21st International Congress on Modelling and Simulation, Gold Coast, Australia, 29 Nov to 4 Dec 2015
www.mssanz.org.au/modsim2015

# Mean Shift Detection for State Space Models

**J. Kuhn** [a b], **M. Mandjes** [b] **and T. Taimre** [a]

[a]School of Mathematics and Physics, The University of Queensland, Australia
[b]Korteweg-de Vries Institute for Mathematics, University of Amsterdam, The Netherlands
*Email: j.kuhn@uq.edu.au*

**Abstract**: In this paper we develop and validate a procedure for testing against a shift in mean in the observa-tions and hidden state sequence of state space models with Gaussian noise. State space models are popular for modelling stochastic networks as they allow to take into account that observations of the true state of a sys-tem may be corrupted by measurement noise (usually, a Gaussian noise process is assumed). Although state space models are very general, they are still relatively tractable in that the true system state can be estimated efficiently by a recursive procedure known as Kalman filtering.

State space models can be regarded as a special type of hidden Markov model. As such, they are a flexible modelling tool that has been found useful, for example, for modelling road networks (Stathopoulos and Karlaftis, 2003) to account for uncertainty in the measurement of travel times. For instance, we may assume that travel times have to be estimated from flow and occupancy data. An increase in the unobserved mean travel time can be caused by traffic congestion; a shift in the mean value of the observations on the other hand could indicate a bias of the sensors. State space models can also be used to model communication networks: Suppose that the current *state* of a channel (e.g. measured by the probability of packet loss) is not observed directly, but has to be inferred from the received package flow. A change in the mean value of the hidden state sequence, or a change in the mean value of the received package flow can deteriorate the performance of the network if it remains unrecognised.

This motivates us to investigate procedures for testing against a shift in mean in the observations and hidden state sequence of state space models. The objective is to detect a change as quickly as possible while keeping the ratio of false alarms at a pre-specified low level.

Since the observations are generally not independent, in (Basseville and Nikiforov, 1993) a cumulative sum (CUSUM) procedure is applied to the (independent) sequence of innovations, which is obtained as a by-product from Kalman filter estimation of the hidden states. That is, a log-likelihood ratio (LLR) test statistic is used and an alarm is raised as soon as this test statistic exceeds a certain threshold that is assumed to be given. Change point detection for state space models has also been considered in (Lai and Shan, 1999) for the case where the size of the mean shift is unknown, in which case a generalized LLR test can be applied.

In this paper we tackle the question of how the threshold of the sequential LLR test can be chosen when the shift size is assumed to be known. In practice, the latter assumption can be dealt with by realizing that typically there will be a minimum change size that is of interest from an engineering perspective, and that can thus be used as input for the model. Based on this assumption we can identify the appropriate level of the threshold based on approximations of the false alarm probability – essentially the probability that a random walk process exceeds a given threshold on an interval.

A persistent change in the mean value of the observations results in a dynamic change in the mean value of the innovations, which are therefore not identically distributed after the change point. However, it follows from the stability properties of the Kalman filter that under weak conditions the magnitude of the shift converges to a constant. This allows for large-deviations (LD) approximations as well as approximations motivated by a functional central limit theorem (CLT). LD approximations to the false alarm probability have been considered in (Bucklew, 1985; Ellens et al., 2013; Kuhn et al., 2014) for testing i.i.d. and vector autoregressive moving average (VARMA) models. CLT approximations were motivated, for example, in Siegmund (1985). We compare the numerical performance of the tests under both types of limiting regimes with respect to the false alarm probability and the detection delay.

***Keywords:*** *Change point detection, threshold approximation, state space models, Gaussian processes*

## 1 INTRODUCTION

In this paper we investigate procedures for testing against a shift in mean in the observations and hidden state sequence of state space models with Gaussian noise. A change point is detected as soon as the LLR test statistic exceeds a threshold that we can obtain from LD or CLT approximations to the false alarm probability. For an introduction to state space models and Kalman filter estimation refer to Goodwin and Sin (1984). A detailed account of the state of the art in change point detection is provided in Tartakovsky et al. (2014).

The paper is organized as follows. In Section 2 we define the state space model with a change in the mean value of both the hidden and the observed sequence. In Section 3 the testing procedure is explained; that is, we determine the LLR test statistic and show how the threshold can be approximated. Section 4 provides a comparison of the numerical performance of the tests under both types of limiting regimes with respect to the false alarm probability and the detection delay. We conclude in Section 5.

## 2 MODEL AND FRAMEWORK

We consider the following basic state space model of a sequence of observations $(V_t)$, time $t$ being discrete, with a shift in mean at the *change point* $k$:

$$X_{t+1} = AX_t + Y_t + M\,\mathbb{1}_{\{t\geq k\}}\,, \quad V_t = BX_t + Z_t + N\,\mathbb{1}_{\{t\geq k\}}\,. \tag{1}$$

The $d_x$-dimensional process $(X_t)$ represents the unobserved (latent) state of the system, with state transition matrix $A \in \mathbb{R}^{d_x\times d_x}$ that has eigenvalues within the unit circle. The vector of observations $V_t \in \mathbb{R}^{d_v}$ is a linear transformation of $X_t$. The vectors $M$ and $N$ model the shift in mean. For brevity, we assume that the Gaussian white noise processes $Y_t \sim \mathcal{N}(0, Q)$ and $Z_t \sim \mathcal{N}(0, R)$ are independent, and that $A$, $B$, $Q$, $R$, $M$, and $N$, are known.

We assume that observations arrive sequentially one at a time. Therefore, we consider testing shifting windows of $n$ observations with the objective to detect whether at some point within the current window a change in mean has occurred. To this end, we define a hypothesis test; the alternative hypothesis is essentially the union of hypotheses $H_1(k)$: *A change in mean occurred exactly at time $k$, for a specific $k \in \{1,\dots,n\}$.*

Denote $V_1^t := (V_1,\dots,V_t)$. The minimum variance estimator $\widehat{X}_t = \mathbb{E}[X_t \,|\, V_1^{t-1}]$ for the hidden state $X_t$ can be computed efficiently using the well-known Kalman filter [for details see e.g. Goodwin and Sin (1984)] as

$$\widehat{X}_t = A\widehat{X}_{t-1} + K_{t-1}(V_{t-1} - B\widehat{X}_{t-1})\,, \quad \widehat{X}_0 = x_0\,.$$

where $K_t := A\Sigma_t B'(B\Sigma_t B'+R)^{-1}$ is the *Kalman gain*, and $\Sigma_t = A\Sigma_{t-1}A'+Q-K_{t-1}(B\Sigma_{t-1}B'+R)K_{t-1}'$ is the *state error covariance matrix*. As a by-product the sequence of *innovations* is obtained,

$$\varepsilon_t := V_t - B\,\widehat{X}_t\,.$$

These represent the new information which is not contained in $V_1^{t-1}$. They are independent Gaussian zero-mean vectors with covariance

$$\Omega_t := \mathrm{Cov}(\varepsilon_t) = B\Sigma_t B' + R\,.$$

The persistent change in mean in $X_t$ and $V_t$ results in a dynamic change in the innovations; namely, the shift in mean on $\varepsilon_t$ is (Basseville and Nikiforov, 1993, Eq. (7.2.110))

$$\rho(t,k) = B\left[\psi(t,k) - A\,\zeta(t-1,k)\right] + N,$$

where $\psi(t,k) = A\,\psi(t-1,k) + M$, $\zeta(t,k) = A\,\zeta(t-1,k) + K_t\,\rho(t,k)$, with initial conditions $\psi(k,k) = 0$, $\zeta(k-1,k) = 0$. Thus, we are interested in testing whether there is a change point at *some* $k \in \{1,\dots,n\}$:

$$H_0 : \varepsilon_t \sim \mathcal{N}(0,\,\Omega_{t|t-1}) \quad \text{vs} \quad H_1 : \bigcup_{k=1}^{n} \left[H_1(k) : \varepsilon_t \sim \mathcal{N}\big(\rho(t,k),\,\Omega_{t|t-1}\big)\right]$$

with $t \geq k$. That is, we have to test whether any of the hypotheses $H_1(k)$ holds.

Note that the signature $\rho(t,k)$ of the change on the innovation depends upon both $k$ and $t$ during the transient phase of the Kalman filter. Provided that $\Sigma_t$ – the estimated covariance matrix of $X_t$ – converges to some

matrix $\Sigma$ as $t$ grows large, it can be seen that the Kalman gain $K_t$ converges to $K = \Sigma B'(B\Sigma B' + R)^{-1}$ (Basseville and Nikiforov, 1993, Section 3.2.3.2). For conditions under which this holds see (Goodwin and Sin, 1984, Section 7.3.1.2). The limit $\Sigma$ (if it exists) can be obtained as the solution of the algebraic Riccati equation

$$\Sigma - A\Sigma A' + A\Sigma B'(B\Sigma B' + R)^{-1}B\Sigma A' - Q = 0\,.$$

In this case we obtain (Basseville and Nikiforov, 1993, Eq. (7.2.112)) that asymptotically

$$\rho(t,k) \to B\big(I - A(I-KB)\big)^{-1}M + (I - B\big(I - A(I-KB)\big)^{-1}AK)N =: \rho\,.$$

Then it also holds that $\Omega_t \to B\Sigma B' + R =: \Omega$.

These limiting expressions are useful for obtaining approximations to the false alarm probability; we define a change point detection procedure based on such approximations in Section 3. Moreover, they yield an approximation to the LLR test statistic that can be computed in a recursive manner – in Section 4 we numerically evaluate the test performance when the approximate LLR is used rather than the actual LLR.

## 3 TESTING FOR A CHANGE IN MEAN

We now propose a change point detection test with LLR test statistic $\mathcal{L}_{n\beta}\big(V_1^n\big)$ defined in (2). It will turn out to be convenient to express the change point $k$ via the window size $n$, that is, we write $k = n\beta + 1$, where (throughout the paper) $\beta \in \mathcal{B}_n := \{0/n, 1/n, \ldots, (n-1)/n\}$. In line with (Bucklew, 1985, Ch. VI.E, Eq.(43)) we reject $H_0$ if

$$\max_{\beta\in\mathcal{B}_n} \frac{1}{n}\mathcal{L}_{n\beta}\big(V_1^n\big) = \max_{\beta\in\mathcal{B}_n} \frac{1}{n}\log\frac{q_{n\beta}\big(V_1^n\big)}{p\big(V_1^n\big)} > b, \tag{2}$$

where $p(V_1^n)$ and $q_{n\beta}(V_1^n)$ denote the density functions of $V_1^n$ under $H_0$ and $H_1(n\beta+1)$, respectively; and $b$ is a threshold specified below. To evaluate the LLRs for a particular window, note that the joint likelihood of $V_1^n$ is given by

$$p\big(V_1^n\big) = \prod_{t=1}^{n} p\big(V_t \,|\, V_1^{t-1}\big) = \prod_{t=1}^{n} \frac{1}{\sqrt{(2\pi)^{d_v}\,|\Omega_t|}} \exp\left[-\frac{1}{2}\big(V_t - B\widehat{X}_t\big)\Omega_t^{-1}\big(V_t - B\widehat{X}_t\big)'\right].$$

Thus we have that $p(V_1^n) = \prod_{t=1}^n p(\varepsilon_t)$, where (abusing notation) $p(\cdot)$ denotes the density function corresponding to its argument. Hence, we can write the LLR as

$$\mathcal{L}_{n\beta}(\varepsilon_1^n) = \sum_{t=n\beta+1}^{n} \rho(t, n\beta+1)'\Omega_t^{-1}\varepsilon_t - \frac{1}{2}\rho(t,n\beta+1)'\Omega_t^{-1}\rho(t,n\beta+1)\,. \tag{3}$$

Note that this is not a backward recursion (over $\beta$) because the recursive computation of $\rho(t, n\beta+1)$ proceeds forward. However, for large $n(1-\beta)$ we have

$$\mathcal{L}_{n\beta}(\varepsilon_1^n) \approx \sum_{t=n\beta+1}^{n} \ell(\varepsilon_t) := \sum_{t=n\beta+1}^{n} \rho'\Omega^{-1}\varepsilon_t - \frac{1}{2}\rho'\Omega^{-1}\rho\,. \tag{4}$$

We show numerically in Section 4 that the test performance remains good if the LLR (3) is replaced by the approximate LLR (4). The mean and variance of the asymptotic likelihood increments (under $H_0$) are

$$\mu := \mathbb{E}\big[\ell(\varepsilon_t)\big] = -\frac{1}{2}\rho'\Omega^{-1}\rho\,, \quad \sigma^2 := \mathrm{Var}\left(\ell(\varepsilon_t)\right) = \rho'\Omega^{-1}\rho\,.$$

One would like to choose the threshold $b$ in such a way that false alarm probability is kept at a given low level. That is, $b$ should satisfy

$$\mathbb{P}_0\left(\max_{\beta\in\mathcal{B}_n} \frac{1}{n}\mathcal{L}_{n\beta}\big(\varepsilon_1^n\big) > b\right) = \alpha\,. \tag{5}$$

In the following subsections we discuss how to obtain the threshold based on either LD approximations or, alternatively, CLT approximations to the false alarm probability. The performance of the resulting procedures is compared in Section 4.

### 3.1 Threshold Function based on LD Approximations

Since we wish the false alarm probability to be *small*, we certainly are concerned with a rare event; this motivates us to invoke LD theory. Change point detection procedures based on LD approximations have been considered in (Bucklew, 1985; Ellens et al., 2013; Kuhn et al., 2014) for i.i.d. and VARMA models, yielding a threshold *function* $b(\cdot)$ that depends on the assumed position of the change point under the alternative hypothesis. We now explain how to obtain a threshold function from LD approximations for the state space model. First, note that

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}_0 \left( \max_{\beta\in\mathcal{B}_n} \frac{1}{n} \mathcal{L}_{n\beta}\big(\varepsilon_1^n\big) > b \right) = \max_{\beta\in\mathcal{B}_n} \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}_0 \left( \frac{1}{n} \mathcal{L}_{n\beta}\big(\varepsilon_1^n\big) > b \right)$$

(for details see (Ellens et al., 2013, Section 2)). LD theory suggests that for fixed $\beta$ the false alarm probability can be approximated by $\mathbb{P}_0\left(n^{-1}\mathcal{L}_{n\beta}(\varepsilon_1^n) > b(\beta)\right) \approx \exp\Big(-n\mathcal{I}\big(b(\beta)\big)\Big)$, where $\mathcal{I}$ denotes a function specified below. Recall that we wish the false alarm probability to be kept at a small level $\alpha$. This suggests to pick the threshold function $b$ such that it satisfies

$$\alpha = \exp\Big(-n\mathcal{I}\big(b(\beta)\big)\Big) \tag{6}$$

for all $\beta \in \mathcal{B}_n$. This choice entails that raising a false alarm is essentially equally likely irrespective of the supposed location of the change point within the window, and it is therefore optimal in terms of type II error performance; see (Bucklew, 1985, Ch. VI.E).

Now let us make the above more rigorous. The limiting logarithmic moment-generating function $\Lambda(\lambda)$ associated with the distribution of the LLR is defined as

$$\Lambda(\lambda) := \lim_{n\to\infty} \frac{1}{n(1-\beta)} \log M_{n\beta}(\lambda) := \lim_{n\to\infty} \frac{1}{n(1-\beta)} \log \mathbb{E}_0 \left( e^{\lambda \mathcal{L}_{n\beta}(\varepsilon_1^n)} \right) ; \tag{7}$$

we assume for now that this function exists and is finite for every $\lambda \in \mathbb{R}$. Define $\mathcal{I}$ as the Fenchel–Legendre transform

$$\mathcal{I}\big(b(\beta)\big) := \sup_{\lambda\in\mathbb{R}} \big(\lambda b(\beta) - (1-\beta)\,\Lambda(\lambda)\big)\,.$$

Provided that $\Lambda(\lambda)$ exists for all $\lambda \in \mathbb{R}$, noting that we can rescale as written out in (8), the Gärtner–Ellis theorem (Bucklew, 1985; Dembo and Zeitouni, 1998) yields

$$\lim_{n\to\infty} \frac{1-\beta}{n(1-\beta)} \log \mathbb{P}_0 \left( \frac{1}{n(1-\beta)} \mathcal{L}_{n\beta}(\varepsilon_1^n) - \frac{b(\beta)}{1-\beta} > 0 \right) = -\,\mathcal{I}\big(b(\beta)\big)\,. \tag{8}$$

In accordance with the idea expressed in (6), we choose the threshold function $b(\cdot)$ such that it satisfies

$$-\mathcal{I}(b(\beta)) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}_0 \left( \frac{1}{n} \mathcal{L}_{n\beta}(\varepsilon_1^n) - b(\beta) > 0 \right) = -\,\gamma \tag{9}$$

for some positive $\gamma = -n^{-1}\log\alpha$, across all $\beta \in \mathcal{B}_n$. Asymptotically, as $n \to \infty$, the probability of raising a false alarm within the window is then kept at level $\alpha$.

To be able to obtain $b(\beta)$ from (9), we need to compute the limiting log-moment-generating function $\Lambda(\lambda)$ in more explicit terms (this way we also check that it indeed exists and is finite for all $\lambda$). Because the sequence of innovations is independent, with $k = n\beta + 1$, we can write the associated moment-generating function as

$$M_{n\beta}(\lambda) = \mathbb{E}_0 \left[ \exp\left( \lambda \sum_{t=k}^{n} \log \frac{q\big(V_t \,|\, V_k^{t-1}\big)}{p\big(V_t \,|\, V_k^{t-1}\big)} \right) \right] = \prod_{t=k}^{n} \mathbb{E}_{0,t} \left[ \left( \frac{q(\varepsilon_t)}{p(\varepsilon_t)} \right)^{\lambda} \right] ,$$

where, abusing notation, $p$ and $q$ refer to the distribution of their argument under $H_0$ and $H_1(k)$ respectively, and $\mathbb{E}_{0,t}$ indicates that the expectation is taken with respect to $p(\varepsilon_t)$. As in (Ellens et al., 2013, Section 3) we can evaluate this as $\prod_{t=k}^{n} \exp\left[\frac{\lambda}{2}(\lambda-1)\rho(t,k)'\Omega_{t|t-1}^{-1}\rho(t,k)\right]$. Combining the above, we may take

$\Lambda(\lambda) \approx \frac{\lambda}{2}(\lambda-1)\rho'\Omega^{-1}\rho$ as an approximation for $\Lambda(\cdot)$. With this approximation we can compute a threshold function $b(\beta)$ from

$$\gamma = \sup_{\lambda} \left\{ \lambda b(\beta) + (1-\beta)\frac{\lambda}{2}(1-\lambda)\rho'\Omega^{-1}\rho \right\} =: \mathcal{I}\big(b(\beta)\big)\,. \tag{10}$$

The optimizing $\lambda$ is $1/2 + b(\beta)/\big[(1-\beta)\,\rho'\,\Omega^{-1}\,\rho\big]$, whence from (10) we obtain the desired closed-form expression for $b(\cdot)$:

$$b(\beta) = -\frac{1-\beta}{2}\,\rho'\,\Omega^{-1}\rho \,\pm\, \sqrt{2(1-\beta)\rho'\,\Omega^{-1}\rho\,\gamma}\,. \tag{11}$$

### 3.2 Threshold Function by CLT Arguments

As an alternative, we consider the approximation of the false alarm probability based on CLT arguments. Motivated by Donsker's theorem, we can approximate the false alarm probability (5) by (Siegmund, 1985, Eq. (3.15))

$$\mathbb{P}_0\left(\max_{t\in[0,n]} \sigma B_t + \mu\, t \geq b\right) = 1 - \Phi\left(\frac{b-\mu\, n}{\sigma\sqrt{n}}\right) + e^{\frac{2b\mu}{\sigma^2}}\,\Phi\left(\frac{-b-\mu\, n}{\sigma\sqrt{n}}\right)\,, \tag{12}$$

where $B_t$ is a standard Brownian motion (Wiener process). Then a fixed threshold $b$ (rather than a function as before) can be obtained numerically from setting (12) equal to $\alpha$.

## 4 NUMERICAL COMPARISON

We first illustrate how the testing procedures can be applied for on-line testing of shifting windows of size $n = 50$. We test $150$ observations, the change point occurs at $100$ (i.e., it is first contained in window $51$). The obtained alarm ratios (the relative frequencies of alarms raised in $10,000$ runs) are depicted in Fig. 1. The first window containing the change point is indicated by the vertical line. In this example we use the LD threshold (11) to determine whether a change has occurred, and simulate the asymptotic LLR (4) rather than the actual LLR (3) because this is computationally far more efficient. Despite the use of this approximation, it can be seen that the alarm ratios before the change point (the *false alarm rates*) are close to the significance level $\alpha = 0.01$ as desired, while after the change point they rise quickly to $1$.

Next, in order to gain insight about the impact of cross-correlation, we fix the diagonal entries of $A$ to be $A_{11} = A_{22} = 0.5$, and vary the off-diagonal entries (both are taken to be equal, $A_{12} = A_{21}$). For various shift sizes, we provide the achieved false alarm and detection rates when using thresholds obtained based on LD or CLT approximations. As a benchmark, we compare to the performance obtained when the threshold is simply put to zero. Further, we fix $B = 0.5\, I_2$, $Q = R = I_2$ for different choices of $M$, $N$, and $\alpha$. The resulting shift sizes are depicted in Fig. 2. The values plotted in Figs. 3–4 were obtained by averaging the relative frequencies of false and true alarms obtained over $10,000$ runs. The significance level $\alpha$ is indicated by the horizontal dotted black line.

The LD threshold yields false alarm rates that are consistently close to but slightly above the specified level $\alpha$, while the CLT threshold is conservative overall. The detection rates are good in both cases. The detection rates obtained with the zero threshold are the highest; however, this is due to the fact that its false alarm performance is extremely poor.

It can further be seen that the detection rates depend on the size of $\rho$: a larger change is easier to detect (compare to Fig. 2). The accuracy of the CLT approximations seems to improve when $\rho$ is small. In this case $\rho(t,k)$ is closer to $\rho$, even when $t$ is small; this may explain why the Brownian approximation works better in this case.

## 5 CONCLUSIONS

Numerical experiments (some of which were presented in this paper) indicate that both the LD and the CLT approximations work reasonably well. Using the LD threshold yields false alarm rates that are close to the pre-specified level. The CLT threshold is overall more conservative in terms of false alarm rates; however, the detection rates are still close to those obtained when using the LD threshold.

The fact that the tests performed well in our numerical examples also suggests that the test performance is rather robust with respect to the use of the LLR approximation (4) rather than the actual LLR (3). This is a

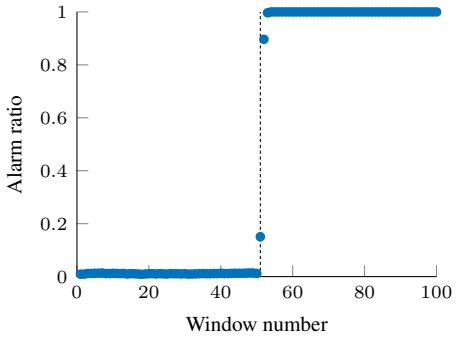


**Figure 1**. Alarm ratios, obtained with the LD threshold (11) for $A = B = 0.5\, I_2$, $M = N = (2,2)'$, $\alpha = 0.01$.

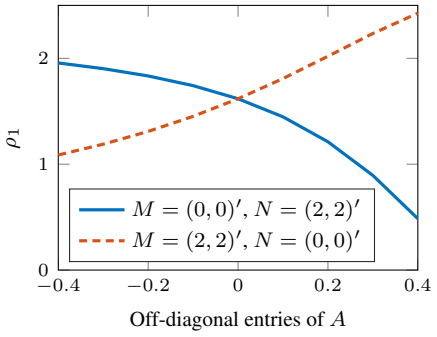


**Figure 2**. Values for the shift size $\rho$ (here $\rho_1 = \rho_2$).

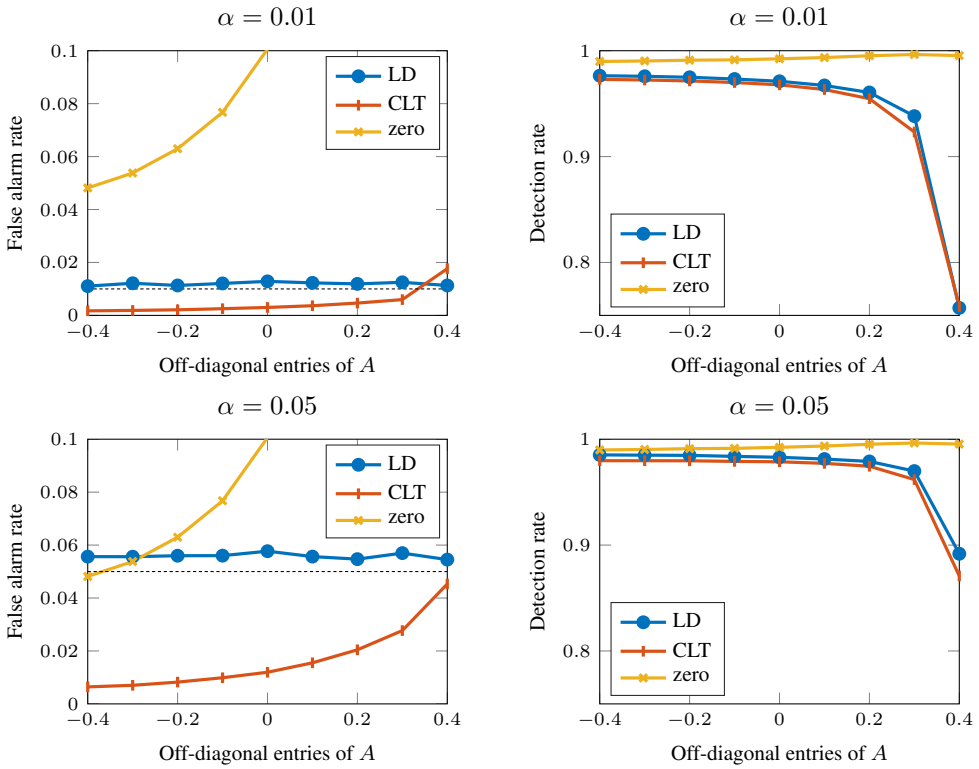


**Figure 3**. False alarm and detection rates with $M = (0,0)'$, $N = (2,2)'$ and $\alpha \in \{0.01, 0.05\}$.

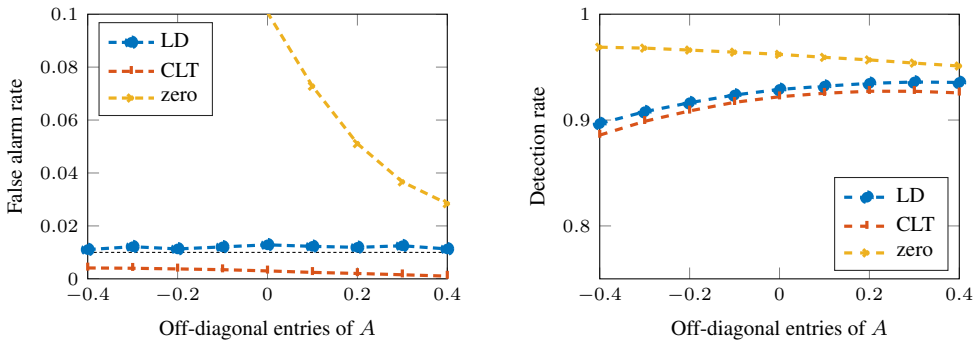


**Figure 4**. False alarm and detection rates with $M = (2,2)'$, $N = (0,0)'$ and $\alpha = 0.01$.

great advantage with respect to computation time, and thus important for applying the proposed procedures for on-line testing.

In future research we will investigate more thoroughly how much can be gained if the LLR (3) is used. Furthermore, it would be interesting to generalize the methods discussed in this paper to situations where the shift size $\rho$ is unknown, as well as to accommodate other types of noise distribution.

**ACKNOWLEDGEMENT**

Julia Kuhn is supported by Australian Research Council (ARC) grant DP130100156.

**REFERENCES**

Basseville, M. and I. Nikiforov (1993). *Detection of Abrupt Changes: Theory and Application.* Englewood Cliffs, Prentice Hall, N.J.

Bucklew, J. (1985). *Large Deviation Techniques in Decision, Simulation, and Estimation*. Wiley, New York.

Dembo, A. and O. Zeitouni (1998). *Large Deviations Techniques and Applications* (2 ed.). Springer-Verlag, New York.

Ellens, W., J. Kuhn, M. Mandjes, and P. Żuraniewski (2013). Changepoint detection for dependent Gaussian sequences. *Submitted,* arXiv:1307.0938.

Goodwin, G. C. and K. S. Sin (1984). *Adaptive Filtering, Prediction and Control*. Information and System Sciences Series. Englewood Cliffs, Prentice Hall, N.J.

Kuhn, J., W. Ellens, and M. Mandjes (2014). Detecting changes in the scale of dependent Gaussian processes: A large deviations approach. In B. Sericola, M. Telek, and G. Horváth (Eds.), *Analytical and Stochastic Modeling Techniques and Applications*, Volume 8499 of *Lecture Notes in Computer Science*, pp. 170–184. Springer International Publishing.

Lai, T. L. (1998). Information bounds and quick detection of parameter changes in stochastic systems. *IEEE Transactions on Information Theory 44*, 2917–2929.

Lai, T. L. and J. Z. Shan (1999). Efficient recursive algorithms for detection of abrupt changes in signals and control systems. *IEEE Transactions on Automatic Control 44*, 952–966.

Siegmund, D. (1985). *Sequential Analysis*. Springer-Verlag, New York.

Stathopoulos, A. and M. G. Karlaftis (2003). A multivariate state space approach for urban traffic flow modeling and prediction. *Transportation Research Part C: Emerging Technologies 11*, 121–135.

Tartakovsky, A., I. Nikiforov, and M. Basseville (2014). *Sequential Analysis: Hypothesis Testing and Changepoint Detection*. Monographs on Statistics & Applied Probability 136. Chapman & Hall/CRC, Boca Raton, Fla.