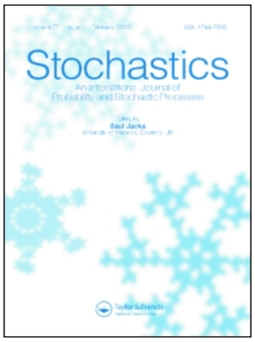


Stochastics

An International Journal of Probability and Stochastic Processes

ISSN: 1744-2508 (Print) 1744-2516 (Online) Journal homepage: http://www.tandfonline.com/loi/gssr20

Quickest detection in the Wiener disorder problem with post-change uncertainty

Heng Yang, Olympia Hadjiliadis & Michael Ludkovski

To cite this article: Heng Yang, Olympia Hadjiliadis & Michael Ludkovski (2017) Quickest detection in the Wiener disorder problem with post-change uncertainty, Stochastics, 89:3-4, 654-685, DOI: 10.1080/17442508.2016.1276908

To link to this article: https://doi.org/10.1080/17442508.2016.1276908

Published online: 09 Jan 2017.

STOCHASTICS, 2017
VOL. 89, NOS. 3–4, 654–685
http://dx.doi.org/10.1080/17442508.2016.1276908

Taylor & Francis
Taylor & Francis Group

Quickest detection in the Wiener disorder problem with post-change uncertainty

Heng Yang[a], Olympia Hadjiliadis[b,c] and Michael Ludkovski[d]

[a]Department of Mathematics, Graduate Center of City University of New York, New York, NY, USA; [b]Department of Mathematics and Computer science, Graduate Center of City University of New York, New York, NY, USA; [c]Department of Mathematics and Statistics, Hunter College, New York, NY, USA; [d]Department of Statistics and Applied Probability, University of California Santa Barbara, Santa Barbara, CA, USA

ABSTRACT

We consider the problem of quickest detection of an abrupt change when there is uncertainty about the post-change distribution. In particular, we examine this problem in the continuous-time Wiener model where the drift of observations changes from zero to a random drift with a prescribed discrete distribution. We set up the problem as a stochastic optimization in which the objective is to minimize a measure of detection delay subject to a constraint on frequency of false alarms. We design a novel composite stopping rule and prove that it is asymptotically optimal of third order under a weighted Lorden's criterion for detection delay. We also develop the strategy to identify the post-change drift and analyze the conditional identification error asymptotically. Our composite rules are based on CUSUM stopping times, as well as their reaction periods, namely the times between the last reset of the CUSUM statistic process and the CUSUM alarm. The established results shed new light on the performance of CUSUM strategies under model uncertainty and offer strong asymptotic optimality results in this framework.

ARTICLE HISTORY
Received 3 July 2015
Accepted 22 December 2016

KEYWORDS
Asymptotic optimality; Wiener disorder problem; change point detection; composite stopping time; CUSUM reaction period; equential analysis

AMS SUBJECT CLASSIFICATIONS
60G40; 62L10; 62L15; 62C20

1. Introduction

The detection problem is concerned with detecting a change in the statistical behaviour of sequential observations by balancing the trade-off between a small detection delay and frequency of false alarms. In the classical (non-Bayesian) formulation, the change point is treated as an unknown but fixed constant, and the post-change behaviour is assumed to be known. Under those assumptions, the resulting min–max problem is optimally solved by the Cumulative Sum (CUSUM) rule, see e.g. [17]. However, the performance of the CUSUM rule is highly sensitive to the signal strength [24]. This is a major practical challenge because typically the post-change signal is uncertain. For example, in radar systems (see [21]) one transmits a pulse, waiting for a potential return signal reflected from a target. One then must decide whether the observations imply the presence of such a target, which induces different signal strengths depending on its identity and properties. In a different context, quickest detection has been applied for identifying infectious disease

CONTACT Heng Yang hyang@gradcenter.cuny.edu

epidemics, where signal strength corresponds to the infectivity parameter which varies widely outbreak-to-outbreak (see [13,15] for example). Arguably, any realistic setup must consider some model uncertainty; this issue remains a major weakness of existing min–max models.

To date, there is scarce literature for designing optimal rules under uncertain signal strength in the non-Bayesian framework. The performance of the CUSUM stopping time in the detection problem with uncertainty is discussed in [27], where it is observed that the CUSUM stopping time is no longer second order asymptotically optimal. In this work, we contribute to this question by designing a *third-order asymptotically optimal* rule for a class of continuous-time quickest detection problems with uncertainty about the post-change distribution. Due to the difficulty to obtain an optimal solution for the complicated detection problem, many studies focus on the asymptotically optimal solutions which are very important from both theoretical and practical points of view (see [8,20] for example). Among the asymptotically optimal results, our work is able to achieve *third-order* asymptotic optimality (see [7]), that is the strongest optimality result to date in any change-point detection model with post-change uncertainty. Specifically, we work with Wiener observations, modelling the post change drift m via a given finite positive discrete distribution that is independent of pre-change observations. As a motivating example, we analyze the binary case, whereby the signal is either weak, represented by a small drift m_1, or strong, represented by a larger drift m_2.

Our main results are threefold. First, we develop a novel family of composite stopping times that combines multiple CUSUM policies along with the CUSUM reaction period (CRP). This family has several desirable behaviours and offers a flexible extension of the classical CUSUM framework. Second, we design and rigorously establish *third-order* asymptotic optimality for the problem of detecting a Wiener disorder with uncertain post-change drift. This is the strongest result to date in any model with post-change uncertainty. Third, we also analyze the question of distinguishing the different values of post-change drift and prove an asymptotic identification result for our composite stopping rules. In combination, these findings constitute a first step towards rigorous treatment of models with random post-change drifts, thereby extending the range of feasible applications of sequential detection.

Measurement of detection performance under model uncertainty is ambiguous. A worst-case analysis essentially reduces to considering the weakest possible signal strength [10]. This may not be the best approach in applications since it leads to increased detection delays in the typical scenario. At the same time, in many problems the decision maker has some idea about likely signal strengths so it is reasonable to specify a distribution for the post-change drift m (and furthermore reasonable to assume independence between the observed signal and the pre-change observations). In contrast, the *timing* of the signal is very difficult to model and the usual Bayesian formulation imposes strong independence assumptions on data vs. change-point that are likely to be violated. Motivated by these considerations, we propose a weighted Lorden's criterion [14] for detecting the unknown constant change point τ. Namely, our problem is that of finding a stopping time to minimize weighted detection delay subject to a frequency of false alarm constraint, with weights given by the probabilities of each of the post-change drifts.

The detection rules we investigate are compositions of CUSUM stopping times with specially chosen threshold parameters. The compositions are based on the CRP, which is defined as the time between the last reset of the CUSUM statistic process and the time at which the CUSUM stopping time draws an alarm. This statistic is related to last passage time distributions and has been studied in the literature (see, for instance, [12,19,28]) in various ways, including the use of the method of enlargement of filtration (see, for instance, [11,23]). We show that the CRP distribution is highly sensitive to post-change drift, offering a probabilistic result of independent interest that highlights new features of CUSUM stopping times.

Since we explicitly fix a (prior) distribution of the post-change drift, our work can be seen as a blend of min–max and Bayesian approaches. In the Bayesian framework, the case of uncertainty in the post-change drift in Wiener observations was considered in [2,3]. The case of uncertainty in post-change parameters has also been studied in Poisson observations in [1,16]. Mixed Wiener and Poisson observations are treated in [5]; see also the recent work on efficient numerical algorithms for the mixed problem in [15]. Our work extends these results to the more conservative/agnostic Lorden framework, while maintaining provable third order asymptotic optimality.

The rest of the paper is organized as follows. In Section 2, we set up the sequential detection problem mathematically and provide a criterion to measure detection delay in our setting. In Section 3, we construct the composite stopping time T_{com}. The main results are in Section 4, where we show that the composite T_{com} is asymptotically optimal of third order as the mean time to the first false alarm increases without bound. Section 4.2 then discusses the identification function associated with T_{com}. We also compare with other commonly used stopping times, such as the generalized likelihood ratio [25] and the mixture likelihood ratio [26] rules, to show that our composite stopping time can provide higher order asymptotic optimality. The details are given in Appendix 2. In Section 5, we use examples to discuss the performance of the composite stopping time. In Section 6, we generalize to the case of three or more values for the post-change drift. Section 7 contains the proofs of properties and facts related to the CRP. All the other proofs are given in Appendix 1.

2. Mathematical setup

2.1. The detection problem

We observe the process $\{Z_t\}_{t\geq 0}$ on a sample space $(\Omega, \mathcal{F})$ with the initial value $Z_0 = 0$. The distribution of the observations may undergo a disorder at the fixed but unknown change time τ.

Without any change point, which formally corresponds to $\tau = +\infty$, the observation process is a standard Brownian motion and its law is given by the Wiener measure P_∞. For any finite τ, we assume that the observation process changes from a standard Brownian motion to a Brownian motion with drift m; that is

$$\mathrm{d}Z_t := \begin{cases} \mathrm{d}W_t & t < \tau \\ m\,\mathrm{d}t + \mathrm{d}W_t & t \geq \tau. \end{cases} \tag{1}$$

The post-change drift m may take values in the finite collection $\{m_1, \ldots, m_N\}$ for some known constants $0 < m_1 < m_2 < \cdots < m_N$. The case of negative drifts can be addressed by similar arguments.

We assume that the probability space supports a uniform random variable U that is independent of $\{Z_t\}_{t=0}^{\infty}$ and define the filtration $\mathcal{G}_t = \sigma(U) \vee \sigma(\{Z_s\}_{s\leq t})$. Note that $\mathcal{G}_0 = \sigma(U)$. This extra enlargement of the natural filtration of Z is to enable randomization.

For each $i = 1, \ldots, N$, we introduce the family of measures $P_\tau^{m_i}$, $\tau \in [0, \infty)$, $i = 1, \ldots, N$, defined on this filtration, such that under $P_\tau^{m_i}$ the drift of Z is zero until τ and m_i for $t \geq \tau$. The uncertainty regarding m is modelled by a probability measure

$$P_\tau^m = \sum_i p_i P_\tau^{m_i},$$

where the weights p_i can be interpreted as the likelihood or relative importance of the case $m = m_i$ with $\sum p_i = 1$. In the special case $N = 2$, m can be viewed as coming from a Bernoulli distribution, taking the value m_1 with probability $p = p_1$ and the value $m_2 > m_1$ with probability $1 - p = p_2$.

Our basic goal is to detect the change point τ by finding a $\mathcal{G}$-stopping time T that balances the trade off between a small detection delay and the constraint on the frequency of false alarm. To this end, we need a measure of detection delay that takes into account the observation path $\{Z_t\}$ and the different values of the post-change drift.

For any $\mathcal{G}$-stopping stopping time T, we define the worst detection delay between the change time τ and its estimator T given the post-change drift $m = m_i$ for $i = 1, \ldots, N$ in the paradigm of Lorden [14]

$$J_i(T) := \sup_{\tau\geq 0} \operatorname*{esssup}_{\omega\in\mathcal{G}_\tau} E_\tau^{m_i}[(T-\tau)^+ | \mathcal{G}_\tau]. \tag{2}$$

Here, we take the essential supremum over all path up to time τ in $\mathcal{G}_\tau$ and take supremum over all possible change time τ. This is co-called 'worst detection delay'. Since m is unknown, we take the average over J_i's according to the weights p_i,

$$J(T) := \sum_{i=1}^{N} p_i J_i(T). \tag{3}$$

The choice $p_i = 1$ reduces to the post-change drift being a known constant m_i.

At the same time, when there is no change, $E_\infty[T]$ gives the mean time to the first false alarm of the $\mathcal{G}$-stopping time T. To control false alarms, we require $E_\infty[T] \geq \gamma$ for some (large) constant $\gamma > 0$. The quickest detection problem can now be represented via

$$\inf_{T\in\mathcal{T}_1} J(T) \quad \text{where} \quad \mathcal{T}_1 := \{\mathcal{G} - \text{stopping time } T\text{: } E_\infty[T] \geq \gamma\}. \tag{P}$$

As usual (see, e.g. [18]), the latter inequality constraint can be reduced to equality $E_\infty[T] = \gamma$. Since for any T with $E_\infty[T] > \gamma$, one may define a randomized rule $\hat{T}$ such that $E_\infty[\hat{T}] = \gamma$ and $\hat{T} \leq T$ whereby $J_i(\hat{T}) \leq J_i(T)$ (namely take $\hat{T} = T \cdot 1_{\{U<\gamma/E_\infty[T]\}}$ where $U \sim U(0,1)$ is independent).

2.2. Lower bound for detection delay

We use the big-O and small-o notations in the usual way, see Appendix 1 for details. Fixing the post-change drift as $m = m_i$, it is well-known that the CUSUM stopping time is optimal for the detection problem (see Theorem 6.12 in Poor and Hadjiliadis [22] for example). This generates the following lower bound on $J(T)$.

Lemma 2.1: *For any $\mathcal{G}$-stopping time $T \in \mathcal{T}_1(\gamma)$, we have the lower bound on the detection delay $J(T)$ as*

$$J(T) \geq LB(\gamma) := \sum_{i=1}^{N} \frac{2p_i}{m_i^2} g\left(f^{-1}\left(\frac{m_i^2\gamma}{2}\right)\right), \tag{4}$$

where

$$g(x) = e^{-x} + x - 1 \quad \text{and} \quad f(x) = e^x - x - 1, \quad x > 0. \tag{5}$$

Moreover, as $\gamma \to \infty$, we have

$$LB(\gamma) = \sum_{i=1}^{N} \frac{2p_i}{m_i^2}\left(\ln\gamma + \ln\frac{m_i^2}{2} - 1\right) + o_\gamma(1) \tag{6}$$

and thus,

$$\lim_{\gamma\to\infty} \frac{LB(\gamma)}{\ln\gamma} = \sum_{i=1}^{N} \frac{2p_i}{m_i^2}. \tag{7}$$

See Appendix 1 for the proof of Lemma 2.1.

3. Construction of T_{com}

In this section we introduce a class of composite stopping times to solve the problem (P). For notational clarity, we first present the case of $N = 2$.

3.1. CUSUM reaction period

We begin by recalling the definition of a CUSUM stopping time with tuning parameter λ as it appears in Hadjiliaids [9] and Hadjiliadis and Moustakides [10]. Consider a process ξ which is Brownian motion with drift M on a probability space $(\Omega, \mathcal{F}, Q)$. For any constants Λ and $K > 0$, a CUSUM stopping time with tuning parameter Λ is defined as

$$T(\xi, \Lambda, K) := \inf\left\{t \geq 0 : y_t \geq K\right\} \tag{8}$$

where the related CUSUM statistic process with tuning parameter Λ is

$$y_t := V_t - \inf_{s\leq t} V_s \quad \text{and} \quad V_t := \Lambda\xi_t - \frac{1}{2}\Lambda^2 t. \tag{9}$$

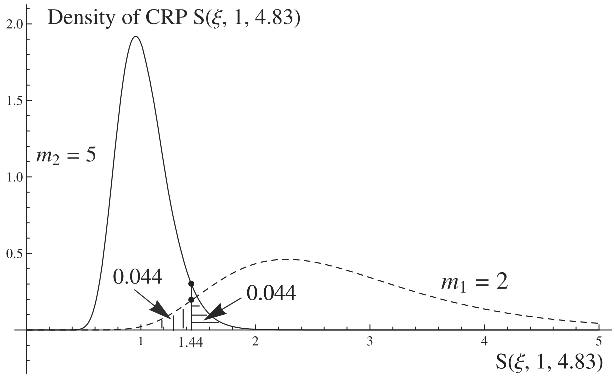


Figure 1. Distribution of the CRP. The solid curve gives the density function of CRP related to Brownian motion with drift 5; the dashed curve gives the density of CRP related to Brownian motion with drift 2.The tuning parameter of CUSUM is 1 and the threshold is 4.83.

Thus, $T(\xi,\Lambda,K)$ is announced as soon as the non-negative CUSUM statistic process y_t hits the threshold K. Corresponding to any $\mathcal{G}$-stopping time $T_K^\Lambda := T(\xi,\Lambda,K)$ of the CUSUM form, there is the last reset time ρ:

$$\rho(\xi,\Lambda,K) := \sup\left\{t \in [0, T_K^\Lambda) : V_t = \inf_{s\leq t} V_s\right\}. \tag{10}$$

The CRP of $T(\xi,\Lambda,K)$ is then defined as

$$S(\xi,\Lambda,K) := T(\xi,\Lambda,K) - \rho(\xi,\Lambda,K). \tag{11}$$

Introduced by Hadjiliadis and Zhang [28], the CRP measures the elapsed time between the last reset when the CUSUM process y_t was zero and the hitting time by y_t of K. Lemma 7.1 gives the explicit density of $S(\xi,\Lambda,K)$ for the above case of ξ being a Brownian motion with drift.

One property of CRP is shown in Figure 1 which illustrates the difference between the distributions of $S(\xi,\Lambda,K)$ for two processes $d\xi_1 = 5dt + dW_t$ and $d\xi_2 = 2dt + dW_t$ with parameters $\Lambda = 1$ and $K = 4.83$. The graph shows that the CRP distribution is highly sensitive to the drift of ξ. In the case of ξ_1, the CRP is likely to be small, and in the case of ξ_2 it is likely to be large. The threshold $b = 1.44$ determines the regions where the respective distribution densities cross-over. Thus, CRP may be used to distinguish different drifts of the observation process.

3.2. *Composite stopping time T_{com}*

We now design a composite CUSUM-based $\mathcal{G}$-stopping stopping time that involves the CRP.

The composite stopping time T_{com} is constructed in two stages. In the first stage, we apply the CUSUM stopping time defined in (8) denoted by

$$T_\nu^\lambda := T(Z,\lambda,\nu)$$

with the parameters $\lambda \in (0, 2m_1)$ and $\nu > 0$, where $Z := \{Z_t\}_{t\geq 0}$ is the observation path. The CRP of the first stage defined in (11) and the last reset time defined in (10) are denoted by

$$S_\nu := S(Z, \lambda, \nu) \quad \text{and} \quad \rho_\nu^\lambda := \rho(Z, \lambda, \nu). \tag{12}$$

Define the reset time shift function θ_s of the process of the observations for a time $s \geq 0$,

$$\theta_s(Z)(t) := Z_{t+s} - Z_s \quad \text{and} \quad \theta_s(Z) := \{Z_{t+s} - Z_s\}_{t\geq 0}.$$

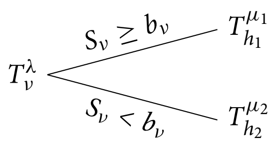
Note that θ_s makes the path re-start from zero at time s. We use θ to define the second-stage $\mathcal{G}$-stopping times $T_{h_1}^{\mu_1}$ and $T_{h_2}^{\mu_2}$ as follows:

$$T_{h_1}^{\mu_1} := T(\theta_{T_\nu^\lambda}(Z), \mu_1, h_1) \quad \text{and} \quad T_{h_2}^{\mu_2} := T(\theta_{T_\nu^\lambda}(Z), \mu_2, h_2), \tag{13}$$

where μ_i's are constants satisfying $0 < \mu_i < 2m_1$ and $h_i > 0$ are the second-stage thresholds for $i = 1, 2$. In the second stage, we apply one of the two stopping times $T_{h_1}^{\mu_1}$ or $T_{h_2}^{\mu_2}$, depending on the value of S_ν from the first stage. In particular, if $S_\nu \geq b_\nu$ for a parameter $b_\nu > 0$, we run the second stage $T_{h_1}^{\mu_1}$. On the other hand, if $S_\nu < b_\nu$, we run the second stage $T_{h_2}^{\mu_2}$. So the composite stopping time T_{com} is defined as

$$T_{\text{com}} := T_\nu^\lambda + \left(1_{\{S_\nu \geq b_\nu\}} T_{h_1}^{\mu_1} + 1_{\{S_\nu < b_\nu\}} T_{h_2}^{\mu_2}\right) = \begin{cases} T_\nu^\lambda + T_{h_1}^{\mu_1} & \text{if } \{S_\nu \geq b_\nu\} \\ T_\nu^\lambda + T_{h_2}^{\mu_2} & \text{if } \{S_\nu < b_\nu\}. \end{cases} \tag{14}$$

The diagram below illustrates the construction of T_{com}.



3.3. Detection delay of T_{com}

The next lemma provides the expressions for expected value of T_{com} under the measures P_∞ and $P_0^{m_i}$.

Lemma 3.1: *For the composite $\mathcal{G}$-stopping time T_{com} defined in (14), we have*

$$\begin{aligned} E_\infty\left[T_{\text{com}}\right] &= F(\lambda, \nu) + P_\infty(S_\nu \geq b_\nu) F(\mu_1, h_1) + P_\infty(S_\nu < b_\nu) F(\mu_2, h_2) \\ E_0^{m_i}[T_{\text{com}}] &= G_i(\lambda, \nu) + P_0^{m_i}(S_\nu \geq b_\nu) G_i(\mu_1, h_1) + P_0^{m_i}(S_\nu < b_\nu) G_i(\mu_2, h_2), \end{aligned} \tag{15}$$

where

$$F(x, y) = \frac{2}{x^2} f(y) \quad \text{and} \quad G_i(x, y) = \frac{2}{(2m_i - x)^2} g\left(\frac{2m_i - x}{x} y\right) \tag{16}$$

for $y > 0$ and f, g are in (5).

See Appendix 1 for the proof of Lemma 3.1.

From the results in Hadjiliadis and Moustakides [10] and Moustakides [18], it can be seen that the detection delay of T_{com} given $m = m_i$ satisfies

$$J_i(T_{\text{com}}) = E_0^{m_i}\left[T_{\text{com}}\right]. \tag{17}$$

Moreover, it is easy to see that $f(x) = e^x + O_x(x)$ on $(-\infty, \infty)$, $g(x) = x - 1 + o_x(1)$ on $(0, \infty)$ and $g(x) = e^{-x} + O_x(x)$ on $(-\infty, 0)$ as $x \to \infty$. It follows that:

$$\begin{aligned}
\lim_{y\to\infty} F(x,y)e^{-y} &= \frac{2}{x^2};\\
\lim_{y\to\infty} G_i(x,y)y^{-1} &= \frac{2}{(2m_i - x)x}, \quad \text{when} \quad \frac{2m_i - x}{x} > 0;\\
\lim_{y\to\infty} G_i(x,y)e^{\frac{2m_i - x}{x}y} &= \frac{2}{(2m_i - x)^2}, \quad \text{when} \quad \frac{2m_i - x}{x} < 0.
\end{aligned} \tag{18}$$

From Equations (5), (15) and (16), we obtain

$$E_\infty\left[T_{\text{com}}\right] = \frac{2}{\lambda^2}e^{\nu} + \frac{2P_\infty(S_\nu \geq b_\nu)}{\mu_1^2}e^{h_1} + \frac{2P_\infty(S_\nu < b_\nu)}{\mu_2^2}e^{h_2} + C(\nu, h_1, h_2), \tag{19}$$

where $C(\nu, h_1, h_2)$ is a linear function of ν, h_1, h_2. Similarly, for $0 < \lambda, \mu_1, \mu_2 < 2m_1$, from Equations (5), (15), (16) and (18), we obtain

$$\begin{aligned}
E_0^{m_i}\left[T_{\text{com}}\right] &= \frac{2}{\lambda(2m_i - \lambda)}\nu - \frac{2}{(2m_i - \lambda)^2}\\
&\quad + P_0^{m_i}(S_\nu \geq b_\nu)\left(\frac{2}{\mu_1(2m_i - \mu_1)}h_1 - \frac{2}{(2m_i - \mu_1)^2}\right)\\
&\quad + P_0^{m_i}(S_\nu < b_\nu)\left(\frac{2}{\mu_2(2m_i - \mu_2)}h_2 - \frac{2}{(2m_i - \mu_2)^2}\right) + c(\nu, h_1, h_2),
\end{aligned} \tag{20}$$

where $c(\nu, h_1, h_2)$ goes to zero as all three variables ν, h_1 and h_2 go to infinity.

In the sequel, we set

$$E_\infty[T_{\text{com}}] = \gamma, \tag{21}$$

and then proceed to discuss the asymptotic behaviour of the detection delay of T_{com} as γ goes to infinity.

From Lemma 3.1 and the results (19) and (20), we get the requirement on the tuning parameters, that is $0 < \mu_i < 2m_1$. To see this, suppose $\max(\nu, h_1, h_2) = h_1$ and $P_\infty(S_\nu \geq b_\nu) \neq 0$. From (19), we can see that the leading term is of order e^{h_1} and so $E_\infty[T_{\text{com}}] = \gamma$ translates into $h_1 = O(\ln \gamma)$ by properly choosing the parameters. If we choose $\mu_2 > 2m_1$, then from (18), $G_1(\mu_2, h_2) = O(e^{h_2(\mu_2 - 2m_1)/\mu_2})$, which may lead to $J(T_{\text{com}}) = O(\gamma)$ and make the delay far away from the lower bound as γ increases. For this reason, we must choose $0 < \mu_i < 2m_1$. For simplicity, we can choose $\mu_1 = \mu_2 = m_1$.

4. Asymptotic optimality of T_{com}

Our objective is to find an asymptotically optimal stopping time for the detection problem (P). We will establish asymptotic optimality of third order (see Fellouris and

Moustakides [7]) for T_{com} constructed above. In this section, we continue to consider the Bernoulli case $N = 2$. Recall that first order asymptotic optimality means that the ratio between the detection delay $J(T)$ and the lower bound $LB(\gamma)$ goes to 1 as $\gamma \to \infty$, while under second order asymptotic optimality the difference between $J(T)$ and $LB(\gamma)$ remains bounded in the same limit. Finally, the strongest third order asymptotic optimality means that the difference between $J(T)$ and $LB(\gamma)$ goes to zero as $\gamma \to \infty$.

4.1. Third order asymptotic optimality of T_{com} in (P)

We begin by describing the conditions on the parameters in the composite stopping time T_{com} that are required to achieve asymptotic optimality. The motivation is that we use CRP S_ν in the first stage as an indicator to whether there is a change or not. When S_ν is small, it is more likely that there is a change and so we would like the first stage to play an important role and do not need a long second stage; when S_ν is large, it is more likely that there is no change and so we need a long second stage to detect the change.

To rigorize this intuition, we need the following results under the no-change measure P_∞ and the measure $P_0^{m_i}$.

Lemma 4.1: *For parameters $\lambda, b_\nu, \nu > 0$ such that b_ν/ν is a positive constant, we have*

$$\lim_{\nu\to\infty} P_\infty(S_\nu < b_\nu) = 0. \tag{22}$$

Lemma 4.2: *For any parameter $\lambda \in (0, 2m_1)$ and $b_\nu, \nu > 0$ such that $b_\nu/\nu = l$ is a positive constant that satisfies*

$$l > \frac{2}{\lambda(2m_1 - \lambda)}, \tag{23}$$

there exists a positive constant $L = L(m_1, \lambda, l)$ such that

$$\lim_{\nu\to\infty} P_\tau^{m_1}\left(S_\nu \ge b_\nu \,\middle|\, \tau < \rho_\nu^\lambda\right) e^{-L\nu} = 0. \tag{24}$$

In particular, we have

$$\lim_{\nu\to\infty} P_0^{m_1}(S_\nu \ge b_\nu) e^{-L\nu} = 0. \tag{25}$$

See Section 7 for the proofs of Lemmas 4.1 and 4.2.

As a remark, the conditional event $\{\tau < \rho_\nu^\lambda\}$ is used to guarantee that the change has happened when the CRP begins, so that the whole time interval on which the CRP is recorded corresponds to the path with drift m_1 or m_2. This condition disappears asymptotically since as $\nu \to \infty$ we have $\rho_\nu^\lambda \to \infty$, while by assumption τ is a fixed constant. Consequently, in the asymptotic regime the condition $\{\tau < \rho_\nu^\lambda\}$ simply means that there exists a change in the lifetime of the observation process.

We notice that Lemma 4.2 also gives the behaviour of $P_\tau^{m_2}(S_\nu \ge b_\nu \mid \tau < \rho_\nu^\lambda)$ if we substitute m_1 with m_2. Thus, Lemmas 4.1 and 4.2 tell us that the value of S_ν can distinguish the two cases under no drift measure P_∞ and under the drift measure $P_0^{m_1}$ or $P_0^{m_2}$.

In order to ensure the third order asymptotic optimality of T_{com}, we will need to make appropriate choices of its parameters. From (19) and (21) , it can be seen that at least one of the thresholds ν, h_1, h_2 will go to infinity as γ goes to infinity. However, to achieve our purpose, we make the parameters ν, h_1, h_2 all go to infinity in the discussion that follows.

We also choose b_ν to be linear in ν, which is based on the fact that $E_0^{m_i}[S_\nu] = O(\nu)$ and $E_\infty[S_\nu] = O(\nu)$ as $\nu \to \infty$ (see Corollary 7.2 in Section 7). Moreover, we choose λ to satisfy

$$0 < \lambda < 2m_1. \tag{26}$$

And for simplicity, we choose $\mu_1 = \mu_2 = m_1$.

We now proceed to define ν, h_2 in terms of h_1. In particular, let $\nu = \nu(h_1)$ be a linear function of h_1 such that

$$\nu = \frac{p_1(m_2 - m_1)^2}{m_1^2 m_2^2} \frac{1}{\frac{p_1}{\lambda(2m_1-\lambda)} + \frac{p_2}{\lambda(2m_2-\lambda)}} h_1 + c \tag{27}$$

where

$$c = \frac{\sum_i \frac{p_i}{(2m_i-\lambda)^2} + \frac{p_2}{m_2^2} \ln \frac{m_2^2}{m_1^2} + \frac{p_2}{(2m_2-m_1)^2} - \frac{p_2}{m_2^2}}{\sum_i \frac{p_i}{\lambda(2m_i-\lambda)}}.$$

It is easy to see that the coefficient of h_1 in (27) is less than 1. This is because $(m_2 - m_1)/m_2 < 1$ and $\lambda(2m_1 - \lambda) \le m_1^2$.

Then by using λ and ν, we choose b_ν as a function of h_1 such that

$$l > \frac{2}{\lambda(2m_1 - \lambda)} \quad \text{and let} \quad b_\nu = l\nu. \tag{28}$$

We also choose the thresholds h_1, h_2 to satisfy the linear condition

$$h_2 = \frac{(2m_2 - m_1)m_1}{m_2^2} h_1. \tag{29}$$

It is easy to see that $(2m_2 - m_1)m_1/m_2^2 \le 1$ and so $h_2 \le h_1$. The value of h_1 is computed from the equation of the false alarm constraint

$$F(\lambda, \nu) + (1 - P_\infty(S_\nu < b_\nu))F(\mu_1, h_1) + P_\infty(S_\nu < b_\nu)F(\mu_2, h_2) = \gamma, \tag{30}$$

where the left hand side represents the average false alarm in (15), and the expression for $P_\infty(S_\nu < b_\nu)$ is given in Lemma 7.1. Note that the left hand side of (30) is a function of h_1 since we specified all of ν, b_ν, h_2 in terms of h_1.

From conditions (27) and (29), as $\gamma \to \infty$, Equation (30) tells us that all ν, h_1 and h_2 go to infinity.

Condition (27) means that the first stage will have a positive contribution to the detection delay. Condition (29) means that when S_ν is large, we run a second stage with a large threshold and so we need to wait more time to announce the change; when S_ν is small, we run the second stage with a small threshold and so T_{com} stops soon. The requirement in (28) comes from Lemmas 4.2 and 4.1. We can easily see that $P_0^{m_1}(S_\nu \ge b_\nu)$ and $P_0^{m_2}(S_\nu \ge b_\nu)$ both go to zero exponentially, while $P_\infty(S_\nu < b_\nu)$ goes to zero as $\gamma \to \infty$. Thus, condition (28) enables us to tell whether there is a change or not.

By the previous choices of parameters, we have the following results.

Lemma 4.3: *For any composite $\mathcal{G}$-stopping time T_{com} satisfying (21), with $\lambda, \mu_i \in (0, 2m_1)$, where the parameters ν, b_ν, h_1 and h_2 all go to infinity as $\gamma \to \infty$, and where b_ν/ν and h_2/h_1 are constants while $\nu/h_1 < 1, h_2/h_1 < 1$, we have*

$$h_1 = \ln \gamma - \ln \frac{2}{\mu_1^2} + C(\gamma). \tag{31}$$

Here, $C(\gamma)$ is a function that $C(\gamma) \to 0$ as $\gamma \to \infty$.

Theorem 4.4: *Let $\mathcal{R}_1(\gamma)$ be the family of composite $\mathcal{G}$-stopping times of the form T_{com} defined in (14), such that $\mu_1 = \mu_2 = m_1$ and the parameters λ, ν, b_ν, h_i satisfy (26)–(30), when the mean time to the first false alarm satisfies (21). Then, as $\gamma \to \infty$, any stopping time in $\mathcal{R}_1(\gamma)$ is asymptotically optimal of third order to detect the change-point in problem (P) in the sense that*

$$\lim_{\gamma \to \infty} \left[J(T_{\text{com}}) - LB(\gamma) \right] = 0, \tag{32}$$

where $J(T_{\text{com}})$ is the detection delay defined in (3); $LB(\gamma)$ is the lower bound of the detection delay given in Lemma 2.1.

See Appendix 1 for the proof.

Theorem 4.4 gives an asymptotically optimal stopping time of third order in the detection problem (P). We provide a powerful stopping time in the detection problem, such that the difference between the resulting detection delay and the lower bound of the detection delay is close to zero once the average false alarm is large enough. However, the theorem does not provide the non-asymptotic guarantees.

4.2. An identification function

In the previous subsection, we choose the parameters in the composite stopping time to make T_{com} be asymptotically optimal of third order in the problem (P). If we additionally want to identify the post-change distribution, it is possible to construct an identification function $\delta_{T_{\text{com}}} \in \mathcal{G}_{T_{\text{com}}}$ taking values in $\{m_1, m_2\}$ to serve the purpose of post-change identification of the drift.

To this end, we employ the CRP of the second stage $T_{h_2}^{\mu_2}$

$$S_{h_2} := S(Z, \mu_2, h_2)$$

and recall S_ν as the CRP of the first stage. The identification function $\delta_{T_{\text{com}}}$ is defined as follows

$$\delta_{T_{\text{com}}} := \begin{cases} m_1 & \text{if } \{S_\nu \geq b_\nu\} \cup \{S_{h_2} \geq b_{h_2}, S_\nu < b_\nu\}; \\ m_2 & \text{if } \{S_{h_2} < b_{h_2}, S_\nu < b_\nu\}. \end{cases} \tag{33}$$

for positive constants b_ν and b_{h_2}.

The idea of the identification function is that we use the values of the CRPs of both stages together as an indicator to whether the post-change drift is large or small. If we do not have a strong reason to claim that the post-change drift is m_2, then we say that the post-change drift is m_1. More specifically, we consider the cases under the measures $P_\tau^{m_i}(\,\cdot \mid \tau < \rho_\nu^\lambda)$. In such cases, both S_{h_2} and S_ν are related to the observation path with a constant drift m_i. When S_ν is large, we say there is no change based on Lemma 4.1.

Otherwise, when S_{h_2} is large, which corresponds to a slow CUSUM reaction in the second stage, we have an indication of a small post-change drift. Conversely, when S_{h_2} is small, we have an indication of a large post-change drift.

We notice that the measure $P_0^{m_i}$ is a special case in which the change occurs before the last reset of the first stage CUSUM statistic process. It is easy to see that under both $P_0^{m_i}$ and $P_\tau^{m_i}(\,\cdot \mid \tau < \rho_\nu^\lambda)$, the observation path has exactly the same drift at the time that the CRP S_ν starts being recorded. This is also true for S_{h_2}.

We use $\rho_{h_2}^{\mu_2}$ to represent the last reset of the second stage CUSUM statistic process. Then we have the following results concerning the exponential decay of the conditional identification errors.

Lemma 4.5: *For any parameter $\mu_2 \in (0, 2m_1)$, when we choose $b_{h_2}/h_2 = l$ to be a positive constant that satisfies*

$$l < \frac{2}{\mu_2(2m_1 - \mu_2)}, \tag{34}$$

there exists a positive constant $L_1 = L_1(m_1, \mu_2, l)$ such that

$$\lim_{h_2\to\infty} P_\tau^{m_1}\left(S_{h_2} < b_{h_2} \,\middle|\, \tau < \rho_{h_2}^{\mu_2}\right) e^{-L_1 h_2} = 0. \tag{35}$$

Lemma 4.6: *For any parameter $\mu_2 \in (0, 2m_2)$, when we choose $b_{h_2}/h_2 = l$ to be a positive constant that satisfies*

$$l > \frac{2}{\mu_2(2m_2 - \mu_2)}, \tag{36}$$

there exists a positive constant $L_2 = L_2(m_2, \mu_2, l)$ such that

$$\lim_{h_2\to\infty} P_\tau^{m_2}\left(S_{h_2} \geq b_{h_2} \,\middle|\, \tau < \rho_{h_2}^{\mu_2}\right) e^{-L_2 h_2} = 0. \tag{37}$$

See Section 7 for the proofs of Lemmas 4.5 and 4.6.

As a remark, the purpose of the condition $\{\tau < \rho_{h_2}^{\mu_2}\}$ is to guarantee that the second-stage CRP is recorded on the path that has a drift m_i. In other words, we consider the case that the change happens before the last reset of the second-stage CUSUM statistic process. Thus, Lemmas 4.5 and 4.6 still hold if the condition $\{\tau < \rho_{h_2}^{\mu_2}\}$ is substituted with $\{\tau < \rho_\nu^\lambda\}$ since $\rho_\nu^\lambda < \rho_{h_2}^{\mu_2}$.

Based on Lemmas 4.5 and 4.6, for simplicity, we choose $\mu_2 = m_1$ to guarantee the existence of the solution of the inequalities (34) and (36).

In Theorem 4.4, we already have the conditions on the parameters $\lambda, \nu, b_\nu, h_1, h_2$. We still need to choose parameter b_{h_2} in the identification function $\delta_{T_{\text{com}}}$. To satisfy the conditions in Lemmas 4.5 and 4.6, we choose a constant l such that

$$\frac{2}{m_1(2m_2 - m_1)} < l < \frac{2}{m_1^2} \quad \text{and let} \quad b_{h_2} = l \cdot h_2. \tag{38}$$

Proposition 4.7: *Let $T_{\text{com}} \in \mathcal{R}_1(\gamma)$ be a composite $\mathcal{G}$-stopping time in Theorem 4.4, and $\delta_{T_{\text{com}}}$ be the associated identification function defined in (33), such that $\mu_1 = \mu_2 = m_1$ and the parameters $\lambda, \nu, b_\nu, b_{h_2}, h_i$ satisfy (26)–(30) and (38). Then, we obtain*

$$\lim_{\gamma\to\infty} P_\tau^{m_i}\left(\delta_{T_{\rm com}} \neq m_i \middle|\, \tau < \rho_\nu^\lambda\right) = 0 \quad \textit{for } i = 1, 2, \tag{39}$$

where ρ_ν^λ is the last reset of the first-stage CUSUM statistic process in (12).

See Appendix 1 for the proof.

As a remark, when τ is a finite constant, the probability of conditional event in (39) goes to 1 as γ increases, i.e. $P_\tau^{m_i}(\tau < \rho_\nu^\lambda) \to 1$ as $\gamma \to \infty$, since the last reset time ρ_ν^λ also goes to infinity. Thus, (39) is equivalent to

$$\lim_{\gamma\to\infty} P_\tau^{m_i}\left(\delta_{T_{\rm com}} \neq m_i \middle|\, \tau < \infty\right) = 0 \quad \text{for } i = 1, 2. \tag{40}$$

Proposition 4.7 gives the performance of the identification function $\delta_{T_{\rm com}}$ under the case that the change happens before the last reset of the first-stage CUSUM statistic process. It provides a way to make an estimate of the post-change drift, with arbitrarily small conditional identification errors as γ grows in the case that the change point has happened before the last reset of the first-stage CUSUM statistic process.

This identification function in fact comes from the construction of the stopping time. But $\delta_{T_{\rm com}}$ may not be the best statistical estimator of the post-change drift alone. Our primary purpose here is to minimize the detection delay of the change point and Proposition 4.7 shows that our composite stopping time can provide additional information about post-change identification with arbitrarily small conditional errors in special cases.

4.3. Comparing $T_{\rm com}$ with other detection rules

In this subsection, we compare the composite stopping time $T_{\rm com}$ with alternative detection rules for the problem (P).

4.3.1. A randomized composite stopping time

In the definition of the composite stopping time $T_{\rm com}$ in (14), instead of using the CRP, we may consider any event A, such that A is known before the second stage. The complement set of A is denoted by A^c. By the strong Markov property of Z, the stopping times $T_{h_1}^{\mu_1}$ and $T_{h_2}^{\mu_2}$ are independent of A.

For example, we may take $A = \{U \leq q\}$, where U is an independent uniform $U(0, 1)$ random variable and q is a constant. Thus, U is a randomization parameter (a 'coin toss') which determines which of $T_{h_1}^{\mu_1}$ and $T_{h_2}^{\mu_2}$ we use in the second stage. In fact, in such an example, T_ν^λ is not necessary.

Thus, we can define a randomized composite stopping time T_{ran} as

$$T_{ran} = 1_A T_{h_1}^{\mu_1} + 1_{A^c} T_{h_2}^{\mu_2}, \tag{41}$$

where 1_A and 1_{A^c} are indicator functions.

To choose its parameters, set $\mu_1 = m_1$. By giving a requirement on μ_2 as

$$\frac{2m_1 m_2}{m_1 + m_2} < \mu_2 < 2m_1 \tag{42}$$

and the fact that $m_2^2/(2m_2m_1 - m_1^2) \geq 1$, we can define a constant value q that is between 0 and 1:

$$q := \frac{1 - \frac{(2m_1-\mu_2)m_2^2}{(2m_2-\mu_2)m_1^2}}{\frac{m_2^2}{(2m_2-m_1)m_1} - \frac{(2m_1-\mu_2)m_2^2}{(2m_2-\mu_2)m_1^2}}. \tag{43}$$

Now, we pick the event A to satisfy

$$P_\infty(A) = P_0^{m_1}(A) = P_0^{m_2}(A) = q. \tag{44}$$

Let h_1, h_2 go to infinity as $\gamma \to \infty$. We take h_2 to be a linear function in h_1 with a ratio

$$h_2 = \frac{(2m_1 - \mu_2)\mu_2}{m_1^2} h_1 + c \tag{45}$$

where

$$c = \frac{\sum_i p_i \left(\frac{2q}{(2m_i-m_1)^2} + \frac{2(1-q)}{(2m_i-\mu_2)^2}\right) + \sum_i \frac{2p_i}{m_i^2}\left(\ln\frac{m_i^2}{2} - 1\right) - \ln\frac{m_1^2}{2q}\sum_i \frac{2p_i}{m_i^2}}{\sum_i \frac{2p_i(1-q)}{(2m_i-\mu_2)\mu_2}}.$$

It is easy to check that the coefficient is less than 1. So for γ large enough, we have $h_2 < h_1$.

Proposition 4.8: *Let $\mathcal{R}_3(\gamma)$ be the family of randomized composite stopping times of the form T_{ran} defined in (41), such that $\mu_1 = m_1$ and μ_2, h_1, h_2, A satisfy (42)–(45) and $E_\infty[T_{ran}] = \gamma$. Then, as $\gamma \to \infty$, any stopping time in $\mathcal{R}_3(\gamma)$ is asymptotically optimal of third order in problem (P) in the sense that*

$$\lim_{\gamma\to\infty} \left[J(T_{ran}) - LB(\gamma)\right] = 0. \tag{46}$$

Comparing to the result in Theorem 4.7, the randomized stopping time T_{ran} can not provide the identification function of the post-change drift to satisfy the constraint. Moreover, although the definition of T_{ran} only involves a single CUSUM alarm, it is not necessary that T_{ran} has a smaller detection delay than $T_{\rm com}$ when they share the same false alarm constraint. In the construction of T_{ran} we have two stopping times $T_{h_1}^{\mu_1}$ and $T_{h_2}^{\mu_2}$. With the same false alarm rate γ, the value of the threshold h_1 in T_{ran} is more likely to be larger than that in $T_{\rm com}$ (this still depends on the values of other parameters and constants), which may lead to a longer time to detect the change.

Such a randomized stopping time T_{ran} is in fact different from $T_{\rm com}$. Because when $A = \{U \leq q\}$, we have $P_0^{m_1}(A) = P_0^{m_2}(A) = P_\infty(A)$. On the other hand, for the CRP S we have $P_0^{m_1}(S \geq a) \neq P_0^{m_2}(S \geq a) \neq P_\infty(S \geq a)$, which provides a method to identify the value of post-change drift.

4.3.2. A generalized likelihood ratio stopping time

The λ-CUSUM stopping time with tuning parameter m_i is based on the maximum likelihood ratio statistic process

$$y_t^{(i)} := \max_{s\leq t} \log \frac{dP_s^{m_i}}{dP_\infty}\Big|\mathcal{G}_t = m_iZ_t - \frac{1}{2}m_i^2 t - \inf_{s\leq t}\left(m_iZ_s - \frac{1}{2}m_i^2 s\right). \tag{47}$$

We may also consider a stopping time related to the generalized likelihood ratio statistic. Discussion on the generalized likelihood ratio statistic in the detection problem can be found in Siegmund and Venkatraman [25].

In this problem, we have the post-change drift to be either m_1 or m_2, with a null hypothesis that there is no change. So the generalized likelihood ratio statistic corresponds to the process of the form

$$y_t^{\max} := \max\{y_t^{(1)}, y_t^{(2)}\}. \tag{48}$$

This leads to the stopping time

$$T_k^{\max} := \inf\{t \geq 0 : \ y_t^{\max} \geq k\} \tag{49}$$

for a positive constant k. Since $\{\max\{y_t^{(1)}, y_t^{(2)}\} > k\} = \{y_t^{(1)} > k\} \cup \{y_t^{(2)} > k\}$, it is easy to see that $T_k^{\max}$ is the minimum of two CUSUM stopping times:

$$T_k^{\max} = T_1^c \wedge T_2^c$$

where T_1^c and T_2^c are the CUSUM stopping times with respect to $y^{(1)}$ and $y^{(2)}$ separately and with the same threshold k.

For each observation path, such a stopping time is stopped at either T_1^c or T_2^c. The minimum stopping time provides a separation on the whole path space. One can show that $T_k^{\max}$ is not third-order asymptotically optimal. In particular, if $m_2 < 2m_1$, $T_k^{\max}$ is not even second-order asymptotically optimal. See Appendix 2 for details.

4.3.3. A mixture of likelihood ratios stopping time

Instead of the statistic (48), we may consider another statistic as the mixture of the likelihood ratio

$$y_t^{\text{mix}} = p_1 e^{m_1 Z_t - \frac{1}{2} m_1^2 t - \inf_{s \leq t}\left(m_1 Z_s - \frac{1}{2} m_1^2 s\right)} + p_2 e^{m_2 Z_t - \frac{1}{2} m_2^2 t - \inf_{s \leq t}\left(m_2 Z_s - \frac{1}{2} m_2^2 s\right)}.$$

And define a stopping time

$$T_d^{\text{mix}} := \inf\{t \geq 0 : \ y_t^{\text{mix}} \geq e^d\} \tag{50}$$

for a positive constant d.

The statistic y^{mix} is the linear combination of exponential form of reflected Brownian motions with different drift and diffusion parameters. Unfortunately, it is hard to represent the explicit expressions of its expectations under measures $P_0^{m_i}$ and P_∞.

Similar to the generalized likelihood ratio stopping time, we can see that T_d^{mix} is not third-order asymptotically optimal in general. In particular, if $m_2 < 2m_1$, T_d^{mix} is not even second-order asymptotically optimal. See Appendix 2 for details.

5. Numerical illustration

We present an example to illustrate the idea of the composite stopping time T_{com} and to see its performance. Theorem 4.7 tells the asymptotic behaviour of the stopping time as the time to first false alarm increases without bound, assuming the conditions on the

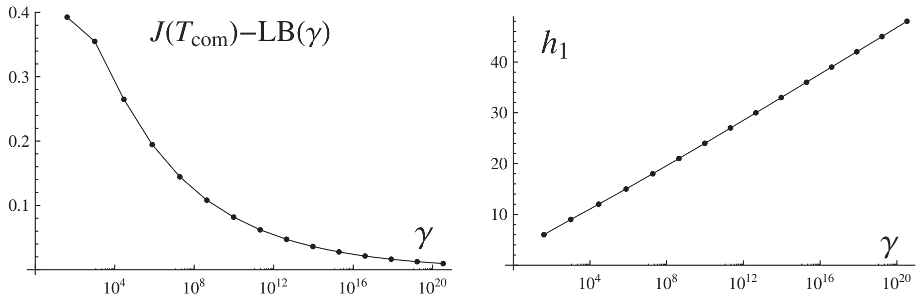


Figure 2. Performance of the composite rule in the case $m_1 = 2, m_2 = 5, p = 0.4$. Left: the difference between detection delay and lower bound; Right: the threshold h_1. Both graphs are plotted against λ with a logarithmic x-axis.

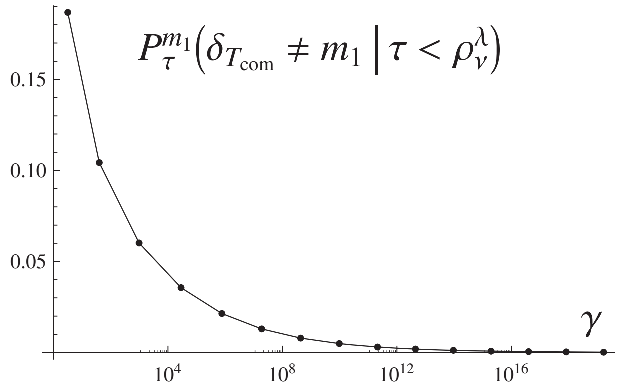


Figure 3. Performance of $\delta_{T_{\text{com}}}$ in the case $m_1 = 2, m_2 = 5, p = 0.4$ under measure $P_\tau^{m_1}$. This graph is plotted against λ with a logarithmic x-axis. The graph of performance of $\delta_{T_{\text{com}}}$ under measure $P_\tau^{m_2}$ has a similar shape.

parameters such as (26) and (28). Since these conditions only specify acceptable ranges, there remains scope for further fine-tuning to optimize performance.

In Figures 2 and 3, we consider the case $m_1 = 2$, $m_2 = 5$ and $p = 0.4$, when changing the value of γ. To evaluate the stopping time in Theorem 4.7, first we select the parameter λ. To avoid the situation of stopping too fast, it is important to guarantee that ν is not too small compared to λ. Thus, we prefer a small value of λ. In this example, we fix $\lambda = m_1/10$. Based on the conditions (27) and (29), the thresholds ν and h_2 are linear functions of h_1. Next, we need to decide the CRP threshold b_ν to satisfy (28). In this example, we choose $b_\nu = 4/(2m_1\lambda - \lambda^2)\nu$. Then, representing h_2, ν, b_ν as functions of h_1, we can solve Equation (30) to get the values of h_1 given γ. For computational convenience, in this example, in fact we choose h_1 as simple integers and then γ is the value that satisfies (30). The parameter b_{h_2} is chosen by equalizing the probabilities $P_\tau^{m_1}(S_{h_2} < b_{h_2} | \tau < \rho_\nu^\lambda) = P_\tau^{m_2}(S_{h_2} > b_{h_2} | \tau < \rho_\nu^\lambda)$.

From Figures 2, 3 and Table 1, we can see that the composite stopping time T_{com} in this example provides good behaviours in both detection and identification. The difference between the detection delay of T_{com} and the lower bound goes to zero, as γ increases and the identification metrics also quickly shrink as γ increases.

Table 1. Example choices of parameters in the case $m_1 = 2, m_2 = 5, p = 0.4$. Here, we have $\lambda = 0.2$, $h_2 = 0.64h_1$. J refers to detection delay $J(T_{\mathrm{com}})$; Diff refers to the difference $J(T_{\mathrm{com}}) - LB(\gamma)$; Rat refers to $J(T_{\mathrm{com}})/LB(\gamma)$; Err1 refers to $P_\tau^{m_1}\left(\delta_{T_{\mathrm{com}}} \neq m_1 \middle| \tau < \rho_\nu^\lambda\right)$; Err2 refers to $P_\tau^{m_2}\left(\delta_{T_{\mathrm{com}}} \neq m_2 \middle| \tau < \rho_\nu^\lambda\right)$.

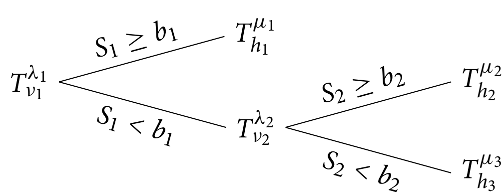
γ	ν	b_ν	h_1	J	LB	$Diff$	Rat	$Err1$	$Err2$
4×10	0.34	1.77	6.0	1.34	0.94	0.39	1.417	0.1868	0.1881
9.8×10^2	0.47	2.45	9.0	2.08	1.72	0.35	1.206	0.1043	0.1051
2.8×10^4	0.6	3.13	12.0	2.82	2.56	0.26	1.104	0.0602	0.0607
7.6×10^5	0.72	3.81	15.0	3.57	3.37	0.19	1.058	0.0356	0.0358
1.9×10^7	0.85	4.5	18.0	4.31	4.17	0.14	1.035	0.0214	0.0215
4.3×10^8	0.98	5.18	21.0	5.05	4.94	0.11	1.022	0.0130	0.0130
9.7×10^9	1.11	5.86	24.0	5.8	5.71	0.08	1.014	0.0079	0.0080
2.1×10^{11}	1.24	6.55	27.0	6.54	6.48	0.06	1.010	0.0049	0.0049
4.5×10^{12}	1.37	7.23	30.0	7.28	7.24	0.05	1.007	0.0030	0.0030
9.4×10^{13}	1.5	7.91	33.0	8.03	7.99	0.04	1.005	0.0019	0.0019
1.9×10^{15}	1.63	8.59	36.0	8.77	8.74	0.03	1.003	0.0012	0.0012
4.0×10^{16}	1.76	9.28	39.0	9.51	9.49	0.02	1.002	0.0007	0.0007
8.2×10^{17}	1.89	9.96	42.0	10.26	10.24	0.02	1.002	0.0005	0.0005
1.7×10^{19}	2.02	10.64	45.0	11.0	10.99	0.01	1.001	0.0003	0.0003
3.4×10^{20}	2.15	11.33	48.0	11.74	11.74	0.01	1.001	0.0002	0.0002

6. Generalizing post-change drift uncertainty

The idea of the composite stopping time in Theorem 4.4 and Proposition 4.7 can be extended to the situation where the post-change drift m takes on more than two values. For simplicity, we discuss in this section the case $N = 3$ so that m is a random variable taking three values $0 < m_1 < m_2 < m_3$.

The idea of quickest detection for such m is to construct a composite stopping time $T_{\mathrm{com}}^{(3)}$ with (up to) 3 stages. We combine two composite stopping times via a binary decision tree and choose the parameters in a backward fashion. The separation in the last (second) composite stopping time is used to distinguish the cases $\{m = m_2\}$ and $\{m = m_3\}$ when there is a change. The separation in the second-to-last (first) composite stopping time is used to distinguish the cases $\{m = m_1\}$ and $\{m > m_1\}$ when there is a change. The identification is done using an appropriate CRP criterion as before. In fact, we design the first node in the tree to help balance the false alarm rate, and ensure its frequency will go to zero when there is a change.

The following diagram illustrates the construction of the stopping time $T_{\mathrm{com}}^{(3)}$. It should be clear that this approach can be extended further to any finite N.



More precisely, we first employ a $\mathcal{G}$-stopping time $T_{\nu_1}^{\lambda_1} := T(Z, \lambda_1, \nu_1)$ of the CUSUM form in (8) with parameters $0 < \lambda_1 < 2m_1$ and $\nu_1 > 0$ and the observation path $Z := \{Z_t\}_{t \geq 0}$. After the first stage, we have the CRP associated with $T_{\nu_1}^{\lambda_1}$ as defined in (11) and the last reset time defined in (10), that are denoted by

$$S_1 := S(Z, \lambda_1, \nu_1) \quad \text{and} \quad \rho_{\nu_1}^{\lambda_1} := \rho(Z, \lambda_1, \nu_1). \tag{51}$$

We choose a threshold $b_1 > 0$ to distinguish the measures P_∞ and $P_0^{m_1}$ as we used in $N = 2$ case.

On the event $\{S_1 \geq b_1\}$, we employ

$$T_{h_1}^{\mu_1} := T\left(\theta_{T_{\nu_1}^{\lambda_1}}(Z), \mu_1, h_1\right),$$

with the parameter $h_1 > 0$ and which helps to do one more stage detection based on the information that the first stage CUSUM statistic process increases slowly.

If $\{S_1 < b_1\}$, assuming that there is a change, we continue to another stage to test between $\{m < m_1\}$ and $\{m \geq m_1\}$ via

$$T_{\nu_2}^{\lambda_2} := T\left(\theta_{T_{\nu_1}^{\lambda_1}}(Z), \lambda_2, \nu_2\right)$$

with parameters $0 < \lambda_2 < 2m_2$ and $\nu_2 > 0$. This stage yields another CRP $S_2 := S_{\nu_2}^{\lambda_2}$. We choose a threshold $b_2 > 0$ to make a condition on S_2. In the case that $\{S_2 \geq b_2\}$, we employ

$$T_{h_2}^{\mu_2} := T\left(\theta_{T_{\nu_1}^{\lambda_1} + T_{\nu_2}^{\lambda_2}}(Z), \mu_2, h_2\right);$$

alternatively if $\{S_2 < b_2\}$, we employ

$$T_{h_3}^{\mu_3} := T\left(\theta_{T_{\nu_1}^{\lambda_1} + T_{\nu_2}^{\lambda_2}}(Z), \mu_3, h_3\right).$$

For the same reason as that of T_{com}, we require $0 < \mu_i < 2m_1$ for $i = 1, 2, 3$. We can choose $\mu_1 = \mu_2 = \mu_3$ for convenience.

Overall, the composite stopping time $T_{\text{com}}^{(3)}$ is

$$T_{\text{com}}^{(3)} := \begin{cases} T_{\nu_1}^{\lambda_1} + T_{h_1}^{\mu_1} & \text{on } B_1 := \{S_1 \geq b_1\}; \\ T_{\nu_1}^{\lambda_1} + T_{\nu_2}^{\lambda_2} + T_{h_2}^{\mu_2} & \text{on } B_2 := \{S_1 < b_1, S_2 \geq b_2\}; \\ T_{\nu_1}^{\lambda_1} + T_{\nu_2}^{\lambda_2} + T_{h_3}^{\mu_3} & \text{on } B_3 := \{S_1 < b_1, S_2 < b_2\}. \end{cases} \tag{52}$$

It is easy to see that three stages $\{T_{\nu_1}^{\lambda_1}, S_1\}$, $\{T_{h_1}^{\mu_1}, T_{\nu_2}^{\lambda_2}, S_2\}$ and $\{T_{h_2}^{\mu_2}, T_{h_3}^{\mu_3}\}$ are independent of each other. Moreover, the above construction also gives an identification stopping time according to

$$\delta^{(3)} := m_i 1_{A_i} \in \mathcal{G}_{T_{\text{com}}^{(3)}}, \tag{53}$$

based on the partition generated by the A_i's for $i = 1, 2, 3$ as follows

$$\begin{cases} A_1 = \{S_1 \geq b_1\} \cup \{S_1 < b_1, S_2 \geq b_2, S_{h_2} \geq b_{h_2}\}; \\ A_2 = \{S_1 < b_1, S_2 \geq b_2, S_{h_2} < b_{h_2}\} \cup \{S_1 < b_1, S_2 < b_2, S_{h_3} \geq b_{h_3}\}; \\ A_3 = \{S_1 < b_1, S_2 < b_2, S_{h_3} < b_{h_3}\}, \end{cases} \tag{54}$$

where b_{h_2} and b_{h_3} are positive constant parameters.

To define the stopping time $T^{(3)}_{\text{com}}$ we need to specify the parameters. For simplicity, we always choose $\mu_1 = \mu_2 = \mu_3 = m_1$. And we choose $\lambda_1 = \lambda_2 := \lambda$ to satisfy

$$0 < \lambda < 2m_1. \tag{55}$$

We prefer to choose the parameters ν_1, ν_2, h_1, h_2, h_3, b_1, b_2, b_{h_2}, b_{h_3} such that they all go to infinity as $\gamma \to \infty$. The generalized construction comes from the case of two drift-values, with very similar parameter requirements.

Let $\nu_1 = \nu_2 := \nu$ be functions of h_1 such that

$$\nu = c_1 h_1 + c_2 \tag{56}$$

where

$$c_1 := \frac{p_1(m_2 - m_1)^2}{m_1^2 m_2^2} \frac{1}{\sum_i \frac{2p_i}{\lambda(2m_i - \lambda)}}$$

and

$$c_2 := \frac{\sum_i \frac{2p_i}{(2m_i-\lambda)^2} + \sum_i \frac{p_i}{(2m_i - m_1)^2} + \sum_i \frac{p_i}{m_i^2}\left(\ln \frac{m_i^2}{2} - 1\right) - \sum_i \frac{p_i}{m_i^2} \ln \frac{m_1^2}{2}}{\sum_i \frac{2p_i}{\lambda(2m_i - \lambda)}}.$$

Then we choose constants l_i for $i = 1, 2$ such that

$$l_1 > \frac{2}{\lambda(2m_1 - \lambda)} \quad \text{and} \quad \frac{2}{\lambda(2m_3 - \lambda)} < l_2 < \frac{2}{\lambda(2m_2 - \lambda)} \tag{57}$$

and choose b_i such that

$$b_1 = l_1 \nu \quad \text{and} \quad b_2 = l_2 \nu. \tag{58}$$

Similarly, for $j = 2, 3$ we choose

$$b_{h_j} = l_j h_j \quad \text{and} \quad \frac{2}{m_1(2m_j - m_1)} < l_j < \frac{2}{m_1(2m_{j-1} - m_1)}. \tag{59}$$

For the thresholds h_2, h_3, we require them to be linear in h_1 as

$$\frac{h_i}{h_1} = \frac{(2m_i - m_1)m_1}{m_i^2} \leq 1 \quad \text{for} \quad i = 2, 3. \tag{60}$$

With the above choices, all parameters are either fixed or expressed in terms of h_1. The latter can be then determined from the false alarm constraint

$$E_\infty\left[T^{(3)}_{\text{com}}\right] = \gamma. \tag{61}$$

Theorem 6.1: *Let $\mathcal{R}^{(3)}(\gamma)$ be the family of composite $\mathcal{G}$-stopping times $T^{(3)}_{\text{com}}$ defined in (52) such that $\mu_1 = \mu_2 = \mu_3 = m_1$, and $\lambda_i, \nu_i, b_i, h_i, b_{h_j}$ satisfy (55)–(61) for $i = 1, 2, 3$ and $j = 2, 3$. Then, as $\gamma \to \infty$, any stopping time in $\mathcal{R}^{(3)}(\gamma)$ is asymptotically optimal of third order in the detection problem (P) in the sense that*

$$\lim_{\gamma\to\infty}\left[J(T^{(3)}_{\text{com}}) - LB(\gamma)\right] = 0 \tag{62}$$

where $LB(\gamma)$ is defined in Lemma 2.1; $J(T^{(3)}_{\text{com}})$ is defined in (3).

Proposition 6.2: *Let $T^{(3)}_{\text{com}} \in \mathcal{R}_3(\gamma)$ be a composite $\mathcal{G}$-stopping time in Theorem 6.1, and $\delta^{(3)}$ be the associated identification function defined in (53), such that $\mu_1 = \mu_2 = \mu_3 = m_1$, and $\lambda_i, \nu_i, b_i, h_i, b_{h_j}$ satisfy (55)–(61) for $i = 1, 2, 3$ and $j = 2, 3$. Then, we obtain*

$$\lim_{\gamma\to\infty} P^{m_i}_{\tau}\left(\delta^{(3)} \neq m_i \middle|\, \tau < \rho^{\lambda_1}_{\nu_1}\right) = 0 \quad \text{for } i = 1, 2, 3, \tag{63}$$

where $\rho^{\lambda_1}_{\nu_1}$ is the last reset of the first stage CUSUM statistic process in (51).

See Appendix for the proof.

7. Properties of CRP

The distribution of the CRP defined in (11) can be derived by specializing the results in Zhang and Hadjiliadis [28], where S^{Λ}_K is called the speed of market crash. The following lemma presents the density function of a CRP where the driver process $\{\xi_t\}_{t\geq 0}$ is a Brownian motion with drift M.

Lemma 7.1: *For $\Lambda \neq 2M$, the CRP S^{Λ}_K, associated with a CUSUM stopping time T^{Λ}_K with tuning parameter Λ and threshold K, has the probability density function*

$$f_{S^{\Lambda}_K}(y) = \sqrt{\frac{2}{\pi}}\frac{\sinh(\delta K)}{\delta\Lambda^3 y^{5/2}} e^{-\frac{\Lambda^2\delta^2}{2}y} \sum_{n=0}^{\infty}\left[(2n+1)^2K^2 - \Lambda^2 y\right] e^{-\frac{(2n+1)^2K^2}{2\Lambda^2 y}}, \tag{64}$$

for $y \in \mathbb{R}_+$, and where

$$\delta := \frac{2M - \Lambda}{2\Lambda}. \tag{65}$$

Proof of Lemma 7.1: In the definition (8), we can rewrite the CUSUM stopping time as

$$T^{\Lambda}_K = \inf\left\{t \geq 0 : \sup_{s\leq t}\left(\mu s - \sigma W_s\right) - \left(\mu t - \sigma W_t\right) \geq K\right\},$$

where $\mu = (\Lambda - 2M)\Lambda/2$ and $\sigma = \Lambda > 0$.

From Section 4.1 in Zhang and Hadjiliadis [28], we have the Laplace transform of the CRP as

$$E^{Q}\left[e^{-aS^{\Lambda}_K}\right] = \frac{C^a_{\delta,\sigma}}{\delta}\frac{\sinh(\delta K)}{\sinh(C^a_{\delta,\sigma}K)}, \tag{66}$$

$$\text{where} \quad \delta = -\frac{\mu}{\sigma^2} \quad \text{and} \quad C^a_{\delta,\sigma} = \sqrt{\delta^2 + \frac{2a}{\sigma^2}} \quad \text{for } a > 0. \tag{67}$$

Our objective is to take the inverse Laplace transform of Equation (66) to obtain the probability density function of S_K^Λ,

$$f_{S_K^\Lambda}(y) = \mathcal{L}_a^{-1}\left[E^Q\left[e^{-aS_K^\Lambda}\right]\right](y) = \frac{\sinh(\delta K)}{\delta}\mathcal{L}_a^{-1}\left[\frac{C_{\delta,\sigma}^a}{\sinh(C_{\delta,\sigma}^a K)}\right](y). \tag{68}$$

Denote

$$z := \frac{\sigma^2}{2K^2}y \quad \text{and} \quad \eta := K^2\delta^2 + \frac{2K^2}{\sigma^2}a = (C_{\delta,\sigma}^a K)^2. \tag{69}$$

By changing variables, we obtain

$$\mathcal{L}_a^{-1}\left[\frac{C_{\delta,\sigma}^a}{\sinh(C_{\delta,\sigma}^a K)}\right](y) = \frac{\sigma^2}{2K^3}e^{-\frac{\delta^2\sigma^2}{2}y}\mathcal{L}_\eta^{-1}\left[\frac{\sqrt{\eta}}{\sinh(\sqrt{\eta})}\right](z). \tag{70}$$

From the series expansion

$$\frac{1}{\sinh(x)} = \frac{2e^{-x}}{1-e^{-2x}} = 2e^{-x}\sum_{n=0}^{\infty}e^{-2nx}, \tag{71}$$

we have

$$\mathcal{L}_\eta^{-1}\left[\frac{\sqrt{\eta}}{\sinh(\sqrt{\eta})}\right](z) = 2\sum_{n=0}^{\infty}\mathcal{L}_\eta^{-1}\left[\sqrt{\eta}e^{-(2n+1)\sqrt{\eta}}\right](z). \tag{72}$$

Then by using formula 3 of Appendix 3 of Borodin and Salminen [4], we obtain

$$\mathcal{L}_\eta^{-1}\left[\sqrt{\eta}e^{-(2n+1)\sqrt{\eta}}\right](z) = \frac{1}{\sqrt{\pi}z^{5/2}}\left(\frac{(2n+1)^2}{4} - \frac{1}{2}z\right)e^{-\frac{(2n+1)^2}{4z}}. \tag{73}$$

Combining (68), (70), (72) and (73), we obtain (64). □

Corollary 7.2: *The expected value of the CRP S_K^Λ is*

$$E[S_K^\Lambda] = \frac{4}{(2M-\Lambda)^2}\left[\coth\left(\frac{2M-\Lambda}{2\Lambda}K\right)\frac{2M-\Lambda}{2\Lambda}K - 1\right], \tag{74}$$

and thus, we obtain

$$\lim_{K\to\infty}\frac{E[S_K^\Lambda]}{K} = \left|\frac{2}{\Lambda(2M-\Lambda)}\right|. \tag{75}$$

The expression (74) follows by differentiating Equation (66) with respect to a and then letting $a = 0$.

Proof of Lemma 4.1: From the density function (64), for $S_\nu := S(Z,\lambda,\nu)$, we can easily get

$$P_\infty(S_\nu < b_\nu) = \frac{2\sqrt{2}\sinh(\nu/2)}{\sqrt{\pi}\,\lambda^3}\int_0^{b_\nu} y^{-5/2}e^{-\frac{1}{8}\lambda^2 y}K(y)\mathrm{d}y, \tag{76}$$

where

$$K(y) := \sum_{n=0}^{\infty}\left[(2n+1)^2\nu^2 - \lambda^2 y\right]e^{-\frac{(2n+1)^2\nu^2}{2\lambda^2 y}}. \tag{77}$$

For large enough ν, on the interval $y \in [0, b_\nu]$, we can obtain

$$K(y) \leq \sum_{n=0}^{\infty} (2n+1)^2 \nu^2 e^{-\frac{(2n+1)^2\nu^2}{2\lambda^2 y}} \leq \nu^2 \int_1^{\infty} x e^{-\frac{x}{2\lambda^2 y}\nu^2} \mathrm{d}x = Cy e^{-\frac{1}{2\lambda^2 y}\nu^2}.$$

Denote

$$B(y) := y^{-\frac{3}{2}} e^{-\frac{1}{8}\lambda^2 y} e^{-\frac{\nu^2}{2\lambda^2 y}}.$$

By using differentiation, as $\nu \to \infty$, we obtain that the maximum of $B(y)$ happens at the position $y_{\max} = (-6 + 2\sqrt{\nu^2 + 9})/\lambda^2 = 2\nu/\lambda^2 + O_\nu(1)$. And $B(y)$ is increasing for $y < y_{\max}$. Thus, for large enough ν (76) leads to

$$P_\infty(S < b) \leq C b_\nu \sinh(\nu/2) B(y_{\max}) = C b_\nu \nu^{-\frac{3}{2}} e^{-\frac{(\lambda^2 y_{\max} - 2\nu)^2}{8\lambda^2 y_{\max}}},$$

where, we use C to represent a generic constant. Since b_ν is linear in ν, we can see that $P_\infty(S_\nu < b_\nu) \leq O(\nu^{-1/2})$ as $\nu \to \infty$, which gives (22). □

Proof of Lemma 4.5: In this proof for simplicity we denote $S := S_{h_2} = S(Z, \mu_2, h_2)$, $b := b_{h_2}$, $\mu = \mu_2$ and $h := h_2$. Under both $P_0^{m_1}$ and $P_\tau^{m_1}(\,\cdot\,|\tau < \rho_{h_2}^{\mu_2})$, the process $\{Z_t\}_{\rho_{h_2}^{\mu_2} \leq t \leq T_{h_2}^{\mu_2}}$ is a Brownian motion with drift m_1. From (66), we obtain the Laplace transform of S

$$E_\tau^{m_1}\left[e^{-aS} \,\middle|\, \tau < \rho_{h_2}^{\mu_2}\right] = \frac{C_{\delta_1,\mu}^a}{\delta_1} \frac{\sinh(\delta_1 h)}{\sinh(C_{\delta_1,\mu}^a h)}, \quad \text{for } a > 0 \tag{78}$$

where

$$\delta_1 := \frac{2m_1 - \mu}{2\mu} \quad \text{and} \quad C_{\delta_1,\mu}^a := \sqrt{\delta_1^2 + \frac{2a}{\mu^2}}. \tag{79}$$

We have $\delta_1 > 0$ when $0 < \mu < 2m_1$. From Chebyshev's inequality, as $h \to \infty$ we know that, for any $a > 0$,

$$P_\tau^{m_1}(S < b \,|\, \tau < \rho_{h_2}^{\mu_2}) \leq e^{ab} E_\tau^{m_1}\left[e^{-aS} \,\middle|\, \tau < \rho_{h_2}^{\mu_2}\right] = O(e^{-r_1(a)h}),$$

where

$$r_1(a) = a\left[\frac{4}{\mu(2m_1 - \mu) + \mu\sqrt{(2m_1 - \mu)^2 + 8a}} - \frac{b}{h}\right]. \tag{80}$$

To guarantee that $\sup_{a \geq 0} r_1(a) > 0$, we require $b/h \leq 2/(\mu(2m_1 - \mu))$, i.e. (34). This is because that the term inside the square brackets in (80) is decreasing in a.

Moreover, fixing b, h, μ, the exponential rate constant in (35) is $L_1 = r_1(a^*)$ where $a^* = \arg\sup_{a \geq 0} r_1(a)$. By differentiation, we can find that the maximum of $r_1(a)$ happens at

$$a^* = \frac{h^2}{2\mu b^2} - \frac{1}{8}(2m_1 - \mu)^2. \tag{81}$$

Then

$$L_1 = r_1(a^*) = \left(\frac{4}{\mu(2m_1-\mu)+2h/b} - \frac{b}{h}\right)\left(\frac{h^2}{2\mu b^2} - \frac{1}{8}(2m_1-\mu)^2\right) > 0.$$

In particular, we obtain (35). □

Proof of Lemmas 4.2 and 4.6: We only need to show Lemma 4.2. Lemma 4.6 follows immediately by substituting the group of parameters $\{\lambda, \nu, m_1\}$ with $\{\mu_2, h_2, m_2\}$. For simplicity, we denote $S := S_\nu$, $b := b_\nu$.

Under $P_\tau^{m_1}(\cdot \mid \tau < \rho_\nu^\lambda)$, the process $\{Z_t\}_{\rho_\nu^\lambda \le t \le T_\nu^\lambda}$ is a Brownian motion with drift m_1. From (66), we obtain the moment generating function of S as

$$E_\tau^{m_1}\left[e^{\theta S} \mid \tau < \rho_\nu^\lambda\right] = \frac{C_{\delta_2,\lambda}^{-\theta}}{\delta_2} \frac{\sinh(\delta_2 \nu)}{\sinh(C_{\delta_2,\lambda}^{-\theta}\nu)}, \tag{82}$$

with radius of convergence $0 < \theta < \frac{1}{2}\delta_2^2\lambda^2$ and

$$\delta_2 = \frac{2m_1 - \lambda}{2\lambda} \quad \text{and} \quad C_{\delta_2,\lambda}^{-\theta} = \sqrt{\delta_2^2 - \frac{2\theta}{\lambda^2}}. \tag{83}$$

We have $\delta_2 > 0$ when $0 < \lambda < 2m_1$. From Chebyshev's inequality, as $\nu \to \infty$ we can get

$$P_\tau^{m_1}(S \ge b \mid \tau < \rho_\nu^\lambda) \le \frac{E_\tau^{m_1}\left[e^{\theta S} \mid \tau < \rho_\nu^\lambda\right]}{e^{\theta b}} = O(e^{-r_2(\theta)\nu}),$$

where

$$r_2(\theta) = \theta\left[\frac{b}{\nu} - \frac{4}{\lambda(2m_1-\lambda) + \lambda\sqrt{(2m_1-\lambda)^2 - 8\theta}}\right]. \tag{84}$$

To guarantee that $\lim_{\theta\to 0^+} r_2(\theta) > 0$, we require $b/\nu \ge 2/(\lambda(2m_1-\lambda))$, i.e. (23). Fixing b, ν, λ and by differentiation, we see that $\theta^* = \arg\sup_{\theta\in[0,\delta_2^2\lambda^2/2]} r_2(\theta)$ satisfies

$$\theta^* = \frac{1}{8}(2m_1-\lambda)^2 - \frac{\nu^2}{2\lambda b^2}. \tag{85}$$

It is easy to check that $0 < \theta^* < \delta_2^2\lambda^2/2$, and

$$L_2 = r_2(\theta^*) = \left(\frac{b}{\nu} - \frac{4}{\lambda(2m_1-\lambda)+2\nu/b}\right)\left(\frac{1}{8}(2m_1-\lambda)^2 - \frac{\nu^2}{2\lambda b^2}\right) > 0.$$

In particular, we obtain (25). □

8. Conclusion

In this paper, we address the Wiener disorder problem with post-change drift uncertainty free of distributional assumptions regarding the disorder time. We model the uncertainty in the drift with a Bernoulli distribution thus giving rise to a blend between Bayesian and min–max frameworks, which naturally arise in multifarious applications (see, e.g. [21]). To address this problem, we design a novel family of composite stopping times that are seen to enjoy third-order asymptotic optimality. Among the asymptotically optimal results in change-point detection problems, our proposed rule is able to achieve the third-order asymptotic optimality, which is the strongest asymptotic optimality to date in any model considering post-change uncertainty in this setup. A remarkable property of such a composite rule is that it can asymptotically distinguish the different values of post-change drift with an adequately controlled asymptotic error. This is achieved by introducing multiple CUSUM-based steps in our composite rule allowing us to record an additional relevant statistic, namely the CRP. The CRP thus extends the role of CUSUM algorithm from a stopping time to an analytic procedure which can produce other finer strategies with inferential power over the post-change drift. This finding constitutes a novel step towards rigorous treatment of models with random post-change drifts, such as the case that the post-change drift is a continuous function of time.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was partially supported by the NSF-DMS-ATD [grant number 1222526], [grant number 1222262]; the PSC-CUNY [grant number 65625-0043]; the CUNY Collaborative [grant number 80209-0921].

References

[1] E. Bayraktar, S. Dayanik, and I. Karatzas, *Adaptive Poisson disorder problem*, Ann. Appl. Probab. 16(3) (2006), pp. 1190–1261.
[2] M. Beibel, *Sequential change-point detection in continuous time when the post-change drift is unknown*, Bernoulli 3(4) (1997), pp. 457–478.
[3] M. Beibel and H.R. Lerche, *Sequential Bayes detection of trend changes*, in *Foundations of Statistical Inference (Shoresh, 2000)*, Contributions to Statistics, Physica, Heidelberg, 2003, pp. 117–130.
[4] A. Borodin and P. Salminen, *Handbook of Brownian Motion – Facts and Formulae*, 2nd ed., Birkhauser Verlag, Basel, 2002.
[5] S. Dayanik, H.V. Poor, and S.O. Sezer, *Multisource Bayesian sequential change detection*, Ann. Appl. Probab. 18(2) (2008), pp. 552–590.
[6] M. Dominé, *First-passage time distribution of a Wiener process with drift concerning two elastic barriers*, J. Appl. Probab. 33 (1996), pp. 164–175.
[7] G. Fellouris and G.V. Moustakides, *Decentralized sequential hypothesis testing using asynchronous communication*, IEEE Trans. Inform. Theory 57(1) (2011), pp. 534–548.
[8] G. Fellouris and G. Sokolov, *Second-order asymptotic optimality in multisensor sequential change detection*, IEEE Trans. Inform. Theory 62(6) (2016), pp. 3662–3675.
[9] O. Hadjiliadis, *Optimality of the 2-CUSUM drift equalizer rules for detecting two-sided alternatives in the Brownian motion model*, J. Appl. Probab. 42(4) (2005), pp. 1183–1193.
[10] O. Hadjiliadis and G.V. Moustakides, *Optimal and asymptotically optimal CUSUM rules for change point detection in the Brownian Motion model with multiple alternatives*, translation in Theory Probab. Appl. 50(1) (2006), pp. 75–85.
[11] T. Jeulin and M. Yor, *Grossissement d'une filtration et semi-martingales: formules explicites* [Magnification of a filtration and semi-martingales: Explicit formulas], in *Séminaire de Probabilités, XII*, C. Dellacherie, P.A. Meyer, and M. Weil, eds., Lecture Notes in Mathematics Vol. 649, Springer-Verlag, Berlin, 1978, pp. 78–97.

[12] C. Kardaras, *On the stochastic behavior of optional processes up to random times*, Ann. Appl. Probab. 25(2) (2015), pp. 429–464.
[13] J. Lin and M. Ludkovski, *Sequential Bayesian inference in hidden Markov stochastic kinetic models with application to detection and response to seasonal epidemics*, Stat. Comput. 24(6) (2014), pp. 1047–1062.
[14] G. Lorden, *Procedures for reacting to a change in distribution*, Ann. Math. Statist. 42(6) (1971), pp. 1897–1908.
[15] M. Ludkovski, *Monte Carlo methods for adaptive disorder problems*, in *Numerical Methods in Finance*, Springer Proceedings in Mathematics Vol. 12, Springer-Verlag, Berlin, 2012, pp. 83–112.
[16] M. Ludkovski and S.O. Sezer, *Finite horizon decision timing with partially observable Poisson processes*, Stoch. Models 28(2) (2012), pp. 207–247.
[17] G.V. Moustakides, *Optimal stopping times for detecting changes in distributions*, Ann. Stat. 14(4) (1986), pp. 1379–1387.
[18] G.V. Moustakides, *Optimality of the CUSUM procedure in continuous time*, Ann. Stat. 32(1) (2004), pp. 302–315.
[19] A. Nikeghbali, *A class of remarkable submartingales*, Stoch. Process. Appl. 116(6) (2006), pp. 917–938.
[20] S. Pergamenchtchikov and A.G. Tartakovsky, *Asymptotically optimal pointwise and minimax quickest change-point detection for dependent data*, Stat. Inference Stoch. Process. (2016) doi:10.1007/s11203-016-9149-x.
[21] H.V. Poor, *An Introduction to signal Detection and Estimation*, Springer-Verlag, Berlin, 1994.
[22] H.V. Poor and O. Hadjiliadis, *Quickest Detection*, Cambridge University Press, Cambridge, 2008.
[23] P.E. Protter, *Stochastic Integration and Differential Equations*, 2nd ed., Springer-Verlag, New York, 2005.
[24] A.N. Shiryaev, *Minimax optimality of the method of cumulative sums (CUSUM) in the case of continuous time*, Russian Mathem. Surv. 51 (1996), pp. 750–751.
[25] D. Siegmund and E.S. Venkatraman, *Using the generalized likelihood ratio statistic for sequential detection of a change-point*, Ann. Stat. 23(1) (1995), pp. 255–271.
[26] Y. Wu, *Inference for Change Point and Post Change Means After a CUSUM Test*, Springer-Verlag, New York, NY, 2004.
[27] H. Yang and O. Hadjiliadis, *Quickest detection with post-change drift uncertainty*, in *Proceedings of the 53rd IEEE Conference on Decision and Control*, Los Angeles, CA, 2014, pp. 6567–6572.
[28] H. Zhang and O. Hadjiliadis, *Drawdowns and the speed of a market crash*, Methodol. Comput. Appl. Probab. 14(3) (2012), pp. 739–752.

Appendix 1.

We give the proofs of lemmas and theorems in this appendix. Recall the definition of the big-O and small-o notation: for any two real-valued functions $a(x)$ and $b(x) \neq 0$ defined on $\mathbb{R}$, we write $a(x) = O_x(b(x))$ if there exists $C > 0$ and x_0 such that

$$0 < \left|\frac{a(x)}{b(x)}\right| \leq C \quad \text{for all} \quad x \geq x_0,$$

and $a(x) = o_x(b(x))$, if

$$\lim_{x\to\infty} \left|\frac{a(x)}{b(x)}\right| = 0.$$

Proof of Lemma 2.1: For $i = 1, \ldots, N$, the conditional detection delay $J_i(T)$ is that of Lorden's criterion[14], and it is known that (see Shiryaev [24] for example), for any $T \in \mathcal{T}_1(\gamma)$

$$J_i(T) \geq \frac{2}{m_i^2} g\left(f^{-1}\left(\frac{m_i^2}{2}\gamma\right)\right),$$

where the right hand side is the detection delay of a CUSUM stopping time with tuning parameter m_i and average false alarm γ, which leads to (4).

To show the asymptotic behaviour, from (5), we can see $f(x) = e^x + O_x(x)$ and $g(x) = x - 1 + o_x(1)$ for $x > 0$ as $x \to \infty$. Thus, as $\gamma \to \infty$,

$$g\left(f^{-1}\left(\frac{m_i^2\gamma}{2}\right)\right) = \ln\gamma + \ln\frac{m_i^2}{2} - 1 + o_\gamma(1)$$

for $i = 1, \ldots, N$. Then, we obtain (6) and (7). □

Proof of Lemma 3.1: From (14), the expectation of T_{com} under any measure P is

$$E[T_{\text{com}}] = E[T_\nu^\lambda] + E[T_{h_1}^{\mu_1}]P(S_\nu \geq b_\nu) + E[T_{h_2}^{\mu_2}]P(S_\nu < b_\nu), \tag{A1}$$

where the independence of S_ν with $\{T_{h_1}^{\mu_1}, T_{h_2}^{\mu_2}\}$ is used in the second and third terms on the right hand side respectively. The expectations of CUSUM stopping time $T_\Upsilon^\Lambda := T(\{Z_t\}_{t\geq 0}, \Lambda, \Upsilon)$ defined in (8) are given by (see Poor and Hadjiliadis [22] for example)

$$\begin{aligned} E_\infty[T_\Upsilon^\Lambda] &= \frac{2}{\Lambda^2}f(\Upsilon) \equiv F(\Lambda, \Upsilon). \\ E_0^{m_i}[T_\Upsilon^\Lambda] &= \frac{2}{(2m_i - \Lambda)^2}g\left(\frac{2m_i - \Lambda}{\Lambda}\Upsilon\right) \equiv G_i(\Lambda, \Upsilon). \end{aligned} \tag{A2}$$

By plugging in two pairs of parameters (λ, ν) and (μ_i, h_i) separately into (A2), we obtain (15). □

Proof of Lemma 4.3: From (15), we have

$$\frac{E_\infty\left[T_{\text{com}}\right]}{e^{h_1}} = \frac{F(\lambda, \nu)}{e^{h_1}} + P_\infty(S_\nu \geq b_\nu)\frac{F(\mu_1, h_1)}{e^{h_1}} + P_\infty(S_\nu < b_\nu)\frac{F(\mu_2, h_2)}{e^{h_1}}. \tag{A3}$$

Due to (18) and $\nu/h_1 < 1$, the first term on the right hand side of (A3) goes to 0 as $\gamma \to \infty$. From Lemma 4.1 and the choice of ν, we have $P_\infty(S_\nu \geq b_\nu) \to 1$ as $\gamma \to \infty$. Condition $h_2/h_1 < 1$ and (18) lead to the third term on the right hand side vanishing asymptotically. Thus, we obtain

$$\lim_{\gamma\to\infty}\frac{E_\infty[T_{\text{com}}]}{e^{h_1}} = \lim_{\gamma\to\infty}P_\infty(S_\nu \geq b_\nu)\lim_{\gamma\to\infty}\frac{F(\mu_1, h_1)}{e^{h_1}} = \frac{2}{\mu_1^2}. \tag{A4}$$

Taking logarithms in (A4), and substituting for $E_\infty[T_{\text{com}}]$ from $E_\infty[T_{\text{com}}] = \gamma$, we obtain

$$\lim_{\gamma\to\infty}\left(\ln\gamma - h_1 - \ln\frac{2}{\mu_1^2}\right) = 0.$$

□

Proof of Theorem 4.4: By substituting $\mu_1 = \mu_2 = m_1$ in Equation (20), as $\gamma \to \infty$, we obtain

$$\begin{aligned} E_0^{m_1}\left[T_{\text{com}}\right] = &\frac{2}{\lambda(2m_1 - \lambda)}\nu - \frac{2}{(2m_1 - \lambda)^2} - \frac{2}{m_1^2} + P_0^{m_1}(S_\nu \geq b_\nu)\frac{2}{m_1^2}h_1 \\ &+ P_0^{m_1}(S_\nu < b_\nu)\frac{2}{m_1^2}h_2 + o_\gamma(1). \end{aligned}$$

From Lemma 4.2 and condition (28), we can see that $P_0^{m_1}(S_\nu \geq b_\nu)$ goes to zero exponentially as $\gamma \to \infty$. So we have $P_0^{m_1}(S_\nu \geq b_\nu)\nu$ goes to zero as $\gamma \to \infty$ too. From (27), ν is linear in h_1. Thus, $P_0^{m_1}(S_\nu \geq b_\nu)h_1$ goes to zero as $\gamma \to \infty$. Therefore as $\gamma \to \infty$, we obtain

$$E_0^{m_1}\left[T_{\text{com}}\right] = \frac{2}{\lambda(2m_1-\lambda)}\nu - \frac{2}{(2m_1-\lambda)^2} - \frac{2}{m_1^2} + \frac{2}{m_1^2}h_2 + o_\gamma(1). \tag{A5}$$

Similarly, applying Lemma 4.2 again and substituting m_1 with m_2, we can deduce that $P_0^{m_2}(S \geq b)h_1$ goes to zero as $\gamma \to \infty$. Thus, as $\gamma \to \infty$, Equation (20) implies

$$E_0^{m_2}\left[T_{\text{com}}\right] = \frac{2}{\lambda(2m_2-\lambda)}\nu - \frac{2}{(2m_2-\lambda)^2} - \frac{2}{(2m_2-m_1)^2} + \frac{2}{m_1(2m_2-m_1)}h_2 + o_\gamma(1). \tag{A6}$$

From condition (29), Lemma 3.1 and Equation (17), as $\gamma \to \infty$, the detection delay of T_{com} becomes

$$J\left(T_{\text{com}}\right) = \left(\frac{2p_1}{\lambda(2m_1-\lambda)} + \frac{2p_2}{\lambda(2m_2-\lambda)}\right)\nu + \left(\frac{2p_1m_1(2m_2-m_1)}{m_1^2m_2^2} + \frac{2p_2}{m_2^2}\right)h_1 - \frac{2p_1}{(2m_1-\lambda)^2} - \frac{2p_2}{(2m_2-\lambda)^2} - \frac{2p_1}{m_1^2} - \frac{2p_2}{(2m_2-m_1)^2} + o_\gamma(1).$$

From condition (27), ν is a linear function of h_1. And from Lemma 4.3, as $\gamma \to \infty$, we have

$$h_1 = \ln\gamma + \ln\frac{m_1^2}{2} + o_\gamma(1).$$

Therefore it follows that as $\gamma \to \infty$

$$J\left(T_{\text{com}}\right) = \left(\frac{2p_1}{m_1^2} + \frac{2p_2}{m_2^2}\right)\ln\gamma - \frac{2p_1}{m_1^2} - \frac{2p_2}{m_2^2} + \frac{2p_1}{m_1^2}\ln\frac{m_1^2}{2} + \frac{2p_2}{m_2^2}\ln\frac{m_2^2}{2} + o_\gamma(1). \tag{A7}$$

Comparing (A7) with Lemma 2.1, we obtain (32). □

Proof of Proposition 4.7: Conditional on $\{\tau < \rho_\nu^\lambda\}$, the observation path segment $\{Z_t\}_{t\geq\rho_\nu^\lambda}$ is the Brownian motion with post-change drift m. So both S_ν and S_{h_2} are CRPs that are recorded on the path of the observation process with drift m.

Using the independence of the two stages in the composite stopping time (14), we obtain

$$P_\tau^{m_1}(\delta_{T_{\text{com}}} \neq m_1|\tau < \rho_\nu^\lambda) = P_\tau^{m_1}(S_{h_2} < b_{h_2}|\tau < \rho_\nu^\lambda)P_\tau^{m_1}(S_\nu < b_\nu|\tau < \rho_\nu^\lambda).$$

From the discussion about the choice of the parameters preceding Theorem 4.4, it follows that both $\nu \to \infty$ and $h_2 \to \infty$ as $\gamma \to \infty$.

From condition (38) and $\mu_2 = m_1$, we can apply Lemma 4.5 to get that $P_\tau^{m_1}(S_{h_2} < b_{h_2}|\tau < \rho_\nu^\lambda) \to 0$ as $\gamma \to \infty$. So we obtain $P_\tau^{m_1}(\delta_{T_{\text{com}}} \neq m_1|\tau < \rho_\nu^\lambda) \to 0$ as $\gamma \to \infty$.

Similarly, from (33), we have

$$P_\tau^{m_2}(\delta_{T_{\text{com}}} \neq m_2|\tau < \rho_\nu^\lambda) = P_\tau^{m_2}(S_{h_2} \geq b_{h_2}|\tau < \rho_\nu^\lambda)P_\tau^{m_2}(S_\nu < b_\nu|\tau < \rho_\nu^\lambda) + P_\tau^{m_2}(S_\nu \geq b_\nu|\tau < \rho_\nu^\lambda).$$

Since condition (28) provides

$$\frac{b_\nu}{\nu} > \frac{2}{\lambda(2m_1-\lambda)} \geq \frac{2}{\lambda(2m_2-\lambda)},$$

Lemma 4.2 leads to $P_\tau^{m_2}(S_\nu \geq b_\nu|\tau < \rho_\nu^\lambda) \to 0$ as $\gamma \to \infty$. From (38) and Lemma 4.6, we can see $P_\tau^{m_2}(S_{h_2} \geq b_{h_2}|\tau < \rho_\nu^\lambda) \to 0$ as $\gamma \to \infty$. It follows that $P_\tau^{m_2}(\delta_{T_{\text{com}}} \neq m_2|\tau < \rho_\nu^\lambda) \to 0$ as $\gamma \to \infty$. Thus, we obtain (39). □

Proof of Proposition 4.8: To define T_{ran} in (41), we replace the event $\{S_\nu \geq b_\nu\}$ in the definition of T_{com} (14) by using a general event A which is independent of the second stage. Using the similar steps as those used in the computation of the detection delay of T_{com} in Lemma 3.1, we will obtain the detection delay of T_{ran} in what follows. Under $P_0^{m_1}$, from condition (45) and $\mu_1 = m_1$, as $\gamma \to \infty$, we obtain

$$E_0^{m_1}[T_{ran}] = \frac{2}{m_1^2}h_1 - \frac{2q}{m_1^2} - \frac{2(1-q)}{(2m_1-\mu_2)^2} + \frac{2(1-q)c}{(2m_1-\mu_2)\mu_2} + o_\gamma(1).$$

Under $P_0^{m_2}$, as $\gamma \to \infty$, we have

$$E_0^{m_2}[T_{ran}] = \frac{2}{m_2^2}h_1 - \frac{2q}{(2m_2-m_1)^2} - \frac{2(1-q)}{(2m_2-\mu_2)^2} + \frac{2(1-q)c}{(2m_2-\mu_2)\mu_2} + o_\gamma(1).$$

Therefore, using the independence of A and the second stage, we obtain the detection delay of T_{ran} as

$$J(T_{ran}) = p_1 E_0^{m_1}[T_{ran}] + p_2 E_0^{m_2}[T_{ran}]. \tag{A8}$$

Moreover, using similar steps as those used in the proof of Lemma 4.3 and (44), as $\gamma \to \infty$, we obtain

$$h_1 = \ln\gamma - \ln\frac{2q}{\mu_1^2} + o_\gamma(1). \tag{A9}$$

Thus, from (A8), (A9) and the condition (45), as $\gamma \to \infty$, we obtain

$$J(T_{ran}) = \left(\frac{2p_1}{m_1^2} + \frac{2p_2}{m_2^2}\right)\ln\gamma + \frac{2p_1}{m_1^2}\left(\ln\frac{m_1^2}{2} - 1\right) + \frac{2p_2}{m_2^2}\left(\ln\frac{m_2^2}{2} - 1\right) + o_\gamma(1).$$

Comparing the above equation with Lemma 2.1 yields (46). □

Proof of Theorem 6.1 and Proposition 6.2: We will develop the proof in the following three steps:

(1) We first compute the asymptotic behaviour of h_1 as $\gamma \to \infty$. The average time to false alarm gives

$$\frac{E_\infty[T_{\text{com}}^{(3)}]}{e^{h_1}} = \frac{F(\lambda_1,\nu_1)}{e^{h_1}} + P_\infty(B_1^c)\frac{F(\lambda_2,\nu_2)}{e^{h_1}} + \sum_{i=1}^{3} P_\infty(B_i)\frac{F(m_1,h_i)}{e^{h_1}}. \tag{A10}$$

From Lemma 4.1, we have $P_\infty(B_1^c) \to 0$, $P_\infty(B_2) \to 0$ and $P_\infty(B_3) \to 0$ as $\gamma \to \infty$. From (56) and (60), we have all terms in (A10) vanishing except the one containing $F(m_1,h_1)$ as $\gamma \to \infty$. Thus, from (18) and Lemma 4.1, we have

$$\lim_{\gamma\to\infty}\frac{E_\infty[T_{\text{com}}^{(3)}]}{e^{h_1}} = \lim_{\gamma\to\infty}\frac{F(m_1,h_1)}{e^{h_1}}\lim_{\gamma\to\infty}P_\infty(B_1) = \frac{2}{m_1^2}.$$

Using the same argument as in the proof of Lemma 4.3, we obtain (31).

(2) We consider the asymptotic behaviour of the identification function $\delta^{(3)}$ in (53) as $\gamma \to \infty$. From Lemma s4.2, 4.5, 4.6, condition (57) and condition (59), we can see that

$$\lim_{\gamma\to\infty} P_\tau^{m_i}(A_i|\tau < \rho_{\nu_1}^{\lambda_1}) = 1 \quad \text{and} \quad \lim_{\gamma\to\infty} P_\tau^{m_i}(A_j|\tau < \rho_{\nu_1}^{\lambda_1}) = 0 \quad \text{for} \quad i \neq j.$$

Therefore, the conditional identification errors behaviour in (63) follows. This proves Proposition 6.2. We now proceed to the last step required to establish Theorem 6.1.

(3) We consider the asymptotic behaviour of the detection delay of $T^{(3)}_{\text{com}}$ as $\gamma \to \infty$. From Equation (16) and the expectations of the CUSUM stopping time in (A2), we obtain

$$E_0^{m_i}[T^{(3)}_{\text{com}}] = G_i(\lambda_1, \nu_1) + P_0^{m_i}(B_1^c)G_i(\lambda_2, \nu_2) + \sum_j P_0^{m_i}(B_j)G_i(\mu_j, h_j). \tag{A11}$$

We choose the parameters $\lambda_1 = \lambda_2 = \lambda$, $\nu_1 = \nu_2 = \nu$ and $\mu_i = m_1$ for $i = 1, 2, 3$. From conditions (57) and (58), Lemmas 4.2, 4.5 and 4.6, it is easy to see that as $\gamma \to \infty$, $P_0^{m_i}(B_1) \to 0$ for $i = 1, 2, 3$; $P_0^{m_1}(B_2) \to 1$, $P_0^{m_2}(B_2) \to 1$, $P_0^{m_3}(B_2) \to 0$; $P_0^{m_1}(B_3) \to 0$, $P_0^{m_2}(B_3) \to 0$, $P_0^{m_3}(B_2) \to 1$. Thus, as $\gamma \to \infty$, we obtain

$$E_0^{m_1}[T^{(3)}_{\text{com}}] = \left(\frac{4}{\lambda(2m_1 - \lambda)}\nu - \frac{4}{(2m_1 - \lambda)^2}\right) + \frac{2}{m_1^2}h_2 - \frac{2}{m_1^2} + o_\gamma(1).$$

And for $i = 2, 3$, as $\gamma \to \infty$, we obtain

$$E_0^{m_i}[T^{(3)}_{\text{com}}] = \left(\frac{4}{\lambda(2m_i - \lambda)}\nu - \frac{4}{(2m_i - \lambda)^2}\right) + \frac{2}{m_1(2m_i - m_1)}h_i - \frac{2}{(2m_i - m_1)^2} + o_\gamma(1).$$

From condition (60) and the choice of ν in (56), as $\gamma \to \infty$, we obtain

$$J(T^{(3)}_{\text{com}}) = \sum_{i=1}^{3} \frac{2p_i}{m_i^2}\ln\gamma + \sum_{i=1}^{3}\frac{2p_i}{m_i^2}\left(\ln\frac{m_i^2}{2} - 1\right) + o_\gamma(1).$$

Thus, from the above equation and Lemma 2.1, we obtain (62). □

Appendix 2.

In this appendix, we discuss the behaviours of the generalized likelihood ratio stopping time and the mixture likelihood ratio stopping time mentioned in Section 4.3.

B.1. A generalized likelihood ratio stopping time

For the generalized likelihood stopping time $T_k^{\max} = T_1^c \wedge T_2^c$ defined in (49), to compute the expectation of $T_k^{\max}$ under $P_0^{m_1}$, we consider

$$\begin{aligned}
\{y_t^{(1)} > k\} &= \left\{W_t > \inf_{s\le t}\left(W_s + \frac{m_1}{2}s\right) + \frac{k}{m_1} - \frac{m_1}{2}t\right\} \\
\{y_t^{(2)} > k\} &= \left\{W_t > \inf_{s\le t}\left(W_s + \left(m_1 - \frac{m_2}{2}\right)s\right) + \frac{k}{m_2} - \left(m_1 - \frac{m_2}{2}\right)t\right\}.
\end{aligned} \tag{B12}$$

If $m_2 < 2m_1$, under $P_0^{m_1}$, both of two process $y^{(1)}$ and $y^{(2)}$ have positive drifts. It is easy to see that

$$\begin{aligned}
&\frac{k}{m_1} - \frac{m_1}{2}t > \frac{k}{m_2} - \left(m_1 - \frac{m_2}{2}\right)t \quad \text{when} \quad t < t_k := \frac{2k}{m_1 m_2}, \\
&\text{and} \quad \inf_{s\le t}\left(W_s + \frac{m_1}{2}s\right) > \inf_{s\le t}\left(W_s + \left(m_1 - \frac{m_2}{2}\right)s\right).
\end{aligned}$$

Then from the right hand sides of (B12), we get $T_2^c < T_1^c$ when $t < t_k$ under $P_0^{m_1}$, which leads to

$$E_0^{m_1}[T_k^{\max}] = E_0^{m_1}[T_2^c 1_{\{t<t_k\}}] + E_0^{m_1}[T_k^{\max} 1_{\{t\ge t_k\}}]. \tag{B13}$$

From $T_k^{\max} < \min\{T_1^c, T_2^c\}$, the second term in (B13) has an upper bound

$$E_0^{m_1}[T_k^{\max} 1_{\{t \geq t_k\}}] \leq \min_i \left(\int_{t_k}^{\infty} t f_i^{m_1}(t)\, \mathrm{d}t \right).$$

where $f_i^{m_1}(t)$ is the density of T_i^c under $P_0^{m_1}$ for $i = 1, 2$. An explicit expression of the density function of T_i^c is given in Domine [6], which gives

$$f_i^{m_1}(t) = e^{\beta_{1i}k - \frac{1}{2}\beta_{1i}^2 m_i^2 t} \frac{m_i^2}{k^2} \sum_{n=1}^{\infty} \frac{\theta_n^2 + \beta_{1i}^2 k^2}{\theta_n^2 + \beta_{1i}^2 k^2 + \beta_{1i} k} \theta_n \sin\theta_n e^{-\frac{1}{2k^2}\theta_n^2 m_i^2 t},$$

where $\beta_{1i} = (2m_1 - m_i)/(2m_i)$ and θ_n are the positive eigenvalues that satisfy the equation $\tan\theta = -\theta/(\beta_{1i}k)$. Since $t_k = 2k/(m_1 m_2)$ is linear in k, by basic computation, we can see that as $k \to \infty$, $E_0^{m_1}[T_k^{\max} 1_{\{t \geq t_k\}}] \leq O_k(k^{-1} e^{-L(m_1, m_2)k})$, where $L(m_1, m_2) > 0$ is a constant. So as $k \to \infty$, when $m_2 < 2m_1$ we obtain

$$E_0^{m_1}[T_k^{\max}] = E_0^{m_1}[T_2^c] + o_k(1). \tag{B14}$$

If $m_2 > 2m_1$, under $P_0^{m_1}$, the drift of $y^{(1)}$ is positive and the drift of $y^{(2)}$ is negative. Then it can be shown that (see Hadjiliadis [9] for example)

$$E_0^{m_1}[T_1^c] \leq E_0^{m_1}[T_k^{\max}] + E_0^{m_1}[T_1^c] P_0^{m_1}(T_2^c < T_1^c)$$
$$E_0^{m_1}[T_2^c] \leq E_0^{m_1}[T_k^{\max}] + E_0^{m_1}[T_2^c] P_0^{m_1}(T_1^c < T_2^c),$$

and thus,

$$\left(E_0^{m_1}[T_1^c]\right)^{-1} \leq \left(E_0^{m_1}[T_k^{\max}]\right)^{-1} \leq \left(E_0^{m_1}[T_1^c]\right)^{-1} + \left(E_0^{m_1}[T_2^c]\right)^{-1}. \tag{B15}$$

From the expectations of CUSUM in (A2), we can see that as $k \to \infty$, $E_0^{m_1}[T_2^c]$ goes to infinity exponentially in k and $E_0^{m_1}[T_1^c]$ goes to infinity linear in k. So as $k \to \infty$, (B15) leads to

$$E_0^{m_1}[T_k^{\max}] = E_0^{m_1}[T_1^c] + o_k(1). \tag{B16}$$

On the other hand, to compute the expectation of $T_k^{\max}$ under $P_0^{m_2}$, we have

$$\{y_t^{(1)} > k\} = \left\{ \left(m_2 - \frac{m_1}{2}\right) t + W_t - \inf_{s \leq t} \left(W_s + \left(m_2 - \frac{m_1}{2}\right) s \right) > \frac{k}{m_1} \right\}$$
$$\{y_t^{(2)} > k\} = \left\{ \frac{m_2}{2} t + W_t - \inf_{s \leq t} \left(W_s + \frac{m_2}{2} s \right) > \frac{k}{m_2} \right\}.$$

Since $m_2 - m_1/2 > 0$, by using the similar arguments to those used to show (B14), as $k \to \infty$, we obtain

$$E_0^{m_2}[T_k^{\max}] = E_0^{m_2}[T_2^c] + o_k(1). \tag{B17}$$

Also under P_∞, we have

$$\{y_t^{(1)} > k\} = \left\{ \left(-\frac{m_1}{2}\right) t + W_t - \inf_{s \leq t} \left(W_s + \left(-\frac{m_1}{2}\right) s \right) > \frac{k}{m_1} \right\}$$
$$\{y_t^{(2)} > k\} = \left\{ \left(-\frac{m_2}{2}\right) t + W_t - \inf_{s \leq t} \left(W_s + \left(-\frac{m_2}{2}\right) s \right) > \frac{k}{m_2} \right\}.$$

A similar argument shows that, as $k \to \infty$,

$$E_\infty[T_k^{\max}] = E_\infty[T_2^c] + o_k(1). \tag{B18}$$

As a combination of previous results, if $m_2 < 2m_1$, from (B14), (B17) and (B18), we can see that the performance of $T_k^{\max}$ is asymptotically the same as the performance of T_2^c. If $E_\infty[T_k^{\max}] = \gamma$, from (B18), we can obtain that $(\ln\gamma)/k \to 1$ as $\gamma \to \infty$. Then from the expectation of CUSUM in (A2), we can see that

$$\lim_{\gamma\to\infty} \frac{J(T_2^c)}{\ln\gamma} = \frac{2p_1}{(2m_1 - m_2)m_2} + \frac{2p_2}{m_2^2}.$$

Thus, $T_k^{\max}$ is not second order asymptotically optimal when $m_2 < 2m_1$.

If $m_2 > 2m_1$, from the results (B16)–(B18), when $E_\infty[T_k^{\max}] = \gamma$, we obtain

$$\lim_{\gamma\to\infty} \left[J(T_k^{\max}) - LB(\gamma)\right] = \frac{2p_1}{m_1^2} \ln \frac{m_2^2}{m_1^2}.$$

Thus, $T_k^{\max}$ is not third order asymptotically optimal when $m_2 > 2m_1$.

B.2. A mixture of likelihood ratios stopping time

For y_t^{mix}, it is easy to see that

$$\min\{y_t^{(1)}, y_t^{(2)}\} \le \log y_t^{\mathrm{mix}} \le \max\{y_t^{(1)}, y_t^{(2)}\} \tag{B19}$$

where $y^{(i)}$ is defined in (47).

In order to analyze the behaviour of T_d^{mix}, we use the similar arguments to those used in the case of the generalized likelihood ratio stopping time. In particular, as $d \to \infty$, we obtain

$$E_\infty[T_2^c] + o_d(1) \le E_\infty[T_d^{\mathrm{mix}}] \le E_\infty[T_1^c] + o_d(1).$$

where T_1^c and T_2^c are the CUSUM stopping times corresponding to $y^{(1)}$ and $y^{(2)}$ respectively and with the same threshold d. Thus, if we let $E_\infty[T_d^{\mathrm{mix}}] = \gamma$, it is easy to see that, as $\gamma \to \infty$,

$$\log\gamma + \log\frac{m_1^2}{2} + o_\gamma(1) \le d \le \log\gamma + \log\frac{m_2^2}{2} + o_\gamma(1).$$

If $m_2 < 2m_1$, under $P_0^{m_1}$ and $P_0^{m_2}$, from (B19) and the behaviour of the generalized likelihood ratio stopping time $T_d^{\max}$, we know that

$$E_0^{m_i}[T_d^{\mathrm{mix}}] \ge E_0^{m_i}[T_d^{\max}] = E_0^{m_i}[T_2^c] + o_\gamma(1).$$

So its performance is asymptotically worse than that of a CUSUM stopping time with parameter m_2, and thus, T_d^{mix} is not second-order asymptotically optimal.

If $m_2 > 2m_1$, since

$$\log y_t^{\mathrm{mix}} \ge \log p_i + y_t^{(i)},$$

we obtain

$$E_0^{m_i}[T_d^{\mathrm{mix}}] \le E_0^{m_i}[T_i'],$$

where T_i' is the CUSUM stopping time with threshold $d - \log p_i$ and parameter m_i. Thus,

$$\lim_{\gamma\to\infty} \left[J(T_d^{\mathrm{mix}}) - LB(\gamma)\right] \le \frac{2p_1}{m_1^2} \ln\frac{m_2^2}{m_1^2} - \frac{2p_1}{m_1^2}\log p_1 - \frac{2p_2}{m_2^2}\log p_2. \tag{B20}$$

And so T_d^{mix} gives second-order asymptotic optimality in this case. But the detection delay function $J(T_d^{\mathrm{mix}})$ is no longer a linear function of p_i while $LB(\gamma)$ is still linear in p_i. Thus, the difference $J(T_d^{\text{mix}}) - LB(\gamma)$ can not cancel the constant term that depends on p_i, and so it can not be zero for any value of p_i. In other words, T_d^{mix} does not have the third order asymptotic optimality in general.