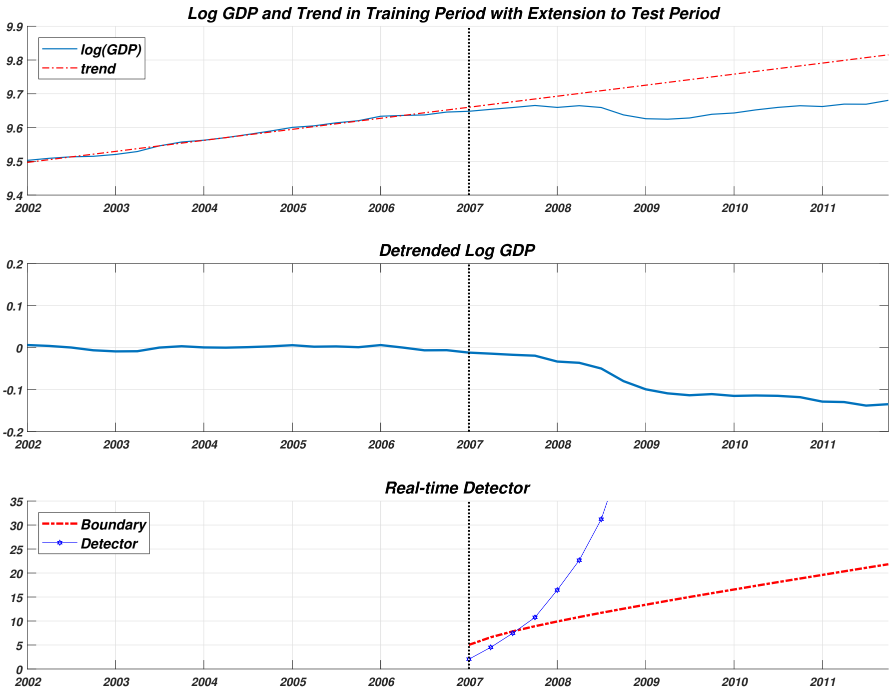


Figure 6.2: Upper panel: log of U.S. real GDP during Q1 2002 to Q4 2011 and the trend in the training period with extension to the test period. Middle Panel: detrended log of U.S. real GDP. Lower Panel: real-time Detector $V_{20}^{(1)}$ versus the boundary function $g_{20}(k)$ in the test period. The vertical dash line indicates the division between the training and test samples.

7. Application 2: monitoring for changes in volatility in the S&P 500

Time series of price–to–fundamentals ratios are commonly used to test for the presence of stock market bubbles. Following Phillips et al. (2011, 2015a), we consider monitoring the price–to–dividend (P/D) ratio of the S&P 500 stock market index. Namely, with P_t denoting the S&P 500 stock price index, and D_t denoting the real S&P 500 stock price index dividend, we monitor the series $P/D_t := P_t/D_t$. The monthly dividend data are computed from the S&P 500 four-quarter totals for the quarter since 1926, with linear interpolation to obtain monthly figures. The specific data we used was downloaded from Professor Robert Shiller's website[3], which is further described in Chapter 26 of Shiller (1992). Since we require that the data are approximately stationary during the training sample, we study the first differenced series $dP/D_t = P/D_t - P/D_{t-1}$. This analysis is then aiming to detect changes in the volatility of the series.

We considered monitoring for changes in volatility during two historical periods: The dot-com bubble, with data spanning from January 1988 to December 1999, and with recent data, spanning from January 2011 to March 2019. We refer to these respective periods as Period 1 and Period 2 below. The series dP/D_t from each of these periods were divided into training and testing sets, and the validity of Assumption 2.1 for each training sample was evaluated by an application of the KPSS test. This information along with some further summary information from the samples relevant for the proposed procedures is contained in Table 7.1. In each case the training samples were found to be reasonably stationary. The estimated GARCH(1,1) parameters suggest that the test sample in Period 1 is non-stationary and the test sample in Period 2 is stationary. Thus, we expect that the detector will go across the boundary in Period 1 but not in Period 2. The monitoring procedure was carried out using the detector $V_M^{(1)}$ and the quadratic spectral kernel K_Q to estimate the long run variance parameter. The boundary function was determined by (5.4) with $q = 0.05$, $\gamma = 0.35$, and the critical value is calculated using the scale transformation in Remark 3.1. The results of the application to each period are detailed in the subsections below.

7.1. Period 1: Jan 1988 to Dec 1999

The training period was set to be from January 1988 to December 1994, constituting 84 observa-

[3] http://www.econ.yale.edu/~shiller/data.htm

Table 7.1: Description of data in each period considered, including the time spans associated with each training and testing sets, the values of KPSS test statistics, and estimated GARCH(1,1) coefficients using Quasi-maximum likelihood estimation.

		Time Span	KPSS Stat	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}+\hat{\beta}$
Period 1	**Training**	Jan 1988 – Dec 1994	0.068	0.053 (0.092)	0.040 (0.046)	0.873 (0.172)	0.913
	Test	Jan 1995 – Dec 1999	0.053	0.185 (0.314)	0.196 (0.146)	0.804 (0.145)	1.000
Period 2	**Training**	Jan 2011 – Dec 2015	0.097	0.642 (0.632)	0.303 (0.254)	0.424 (0.417)	0.727
	Test	Jan 2016 – Mar 2019	0.112	0.402 (0.575)	0.311 (0.311)	0.485 (0.551)	0.796

Note: the setting of the KPSS test used does not include a linear trend component, and the bandwidth is set to $\lfloor 4(N/100)^{1/4} \rfloor$. The numbers in parentheses are the standard error of the estimation. KPSS critical values are 0.347 (10% level), 0.463 (5% level), 0.739 (1% level) in this setting. *** and ** indicate values significant to the 1% and 5% significance level of the asymptotic distribution.

tions. We chose the training period in order to avoid the Black Monday market crash in October 1987. We then took the test period to be from January 1995 to December 1999, which contains the formation of the infamous dot-com bubble. The upper panel of Figure 7.1 shows a plots of the series dP/D_t during this period. It is observed that the dP/D_t is stable during the training period. The dP/D_t begins to become volatile from January 1997, with a clear growing volatility till the end of 1999.

The lower panel of Figure 7.1 shows the trajectory of the detector $V_{84}^{(1)}$ against the boundary function $g_{84}(k)$. The detector exceeds the boundary in January 1997 for the first time, indicating the presence of a change on that day. This date was approximately three years before the traditionally accepted date of the bursting of the dot-com bubble early in the year 2000. This appears to be similar to Example 2.5 modelling a GARCH(1,1) that changes from stationarity to non-stationarity.

7.2. Period 2: Jan 2011 to Mar 2019

The upper panel of Figure 7.2 plots the time series trajectories of the fist order difference of

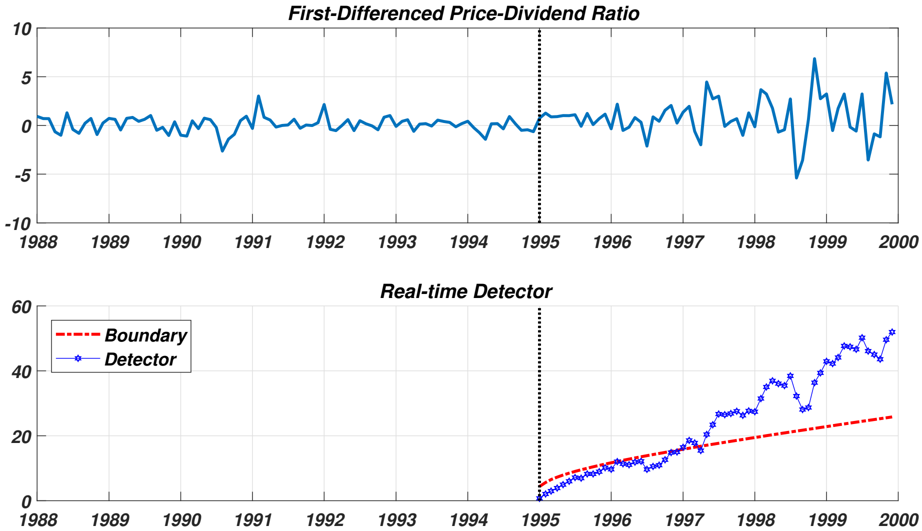


Figure 7.1: Upper panel: Price-dividend ratio of the S&P 500 during 1995 Jan to 1999 Dec. Lower Panel: Real-time Detector $V_{84}^{(1)}$ versus the boundary function $g_{84}(k)$ in the test period. The vertical dash line indicates the division between the training and test samples.

the P/D ratio during the second example period. The training period is set to be from January 2011 to December 2015, containing 60 observations. The choice of the initial date of the training sample is meant to avoid the subprime mortgage crisis in 2008. Except for the outlier at August 2011, the dP/D_t is relatively stable in the training sample. The test period we took to be from January 2016 to March 2019. The dP/D_t seems to have the same level of volatility as the training period. As in the previous example, we present the trajectory of the detector $V_{60}^{(1)}$ and the boundary function $g_{60}(k)$ in the lower panel of Figure 7.2. The detector never crossed the boundary function during the testing period, indicating that no change would have been detected during the testing period. This result can be interpreted to mean that the increasing trend visible in the P/D ratio in the testing sample is consistent, at level 5% with the fluctuations observed in the P/D ratio in the test sample.

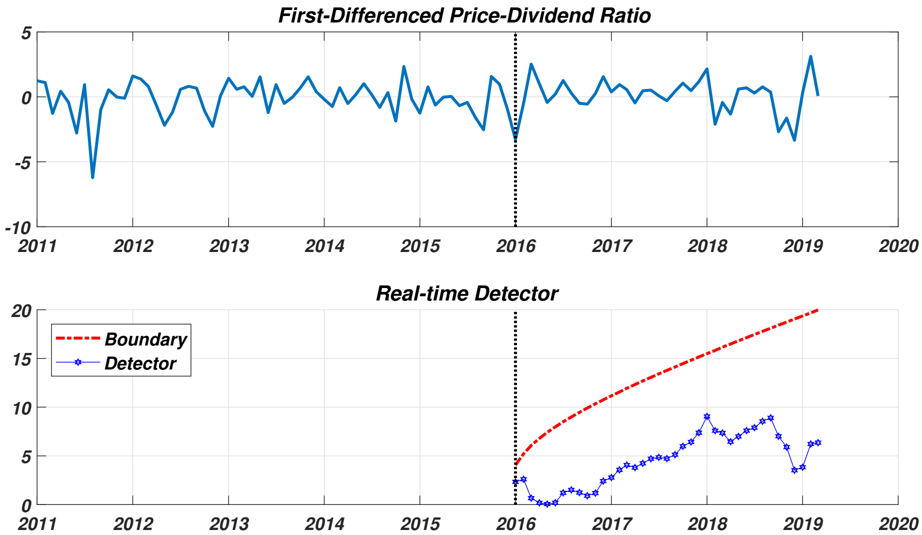


Figure 7.2: Upper panel: Price-dividend ratio of the S&P 500 during 2011 Jan to 2019 Mar. Lower Panel: Real-time Detector $V_{60}^{(1)}$ versus the boundary function $g_{60}(k)$ in the test period. The vertical dash line indicates the division between the training and test samples.

8. Proofs of Technical Results

We start with a simple lemma.

Lemma 8.1. *If Assumption 3.4 holds, then for all $0 \le \gamma < 1/2$ and $x > 0$ we have that*

$$\lim_{a\to 0}\limsup_{M\to\infty} P\left\{ M^{\gamma-1/2} \max_{1\le k\le aM} k^{-\gamma}\left|\sum_{\ell=1}^{k}(X_\ell - EX_\ell)\right| > x\right\} = 0.$$

Proof. Let ln denote the logarithm function to base 2 and $e_i = X_i - EX_i$. It is easy to see that

$$\begin{aligned}\max_{1\le k\le aM} k^{-\gamma}\left|\sum_{\ell=1}^{k} e_\ell\right| &\le \max_{1\le i\le \ln(aM)+1}\ \max_{2^{i-1}<k\le 2^i} k^{-\gamma}\left|\sum_{\ell=1}^{k} e_\ell\right| \\ &\le \max_{1\le i\le \ln(aM)+1}\ \max_{2^{i-1}<k\le 2^i} 2^{-(i-1)\gamma}\left|\sum_{\ell=1}^{k} e_\ell\right|.\end{aligned}$$

Assumption 3.4 and Móricz et al. (1982) yield that

$$E\left(\max_{1\le k\le 2^i}\left|\sum_{\ell=1}^{k} e_\ell\right|\right) \le C_1 2^{i\nu/2}$$

with some constant C_1. Hence by Markov's inequality we have

$$\begin{aligned}P\left\{M^{\gamma-1/2}\max_{1\le k\le aM} k^{-\gamma}\left|\sum_{\ell=1}^{k} e_\ell\right| > x\right\} &\le \sum_{i=1}^{\ln(aM)+1} P\left\{\max_{2^{i-1}<k\le 2^i} k^{-\gamma}\left|\sum_{\ell=1}^{k} e_\ell\right| > xM^{1/2-\gamma}\right\} \\ &\le \sum_{i=1}^{\ln(aM)+1} P\left\{\max_{1\le k\le 2^i}\left|\sum_{\ell=1}^{k} e_\ell\right| > x2^{(i-1)\gamma}M^{1/2-\gamma}\right\} \\ &\le C_1 x^{-\nu} M^{(\gamma-1/2)\nu}\sum_{i=1}^{\ln(aM)+1} 2^{-(i-1)\gamma\nu}2^{i\nu/2} \\ &\le C_1\frac{2^{1/2}}{2^{(-\gamma+1/2)\nu}} a^{(1/2-\gamma)\nu},\end{aligned}$$

completing the proof. □

Proof of Theorem 3.1. It follows from Assumptions 2.1 and H_0 that for all $a > 0$ that

$$\max_{aM\le k\le T}\frac{Z_M^{(1)}(k)}{M^{1/2}(1+k/M)f(k/(k+M))} \xrightarrow{\mathcal{D}} \sup_{a\le t\le \theta/(1-\theta)}\frac{\sigma|W_2(t)-tW_1(1)|}{(1+t)f(t/(1-t))}, \tag{8.1}$$

where W_1, and W_2 are independent Wiener processes. Using Assumption 3.1 and the law of the iterated logarithm for W_2 we get that

$$\lim_{a\to 0}\max_{0<t\le a}\frac{|W_2(t)|}{(1+t)f(t/(1-t))} = 0 \quad \text{a.s.} \tag{8.2}$$

and

$$\lim_{a\to 0}\max_{0<t\le a}\frac{|tW_1(t)|}{(1+t)f(t/(1-t))}=0 \quad \text{a.s.} \tag{8.3}$$

Lemma 8.1 yields that

$$\lim_{a\to 0}\limsup_{M\to\infty}P\left\{\max_{1\le k\le aM}\frac{|S_{M,2}(k)|}{M^{1/2}(1+k/M)f(k/(k+M))}>x\right\}=0 \quad \text{for all } \ x \tag{8.4}$$

and

$$\lim_{a\to 0}\limsup_{M\to\infty}P\left\{\max_{1\le k\le aM}\frac{(k/M)|S_{M,1}(k)|}{M^{1/2}(1+k/M)f(k/(k+M))}>x\right\}=0 \quad \text{for all } \ x. \tag{8.5}$$

Since we can choose a as small as we wish in (8.1), we obtain from (8.2)–(8.5) that

$$\max_{1\le k\le T}\frac{Z_M^{(1)}(k)}{M^{1/2}(1+k/M)f(k/(k+M))}\xrightarrow{\mathfrak{D}}\sup_{0\le t\le\theta/(1-\theta)}\frac{\sigma|W_2(t)-tW_1(1)|}{(1+t)f(t/(1+t))}. \tag{8.6}$$

It is well known that

$$\{W_2(t)-tW_1(1),0\le t<\infty\}\stackrel{\mathfrak{D}}{=}\{(1+t)W(t/(1+t)),0\le t<\infty\}, \tag{8.7}$$

where W denotes a Wiener process. Hence

$$\sup_{0\le t\le\theta/(1-\theta)}\frac{|W_2(t)-tW_1(1)|}{(1+t)f(t/(1+t))}\stackrel{\mathfrak{D}}{=}\sup_{0\le t\le\theta}\frac{|W(t)|}{f(t)},$$

and therefore (3.5) follows from (8.6) and Assumption 3.3.

Since

$$\begin{aligned}\max_{1\le k\le aM}k^{-\gamma}\left|\frac{1}{k}\sum_{\ell=1}^{k}\left(S_{M,2}(\ell)-\frac{\ell}{M}S_{M,1}(M)\right)\right| &\le \max_{1\le k\le aM}k^{-\gamma}\max_{1\le \ell\le k}\left|S_{M,2}(\ell)-\frac{\ell}{M}S_{M,1}(M)\right|\\ &\le \max_{1\le k\le aM}\max_{1\le \ell\le k}\ell^{-\gamma}\left|S_{M,2}(\ell)-\frac{\ell}{M}S_{M,1}(M)\right|\\ &\le \max_{1\le \ell\le aM}\ell^{-\gamma}\left|S_{M,2}(\ell)-\frac{\ell}{M}S_{M,1}(M)\right|,\end{aligned}$$

by (8.4) and (8.5) we have

$$\lim_{a\to 0}\limsup_{M\to\infty}P\left\{\max_{1\le k\le aM}\frac{Z_M^{(2)}(k)}{M^{1/2}(1+k/M)f(k/(k+M))}>x\right\}=0 \quad \text{for all } \ x. \tag{8.8}$$

Using again Assumptions 2.1 and H_0 we get that □

$$\begin{aligned}&\max_{aM\le k\le T}\frac{Z_M^{(2)}(k)}{M^{1/2}(1+k/M)f(k/(k+M))}\\ &\qquad\xrightarrow{\mathfrak{D}}\sup_{a\le t\le\theta/(1-\theta)}\frac{\sigma|W_2(t)-tW_1(1)-(1/t)\int_0^t(W_2(u)-uW_1)du|}{(1+t)f(t/(1+t))},\end{aligned} \tag{8.9}$$

where W_1 and W_2 are independent Wiener processes. It follows from (8.2) and (8.3) that

$$\lim_{a\to 0}\sup_{0<t\le a}\frac{|W_2(t)-tW_1(1)-(1/t)\int_0^t(W_2(u)-uW_1)du|}{(1+t)f(t/(1+t))}=0 \quad \text{a.s.} \tag{8.10}$$

Putting together (8.8)–(8.10) we conclude

$$\begin{aligned}\max_{1\le k\le T}&\frac{Z_M^{(2)}(k)}{M^{1/2}(1+k/M)f(k/(k+M))} \\ &\xrightarrow{\mathcal{D}} \sup_{0\le t\le \theta/(1-\theta)}\frac{\sigma|W_2(t)-tW_1(1)-(1/t)\int_0^t(W_2(u)-uW_1)du|}{(1+t)f(t/(1+t))},\end{aligned} \tag{8.11}$$

and therefore (3.6) follows from Assumption 3.2 and (8.7). □

8.1. Proof of Theorems 4.1–4.5

It is assumed in Section 4 that X_i is constant and therefore we can assume without loss of generality that $EX_i = 0$.

We start with the proof of Theorem 4.1. The recursion in (2.2) can be solved explicitly and we get

$$b_t=\sum_{\ell=1}^{t-1}\rho_M^{\ell}\epsilon_{t-\ell}+\rho_M^t b_0, \quad 1\le t\le B,$$

where $b_0=\eta_{k^*}$. First we establish an upper bound for b_t.

Lemma 8.2. *If* (2.2), (2.3) *and Assumption 4.1 hold, then we have that*

$$\max_{1\le t\le B} t^{-1/\kappa}|b_t| = O\left(M/a_M\right) \quad \textit{a.s.}$$

Proof. It is well known that Assumption 4.1 implies that

$$|\epsilon_t| = o\left(t^{1/\kappa}\right) \quad \text{a.s.}$$

(cf. Chow and Teicher (1988)) and therefore

$$\max_{1\le t\le B} t^{-1/\kappa}|b_t| \le \max_{1\le t\le B} t^{-1/\kappa}|\epsilon_t|\sum_{\ell=1}^{\infty}\left(1-\frac{a_M}{M}\right)^{\ell}.$$

According to (2.3)

$$\sum_{\ell=1}^{\infty}\left(1-\frac{a_M}{M}\right)^{\ell} = O\left(M/a_M\right),$$

the lemma is proven since

$$\max_{1\le t\le B}|\rho_M^t b_0| = O(1) \quad \text{a.s.}$$

□

Lemma 8.3. *If* (2.2), (2.3) *and Assumption 4.1 hold, then one can define a Wiener process* W *such that*

$$\max_{1\le t\le B} t^{-1/\kappa}\left|(1-\rho_M)\sum_{s=1}^t b_s - (E\epsilon_0^2)^{1/2}W(t)\right| = O(1) \quad a.s.$$

Proof. Computing the sum of both sides of equation (2.2) we obtain that

$$\sum_{s=1}^t b_s = \rho_M\sum_{s=1}^t b_s + \sum_{s=1}^t \epsilon_s,$$

and therefore

$$(1-\rho_M)\sum_{s=1}^t b_s = \sum_{s=1}^t \epsilon_s + b_0 - b_t. \tag{8.12}$$

By the Komlós–Major–Tusnády approximation (cf. Csörgő and Révész (1981)) one can define a Wiener process $W(t), t\ge 0$ such that

$$\left|\sum_{s=1}^t \epsilon_s - (E\epsilon_0^2)^{1/2}W(t)\right| = o(t^{1/\kappa}) \quad \text{a.s.} \tag{8.13}$$

Now the result follows from Lemma 8.3, (8.12) and (8.13). □

Proof of Theorem 4.1. We consider the detector $V_M^{(1)}(k)$ only, the behaviour the detector $V_M^{(2)}(k)$ can be analysed similarly. In order to establish parts (i) and (ii), let

$$J_M = \begin{cases} k^*+2, & \text{in case of part (i)},\\ JM^{(1-2\gamma)/(3-2\gamma)} & \text{in case of part (ii)},\end{cases}$$

for some constant $J>0$. We note that

$$P\{\tau_M > J_M\} = P\left\{\max_{1\le k\le J_M} V_M^{(1)}(k)/(M^{1/2}(1+k/M)f(k/(k+M))) < c\right\}$$

and according to Theorem 3.1

$$U_M = \max_{1\le k\le k^*} V_M^{(1)}(k)/(M^{1/2}(1+k/M)f(k/(k+M))) = o_P(1). \tag{8.14}$$

We observe that

$$\lim_{M\to\infty} P\{\tau_M > J_M\}$$
$$= \lim_{M\to\infty} P\left\{\max\left(U_M, \max_{k^*+1\le k\le J_M}\left|\frac{k}{M}\sum_{\ell=1}^M X_\ell - \sum_{\ell=M+1}^{k^*+M} X_\ell - \sum_{\ell=M+k^*+1}^{M+k} X_\ell\right|\frac{1}{M^{1/2}(k/M)^\gamma}\right) < c\delta_f\sigma\right\}.$$

It is easy to see that

$$\max_{k^*+1\le k\le J_M}\left|\frac{k}{M}\sum_{\ell=1}^M X_\ell\right|\frac{1}{M^{1/2}(k/M)^\gamma} = O_P(1)\max_{k^*+1\le k\le J_M}\left(\frac{k}{M}\right)^{1-\gamma} = o_P(1) \tag{8.15}$$

and similarly

$$\max_{k^*+1\le k\le J_M}\left|\sum_{\ell=M+1}^{k^*+M} X_\ell\right|\frac{1}{M^{1/2}(k/M)^\gamma}=o_P(1). \tag{8.16}$$

It follows from (8.12) that

$$\begin{aligned}
&\max_{k^*+1\le k\le J_M}\left|\sum_{\ell=M+k^*+1}^{k+M} X_\ell\right|\frac{1}{M^{1/2}(k/M)^\gamma}\\
&=\frac{1}{1-\rho_M}\max_{k^*+1\le k\le J_M}\left|\sum_{\ell=1}^{k-k^*}\epsilon_\ell+b_0-b_{k-k^*}\right|\frac{1}{M^{1/2}(k/M)^\gamma}\\
&=\frac{M^{1/2+\gamma}}{a_M}\max_{k^*+1\le k\le J_M}\left|\sum_{\ell=1}^{k-k^*}\epsilon_\ell+b_0-b_{k-k^*}\right|k^{-\gamma}\\
&\ge\frac{M^{1/2+\gamma}}{a_M}\left|\sum_{\ell=1}^{J_M-k^*}\epsilon_\ell+b_0-b_{J_M-k^*}\right|J_M^{-\gamma}\\
&=\frac{M^{1/2+\gamma}}{a_M J_M^\gamma}\left|Z_M+b_0\right|,
\end{aligned} \tag{8.17}$$

where

$$Z_M=\sum_{\ell=1}^{J_M-k^*}\epsilon_\ell-\sum_{\ell=1}^{J_M-k^*-1}\rho_M^\ell\epsilon_{J_M-k^*-\ell}=\sum_{\ell=1}^{J_M-k^*}z_{\ell,M}$$

with

$$z_{\ell,M}=(1-\rho_M^\ell)\epsilon_{J_M-k^*-\ell}.$$

We note that the $z_{j,\ell}$'s, $1\le\ell\le J_M$ are independent mean zero random variables, and so

$$E\left(\sum_{\ell=1}^{J_M-k^*-1}z_{\ell,M}\right)^2=\sum_{\ell=1}^{J_M-k^*-1}\sigma_\epsilon^2(1-\rho_M^\ell)^2.$$

In case of part (i) with $J_M=k^*+2$, it is clear that since $a_M/M\to 0$, $Z_M\xrightarrow{P}0$. Further since $P(b_0=0)=0$, the result follows from (8.16) since $a_M/M^{(1/2+\gamma)}\to 0$.

In case of part (ii), where we take $J_M=JM^{(1-2\gamma)/(3-2\gamma)}$, we note that due to (4.7) for M sufficiently large we have that

$$\begin{aligned}
J_M\frac{a_M}{M}&\le JM^{(1-2\gamma)/(3-2\gamma)-1}M^{1/2+\gamma}\\
&=JM^{-(2\gamma-1)^2/2(3-2\gamma)}\to 0,
\end{aligned} \tag{8.18}$$

as $M \to \infty$. Furthermore, we have by the mean value theorem that

$$0 \leq 1 - \rho_M^\ell = e^{p_{\ell,M}}(-\ell \log(\rho_M)), \text{ with } \ell \log(\rho_M) \leq p_{\ell,M} \leq 0. \tag{8.19}$$

It now follows from a two term Taylor series expansion for $\log(1-x)$ about zero and (8.18) that

$$\max_{1 \leq \ell \leq J_M} |\ell \log(\rho_M)| \to 0, \text{ as } M \to \infty.$$

It follows then from (8.19) that

$$\max_{1 \leq \ell \leq J_M} \left| \frac{1 - \rho_M^\ell}{\ell(a_M/M)} - 1 \right| \to 0, \text{ as } M \to \infty,$$

from which we obtain that

$$\left(\sum_{\ell=1}^{J_M - k^* - 1} \sigma_\epsilon^2 (1 - \rho_M^\ell)^2 \right) \left(\sum_{\ell=1}^{J_M - k^* - 1} \sigma_\epsilon^2 \ell^2 (a_M/M)^2 \right)^{-1} \to 1,$$

as $M \to \infty$.

This then establishes that

$$\lim_{M \to \infty} \left(E \left(\sum_{\ell=1}^{J_M - k^* - 1} z_{\ell,M} \right)^2 \right)^{1/2} \left(\sigma_\epsilon^2 J_M^3 (a_M/M)^2/3 \right)^{-1/2} = 1$$

It may be shown in a similar fashion that

$$\left(\sum_{\ell=1}^{J_M - k^* - 1} E|z_{\ell,M}|^\kappa \right)^{1/\kappa} = O\left(\frac{a_M}{M} J_M^{1+1/\kappa} \right).$$

So using Lyapunov's theorem for arrays of independent and identically distributed random variables (see pg. 126 of Petrov (1995)) we obtain that

$$\left(\frac{\sigma^2}{3} J_M^3 \left(\frac{a_M}{M} \right)^2 \right)^{-1/2} \sum_{\ell=1}^{J_M - k^*} z_{\ell,M} \stackrel{\mathcal{D}}{\to} \mathcal{N},$$

where $\mathcal{N}$ stands for a standard normal random variable. Observing that

$$\frac{M^{1/2+\gamma}}{a_M} J_M^{-\gamma} \frac{a_M}{M} J_M^{3/2} = J^{(1-2\gamma)/(3-2\gamma)}$$

and J can be as large as we wish, (ii) is established.

(iii) First we consider τ_M based on $V_M^{(1)}(k)$. We can assume that $xA_M \geq k^* + 1$. We note

$$P\{\tau_M > xA_M\} = P\left\{ \max_{1 \leq k \leq xA_M} V_M^{(1)}(k)/(M^{1/2}(1 + k/M) f(k/(k+M))) < c \right\}.$$

Next we write

$$\lim_{M\to\infty} P\{\tau_M > xA_M\}$$
$$= \lim_{M\to\infty} P\Bigg\{\max\Bigg(U_M, \max_{k^*+1\le k\le xA_M}\Bigg|\frac{k}{M}\sum_{\ell=1}^{M} X_\ell - \sum_{\ell=M+1}^{k^*+M} X_\ell$$
$$- \sum_{\ell=M+k^*+1}^{M+k} X_\ell\Bigg|\frac{1}{M^{1/2}(k/M)^\gamma}\Bigg) < c\delta_f\sigma\Bigg\}$$

with U_M of (8.14). Similarly to (8.15) and (8.16)

$$\max_{k^*1\le k\le xA_M}\left|\frac{k}{M}\sum_{\ell=1}^{M} X_\ell\right|\frac{1}{M^{1/2}(k/M)^\gamma} = o_P(1)$$

and

$$\max_{k^*+1\le k\le xA_M}\left|\sum_{\ell=M+1}^{k^*+M} X_\ell\right|\frac{1}{M^{1/2}(k/M)^\gamma} = o_P(1)$$

for all $x > 0$. Thus by Lemma 8.3 we have

$$\frac{1}{A_M}\max_{k^*+1\le k\le xA_M}\frac{V_M^{(1)}(k)}{g_M(k)} = \frac{|\sum_{i=k^*}^{k+k^*} X_i|}{\sigma c\delta_f A_M M^{1/2}(k/M)^\gamma}(1+o_P(1))$$
$$= \max_{0\le u\le xA_M}(E\epsilon_0^2)^{1/2}\frac{|W(u)|}{u^\gamma}\frac{M^{\gamma-1/2}}{\sigma c\delta_f(1-\rho_M)A_M}(1+o_P(1)).$$

with a suitably chosen Wiener process W. By the scale transformation of the Wiener process we have

$$\max_{0\le u\le xA_M}\frac{|W(u)|}{u^\gamma} \stackrel{\mathcal{D}}{=} x^{1/2-\gamma}A_M^{1/2-\gamma}\max_{0\le t\le 1}\frac{|W(t)|}{t^\gamma},$$

and by the choice of A_M and ρ_M

$$A_M^{1/2-\gamma}\frac{M^{\gamma-1/2}}{(1-\rho_M)A_M} = 1, \tag{8.20}$$

and therefore (4.13) is proven.

Following the proof of 4.13 one can verify that

$$\frac{1}{A_M}\max_{1\le k\le xA_M}\frac{V_M^{(2)}(k)}{g_M(k)} = \frac{|\sum_{i=k^*}^{k+k^*} X_i - \frac{1}{k}|\sum_{\ell=k^*}^{k}\sum_{i=k^*}^{\ell} X_\ell||}{\sigma c\delta_f A_M M^{1/2}(k/M)^\gamma}(1+o_P(1))$$
$$= \max_{0\le u\le xA_M}(E\epsilon_0^2)^{1/2}\frac{|W(u) - \frac{1}{u}\int_0^u W(y)dy|}{u^\gamma}\frac{M^{\gamma-1/2}}{\sigma c\delta_f(1-\rho_M)A_M}(1+o_P(1)).$$

Using again the scale transformation of the Wiener process we conclude

$$\max_{0\le u\le xA_M}\frac{|W(u)-\frac{1}{u}\int_0^u W(y)dy|}{u^\gamma}\stackrel{\mathcal{D}}{=}(xA_M)^{1/2-\gamma}\sup_{0\le t\le 1}t^{-\gamma}\left|W(t)-\frac{1}{t}\int_0^t W(u)dy\right|,$$

so the last part of the theorem follows from (8.20).

Proof of Theorem 4.2. Proof of Theorem 4.2, we follow the proof of Theorem 4.1. Since the calculations in (8.14)-(8.16) only involve random variables before the change, they remain true under (4.15). Clearly, (8.17) also holds. As before, if $J_M = k^*+2$, $Z_M \stackrel{P}{\to} 0$ on account of $a_M/M \to 0$. Observing that in the present case we also have $a_M/M^{(1/2+\gamma)} \to 0$, the proof of part (i) is established.

The proof of the next two cases are based on the following modification of Lemma 8.1. Assume that $P(b_0 = 0) = 1$ and let $C > 0$. Then we have

$$\max_{1\le t\le CM/a_M} t^{-1/\kappa}|b_t| = O(M/a_M) \quad a.s. \tag{8.21}$$

According to the proof of Lemma 8.1,

$$\max_{1\le t\le CM/a_M} t^{-1/\kappa}|b_t| \leqslant \max_{1\le t\le CM/a_M} t^{-1/\kappa}|b_t| \sum_{\ell=1}^{CM/a_M}\left(1+\frac{a_M}{M}\right)^\ell.$$

Using (4.15) and observing that

$$\sum_{\ell=1}^{CM/a_M}\left(1+\frac{a_M}{M}\right)^\ell \leqslant \sum_{\ell=1}^{CM/a_M} e^{\ell a_M/M} \leqslant \int_1^{CM/a_M+1} e^{xa_M/M}dx \leqslant \frac{M}{a_M}\int_0^{C+a_M/M} e^x dx,$$

(8.21) is proven. Hence we can repeat the calculations in the proof of Theorem 4.1 as long as $J_M a_M/M$ is bounded. Hence, we need to establish the last part of Theorem 4.2. Let

$$J_M = C\left(\frac{M}{a_M}\right)\log M$$

Following the calculations in (8.17) we get that on account of $b_0 = 0$ that

$$\max_{k^*+1\leqslant k\leqslant J_M}\left|\sum_{\ell=M+k^*+1}^{k+M} X_\ell\right|\frac{1}{M^{1/2}(k/M)^\gamma} \geqslant \frac{M^{1/2+\gamma}}{a_M J_M^\gamma}\left|\sum_{\ell=1}^{J_M-k^*}\varepsilon_\ell - b_{J_M-k^*}\right|.$$

Next we note that

$$\left|\sum_{\ell=1}^{J_M-k^*}\varepsilon_l\right| = O_p(J_M^{1/2}).$$

For any $T = T_M$

$$\begin{aligned} b_T &= \sum_{\ell=1}^{T-1} \left(1+\frac{a_M}{M}\right)^{\ell} \varepsilon_{T-\ell} \\ &= \left(1+\frac{a_M}{M}\right)^{T} \sum_{\ell=1}^{T-1} \left(1+\frac{a_M}{M}\right)^{\ell-T} \varepsilon_{T-\ell} \\ &= \left(1+\frac{a_M}{M}\right)^{T} \sum_{s=1}^{T-1} \left(\frac{1}{1+\frac{a_M}{M}}\right)^{s} \varepsilon_{s} \\ &= \left(1+\frac{a_M}{M}\right)^{T} \sum_{s=1}^{T-1} \left(1-\frac{C_M}{M}\right)^{s} \varepsilon_{s}, \end{aligned}$$

where $C_M = \left(1+\frac{a_M}{M}\right) a_M$ and therefore $C_M \to \infty$ and $C_M/M \to \infty$. Let

$$S(u) = \sum_{\ell=1}^{u} \varepsilon_\ell \text{ and } S(0) = 0.$$

By Abel's summation formula

$$\begin{aligned} \sum_{s=1}^{T-1} \left(1-\frac{C_M}{M}\right)^{s} \varepsilon_s &= \sum_{s=1}^{T-1} \left(1-\frac{C_M}{M}\right)^{s} [S(s)-S(s-1)] \\ &= \sum_{s=1}^{T-2} S(t) \left[\left(1-\frac{C_M}{M}\right)^{t} - \left(1-\frac{C_M}{M}\right)^{t-1}\right] + \left(1-\frac{C_M}{M}\right)^{T-1} S(T-1) \\ &= -\frac{C_M}{M} \sum_{s=1}^{T-2} S(t) \left(1-\frac{C_M}{M}\right)^{t-1} + \left(1-\frac{C_M}{M}\right)^{T-1} S(T-1). \end{aligned}$$

Using now (8.13) we get that

$$\sum_{s=1}^{T-2} \left|S(t) - (E\varepsilon_0^2)^{1/2} W(t)\right| \left(1-\frac{C_M}{M}\right)^{t-1} \overset{a.s.}{=\!=} O\left(\sum_{t=1}^{\infty} t^{1/\kappa} e^{-tC_M/M}\right)$$

and

$$\sum_{t=1}^{\infty} t^{1/\kappa} e^{-tC_M/M} = O\left(\left(\frac{M}{C_M}\right)^{1+1/\kappa}\right).$$

Thus we get with probability one that

$$\left(1+\frac{a_M}{M}\right)^{-T} b_T = \sum_{\ell=1}^{T-1} \left(1-\frac{C_M}{M}\right)^{s} \left(E\varepsilon_0^2\right)^{1/2} n_\ell + O\left(\left(\frac{M}{a_M}\right)^{1/\kappa}\right) + O\left(T^{1/\kappa} e^{-Ta_M/M}\right).$$

where $\{n_i, i \geqslant 1\}$ are independent standard normal random variables.

Observing that $J_M^{1/\kappa} e^{-J_M a_M/M} \to 0$ and

$$\left(\frac{M}{a_M}\right)^{-1/2} \sum_{\ell=1}^{T-1} \left(1-\frac{C_M}{M}\right)^{s} \left(E\varepsilon_0^2\right)^{1/2} n_\ell \overset{\mathcal{D}}{\to} N(0,\tau^2)$$

with some $\tau > 0$ and $N(0, \tau^2)$ is a normal random variable with zero mean and variance τ^2. It is easy to see that

$$\frac{M^{1/2+\gamma}}{a_M J_M^\gamma}\left(1+\frac{a_M}{M}\right)^{J_M}\left(\frac{M}{a_M}\right)^{1/2} \to \infty$$

assuming that C is large enough, the proof of Theorem 4.2 is completed.

Lemma 8.4. *If* (2.4), (4.17)–(4.20) *and Assumption 4.1 hold, then we have that*

$$\max_{1\le t\le B} t^{-1/\kappa}|b_t| = O\left(M/a_M\right) \quad a.s. \tag{8.22}$$

and there is a Wiener process $W(u), u \ge 0$ *such that*

$$\max_{1\le t\le B}\left|(1-\rho_M)\sum_{s=1}^t b_s - (E\epsilon_0^2)^{1/2}(1+\alpha_1+\cdots+\alpha_q)W(t)\right| = O(1) \quad a.s. \tag{8.23}$$

Proof. Let

$$\boldsymbol{\eta}_t = (b_t, b_{t-1}, \ldots, b_{t-p})^\top, \quad \boldsymbol{\epsilon}_t = \left(\epsilon_t + \sum_{\ell=1}^q \alpha_\ell \epsilon_{t-\ell}, 0, \ldots, 0\right)^\top \in R^{p+1}$$

and

$$\mathbf{A} = \begin{pmatrix} \beta_1 & \beta_2 & \beta_3 & \ldots & \beta_{p-1} & \beta_p \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 0 \end{pmatrix}.$$

Now (2.4) can be written as

$$\boldsymbol{\eta}_t = \mathbf{A}\boldsymbol{\eta}_{t-1} + \boldsymbol{\epsilon}_t, \quad t = 1, 2, \ldots$$

and therefore

$$\boldsymbol{\eta}_t = \mathbf{A}^t\boldsymbol{\eta}_0 + \sum_{\ell=0}^{t-1}\mathbf{A}^\ell\boldsymbol{\epsilon}_{t-\ell}.$$

We note that that $\det(\mathbf{A} - t\mathbf{I}_n) = t^p\varphi(t)$ ($\mathbf{I}_n$ denotes the $n \times n$ identity matrix) and therefore the eigenvalues of $\mathbf{A}$ are $1/r_{1,M}, 1/r_{2,M}, \ldots, 1/r_{p,M}$. Since the eigenvalues of $\mathbf{A}$ are distinct, we can find a nonsingular matrix $\mathbf{T}$ such that

$$\mathbf{A} = \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}^{-1},$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix with $1/r_{i,M}$ in the diagonal (cf. Exercise 7.32 on page 171 in Abadir and Magnus (2005)). Hence

$$\mathbf{A}^t = \mathbf{T}\boldsymbol{\Lambda}^t\mathbf{T}^{-1}. \tag{8.24}$$

Let $\|\cdot\|$ denote the Euclidean norm of vectors and matrices. It follows from assumptions (4.17)–(4.20) and (8.24) that

$$\|\mathbf{A}^t\boldsymbol{\eta}_0\| \leq C_1\left(1-\frac{a_M}{M}\right)^t \|\boldsymbol{\eta}_0\|$$

with some constant C_1. Hence we obtain the decomposition

$$b_t = \sum_{s=0}^{t} w_s\epsilon_{t-s} + z_t, \quad |w_s| \leq C_2(1-a_M/M)^s \ \text{ and } \ |z_t| \leq C_3(1-a_M/M)^t|b_0|, \tag{8.25}$$

where C_2 and C_3 are constants. By the mean value theorem, $1-\rho_M = 1-(\beta_1+\cdots\beta_p) = \varphi(1)-\varphi(r_{1,M}) = \varphi'(\xi_M)(1-r_{1,M}) = -\varphi'(\xi_M)a_M/M$, with $\xi_M \in [1, 1+a_M/M]$. Clearly then $\xi_M \to 1$ as $M \to \infty$, and since all roots other than $r_{1,M}$ of φ are bounded away from one, $\lim_{M\to\infty}\varphi'(1) \neq 0$. Hence $1-\rho_M = O(a_M/M)$.

It follows from (2.4) that

$$\begin{aligned}(1-\rho_M)\sum_{s=1}^{t} b_s = (1+\alpha_1+\cdots+\alpha_q)&\sum_{s=1}^{t}\epsilon_t \\ &+\sum_{\ell=1}^{p}\beta_\ell[(b_0+\ldots+b_{-\ell+1})-(b_t+b_{t-1}+\ldots+b_{t-\ell+1})] \\ &+\sum_{\ell=1}^{q}\alpha_\ell[(\epsilon_0+\ldots+\epsilon_{-\ell+1})-(\epsilon_t+\ldots+\epsilon_{t-\ell+1})].\end{aligned} \tag{8.26}$$

Due to (8.25) and (8.26), the results (8.22) and (8.23) can be established along the lines of the proofs of Lemmas 8.2 and 8.3, respectively. □

Proof of Theorem 4.3: Theorem 4.3 now follows as Theorem 4.1 with Lemma 8.4 replacing Lemmas 8.2 and 8.3. □

We now turn to the proof of Theorem 4.4, which utilizes the following lemmas:

Lemma 8.5. *If* (4.21)*,* (4.22)*,* (4.24) *and Assumptions 2.1 hold,* $J_M \to \infty$*, then we have that*

$$\frac{1}{\sigma_\epsilon J_M^{5/2}}\left|\sum_{t=1}^{J_M} b_t\right| \ \stackrel{\mathcal{D}}{\to}\ \left|\int_0^1\int_0^u W(s)dsdu\right|,$$

where W denotes a Wiener process.

Proof. We follow the proof of Lemma 8.4. Let $\boldsymbol{\eta}_t = (b_t, b_{t-1})^\top, \boldsymbol{\epsilon}_t = (\epsilon_t, 0)^\top$ and

$$\mathbf{A} = \begin{pmatrix} 2\beta & -\beta^2 \\ 1 & 0 \end{pmatrix}.$$

Now the recursion in (4.21) can be written as

$$\boldsymbol{\eta}_t = \mathbf{A}\boldsymbol{\eta}_{t-1} + \boldsymbol{\epsilon}_t, \;\; t \ge 1 \quad \text{and therefore} \quad \boldsymbol{\eta}_t = \mathbf{A}^t \boldsymbol{\eta}_0 + \sum_{\ell=0}^{t-1} \mathbf{A}^\ell \boldsymbol{\epsilon}_{t-\ell},$$

It is easy to see that β is the eigenvalue of $\mathbf{A}$ with algebraic multiplicity 2 and geometric multiplicity 1. Hence $\mathbf{A}$ cannot be diagonalized but in can be written in Jordan form (cf. Abadir and Magnus (2005)). There is a non–singular matrix $\mathbf{T}$ such that

$$\mathbf{A} = \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}^{-1}$$

with

$$\boldsymbol{\Lambda} = \begin{pmatrix} \beta & 1 \\ 0 & \beta \end{pmatrix} \quad \text{and} \quad \mathbf{T} = \begin{pmatrix} \beta & \beta+1 \\ 1 & 1 \end{pmatrix}$$

(cf. Exercise 7.90 in Abadir and Magnus (2005)). Hence

$$\mathbf{A}^t = \mathbf{T}\boldsymbol{\Lambda}^t\mathbf{T}^{-1}$$

and using mathematical induction one can show that

$$\boldsymbol{\Lambda} = \begin{pmatrix} \beta^t & t\beta^{t-1} \\ 0 & \beta^t \end{pmatrix}, \quad t = 1, 2, \ldots$$

Hence we get that

$$b_t = (1+t)\beta^t b_0 - t\beta^{t+1} b_{-1} + \sum_{s=0}^{t-1} (s+1)\beta^s \epsilon_{t-s}.$$

Let $J = J_M$. It is easy to see that

$$\max_{1 \le t \le J} \left| (1+t)\beta^t b_0 - t\beta^{t+1} b_{-1} \right| = O_P(J). \tag{8.27}$$

Let $\underline{S}(0) = 0, \; \underline{S}(\ell) = \epsilon_1 + \epsilon_2 + \ldots + \epsilon_\ell, \ell \ge 1$. It is easy to see that for all $t \ge 1$ we have by Abel's summation formula that with $u_\ell = (t - \ell + 1)\beta^{t-\ell+1}, 1 \le \ell \le t$

$$\begin{aligned}
\sum_{s=0}^{t-1} (s+1)\beta^s \epsilon_{t-u} &= \sum_{\ell=1}^{t} (\ell+1)\beta^{t-\ell+1}\epsilon_\ell \\
&= u_t \underline{S}(t) + \sum_{\ell=1}^{t-1} \underline{S}(\ell)(u_\ell - u_{\ell+1}) \\
&= \beta \underline{S}(t) + \sum_{\ell=1}^{t-1} \underline{S}(\ell)\beta^{t-\ell} + \sum_{s=1}^{t-1} \underline{S}(\ell)(\beta^{t-\ell} - \beta^{t-\ell+1}) \\
&= \sum_{\ell=1}^{t-1} \underline{S}(\ell) + \beta \underline{S}(t) + \sum_{\ell=1}^{t-1} \underline{S}(\ell)(\beta^{t-\ell} - 1) + \sum_{s=1}^{t-1} \underline{S}(\ell)(\beta^{t-\ell} - \beta^{t-\ell+1}).
\end{aligned}$$

It follows from Assumption 4.1 and the weak convergence of partial sums that

$$\max_{1\le t\le J} |\beta \underline{S}(t)| = O_P(J^{1/2}). \tag{8.28}$$

Let

$$\underline{S}_J(u) = \begin{cases} 0, \text{ if } 0 \le u < 2/J \\ \dfrac{1}{\sigma_\epsilon J^{3/2}} \displaystyle\sum_{\ell=1}^{\lfloor Ju \rfloor -1} \underline{S}(\ell), \text{ if } 2/J \le u \le 1. \end{cases}$$

Using again Assumption 4.1 and the weak convergence of partial sums we get that

$$\underline{S}_J(u) \overset{\mathcal{D}[0,1]}{\longrightarrow} \int_0^u W(s)ds, \tag{8.29}$$

where W denotes a Wiener process. Using (4.24) we conclude that

$$\max_{2\le t\le J} \left| \sum_{\ell=1}^{t-1} \underline{S}(\ell)(\beta^{t-\ell} - 1) \right| \le c \frac{a_M}{M} \max_{2\le t\le J} \sum_{\ell=1}^{t-1} (t-\ell)|\underline{S}(\ell)|$$

and

$$\max_{2\le t\le J} \left| \sum_{\ell=1}^{t-1} \underline{S}(\ell)(\beta^{t-\ell} - \beta^{t-\ell+1}) \right| \le c \frac{a_M}{M} \max_{2\le t\le J} \sum_{\ell=1}^{t-1} |\underline{S}(\ell)|$$

with some constant c. Similarly to (8.29) one can show that

$$\frac{1}{\sigma_\epsilon J^{5/2}} \max_{2\le t\le J} \sum_{\ell=1}^{t-1} (t-\ell)|\underline{S}(\ell)| \overset{\mathcal{D}}{\to} \max_{0\le u\le 1} \int_0^u (u-s)|W(s)|ds$$

and

$$\frac{1}{\sigma_\epsilon J^{3/2}} \sum_{\ell=1}^{J-1} |\underline{S}(\ell)| \overset{\mathcal{D}}{\to} \int_0^1 |W(s)|ds$$

and therefore by (4.24) we obtain that

$$\max_{2\le t\le J} \left| \sum_{\ell=1}^{t-1} \underline{S}(\ell)(\beta^{t-\ell} - 1) \right| = o_P(J^{3/2}) \tag{8.30}$$

and

$$\max_{2\le t\le J} \left| \sum_{\ell=1}^{t-1} \underline{S}(\ell)(\beta^{t-\ell} - \beta^{t-\ell+1}) \right| = o_P(J^{3/2}). \tag{8.31}$$

Let

$$b_J(u) = \frac{1}{\sigma_\epsilon J^{3/2}} b_{\lfloor Ju \rfloor} \quad 0 \le u \le 1.$$

Putting together (8.27)–(8.31) we get that

$$b_J(u) \overset{\mathcal{D}[0,1]}{\longrightarrow} \int_0^u W(s)ds,$$

where W is a Wiener process. Thus Lemma 8.5 is implied by the continuous mapping theorem.

□

Proof of Theorem 4.4. First we note that (4.25) implies that $J_M \to \infty$. According to the proof Theorem 4.1, we need to show only that

$$\frac{\left|\sum_{t=1}^{J_M} b_t\right|}{M^{1/2}((k^* + J_M)/M)^\gamma} \xrightarrow{P} \infty. \tag{8.32}$$

Putting together Lemma 8.5 with (4.3) and (4.25), we obtain immediately 8.32 for the detector $V_M^{(1)}(k)$. Similar arguments can be used in case of $V_M^{(2)}(k)$. □

Proof of Remark 4.1. It follows as in the above calculations that

$$\lim_{M\to\infty} P(\tau_M > xM^{1/5}) = \lim_{M\to\infty} P\left(\max_{1\le k\le xM^{1/5}-k^*}\left|M^{-1/2}\sum_{t=1}^{k} b_t\right| < c\sigma\delta_f\right).$$

As in Lemma 8.5, we have that

$$M^{-1/2}\sum_{t=1}^{\lfloor uM^{1/5}\rfloor} b_t \overset{\mathcal{D}[0,1]}{\longrightarrow} \sigma_\epsilon \int_0^u \int_0^z W(y)dydz,$$

from which we obtain that

$$\lim_{M\to\infty} P\left(\max_{1\le k\le xM^{1/5}-k^*}\left|M^{-1/2}\sum_{t=1}^{k} b_t\right| < c\sigma\delta_f\right) = P\left(\sup_{0\le u\le x}\left|\int_0^u\int_0^z W(y)dydz\right| < cx^{-5/2}\frac{\delta_f\sigma}{\sigma_\epsilon}\right).$$

By changing variables in the integral and applying the scale transofmration of the Wiener process, we get that

$$\sup_{0\le u\le x}\left|\int_0^u\int_0^z W(y)dydz\right| \xrightarrow{\mathcal{D}} x^{-5/2}\sup_{0\le u\le 1}\left|\int_0^u\int_0^z W(y)dydz\right|,$$

from which the result follows

Lemma 8.6. *If* (2.6), (4.27)–(4.31) *hold,* $k = k_M$ *and*

$$\phi_M = O(1/k_M),$$

then we have that

$$\left(\frac{\phi_M}{k_M}\right)^{1/2}\sum_{t=1}^{k_M} b_t \xrightarrow{\mathcal{D}} N(0,\omega^2),$$

where $N(0,\omega^2)$ *is a normal random variable with zero mean and variance* ω^2*.*

Proof. We recall that $\phi = \phi_M = 1 - (\alpha + \beta)$. According to Corollary 3.1 in Hall and Heyde (1980) we need to show only that

$$\frac{\phi}{k}\sum_{i=1}^{k}\sigma_i^2 \xrightarrow{P} \omega^2 \tag{8.33}$$

and

$$\frac{\phi^2}{k^2}\sum_{i=1}^{k}\sigma_i^4 \xrightarrow{P} 0 \tag{8.34}$$

as $M \to \infty$. It follows from (2.6) that

$$\sigma_i^2 = \omega + (\alpha\epsilon_{i-1} + \beta)\sigma_{i-1}^2, i \geq 1 \tag{8.35}$$

and therefore

$$\sigma_i^2 = \omega\sum_{\jmath=0}^{i-1}\prod_{\ell=1}^{j}(\alpha\epsilon_{i-\ell}^2 + \beta) + \sigma_0^2\prod_{\ell=0}^{j}(\alpha\epsilon_{i-\ell}^2 + \beta)$$

($\prod_\emptyset = 1$). Hence

$$E\sigma_i^2 = \omega\frac{1 - (\alpha + \beta)^{i-1}}{\phi} + (\alpha + \beta)^i E\sigma_0^2. \tag{8.36}$$

Squaring (8.35), taking the expected value of both sides we conclude that with $z_i = \omega^2 + 2\omega(\alpha + \beta)E\sigma_{i-1}^2$ and $z = E(\alpha\epsilon_{i-1}^2 + \beta)^2$ that

$$\begin{aligned} E\sigma_i^4 =& \omega^2 + 2\omega(\alpha + \beta)E\sigma_{i-1}^2 + E(\alpha\epsilon_{i-1}^2 + \beta)^2 E\sigma_{i-1}^4 \\ =& z_i + zE\sigma_{i-1}^4, \end{aligned}$$

which yields

$$E\sigma_i^4 = \sum_{\ell=1}^{i-1} z^{\ell-1}z_{i-\ell} + z^i E\sigma_0^4. \tag{8.37}$$

By (8.36) we have that

$$z_{i-\ell} \leq \omega^2 + \frac{2\omega^2}{\phi} + (\alpha + \beta)^{i-\ell}E\sigma_0^2$$

and therefore (8.37) yields

$$E\sigma_i^4 \leq \left(\omega^2 + \frac{2\omega^2}{\phi} + E\sigma_0^2\right)\frac{z(1 - z^{i-2})}{1 - z} + z^{i-1}E\sigma_0^4. \tag{8.38}$$

Observing that

$$E(\alpha\epsilon_{i-1}^2 + \beta)^2 = (\alpha + \beta)^2 + \alpha^2 E(\epsilon_{i-1}^2 - 1)^2 = 1 - 2\phi_m + o(a_M/M), \text{ as } M \to \infty,$$

it follows from (8.38) there is a constant C such that

$$E\sigma_i^4 \leq C/\phi_M^2. \tag{8.39}$$

Using (8.39) we conclude that

$$\frac{\phi^2}{k^2}\sum_{i=1}^{k} E\sigma_i^4 \leq \frac{C}{k},$$

so Markov's inequality implies (8.34). We rewrite (8.35) as

$$\sigma_i^2 = \omega + (\alpha+\beta)\sigma_{i-1}^2 + \alpha(\epsilon_{i-1}^2 - 1)\sigma_{i-1}^2$$

we get

$$\frac{\phi}{k}\sum_{i=1}^{k}\sigma_i^2 = \omega + \frac{1}{k}(\alpha+\beta)\sigma_0^2 - \frac{1}{k}(\alpha+\beta)\sigma_k^2 + \frac{1}{k}\sum_{i=1}^{k}\alpha(\epsilon_{i-1}^2 - 1)\sigma_{i-1}^2.$$

Using (8.39) we obtain that

$$\begin{aligned} E\left(\frac{1}{k}\sum_{i=1}^{k}\alpha(\epsilon_{i-1}^2 - 1)\sigma_{i-1}^2\right)^2 &= \frac{1}{k^2}\sum_{i=1}^{k}\alpha^2 E(\epsilon_{i-1}^2 - 1)^2 E\sigma_{i-1}^4 \\ &= o\left(\frac{1}{k}\frac{a_M}{M}\frac{1}{\phi_M^2}\right), \end{aligned}$$

so by Chebyshev's inequality we have

$$\frac{1}{k}\sum_{i=1}^{k}\alpha(\epsilon_{i-1}^2 - 1)\sigma_{i-1}^2 = o_P(1).$$

Combining (8.36) with Markov's inequality we conclude

$$\frac{1}{k}\sigma_k^2 = o_P(1),$$

which completes the proof of (8.33).

Proof of Theorem 4.5. Let

$$k_M = C\left(\frac{M}{a_M} + \left(\frac{a_M}{M^{2\gamma}}\right)^{1/(1-2\gamma)}\right).$$

We note that k_M satisfies the assumption of Lemma 8.6. It follows from Lemma 8.6 that for $i = 1, 2$ and all $x > 0$

$$\lim_{C\to\infty}\liminf_{M\to\infty} P\left\{V^{(i)}(k_M) > x\right\} = 1,$$

establishing Theorem 4.5. □

Acknowledgement

We are grateful to the Editor (Oliver Linton), one anonymous Associate Editor and two Referees for their helpful insights and remarks, which have improved the manuscript significantly. The research of Zhenya Liu was supported by the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China [17XNQJ01]. The research of Gregory Rice was partially supported by the discovery grant from the Natural Sciences and Engineering Research Council of Canada.

References

[1] Abadir, K.M. and Magnus, J.R.: *Matrix Algebra.* Cambridge University Press, Cambridge, U.K., 2005.

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

[3] Aue, A., Hörmann, S. Horváth, L. and Hušková, M.: Dependent functional linear models with applications to monitoring structural change. *Statistica Sinica* **24**(2014), 1043–1073 (with online supplemen).

[4] Aue, A., Hörmann, S., Horváth, L., Hušková, M. and Steinebach, J.: Sequential testing for the stability of portfolio betas. *Econometric Theory* **28**(2012), 804–837.

[5] Aue, A. and Horváth, L.: A limit theorem for mildly explosive autoregression with stable errors. *Econometric Theory* **23**(2006) 201–220.

[6] Aue, A. and Horváth, L.: Structural breaks in time series. *Journal of Time Series Analysis* **23**(2013), 1–16.

[7] Berkes, I., Hörmann, S. and Horváth, L.: The functional central limit theorem for a family of GARCH observations with applications. *Statistics & Probability Letters* **78**(2008), 2725–2730.

[8] Berkes, I., Horváth, L. and Kokoszka, P.: Near–integrated GARCH sequences. *Journal of Applied Probability* **15**(2005), 890–913.

[9] Billingsley, P.: *Convergence of Probability Measures.* Wiley, New York, 1968.

[10] Bollerslev, T.: Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* **31**(1986), 307–327.

[11] Busetti, F., and Taylor. A.: Tests of Stationarity Against a Change in Persistence, *Journal of Econometrics* **123**(2004) 33–66.

[12] Bradley, R.C.: *Introduction to strong mixing conditions.* Volumes 1–3. Kendrick Press, Heber City, UT, 2007.

[13] Caner, M. and Kilian, L.: Size distortions of tests of the null hypothesis of stationarity: evidence and implications for the PPP debate. *Journal of International Money and Finance* **20**(2001), 639–657.

[14] Cochrane, J.: A critique of the application of unit root tests, *Journal of Economic Dynamics and Control*, (15), (1991), 275– 284

[15] Chow, Y.S. and Teicher, H.: *Probability Theory: Independence, Interchangeability, Martingales.* Springer Verlag, 1988.

[16] Chu, C.–S.J., Stinchcombe, J. and White, H.: Monitoring structural change. *Econometrica* **64**(1996), 1045–1065.

[17] Csörgő, M. and Révész, P.: *Strong Approximations in Probability and Statistics.* Academic Press, 1981.

[18] Cheung, Y. W. Chinn, M. D.: Further investigation of the uncertain unit root in GNP. *Journal of Business & Economic Statistics*, **15**(1997), 68–73.

[19] Diebold, F. X. and Senhadji, A. S.: The uncertain unit root in real GNP: Comment. *The American Economic Review*, **86**(1996)., 1291–1298.

[20] Dedecker, J., Doukhan, P., Lang, G., León R., José, R. Louhichi, S. and Prieur, C.: *Weak Dependence: with Examples and Applications.* Lecture Notes in Statistics, 190. Springer, New York, 2007.

[21] Engle, R.F.: Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**(1982), 987–1007.

[22] Engle, R.F. and Bollerslev, T.: Modelling the persistence of conditional variances. *Econometric Reviews* **5**(1986), 1–50.

[23] Francq, C. and Zakoian, J–M.: *GARCH Models.* Wiley, New York, 2010.

[24] Giraitis, L., Kokoszka, P.S., Leipus, R. and Teyssiére, G.: Rescaled variance and related tests for long memory in volatility and levels. *Journal of Econometrics* **112**(2003), 265–294.

[25] Gürkaynak, R.: Econometric tests of asset price bubbles: taking stock, *Journal of Economic Surveys*, (2005) 166–186.

[26] Hafner, C. M.: Testing for Bubbles in Cryptocurrencies with Time-Varying Volatility. *Journal of Financial Econometrics*, forthcoming.

[27] Hall, P. and Heyde, C.C.: *Martingale Limit Theory and Its Application.* Academic Press, New York, 1980.

[28] Hoga, Y.: Monitoring multivariate time series. *Journal of Multivariate Analysis* **155**(2017), 105–121.

[29] Hong, S. Y., Linton, O., and Zhang, H. J.: An investigation into multivariate variance ratio statistics and their application to stock market predictability. *Journal of Financial Econometrics* **15(2)** (2017), 173–222.

[30] Homm, U. and Breitung, J.: Testing for Speculative Bubbles in Stock Markets: A Comparison of Alternative Methods, *Journal of Financial Econometrics*,**10** (2012), 198–231.

[31] Horváth, L., Horváth, Zs., Hušková, M.: Ratio tests for change point detection. In: *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor P. K. Sen.* IMS, 2008, pp. 293—304.

[32] Horváth, L., Hušková, M., Kokoszka, P. Steinebach, J.: Monitoring changes in linear models. *Journal of Statistical Planning and Inference* **126**(2004) 225–251.

[33] Horváth, L. and Rice, G.: Extensions of some classical methods in change point analysis (with discussions) *TEST* **23**(2014), 219–290.

[34] Hull, J.C.: *Options, Futures and Other Derivatives.* Prentice Hall, Upper Saddle River, Nj, 2000.

[35] Kim, J.–Y.: Detection of change in persistence of a linear time series. *Journal of Econometrics* **95**(2000), 97–116.

[36] Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y.: Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? *Journal of Econometrics* **54**(1992), 159–178.

[37] Lee, J.H. and Phillips, P.C.B.: Asset pricing with financial bubble risk. *Journal of Empirical Finance* **38**(2016), 590-622.

[38] Liu, W. and Wu, W.B.: Asymptotics of spectral density estimates. *Econometric Theory* **26**(2010), 1218–1245.

[39] Linton, O.: *Financial Econometrics: Models and Methods.* Cambridge University Press, Cambridge, 2019.

[40] Murray, C. J. and Nelson, C. R.: The uncertain trend in US GDP. *Journal of Monetary Economics*, **46**(2000), 79-95.

[41] Michaelides, P.G., Tsionas, E.G. and Konstantakis, K.N.: *Financial bubble detection: a non–linear method with application to S&P 500.* MPRA Paper No. 74477, 2014. `https://mpra.ub.uni-muenchen.de/74477/`

[42] Müller, U.K.: Size and power of tests of stationarity in highly autocorrelated time series. *Journal of Econometrics* **128**(2005), 195–213.

[43] Nelson, C. R. and Plosser, C. R.: Trends and random walks in macroeconmic time series: some evidence and implications. *Journal of Monetary Economics*, **10**(1982), 139-162.

[44] V.V. Petrov, *Limit Theorems of Probability Theory*, Clarendon Press, Oxford, 1995.

[45] Phillips, P.C.B. and Magdalinos, T.: Limit theory for moderate deviations from a unit root. *Journal of Econometrics* **136**(2007) 115—130.

[46] Perron, P.: The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, (1989) 1361–1401.

[47] Phillips, P.C.B., Shi, S. and Yu. J.: Specification sensitivity in right–tailed unit root testing for explosive behaviour. *Oxford Bulletin of Economics and Statistics* **76**(2014), 15–333.

[48] Phillips, P.C.B., Shi, S. and Yu, J.:Testing for multiple bubbles: historical episodes of exuberance and collepse. in the S&P 500. *International Economic Review* **56**(2015a), 1043–1177.

[49] Phillips, P.C.B., Shi, S. and Yu, J.: Testing for multiple bubbles: limit theory of real–time detectors. *International Economic Review* **56**(2015b), 1079–1133.

[50] Phillips, P.C.B. and Solo, V.: Asymptotics for linear processes. *Annals of Statistics* **20**(1992), 971—1001.

[51] Phillips, P.C.B. and Yu, J: Dating the timeline of financial bubbles during the subprime crisis. *Quantitative Economics* **2**(2011), 455–491.

[52] Móricz, F.A., Serfling, R.J. and Stout, W.F.: Moment and probability bounds with quasi–superadditive structure for the maximum partial sum. *Annals of Probability* **4**(1982), 1032–1040.

[53] Shiller, R. J.: *Market Volatility.* MIT press, 1992.

[54] Shorack, G.R. and Wellner, J.A.: *Empirical Processes with Applications to Statistics.* Wiley, 1986.

[55] Taniguchi, A. and Kakizawa, Y.: *Asymptotic Theory of Statistical Inference for Time Series.* Springer, New York, 2000.

[56] Wu, W.B. and Woodroofe, M.: Martingale approximations for sums of stationary processes. *Annals of Probability* **32**(2004), 1674–1690.

[57] Zivot, E., and D.W.K. Andrews.: Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis, *Journal of business & economic statistics*, **20** (2002), 25–44.