Statistics and Probability Letters 184 (2022) 109385

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

An alternative sequential method for the state estimation of a partially observed SETAR(1) process

Paula Milheiro-Oliveira

Universidade do Porto, Faculdade de Engenharia and CMUP, P-4200-465 PORTO, Portugal

ARTICLE INFO

Article history:
Received 25 June 2021
Received in revised form 16 November 2021
Accepted 22 January 2022
Available online 2 February 2022

Keywords:
Threshold models
Partial observations
Nonlinear filtering
Hypothesis testing

ABSTRACT

We discuss a novel sequential test based on the quadratic variation of the observations to decide which regime governs the dynamics of the non-observed process in a filtering problem with small observation noise. The non-observed state process is a self-exciting threshold autoregressive process of order one (SETAR(1)) with two regimes. The observation function is not one-to-one. The proposed procedure performs well and may be competitive in some applications.

1. Introduction

Problems of estimating a non-observed threshold autoregressive-type nonlinear process arise in practice in a diversity of engineering and financial applications. Real-world examples can be found in volatility extraction from economic and financial time series (e.g. Tong and Yeung, 1991) such as exchange rates, indices, for instance GNP and stock prices, and prediction. Common target tracking problems, for example, also fall in the category of problems under investigation here (Bar-Shalom and Li, 1996; Liu and Zhang, 2001). When the connection between the non-observed process and the noisy observations leans on an observation function that is one-to-one, the estimation problem (filtering) is not difficult to solve (e.g. Gelb, 1974). Additional difficulties come on the scene otherwise and accurate estimators (filters) cannot generally be found. The problem in discrete time was first treated in Fleming and Zhang (1991) and Fleming et al. (1991) where statistical hypothesis tests were proposed, namely a quadratic variation test (QVT) and a likelihood ratio test (LRT), aiming at taking a decision to choose among competing filters on successive time instants. For such tests to perform a detectability assumption must hold true. Later, Milheiro-Oliveira and Roubaud (1995) adapted the LRT test to the context of a particular detectability assumption not covered by previous authors. See also Fleming and Zhang (1992) for an exploration of the related numerical results. The continuous-time analogues appeared earlier in Fleming et al. (1988a) and Roubaud (1993, 1995) for the piecewise linear case, in Fleming and Pardoux (1989) for the nonlinear piecewise monotone case, and were later generalized in Zhang (1998) and Wang et al. (2006) to the case of a diffusion term depending on an unknown Markov chain and to hybrid systems. The problem studied in Tong and Yeung (1991) is different: tests for nonlinearity in time series are adapted and extended to cope with partially observed series with unknown parameters. Another related but different problem is treated in Brouste et al. (2020), where a likelihood ratio test for a change in the mean-reverting parameter of a first-order AR(1) with stationary Gaussian noise is proposed. Although the model investigated is more general in that the noise may not be white and model parameters are also unknown, Brouste et al. (2020) do not consider partially observed time series. Related work on partially observed discrete dynamical systems (PODDS) investigates state

E-mail address: poliv@fe.up.pt.

https://doi.org/10.1016/j.spl.2022.109385

estimation in a class of piecewise models for which the state space is finite (see Imani and Ghoreishi, 2021 and references therein) which is not the case in this paper. We should also refer recent research on data science that uses state space description of the dynamics and explicit state space estimation methods of the type presented here to establish self-supervised learning algorithms. An example is the adoption of mass–spring systems representation for robot manipulation of deformable linear objects (Yan et al., 2020).

To be more specific, let us consider the following discrete time piecewise linear model

$$\begin{cases} X_{k+1} = X_k + \varepsilon\, b(X_k) + \sqrt{\varepsilon}\,\sigma(X_k)\,u_k \\ Y_k = h(X_k) + \sqrt{\varepsilon}\,v_k\,, \end{cases} \qquad k = 0, 1, \ldots, K \tag{1}$$

with

(H_1) $b(x) = B_- x \mathbf{1}_{\{x<0\}} + B_+ x \mathbf{1}_{\{x\geq 0\}}$, $\sigma(x) = \sigma_- x \mathbf{1}_{\{x<0\}} + \sigma_+ x \mathbf{1}_{\{x\geq 0\}}$, $h(x) = H_- x \mathbf{1}_{\{x<0\}} + H_+ x \mathbf{1}_{\{x\geq 0\}}$, $B_-, B_+, \sigma_-, \sigma_+, H_-$ and H_+ being parameters of the model, assumed known;
(H_2) $\{u_k\}_k$, $\{v_k\}_k$ are standard Gaussian independent white noises;
(H_3) X_0 is a random variable with law $\mathcal{N}(\mu_0, \sigma_0^2)$ and is independent of the noises $\{u_k\}_k$ and $\{v_k\}_k$;
(H_4) $\varepsilon > 0$ is a small parameter,

all processes taking values in $\mathbb{R}$. The process $\{X_k\}_k$ is nonlinear and is called a Self-Exciting Threshold Autoregressive process of order 1 (SETAR(1)).

Note that this model corresponds to an Euler discretization of a piecewise linear continuous-time model with time step equal to ε. Constant terms are not included in the model since, by change of origin, one can always reduce any piecewise linear model to the above case.

Assuming that the process $\{X_k\}_k$ is not observed but observations are available corresponding to realizations of the process $\{Y_k\}_k$, our goal is to solve the problem of estimating X_k, for each time instant t_k, based on the observations available until t_k, $Y_1, Y_2, \ldots, Y_k$. It is well known that the best estimate of X_k in the mean square sense is $\hat{X}_k = E[X_k|\mathcal{Y}_0^k]$, where $\mathcal{Y}_0^k$ is the σ-algebra of the observations until time t_k. If model (1) was linear it would be possible to obtain closed formulas for $\hat{X}_k$ and $E[(X_k - \hat{X}_k)^2 \mid \mathcal{Y}_0^k]$, thus completely characterizing the conditional distribution of X_k given $\mathcal{Y}_0^k$, which would be Gaussian (Kalman and Bucy, 1961). Since model (1) is nonlinear, the estimation requires studying the just mentioned conditional distribution. Unfortunately, computation of the moments and their behavior along time is rather complex and difficult to perform in real time. For this reason it is frequent in engineering and financial applications to approximate the solution by using suboptimal estimators. The Extended Kalman Filter (EFK) is a rather popular estimator for this purpose (we refer for instance to Gelb, 1974). If $H_+H_- < 0$, i.e. the function h is not one-to-one, although it is easy to estimate $h(X_k)$ the same is not true for X_k. The EKF does not necessarily converge to the optimal solution, the difficulty being the identification of whether $x_k > 0$ or $x_k < 0$ (the sign of X_k). In this paper we are interested in this last case. In situations where the conditional variance is small the problem can indeed be solved. That is the main reason to set a "detectability assumption" which can be one of the following:

(HD) $H_-^2\sigma_-^2 = H_+^2\sigma_+^2$ and $B_+ \neq B_-$;
(HD') $H_-^2\sigma_-^2 \neq H_+^2\sigma_+^2$.

The problem under assumption (HD') is studied in Fleming et al. (1991). In the present paper, we study the problem under assumption (HD). Without loss of generality we will assume that

(H_5) $H_+ > 0, H_- < 0$.

We will also assume that

(H_6) $B_+ < 0\,, B_- < 0\,.$

Assumption (H_6) will be necessary in Section 3, when the existence of a stationary solution of system (1) is assumed. Although assumption (H_6) appears to be restrictive, this is in fact not the case. If e.g. $B_+ > 0\,, B_- < 0$ the trajectories, once they escape to the positive side of the plane, will explode if given enough time. In this region, system (1) behaves almost like a linear system thus, again, an accurate solution to the problem is well known (Kalman and Bucy, 1961). Assumption (H_6) corresponds to the situation where the non-observed process is most likely to cross 0, thus creating difficulties in the identification of its sign.

The paper is organized as follows: in Section 2 we explain the steps that should be followed in order to solve the filtering problem; in Section 3 we propose and justify the use of a sequential test based on the quadratic variation of the observation for decision on the sign of the non-observed process, which is our main contribution to the subject; finally, in Section 4 we briefly present some simulation results.

The following definition will be used in the sequel:

Definition 1.1. Consider a stochastic process $\{\xi_k\}_k$ with values in $\mathbb{R}$. We will write $\xi_k = \mathcal{O}(\varepsilon^p)$, with $p > 0$, when, for some $q_1, q_2 \geq 0$, $\exists\, C_1, C_2, C_3 > 0 \;:\; \forall_{k\geq 0}\; E[|\xi_k|] \leq \frac{C_1}{\varepsilon^{q_1}} e^{-C_3 k/\varepsilon^{q_2}} + C_2\varepsilon^p$. In this situation the process $\{\xi_k\}_k$ is usually said to converge to zero with rate of order ε^p, when ε converges to zero.

2. Solving the filtering problem

The described filtering problem can be solved by performing the following steps on the observed time interval:

1. apply a test for detection of zero crossings;
2. for each time interval with no zero crossings detected apply a test to decide in which region the non-observed state is evolving (i.e. $X < 0$ or $X > 0$);
3. for those detected time intervals use the Kalman–Bucy Filter (KF) associated with either the linear system

$$(+) \qquad \begin{cases} X_{k+1} &= (1+B_+\varepsilon)X_k + \sqrt{\varepsilon}\sigma_+ u_k \\ Y_k &= H_+X_k + \sqrt{\varepsilon}v_k \end{cases}$$

or the linear system

$$(-) \qquad \begin{cases} X_{k+1} &= (1+B_-\varepsilon)X_k + \sqrt{\varepsilon}\sigma_- u_k \\ Y_k &= H_-X_k + \sqrt{\varepsilon}v_k \end{cases}$$

according to the region that was found in the previous step.

When ε converges to zero, if the tests perform well, the estimation procedure as set out should converge to the optimal estimator. Indeed, the two KFs have a "short memory" (old values of the observation are essentially not used), therefore the effect of the initial condition vanishes. As a consequence, on the time intervals in which the state $\{X_k\}$ remains away from zero, if we know its sign, the corresponding KF, which is an exact finite dimensional filter in the linear case, provides an accurate approximation of the optimal filter. After a certain time, the conditional law has only one significant peak and is approximately Gaussian. The same argument is used in Roubaud (1995)[p. 165] for the continuous-time analog; see also Pardoux and Roubaud (1989) [Theorem 5.1].

Regarding step 1, that is, the detection of zero crossings, two types of tests can be adopted: one test based on the observations or one test based on the output of the KFs associated with both linear models $(-)$ and $(+)$. The test based on the observations is proposed in Fleming et al. (1991) under assumption (HD') and one can conclude without much effort that it can still be used under assumption (HD). The test based on the output of the KFs is proposed in Milheiro-Oliveira and Roubaud (1995). In this paper, a test of the first type, that is a zero crossings test based on the observations, will be adopted and can be described as follows (see Fleming et al. (1991) for details). With c given by[1] $c = \frac{\Phi^{-1}(1-\alpha)\sqrt{\varepsilon}}{|H_- - H_+|} \max\left(\sqrt{H_+^2H_-^2\sigma_-^2 + H_-^2 + H_+^2}, \sqrt{H_+^2H_-^2\sigma_+^2 + H_-^2 + H_+^2}\right)$, check if $\{|Y_k| \geq c\,,\ k = i_0, i_0+1, \ldots, i_0+m\}$ occurs for some i_0 and m. If true decide that no zero crossings exist on $\{i_0, i_0+1, \ldots, i_0+m\}$. Thus we will assume that x_k may have changed sign only when the observations drop below level c, in absolute value.

Let us now focus on step 2. For the decision on the sign of X_k, likelihood ratio tests are proposed in the literature (see Milheiro-Oliveira and Roubaud, 1995; see Fleming et al. (1988b) for the analogue under detectability assumption (HD')). Under detectability assumption (HD') another type of test is proposed in Fleming et al. (1991), a test based on the quadratic variation of the observations. The aim of the present paper is to adapt this type of test to assumption (HD), the general idea of the test and the techniques involved being similar to those in Fleming et al. (1991). As usual, a test statistic S_n is designed in such a way that the alternative represented by system $(-)$ corresponds to an event "$S_n \leq -l_1$" and the alternative represented by system $(+)$ corresponds to an event "$S_n \geq l_2$", for given constants l_1 and l_2. In order to perform this kind of test, formulas for l_1 and l_2 are needed. The test is sequential in the sense that the test statistic S_n evolves with time as observations are collected. In the next Section, approximated formulas are derived for l_1 and l_2 as well as for the expected times to reach a decision.

3. A quadratic variation based test for decision on the sign of the non-observed process

As mentioned in the previous Section, it is our purpose to build a test based on the quadratic variation of the observation process $\{Y_k\}_k$ which enables us to decide on the sign of the non-observed process $\{X_k\}_k$, as observations are being provided. In the design of the sequential test, the alternatives to be tested are the occurrence of one of the following events: $A_+ = \{x_k > 0\,;\ k = i_0, i_0+1, \ldots, i_0+m\}$, $A_- = \{x_k < 0\,;\ k = i_0, i_0+1, \ldots, i_0+m\}$, where the interval $[i_0, i_0+m]$ is defined in the first step of the procedure presented in Section 2, that is, it is an interval where one assumes that no zero crossings of the non-observed process $\{X_k\}_k$ occur. Therefore, for fixed i_0 and m, we consider the null hypothesis $H_0 :\ b(x_k) = B_+\,,\ \sigma(x_k) = \sigma_+\,,\ h(x_k) = H_+$, for $k = i_0, \ldots, i_0+m$ and the alternative $H_1 :\ b(x_k) = B_-\,,\ \sigma(x_k) = \sigma_-\,,\ h(x_k) = H_-$, for $k = i_0, \ldots, i_0+m$.

In order to define the test statistic, the following notation is introduced:

$$\Delta_k^+ \stackrel{def}{=} Y_{k+1} - (1+B_+\varepsilon)Y_k\,, \quad \Delta_k^- \stackrel{def}{=} Y_{k+1} - (1+B_-\varepsilon)Y_k.$$

[1] Φ denotes the standard normal distribution function.

Notice that under H_0 one can write that

$$\Delta_k^+ = H_+\sigma_+\sqrt{\varepsilon}\,u_k + \sqrt{\varepsilon}\,v_{k+1} - (1+B_+\varepsilon)\sqrt{\varepsilon}\,v_k\,, \tag{2}$$

$$\Delta_k^- = H_+(B_+ - B_-)\varepsilon X_k + H_+\sigma_+\sqrt{\varepsilon}\,u_k + \sqrt{\varepsilon}\,v_{k+1} - (1+B_-\varepsilon)\sqrt{\varepsilon}\,v_k \tag{3}$$

while, under H_1, one writes

$$\Delta_k^+ = H_-(B_- - B_+)\varepsilon X_k + H_-\sigma_-\sqrt{\varepsilon}\,u_k + \sqrt{\varepsilon}\,v_{k+1} - (1+B_+\varepsilon)\sqrt{\varepsilon}\,v_k\,,$$
$$\Delta_k^- = H_-\sigma_-\sqrt{\varepsilon}\,u_k + \sqrt{\varepsilon}\,v_{k+1} - (1+B_-\varepsilon)\sqrt{\varepsilon}\,v_k\,.$$

It is easy to conclude that, under hypothesis H_0, one has $\Delta_k^+ \sim \mathcal{N}(0, \Gamma_+^2\varepsilon)$ and, under hypothesis H_1, one has $\Delta_k^- \sim \mathcal{N}(0, \Gamma_-^2\varepsilon)$ with

$$\Gamma_+^2 \stackrel{def}{=} H_+^2\sigma_+^2 + 1 + (1+B_+\varepsilon)^2\,, \quad \Gamma_-^2 \stackrel{def}{=} H_-^2\sigma_-^2 + 1 + (1+B_-\varepsilon)^2\,. \tag{4}$$

Notice that assumption (HD) implies $\Gamma_+^2 - \Gamma_-^2 = \mathcal{O}(\varepsilon)$. More precisely

$$\Gamma_+^2 - \Gamma_-^2 = \varepsilon(B_+ - B_-)(2 + (B_+ + B_-)\varepsilon)\,. \tag{5}$$

Notice also that, from (2)–(4), under H_0 we can easily write that $\frac{1}{\sqrt{\varepsilon}}\Delta_k^- = \Gamma_+ w_k^+ + \mathcal{O}(\sqrt{\varepsilon})$ and $\frac{1}{\sqrt{\varepsilon}}\Delta_k^+ = \Gamma_+ w_k^+$, with $w_k^+ = [H_+\sigma_+ u_k + v_{k+1} - (1+B_+\varepsilon)v_k]\,/\,\Gamma_+$. The processes $\{w_k^+ : k \text{ odd}\}$ and $\{w_k^+ : k \text{ even}\}$ are both standard Gaussian white noises.

Let us now define the process $Z_k = \frac{1}{\varepsilon}\ln\frac{\Gamma_-}{\Gamma_+} + \frac{1}{2\varepsilon}\left(\frac{\Delta_k^{-2}}{\Gamma_-^2} - \frac{\Delta_k^{+2}}{\Gamma_+^2}\right)$ and finally the test statistic

$$S_n = \sum_{k=i_0}^{i_0+n} Z_k\,, \text{ for } 0 \le n \le m\,. \tag{6}$$

The decision rule of the test is as follows: accumulate S_n until either "$S_{N^*} \le -l_1$" or "$S_{N^*} \ge l_2$" occurs or until the end of the time interval $[i_0, i_0+m]$ previously defined is reached; if "$S_{N^*} \le -l_1$" occurs then hypothesis H_1 is kept; if "$S_{N^*} \ge l_2$" occurs then H_0 is kept; otherwise no decision is taken. The variable N^* is a stopping time: $N^* = \inf(\{n \in \mathbb{N} : S_n \le -l_1 \text{ or } S_n \ge l_2\}, m)$. Having in mind that formulas for l_1 and l_2 must be provided let us first rewrite the statistic S_n as

$$S_n = \frac{n+1}{\varepsilon}\ln\frac{\Gamma_-}{\Gamma_+} + \frac{1}{2\varepsilon^2}\sum_{k=i_0}^{i_0+n}\left(\frac{\Delta_k^{-2}}{\Gamma_-^2} - \frac{\Delta_k^{+2}}{\Gamma_+^2}\right)\,. \tag{7}$$

Let us consider the error probabilities

$$p_+ = P(\text{“rejecting } H_0\text{” } |H_0)\,, \quad p_- = P(\text{“rejecting } H_1\text{” } |H_1)\,. \tag{8}$$

Approximate formulas for these probabilities are proposed in Proposition 3.1. Approximate values for the two constants l_1 e l_2 will stem from solving (9) and (10) for fixed values of p_+ and p_-.

Proposition 3.1. *The probabilities p_+ and p_- defined in (8) are approximately given by*

$$p_+ = \frac{1 - e^{-\theta_+ l_2}}{e^{\theta_+ l_1} - e^{-\theta_+ l_2}} \tag{9}$$

and

$$p_- = \frac{1 - e^{\theta_- l_1}}{e^{-\theta_- l_2} - e^{\theta_- l_1}}\,, \tag{10}$$

with

$$\theta_+ = \frac{\left(-4|B_+| + H_+^2\sigma_+^2(B_+ - B_-)\right)(H_+^2\sigma_+^2 + 2)^3}{2|B_+|(B_+ - B_-)\left(12 + 12H_+^2\sigma_+^2 + 5H_+^4\sigma_+^4 + H_+^6\sigma_+^6\right)}\,, \tag{11}$$

$$\theta_- = \frac{\left(-4|B_-| - H_-^2\sigma_-^2(B_+ - B_-)\right)(H_-^2\sigma_-^2 + 2)^3}{2|B_-|(B_+ - B_-)\left(12 + 12H_-^2\sigma_-^2 + 5H_-^4\sigma_-^4 + H_-^6\sigma_-^6\right)}\,. \tag{12}$$

The main arguments used to derive these approximate formulas are now presented. They rest on Lemmas 3.1 to 3.3 given below.

Lemma 3.1. *Let us assume that H_0 holds. Then one has $X_k^4 = \mathcal{O}(1)$.*

Proof. Using (1) and some elementary properties of the Gaussian law it is immediate to conclude that $E[X_k^4] = (1+B_+\varepsilon)^4 E[X_{k-1}^4] + 6(1+B_+\varepsilon)^2\sigma_+^2\varepsilon E[X_{k-1}^2] + 3\sigma_+^4\varepsilon^2$. Applying Holder's inequality to the term on $E[X_{k-1}^2]$ the upper bound $E[X_k^4] \leq [(1+B_+\varepsilon)^4(1+|B_+|\varepsilon)]^k E[X_0^4] + \left(\frac{3}{|B_+|}+\varepsilon\right)\frac{3\sigma_+^4}{2|B_+|}$ is obtained since, as $(H6)$ holds, for ε small enough one has $1-(1+B_+\varepsilon)^4(1+|B_+|\varepsilon) \geq 2|B_+|\,\varepsilon$. The term on $E[X_0^4]$, representing the effect of the initial condition, vanishes exponentially with time given that $(1+B_+\varepsilon)^4(1+|B_+|\varepsilon) < 1$ (see also Amendola et al., 2006). □

Let us define a new process $\{\eta_k^+\}_k$ by

$$\eta_k^+ \overset{def}{=} \frac{1}{\varepsilon^2}\left(\frac{\Delta_k^{-2}}{\Gamma_-^2} - \frac{\Delta_k^{+2}}{\Gamma_+^2}\right) - \frac{H_+^2\sigma_+^2(B_+-B_-)^2}{2|B_+|\Gamma_-^2}\,. \tag{13}$$

In the sequel we will use the following notation for $\{\eta_k\}_k$: the upper script "+ " (respectively "-") in a variable means that the variable will be considered under hypothesis H_0 (respectively H_1). The test statistic S_n given in (7) can be easily rewritten depending on $\{\eta_k^+\}_k$, under the hypothesis that H_0 holds. On the other hand, one can derive a set of properties for $\{\eta_k^+\}_k$ that explain how this process behaves approximately. These are established in the next Lemma.

Lemma 3.2. *The following properties hold for process $\{\eta_k^+\}_k$:*

(i) $\eta_k^+ - \dfrac{H_+^2\sigma_+^2(B_+-B_-)^2}{4\Gamma_-^2}\varepsilon = \mathcal{O}(\varepsilon^2)$;

(ii) $var\,(\eta_k^+) = \dfrac{4(B_+-B_-)^2}{\Gamma_+^4\Gamma_-^4}\left(4+8H_+^2\sigma_+^2+5H_+^4\sigma_+^4+H_+^6\sigma_+^6\right)+\mathcal{O}(\sqrt{\varepsilon})$;

(iii) $\{\eta_{2k-1}^+\}_k$, $\{\eta_{2k}^+\}_k$ *both are approximately non correlated sequences;*

(iv) $\eta_{2k-1}^+\eta_{2k}^+ - \dfrac{8(B_+-B_-)^2}{\Gamma_+^4\Gamma_-^2} = \mathcal{O}(\varepsilon)$.

This means that the sequences $\{\eta_{2k-1}^+\}_k$, $\{\eta_{2k}^+\}_k$ behave approximately like mutually correlated noises with correlation vanishing with time lag.

Proof.

i. Since, under hypothesis H_0, Δ_k^+ has zero mean with variance $\Gamma_+^2\varepsilon$ then

$$E[\eta_k^+] = \frac{E[\Delta_k^{-2}]-\Gamma_-^2\varepsilon}{\Gamma_-^2}\frac{1}{\varepsilon^2} - \frac{H_+^2\sigma_+^2(B_+-B_-)^2}{2\Gamma_-^2|B_+|} \tag{14}$$

where $E[\Delta_k^{-2}]-\Gamma_-^2\varepsilon = H_+^2(B_+-B_-)^2\varepsilon^2\,E[X_k^2]$ as a consequence of (3) and (4). Using the fact that $E[X_k^2] = (1+B_+\varepsilon)^{2k}\,E[X_0^2]+\sigma_+^2\frac{1-(1+B_+\varepsilon)^{2k}}{|B_+|(2+B_+\varepsilon)}$ and some trivial calculations that bring in assumptions (HD) and $(H6)$, we obtain $E[\eta_k^+] = \frac{H_+^2(B_+-B_-)^2}{\Gamma_-^2}\left[(1+B_+\varepsilon)^{2k}\left(E[X_0^2]-\frac{\sigma_+^2}{|B_+|(2+B_+\varepsilon)}\right)+\frac{\sigma_+^2}{2(2+B_+\varepsilon)}\,\varepsilon\right]$. From this expression we derive the desired result since $(1+B_+\varepsilon)^{2k}$ vanishes exponentially fast. □

ii. Using the definition (13) of η_k^+ and (14) we find that

$$var\,(\eta_k^+) = \left\{E\left[\left(\frac{\Delta_k^{-2}-E[\Delta_k^{-2}]}{\Gamma_-^2}\right)^2\right]+E\left[\left(\frac{\Delta_k^{+2}-\Gamma_+^2\varepsilon}{\Gamma_+^2}\right)^2\right]\right.$$
$$\left.-2E\left[\frac{\Delta_k^{-2}-E[\Delta_k^{-2}]}{\Gamma_-^2}\,\frac{\Delta_k^{+2}-\Gamma_+^2\varepsilon}{\Gamma_+^2}\right]\right\}\frac{1}{\varepsilon^4}\,. \tag{15}$$

From (3), since assumption (H_2) holds and $\Delta_k^+\sim\mathcal{N}(0,\Gamma_+^2\varepsilon)$, we can write that

$$E\left[\left(\Delta_k^{-2}-E[\Delta_k^{-2}]\right)^2\right] = 2\left[H_+^4\sigma_+^4+1+(1+B_-\varepsilon)^4\right]\varepsilon^2$$
$$+4\left[H_+^2\sigma_+^2+H_+^2\sigma_+^2(1+B_-\varepsilon)^2+(1+B_-\varepsilon)^2\right]\varepsilon^2+\mathcal{O}(\varepsilon^3)$$
$$E\left[\left(\Delta_k^{+2}-\Gamma_+^2\varepsilon\right)^2\right] = 2\Gamma_+^4\varepsilon^2$$
$$E\left[\left(\Delta_k^{-2}-E[\Delta_k^{-2}]\right)\left(\Delta_k^{+2}-\Gamma_+^2\varepsilon\right)\right] = 2\left[H_+^4\sigma_+^4+1+(1+B_-\varepsilon)^2(1+B_+\varepsilon)^2\right]\varepsilon^2$$
$$+4\left[H_+^2\sigma_+^2+H_+^2\sigma_+^2(1+B_-\varepsilon)(1+B_+\varepsilon)+(1+B_-\varepsilon)(1+B_+\varepsilon)\right]\varepsilon^2+\mathcal{O}(\varepsilon^{5/2})\,.$$

The result comes out by replacing these expressions in (15) followed by a sequence of elementary calculations which use the fact that (*HD*) holds. □

iii. Applying the same arguments used before in the proof of item *ii.*, we can rewrite (13) as $\eta_k^+ - E\left[\eta_k^+\right] = \frac{1}{\Gamma_-^2\Gamma_+^2}\left[\left(\Gamma_+^2-\Gamma_-^2\right)\left(H_+^2\sigma_+^2(u_k^2-1)+(v_{k+1}^2-1)\right)+\left(-\Gamma_+^2(1+B_-\varepsilon)^2+\Gamma_-^2(1+B_+\varepsilon)^2\right)(v_k^2-1)\right]\frac{1}{\varepsilon}+\tau_k+\mathcal{O}(\varepsilon)$, with $\tau_k = \tau(u_k, v_k, v_{k+1}, X_k)$ being a random variable given by a sum of products all involving at least one odd power of the noises in such a way that its expected value is zero and τ_k, τ_{k+2} are almost uncorrelated. This expression highlights the fact that the correlation between η_k^+ and η_{k+2}^+ is of low order. □

iv. Applying the same arguments used before in the proof of item *iii.*, we find that $cov[\eta_k^+,\eta_{k+1}^+] = \frac{1}{\Gamma_-^4\Gamma_+^4}\left(\Gamma_+^2-\Gamma_-^2\right)\left[-\Gamma_+^2(1+B_-\varepsilon)^2+\Gamma_-^2(1+B_+\varepsilon)^2\right]E[(v_{k+1}^2-1)^2]\frac{1}{\varepsilon^2}+\mathcal{O}(\varepsilon)$. Using (5) and, as a consequence, the fact that $-\Gamma_+^2(1+B_-\varepsilon)^2+\Gamma_-^2(1+B_+\varepsilon)^2 = 2\Gamma_-^2(B_+-B_-)\varepsilon+\mathcal{O}(\varepsilon^2)$ ends the proof. □

Resorting to Lemmas 3.1 and 3.2, an approximation of the process $\{\varepsilon S_n\}_n$ by a diffusion process can be established. This is done in the following lemma:

Lemma 3.3. *Let us assume that H_0 holds, consider the statistic S_n given by (6) and take $t = \varepsilon n$. Then εS_n can be approximated by ζ_t as $n \to +\infty$, with $\{\zeta_t\}_t$ being the diffusion process solution of the stochastic differential equation*

$$d\zeta_t = \gamma_+^2\theta_+ dt + \sqrt{2\varepsilon}\,\gamma_+ dW_t^+\,,\quad \zeta_0 = 0 \tag{16}$$

where $\{W_t^+\}_t$ is a standard Wiener process under H_0, θ_+ is given by (11) and

$$\gamma_+ = \frac{|B_+-B_-|\sqrt{12+12H_+^2\sigma_+^2+5H_+^4\sigma_+^4+H_+^6\sigma_+^6}}{\sqrt{2}(H_+^2\sigma_+^2+2)^2}\,. \tag{17}$$

Proof. Let us define processes $W_t^o(n) = \frac{\sqrt{2\varepsilon}}{\sigma_{\eta^+}}\left(\eta_1^+ + \eta_3^+ + \cdots + \eta_{2[n/2]-1}^+\right)$ and $W_t^e(n) = \frac{\sqrt{2\varepsilon}}{\sigma_{\eta^+}}\left(\eta_2^+ + \eta_4^+ + \cdots + \eta_{2[n/2]}^+\right)$ with $\sigma_{\eta^+}^2 = var[\eta^+]$. As $n \to +\infty$ the following weak convergence holds (Billingsley, 1999, Theor. 14.1): $W_t^o(n) \longrightarrow W_t^o$ and $W_t^e(n) \longrightarrow W_t^e$ where $\{W_t^o\}_t$ and $\{W_t^e\}_t$ are Wiener processes. Then using Lemma 3.2 the process $W_t^+ = \frac{W_t^o+W_t^e}{C_+}$, with $C_+^2 = 2\left(1+\frac{16}{\sigma_{\eta^+}^2}\frac{(B_+-B_-)^2}{\Gamma_+^4\Gamma_-^2}\right)$, is a standard Wiener process. From (7) we obtain under H_0: $\varepsilon S_n = (n+1)\varepsilon\gamma_+^2\theta_+ + \sqrt{2\varepsilon}\,\gamma_+\frac{W_t^o(n)+W_t^e(n)}{C_+}$, with $\gamma_+ = \frac{1}{4}\sigma_{\eta^+}C_+$ and θ_+ given by (11). From here we derive the approximation (Fleming and Rishel, 2012, chap. V-Sec. 5). □

Proof of Proposition 3.1. The proof closely follows (Fleming et al., 1991)[p. 1183]. Take $p_+(z) = P_{\zeta_0=z}\{\zeta_{T_+^*} = -\varepsilon l_1\,|A_+\}$ where $T_+^* = \inf\{t\ :\ \zeta_t \le -\varepsilon l_1\ or\ \zeta_t \ge \varepsilon l_2\}$ is a stopping time. Applying Dynkin's formula (see details in Fleming and Rishel, 2012[chap.V-7.1]) the following ODE for $p_+(z)$ is obtained: $\begin{cases}\varepsilon p_+'' + \theta_+ p_+' = 0\\ p_+(-\varepsilon l_1) = 1\,,\ p_+(\varepsilon l_2) = 0\end{cases}$. The error probability that we were looking for, defined in (8), is $p_+ = p_+(0) = \frac{1-e^{-\theta_+ l_2}}{e^{\theta_+ l_1}-e^{-\theta_+ l_2}}$. A similar reasoning under hypothesis H_1 gives $\begin{cases}\varepsilon p_-'' + \theta_- p_-' = 0\\ p_-(-\varepsilon l_1) = 0\,,\ p_-(\varepsilon l_2) = 1\end{cases}$ and allows the computation of the error probability p_-. □

Results by Fleming and Rishel (2012) enable also the computation of the expected times to reach a decision under hypothesis H_0 and H_1, respectively:

$$E[T_+^*] = \varepsilon\left(\frac{l_2}{\gamma_+^2\theta_+} - p_+\frac{l_1+l_2}{\gamma_+^2\theta_+}\right)\,,\ E[T_-^*] = \varepsilon\left(-\frac{l_1}{\gamma_-^2\theta_-} - p_-\frac{l_1+l_2}{\gamma_-^2\theta_-}\right)\,. \tag{18}$$

Note that using the expressions of γ_- and (12) gives $\gamma_-^2\theta_- = \frac{-4|B_-|(B_+-B_-)-H_-^2\sigma_-^2(B_+-B_-)^2}{4|B_-|(H_-^2\sigma_-^2+2)}$. The formula for the expected time $E[T_+^*]$ can be derived by solving the ODE $\begin{cases}\varepsilon\gamma_+^2 g_+'' + \gamma_+^2\theta_+ g_+' + 1 = 0\\ g_+(-\varepsilon l_1) = g_+(\varepsilon l_2) = 0\end{cases}$, and taking $E[T_+^*] = g_+(0)$. A similar argument can be used to obtain the expression of $E[T_-^*]$.

4. Simulation experiments

Monte Carlo simulations are done in order to observe the behavior of the proposed test when deciding on H_0 against H_1. The simulations run for $M = 200$ Monte Carlo replications with the initial condition $X_0 \sim \mathcal{N}(-5; 0.1^2)$, and in the time interval [0, 100]. Illustrative examples are presented in Table 1 as well as the results obtained by applying the tests presented in the previous Sections with $p_+ = p_- = 5\%$. We also registered the percentage of monotony intervals that have been incorrectly detected in the simulations, denoted p_*.

Example 2 and 3 illustrate, respectively, the influence of the drift and of the signal to noise ratio (here represented by $|H_+|$ and $|H_-|$) on the test performance. We would intuitively expect a decrease in the waiting times when $|B_+ - B_-|$, $|B_+|$

Table 1
Numerical results of the tests (p_* = percentage of wrong decisions).

	Parameters	l_1, l_2	p_*	$E[T_-]$, $E[T_+]$
Example 1	$B_+ = -0.05, B_- = -1, H_+ = \sigma_+ = \sigma_- = 1, H_- = -1$	0.42, 1.10	4.8%	0.25, 0.17
Example 2	$B_+ = -0.25, B_- = -5, H_+ = \sigma_+ = \sigma_- = 1, H_- = -1$	2.14, 5.69	4.5%	1.27, 0.86
Example 3	$B_+ = -0.25, B_- = -1, H_+ = 4, \sigma_+ = 0.05, H_- = -2, \sigma_- = 1$	0.53, 0.60	5.1%	2.68, 2.16

and $|B_-|$ increase, an effect that we do not observe in the examples. The comparison of the results with those in Milheiro-Oliveira and Roubaud (1995) highlights the fact that the decision test proposed in this paper may have some advantage over those existing in the literature in that it may be faster in reaching a decision, depending on the particular problem that one wants to solve.

We should stress that although the derivation of formulas (9)–(12) involves a considerable amount of computations, implementation of the proposed test itself presents no real challenge, as the complexity of the algorithm is similar to that in Milheiro-Oliveira and Roubaud (1995) or in Fleming et al. (1991).

Acknowledgments

This research was partially supported by CMUP, Portugal (UID/MAT/00144/2019) funded by FCT, Portugal (Pt) with National and European structural funds through FEDER, under partnership agreement PT2020, and by project STRIDE (NORTE-01-0145-FEDER-000033) supported by NORTE2020. The author is very grateful to the reviewers for their anonymous contributions.

References

Amendola, A., Niglio, M., Vitale, C., 2006. The moments of SETARMA models. Statist. Probab. Lett. 76 (6), 625–633.
Bar-Shalom, Y., Li, X.-R., 1996. Estimation And Tracking: Principles, Techniques, And Software. Artech House.
Billingsley, P., 1999. Convergence Of Probability Measures, second ed. In: Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons Inc..
Brouste, A., Cai, C., Soltane, M., Wang, L., 2020. Testing for the change of the mean-reverting parameter of an autoregressive model with stationary Gaussian noise. Stat. Inference Stoch. Process. 23 (2), 301–318.
Fleming, W.H., Ji, D., Pardoux, E., 1988a. Piecewise linear filtering with small observation noise. In: Bensoussan, A., Lions, J.L. (Eds.), Anal. Optim. Syst.. Springer Berlin Heidelberg, pp. 725–739.
Fleming, W., Ji, D., Salame, P., Zhang, Q., 1988b. Piecewise Monotone Filtering in Discrete-Time with Small Observation Noise. Technical Report, Brown Univ, Providence, RI.
Fleming, W.H., Ji, D., Salame, P., Zhang, Q., 1991. Piecewise monotone filtering in discrete-time with small observation noise. IEEE Trans. Automat. Control 36 (10), 1181–1186.
Fleming, W.H., Pardoux, E., 1989. Piecewise monotone filtering with small observation noise. SIAM J. Control Optim. 27 (5), 1156–1181.
Fleming, W.H., Rishel, R.W., 2012. Deterministic And Stochastic Optimal Control. In: Stochastic Modelling and Applied Probability, Springer.
Fleming, W.H., Zhang, Q., 1991. Nonlinear filtering with small observation noise: piecewise monotone observations. In: Mayer-Wolf, E., Merzbach, E., Shwartz, A. (Eds.), Stoch. Anal.. Academic Press, pp. 153–168.
Fleming, W.H., Zhang, Q., 1992. Piecewise monotone filtering with small observation noise: Numerical simulations. In: Karatzas, I., Ocone, D. (Eds.), Appl. Stoch. Anal.. Springer, pp. 108–120.
Gelb, A., 1974. Applied Optimal Estimation. MIT Press.
Imani, M., Ghoreishi, S.F., 2021. Partially-observed discrete dynamical systems. In: 2021 Am. Control Conf.. pp. 310–315.
Kalman, R.E., Bucy, R.S., 1961. New results in linear filtering and prediction theory. J. Basic Eng. 83 (1), 95–108.
Liu, R.H., Zhang, Q., 2001. Nonlinear filtering: A hybrid approximation scheme. IEEE Trans. Aerosp. Electron. Syst. 37 (2), 470–480.
Milheiro-Oliveira, P.M., Roubaud, M.C., 1995. Discrete-time piecewise linear filtering with small observation noise. IEEE Trans. Automat. Control 40 (12), 2149–2152.
Pardoux, E., Roubaud, M.C., 1989. General piecewise linear filtering problems with small observation noise. In: Proc. 28th IEEE Conf. Decis. Control. pp. 232–235.
Roubaud, M.C., 1993. An approximate filter for a partially observed piecewise linear system with small observation noise. Systems Control Lett. 20 (3), 239–247.
Roubaud, M.C., 1995. Filtrage linéaire par morceaux avec petit bruit d'observation. Appl. Math. Optim. 32 (2), 163–194.
Tong, H., Yeung, I., 1991. On tests for self-exciting threshold autoregressive-type non-linearity in partially observed time series. J. R. Stat. Soc. Ser. C 40 (1), 43–62.
Wang, J.W., Zhang, Q., Yin, G., 2006. Two-time-scale hybrid filters: Near optimality. SIAM J. Control Optim. 45 (1), 298–319.
Yan, M., Zhu, Y., Jin, N., Bohg, J., 2020. Self-supervised learning of state estimation for manipulating deformable linear objects. IEEE Robot. Autom. Lett. 5 (2), 2372–2379.
Zhang, Q., 1998. Nonlinear filtering and control of a switching diffusion with small observation noise. SIAM J. Control Optim. 36 (5), 1638–1668.