Ann. Inst. Statist. Math.
Vol. 57, No. 3, 553–573 (2005)

TEST FOR PARAMETER CHANGE BASED ON THE ESTIMATOR MINIMIZING DENSITY-BASED DIVERGENCE MEASURES

SANGYEOL LEE AND OKYOUNG NA

Department of Statistics, Seoul National University, Seoul 151-742, Korea

(Received May 26, 2003; revised March 30, 2004)

Abstract. In this paper we consider the problem of testing for a parameter change based on the cusum test proposed by Lee *et al.* (2003, *Scandinavian Journal of Statistics*, **30**, 781–796). The cusum test statistic is constructed via employing the estimator minimizing density-based divergence measures. It is shown that under regularity conditions, the test statistic has the limiting distribution of the sup of standard Brownian bridge. Simulation results demonstrate that the cusum test is robust when outliers exist.

Key words and phrases: Test for parameter changes, cusum test, density-based divergence measures, robust property, weak convergence, Brownian bridge.

1. Introduction

The problem of testing for parameter changes in statistical models has a long history. It originally started in the quality control context and then has been extended to various areas such as economics, finance, medicine, and seismic signal analysis. Since the paper of Page (1955), there have been published a vast amount of articles. For a general review of the change point problem, see Csörgő and Horváth (1997) and the papers therein. In iid samples, the parametric approach based on the likelihood was taken by many authors (cf. Chan and Gupta (2000)). However, the parametric approach is not proper when no assumptions are imposed on the underlying distribution of observations. For instance, no parametric approaches are directly applicable to the test for changes in the autocorrelations of stationary time series. To overcome such a problem, Lee *et al.* (2003) devised a cusum test in the same spirit of Inclán and Tiao (1994). The idea of the cusum test is the same as the one for the mean and variance change, but it includes a large number of other cases, such as the autoregressive coefficient in the random coefficient autoregressive models and the ARCH parameters. The cusum test has merit that it can test the existence of change points and, at the same time, allocate their locations. Furthermore, one can employ any estimators in construction of the cusum test as long as they satisfy certain regularity conditions. For instance, when there is a concern about outliers, a robust estimator can be utilized.

Recently, Basu *et al.* (1998) (BHHJ in the sequel) introduced a new estimation procedure minimizing a density-based divergence measures, called density power divergences. Compared to other density-based methods, such as Beran (1977), Tamura and Boos (1986) and Simpson (1987), which use the Hellinger distance, and Basu and Lindsey (1994) and Cao *et al.* (1995), the new method has an advantage of requiring no smoothing methods. In this case, one can avoid the drawbacks and difficulties, like the selection of bandwidth, that necessarily follow from the kernel smoothing method. In their paper, BHHJ demonstrated that some of the estimators possess strong robust properties with little loss in asymptotic efficiency relative to maximum likelihood estimator (MLE). Therefore, their estimator can be viewed as a good alternative to the MLE in terms of both efficiency and robustness. Seemingly, this result can be reflected in designing a robust cusum test.

In fact, Lee and Park (2001) considered a robust cusum test for the variance change in linear processes based on a trimming method, and demonstrated that it is necessary to use a robust method to prevent outliers from damaging the test procedure. Motivated by the viewpoint that the same phenomenon is anticipated to occur in other situations, we here consider a robust cusum test for a general parameter case. More precisely, we concentrate on the cusum test for parameter changes based on the BHHJ estimator. Despite the estimation method of BHHJ was restricted to iid samples, one can naturally extend the result to correlated observations. Thus in our set-up, we assume that the observations are correlated and satisfy a strong mixing condition in the sense of Rosenblatt. The organization of this paper is as follows. In Section 2, we explain how to construct the cusum test using the BHHJ estimator, and show that the test statistic converges weakly to the sup of a standard Brown bridge under mild conditions. In Section 3, we perform a simulation study and compare the two tests based on the BHHJ estimator and the MLE, respectively. In Section 4, we provide the proofs.

2. Main result

Consider a parametric family of models $\{F_\theta\}$, indexed by the unknown parameter $\theta \in \Theta \subset R^m$, possessing densities $\{f_\theta\}$ with respect to Lebesgue measure, and let $\mathcal{G}$ be the class of all distributions having densities with respect to Lebesgue measure. Aimed at estimating the unknown parameter θ, BHHJ introduced a family of density power divergences d_α, $\alpha \geq 0$;

$$d_\alpha(g, f) := \begin{cases} \int \{f^{1+\alpha}(z) - (1 + \frac{1}{\alpha})g(z)f^\alpha(z) + \frac{1}{\alpha}g^{1+\alpha}(z)\}dz, & \alpha > 0 \\ \int g(z)(\log g(z) - \log f(z))dz, & \alpha = 0, \end{cases}$$

where g and f are density functions, and defined the minimum density power divergence functional $T_\alpha(\cdot)$ by the requirement that for every G in $\mathcal{G}$,

$$d_\alpha(g, f_{T_\alpha(G)}) = \min_{\theta \in \Theta} d_\alpha(g, f_\theta),$$

where g is the density of G. Notice that if G belongs to $\{F_\theta\}$, $T_\alpha(G) := \theta_\alpha = \theta$ for some $\theta \in \Theta$. In this case, given random sample $X_1, \ldots, X_n$ with unknown density g, the minimum density power divergence estimator based on is defined by

$$\hat{\theta}_{\alpha,n} = \arg\min_{\theta \in \Theta} H_{\alpha,n}(\theta), \tag{2.1}$$

where $H_{\alpha,n}(\theta) = n^{-1} \sum_{t=1}^n V_\alpha(\theta; X_t)$ and

$$V_\alpha(\theta; x) := \begin{cases} \int f_\theta^{1+\alpha}(z)dz - (1 + \frac{1}{\alpha})f_\theta^\alpha(x), & \alpha > 0 \\ -\log f_\theta(x), & \alpha = 0. \end{cases}$$

BHHJ showed that the estimator has robust features against outliers but still possesses the efficiency that the MLE has when the true density belongs to the parametric family $\{F_\theta\}$. Our objective here is to test the constancy of the unknown parameter θ through the cusum test, based on the minimum density power divergence estimator, introduced by Lee *et al.* (2003). For this task, we set up the null and alternative hypotheses:

$$\begin{aligned} &H_0 : \theta_\alpha \text{ does not change over } X_1, \ldots, X_n. \quad \text{vs.} \\ &H_1 : \text{not } H_0, \end{aligned}$$

where n denotes the sample size.

When the random sample is iid with distribution G, BHHJ showed that under H_0 and the conditions with $r = 3$ below, $\hat{\theta}_{\alpha,n}$ is weakly consistent for $\theta_\alpha = T_\alpha(G)$ and $\sqrt{n}(\hat{\theta}_{\alpha,n} - \theta_\alpha)$ is asymptotically normal with mean zero vector. However, in order to achieve the asymptotic distribution of the cusum test statistic, we need the strong consistency of the estimator. Towards this end, we assume that the following conditions hold under the null hypothesis.

A1. The distribution F_θ and G have common support, so that the set $\mathcal{X}$ on which the densities are greater than zero is independent of θ.

A2. There is an open set ϑ of the parameter space Θ containing the best fitting parameter θ_α such that for all $x \in \mathcal{X}$, and all $\theta \in \vartheta$, the density $f_\theta(x)$ has continuous partial derivatives of order $r(\geq 0)$ with respect to θ and

$$E\left|\frac{\partial^j f_\theta(X)}{\partial\theta_{i_1}\cdots\partial\theta_{i_j}}\right| < \infty, \qquad 0 \leq j \leq r.$$

A3. The integral $\int f_\theta^{1+\alpha}(z)dz$ can be differentiated r-times ($r \geq 0$) with respect to θ, and the derivative can be taken under the integral sign.

A4. For each $1 \leq i_1, \ldots, i_r \leq m$, there exist functions $M_{\alpha,i_1\cdots i_r}(x)$ with $EM_{\alpha,i_1\cdots i_r}(X) < \infty$ such that

$$\left|\frac{\partial^r V_\alpha(\theta;x)}{\partial\theta_{i_1}\cdots\partial\theta_{i_r}}\right| \leq M_{\alpha,i_1\cdots i_r}(x)$$

for all $\theta \in \vartheta$ and $x \in \mathcal{X}$.

A5. There exists a nonsingular matrix J_α, defined by

$$J_\alpha := \frac{1}{1+\alpha}E\left(\frac{\partial^2 V_\alpha(\theta_\alpha;X)}{\partial\theta^2}\right).$$

A6. Under H_0, $\{X_t\}$ is ergodic and strictly stationary.

In BHHJ, the random observations are restricted to iid cases, but in the present paper we assume that the random sample is taken from an ergodic and strictly stationary process $\{X_t; t = 1, 2, \ldots\}$. As a special case, one can consider the time series contaminated by some outliers, namely, $X_t = (1 - p_t)X_{o,t} + p_t X_{c,t}$, where p_t are iid r.v.'s with $0 \leq p_t \leq 1$, $\{X_{o,t}\}$ is a strong mixing process with the density f_θ, the contaminating process $\{X_{c,t}\}$ is a sequence of iid r.v.'s with $EX_{c,t} = 0$ and $EX_{c,t}^2 < \infty$, and $\{p_t\}$, $\{X_{o,t}\}$ and $\{X_{c,t}\}$ are all independent. Note that the X_t follows an I.O. (innovation outlier) model when $p_t = 0$ for all t and $X_{o,t}$ follow a heavy-tailed non-Gaussian distribution, and an A.O. (additive outlier) model when $p_t = 1/2$ for all t and $X_{c,t}$ has an appropriate distribution (cf. Fox (1972) and Denby and Martin (1979)). It also indicates a S.O. (substitutive outlier) model when p_t are Bernoulli r.v.'s (cf. Bustos (1982)). In this case, if there are no changes in the underlying density of $X_{o,t}$, the θ_α does not change over $X_1, \ldots, X_n$.

In what follows, we assume that $\theta_\alpha = T_\alpha(G)$ exists and is unique, and keep the same definition for the estimator $\hat{\theta}_{\alpha,n}$ as in the iid case. In fact, the estimator is obtained by solving the estimating equations

$$U_{\alpha,n}(\theta) = (1+\alpha)^{-1} \cdot \frac{\partial H_{\alpha,n}(\theta)}{\partial \theta} = 0.$$

Then we have the following result, proof of which is presented in Section 4.

THEOREM 2.1. (Strong consistency) *Suppose that H_0 holds and conditions* A1–A4 *and* A6 *hold for some nonnegative integer r. Then there exists a sequence $\{\hat{\theta}_{\alpha,n}\}$, such that*

$$P\{\hat{\theta}_{\alpha,n} \to \theta_\alpha, \ \text{as } n \to \infty\} = 1. \tag{2.2}$$

Also, if $V_\alpha(\theta; x)$ is differentiable with respect to θ, then we have

$$U_{\alpha,n}(\hat{\theta}_{\alpha,n}) = 0 \tag{2.3}$$

for sufficiently large n.

Next, we derive a functional central limit theorem for $\hat{\theta}_{\alpha,n}$. Assume that conditions A1–A5 hold with $r = 3$. Since (2.3) holds for the minimum density power divergence estimator $\{\hat{\theta}_{\alpha,n}\}$, by expanding the vector $U_{\alpha,n}(\hat{\theta}_{\alpha,n})$ in a Taylor series about θ_α, we have

$$0 = U_{\alpha,n}(\hat{\theta}_{\alpha,n}) = U_{\alpha,n}(\theta_\alpha) - R_{\alpha,n}(\hat{\theta}_{\alpha,n} - \theta_\alpha),$$

where $R_{\alpha,n}$ is the $m \times m$ matrix whose (i,j)-th component is

$$R^{ij}_{\alpha,n} := -\frac{1}{1+\alpha}\left\{ \frac{\partial^2 H_{\alpha,n}(\theta_\alpha)}{\partial\theta_i\partial\theta_j} + \frac{1}{2}\sum_{k=1}^{m} \frac{\partial^3 H_{\alpha,n}(\theta^*_{\alpha,n})}{\partial\theta_i\partial\theta_j\partial\theta_k}(\hat{\theta}_{\alpha,n,k} - \theta_{\alpha,k}) \right\} \tag{2.4}$$

for some point $\theta^*_{\alpha,n} = \theta_\alpha + u(\hat{\theta}_{\alpha,n} - \theta_\alpha)$, $u \in [0,1]$. Therefore, we have

$$\hat{\theta}_{\alpha,n} - \theta_\alpha = J_\alpha^{-1} U_{\alpha,n}(\theta_\alpha) + \Delta_{\alpha,n},$$

where $\Delta_{\alpha,n} = J_\alpha^{-1}(J_\alpha - R_{\alpha,n})(\hat{\theta}_{\alpha,n} - \theta_\alpha)$, and consequently

$$\frac{[ns]}{\sqrt{n}}(\hat{\theta}_{\alpha,[ns]} - \theta_\alpha) = J_\alpha^{-1} \cdot \frac{[ns]}{\sqrt{n}} U_{\alpha,[ns]}(\theta_\alpha) + \frac{[ns]}{\sqrt{n}}\Delta_{\alpha,[ns]}. \tag{2.5}$$

Suppose that there exists a positive definite and symmetric matrix K_α, such that

$$\frac{[ns]}{\sqrt{n}} U_{\alpha,[ns]}(\theta_\alpha) \Rightarrow K_\alpha^{1/2} W_m(s) \tag{2.6}$$

in the $D^m[0,1]$ space (cf. Lemma 4.2), where W_m denotes a m-dimensional standard Brownian motion. Furthermore,

$$\max_{1\le k\le n} \frac{k}{\sqrt{n}}\|\Delta_{\alpha,k}\| = o_P(1) \tag{2.7}$$

(cf. Lemma 4.4). Then, from this, (2.5) and (2.6), we obtain the following result.

THEOREM 2.2. (functional central limit theorem) *Assume that conditions* A1–A6 *hold with* $r = 3$. *In addition, suppose that*

1. $\{X_t\}$ *is strong mixing with mixing order* $\beta(\cdot)$ *of size* $-\gamma/(\gamma-2)$ *for* $\gamma > 2$, *i.e.*, $\sum_{n=1}^{\infty}\beta(n)^{1-2/\gamma} < \infty$.
2. $E|\partial V_\alpha(\theta_\alpha; X)/\partial\theta_i|^\gamma < \infty$ *for* $i = 1,\dots,m$.
3. $nK_{\alpha,n} \to K_\alpha$ *for some positive definite and symmetric matrix* K_α, *where* $K_{\alpha,n}$ *is the covariance matrix of* $U_{\alpha,n}(\theta_\alpha)$.

Then, under H_0, *we have*

$$\frac{[ns]}{\sqrt{n}}(\hat\theta_{\alpha,[ns]} - \theta_\alpha) \Rightarrow J_\alpha^{-1}K_\alpha^{1/2}W_m(s).$$

Remark. The proof of Theorem 2.2 is provided in Section 4. As seen in the proof, the strong mixing condition is very essential. However, in Theorem 2.1, we did not need the mixing condition.

Now utilizing Theorem 2.2, we construct the cusum test for testing H_0 vs. H_1.

THEOREM 2.3. *Define*

$$T_{\alpha,n}^0 := \max_{m\le k\le n} \frac{k^2}{n}(\hat\theta_{\alpha,k} - \hat\theta_{\alpha,n})' J_\alpha K_\alpha^{-1} J_\alpha(\hat\theta_{\alpha,k} - \hat\theta_{\alpha,n}). \tag{2.8}$$

Suppose that the conditions of Theorem 2.2 *hold. Then, under* H_0,

$$T_{\alpha,n}^0 \Rightarrow \sup_{0\le s\le 1}\|W_m^o(s)\|^2,$$

where W_m^o *denotes an* m*-dimensional Brownian bridge. We reject* H_0 *if* $T_{\alpha,n}^0$ *is large.*

Since J_α and K_α are unknown, we should replace them by consistent estimators $\hat J_\alpha$ and $\hat K_\alpha$. First, note that

$$J_\alpha = \int u_{\theta_\alpha}(z)u_{\theta_\alpha}(z)' f_{\theta_\alpha}^{1+\alpha}(z)dz + \int (i_{\theta_\alpha}(z) - \alpha u_{\theta_\alpha}(z)u_{\theta_\alpha}(z)')(g(z) - f_{\theta_\alpha}(z)) f_{\theta_\alpha}^{\alpha}(z)dz,$$

where $u_\theta(z) = \partial \log f_\theta(z)/\partial\theta$ and $i_\theta(z) = -\partial u_\theta(z)/\partial\theta$. Therefore, if we put

$$\begin{aligned}\hat J_\alpha = & \int \{(1+\alpha)u_{\hat\theta_{\alpha,n}}(z)u_{\hat\theta_{\alpha,n}}(z)' - I_{\hat\theta_{\alpha,n}}(z)\} f_{\hat\theta_{\alpha,n}}^{1+\alpha}(z)dz \\ & + \frac{1}{n}\sum_{t=1}^{n}\{I_{\hat\theta_{\alpha,n}}(X_t) - \alpha u_{\hat\theta_{\alpha,n}}(X_t)u_{\hat\theta_{\alpha,n}}(X_t)'\} f_{\hat\theta_{\alpha,n}}^{\alpha}(X_t),\end{aligned}$$

then one can show that $\hat{J}_\alpha$ converges to J_α almost surely (cf. Lemma 4.5).

Now, we estimate K_α. Under the assumptions of Theorem 2.2,

$$K_\alpha = \sum_{k=-\infty}^{\infty} \frac{1}{(1+\alpha)^2} \operatorname{Cov}\left(\frac{\partial V_\alpha(\theta_\alpha, X_0)}{\partial\theta}, \frac{\partial V_\alpha(\theta_\alpha, X_k)}{\partial\theta}\right)$$

exists due to Theorem 1.5 in Bosq ((1996), p. 32). Assuming that

K1. $\sum_{n=1}^{\infty} \beta(n)^{1/3} < \infty$;

K2. $E|\partial V_\alpha(\theta_\alpha; X)/\partial\theta_i|^6 < \infty$ for $i = 1, \ldots, m$;

K3. there exists a function $M_\alpha(x)$ with $EM_\alpha(X)^2 < \infty$ such that $\sum_{i,j=1}^{m} |\partial^2 V_\alpha(\theta; x)/\partial\theta_i \partial\theta_j| \leq M_\alpha(x)$ for all $\theta \in \Theta$ and $x \in \mathcal{X}$,

one can show that $\hat{K}_\alpha \to K_\alpha$ in probability (cf. Lemma 4.8), where

$$\hat{K}_\alpha = \sum_{k=-h_n}^{h_n} \frac{1}{n(1+\alpha)^2} \sum_{t=1}^{n-k} \frac{\partial V_\alpha(\hat{\theta}_{\alpha,n}; X_t)}{\partial\theta} \cdot \frac{\partial V_\alpha(\hat{\theta}_{\alpha,n}; X_{t+k})}{\partial\theta}'$$

and $\{h_n\}$ is a sequence of positive integers such that

$$h_n \to \infty \quad \text{and} \quad h_n/\sqrt{n} \to 0. \tag{2.9}$$

Combining the above convergence results and Theorem 2.3, we establish the following.

THEOREM 2.4. *Define the test statistic* $T_{\alpha,n}$ *by*

$$T_{\alpha,n} := \max_{m \leq k \leq n} \frac{k^2}{n} (\hat{\theta}_{\alpha,k} - \hat{\theta}_{\alpha,n})' \hat{J}_\alpha \hat{K}_\alpha^{-1} \hat{J}_\alpha (\hat{\theta}_{\alpha,k} - \hat{\theta}_{\alpha,n}). \tag{2.10}$$

Suppose that the conditions of Theorem 2.2 *and* K1–K3 *hold. Then, under* H_0,

$$T_{\alpha,n} \Rightarrow \sup_{0 \leq s \leq 1} \|W_m^o(s)\|^2.$$

We reject H_0 *if* $T_{\alpha,n}$ *is large.*

Remark. If $\{X_t\}$ is the $M(\geq 0)$-dependent process, then we can use $h_n = M$ instead of (2.9).

3. Simulation results

In this section, we evaluate the performance of the test statistic $T_{\alpha,n}$ in Theorem 2.4 through a simulation study. In particular, we consider the situation that an exponential distribution $f_\theta(x) = \theta \exp(-\theta x)$ is fitted to the iid observations $X_1, \ldots, X_n$ with density g. The empirical sizes and powers are calculated at a nominal level of 0.1 for $\alpha = 0.0$, 0.05, 0.1, 0.15, 0.2, 0.25, 0.5 and 1.0. First, we consider the case that $g(x) = f_\theta(x)$. The empirical sizes are calculated with sets of 200, 300 and 500 observations generated from f_1. The figures in the 'size' row of Table 1 indicate the proportion of the number of rejections of the null hypothesis H_0, under which no parameter changes are assumed

Table 1. Empirical sizes and powers of $T_{\alpha,n}$ for the exponential model.

		α							
	n	0.0	0.05	0.1	0.15	0.2	0.25	0.5	1.0
size	200	0.143	0.141	0.144	0.148	0.148	0.139	0.144	0.146
	300	0.113	0.112	0.106	0.113	0.115	0.115	0.129	0.146
	500	0.113	0.117	0.114	0.108	0.109	0.107	0.106	0.110
power:	200	0.976	0.983	0.984	0.985	0.984	0.982	0.960	0.901
$\Delta = 2$	300	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.983
	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999

Table 2. Empirical sizes and powers of $T_{\alpha,n}$ for the exponential model with outliers.

			α							
	μ_V	n	0.0	0.05	0.1	0.15	0.2	0.25	0.5	1.0
size	5	200	0.149	0.117	0.114	0.126	0.125	0.126	0.121	0.127
		300	0.120	0.102	0.102	0.094	0.097	0.100	0.107	0.116
		500	0.131	0.119	0.108	0.099	0.100	0.106	0.120	0.134
	10	200	0.189	0.111	0.108	0.116	0.120	0.127	0.130	0.129
		300	0.174	0.111	0.104	0.102	0.109	0.110	0.130	0.138
		500	0.142	0.095	0.096	0.108	0.114	0.120	0.137	0.136
	20	200	0.312	0.120	0.088	0.095	0.108	0.117	0.133	0.151
		300	0.251	0.103	0.085	0.093	0.106	0.120	0.133	0.135
		500	0.204	0.112	0.102	0.116	0.118	0.124	0.140	0.132
power:	5	200	0.105	0.278	0.516	0.695	0.778	0.844	0.882	0.830
$\Delta = 2$		300	0.208	0.493	0.768	0.909	0.959	0.980	0.982	0.964
		500	0.428	0.832	0.979	0.996	1.000	1.000	1.000	0.998
	10	200	0.073	0.074	0.381	0.667	0.799	0.862	0.894	0.849
		300	0.049	0.177	0.634	0.884	0.955	0.983	0.985	0.969
		500	0.069	0.400	0.892	0.994	1.000	1.000	1.000	0.999
	20	200	0.198	0.030	0.397	0.755	0.872	0.914	0.916	0.849
		300	0.147	0.048	0.624	0.934	0.977	0.984	0.980	0.966
		500	0.121	0.151	0.882	0.997	1.000	1.000	1.000	0.998

to occur, out of 1000 repetitions. To examine the power, we consider the alternative hypothesis

$$\begin{aligned} H_1 : X_t &\sim f_1, \qquad t = 1, \ldots, [n/2], \\ X_t &\sim f_\Delta, \qquad t = [n/2] + 1, \ldots, n. \end{aligned}$$

For $\Delta = 2$, $n = 200$, 300, 500, the number of rejections of the null hypothesis are calculated out of 1000 repetitions. The results are summarized in Table 1. From the results, we can see that $T_{\alpha,n}$ does not produce severe size distortions and has good powers, close to 1 for all α except for $\alpha = 1.0$. This implies that the estimator with $\alpha > 0$, not very close to 1, does not damage the test even when there are no outliers.

In order to see how $T_{\alpha,n}$ performs for the data with outliers, we consider the mixture

Table 3. Empirical sizes and powers of $T_{\alpha,n}$ for the Gaussian AR(1) process.

			α							
	ϕ	n	0.0	0.05	0.1	0.15	0.2	0.25	0.5	1.0
size	0.1	200	0.084	0.086	0.085	0.087	0.085	0.088	0.083	0.096
		300	0.088	0.088	0.092	0.091	0.092	0.092	0.093	0.094
		500	0.077	0.076	0.076	0.079	0.082	0.082	0.079	0.081
	0.5	200	0.082	0.083	0.079	0.078	0.076	0.075	0.075	0.082
		300	0.060	0.060	0.062	0.059	0.058	0.062	0.068	0.078
		500	0.075	0.075	0.073	0.069	0.073	0.076	0.078	0.082
	0.8	200	0.092	0.086	0.086	0.084	0.083	0.086	0.084	0.092
		300	0.114	0.109	0.106	0.102	0.106	0.100	0.102	0.101
		500	0.114	0.114	0.113	0.107	0.107	0.107	0.103	0.107
power:	0.1	200	1.000	1.000	1.000	1.000	0.999	0.999	0.996	0.983
$\Delta = 1$		300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.5	200	0.871	0.872	0.862	0.860	0.859	0.855	0.804	0.700
		300	0.980	0.977	0.976	0.973	0.972	0.971	0.966	0.936
		500	1.000	1.000	0.999	0.999	0.999	0.999	0.998	0.995
	0.8	200	0.391	0.378	0.367	0.368	0.357	0.353	0.307	0.244
		300	0.459	0.453	0.449	0.443	0.441	0.435	0.402	0.355
		500	0.659	0.659	0.657	0.654	0.648	0.641	0.615	0.557

distribution $g(z) = (1-\epsilon)f_\theta(z) + \epsilon h(z)$, where $h(z)$ represents a contaminating distribution. For $h(z)$, we consider the exponential distribution with mean μ_V larger than 1. The empirical sizes and powers are calculated out of 1000 repetitions for n = 200, 300, 500, μ_V = 5, 10, 20, $\epsilon = 0.1$ and $\Delta = 2$. Table 2 shows that $T_{\alpha,n}$ with $\alpha = 0$ (the test based on the maximum likelihood estimator) has severe size distortions and produces very low powers. However, the test with $\alpha \geq 0.1$ cures the drawback remarkably. This result demonstrates the validity of $T_{\alpha,n}$ with $\alpha \geq 0.1$.

Next, we consider the mean change test in time series data. Let $\{X_{o,t}\}$ be an AR(1) process: $X_{o,t}-\mu = \phi(X_{o,t-1}-\mu)+\epsilon_t$, where ϵ_t are iid standard normal random variables. First, we handle the case that $X_t = X_{o,t}$ and they are not contaminated by outliers. We fit a normal distribution to $X_1, \ldots, X_n$, and perform the cusum test for μ utilizing the estimators $\hat{\mu}_{\alpha,k}$, $k = 1, \ldots, n$. Under the null hypothesis H_0, μ is assumed to be 0 for all observations. For the alternative, we consider

$$\begin{aligned} H_1 : \mu &= 0, \qquad t = 1, \ldots, [n/2], \\ \mu &= \Delta, \qquad t = [n/2]+1, \ldots, n. \end{aligned}$$

The empirical sizes are calculated for n = 200, 300, 500, ϕ = 0.1, 0.5, 0.8, $\Delta = 1$ and $h_n = n^{1/3}$ are used for $\hat{K}$. In each simulation, 100 initial observations are discarded to remove initialization effects. The results, presented in Table 3, show that $T_{\alpha,n}$ does not have severe size distortions and has good powers for all $\alpha \geq 0$ as in the previous case. As expected, we can see that the power increases as n increases and ϕ decreases to 0.

Tables 4 and 5 summarize the empirical sizes and powers of $T_{\alpha,n}$ when outliers are involved in observations. Here we assume that $X_{o,t}$ are contaminated by the outliers

Table 4. Empirical sizes of $T_{\alpha,n}$ for the Gaussian AR(1) process with outliers.

			α							
ϕ	μ_V	n	0.0	0.05	0.1	0.15	0.2	0.25	0.5	1.0
0.1	0	200	0.061	0.075	0.079	0.081	0.078	0.081	0.090	0.105
		300	0.077	0.094	0.095	0.084	0.086	0.085	0.086	0.114
		500	0.081	0.083	0.090	0.086	0.086	0.087	0.081	0.107
	5	200	0.058	0.071	0.077	0.078	0.080	0.081	0.079	0.120
		300	0.075	0.082	0.086	0.075	0.078	0.083	0.095	0.124
		500	0.074	0.081	0.079	0.084	0.083	0.091	0.096	0.123
	10	200	0.073	0.087	0.107	0.099	0.098	0.099	0.108	0.154
		300	0.081	0.080	0.091	0.096	0.090	0.094	0.103	0.158
		500	0.080	0.081	0.088	0.090	0.096	0.098	0.103	0.148
0.5	0	200	0.070	0.065	0.061	0.069	0.074	0.074	0.073	0.082
		300	0.074	0.078	0.074	0.076	0.076	0.066	0.058	0.076
		500	0.075	0.080	0.086	0.090	0.086	0.085	0.083	0.097
	5	200	0.081	0.090	0.085	0.074	0.071	0.070	0.063	0.099
		300	0.069	0.073	0.067	0.074	0.072	0.071	0.065	0.097
		500	0.096	0.090	0.095	0.096	0.095	0.095	0.084	0.118
	10	200	0.069	0.085	0.078	0.079	0.081	0.076	0.065	0.121
		300	0.073	0.088	0.079	0.074	0.073	0.065	0.068	0.134
		500	0.070	0.082	0.076	0.069	0.074	0.077	0.086	0.145
0.8	0	200	0.094	0.111	0.110	0.111	0.117	0.117	0.121	0.128
		300	0.096	0.099	0.105	0.102	0.099	0.098	0.102	0.110
		500	0.091	0.095	0.097	0.091	0.092	0.090	0.091	0.104
	5	200	0.093	0.107	0.106	0.111	0.110	0.110	0.105	0.126
		300	0.099	0.096	0.109	0.106	0.110	0.112	0.114	0.126
		500	0.101	0.115	0.115	0.102	0.103	0.103	0.100	0.118
	10	200	0.089	0.094	0.091	0.096	0.103	0.104	0.110	0.148
		300	0.092	0.106	0.111	0.119	0.121	0.114	0.112	0.161
		500	0.097	0.099	0.099	0.098	0.099	0.104	0.116	0.156

$X_{c,t}$, which are iid $N(\mu_V, 10^2)$, and that the observed r.v.'s follow the model $X_t = (1-p_t)X_{o,t} + p_t X_{c,t}$, where p_t are iid Bernoulli r.v.'s with success probability $p = 0.1$. It is assumed that $\{p_t\}$, $\{X_{o,t}\}$ and $\{X_{c,t}\}$ are all independent. The empirical sizes and powers based on the X_t's are calculated out of 1000 repetitions for n = 200, 300, 500, $\phi = 0.1$, 0.5, 0.8, $\mu_V = 0$, 5, 10 and $\Delta = 1$. Tables 4 and 5 show that the sizes are not severely distorted. However, we can see that the test with $\alpha = 0.0$ produce low powers. The result also demonstrates the validity of our test.

4. Proofs

Throughout this section, we establish the lemmas under the null hypothesis. To prove Theorem 1.1, we need the following lemma.

LEMMA 4.1. *Let $X_1, X_2, \ldots$ be strictly stationary and ergodic. If*

Table 5. Empirical powers of $T_{\alpha,n}$ for the Gaussian AR(1) process with outliers.

			α							
ϕ	μ_V	n	0.0	0.05	0.1	0.15	0.2	0.25	0.5	1.0
0.1	0	200	0.446	0.749	0.871	0.927	0.949	0.972	0.992	0.999
		300	0.606	0.875	0.954	0.980	0.986	0.996	0.999	1.000
		500	0.818	0.980	0.996	0.999	1.000	1.000	1.000	1.000
	5	200	0.422	0.678	0.825	0.898	0.933	0.959	0.991	0.996
		300	0.537	0.834	0.940	0.979	0.992	0.995	1.000	1.000
		500	0.736	0.961	0.997	1.000	1.000	1.000	1.000	1.000
	10	200	0.311	0.525	0.714	0.821	0.894	0.927	0.982	0.994
		300	0.394	0.679	0.831	0.930	0.968	0.981	0.999	1.000
		500	0.578	0.884	0.971	0.995	0.997	0.998	1.000	1.000
0.5	0	200	0.355	0.560	0.652	0.728	0.755	0.785	0.818	0.841
		300	0.492	0.739	0.859	0.895	0.912	0.926	0.946	0.947
		500	0.708	0.922	0.969	0.987	0.992	0.996	0.999	0.999
	5	200	0.328	0.503	0.629	0.705	0.733	0.765	0.819	0.855
		300	0.475	0.705	0.813	0.871	0.901	0.920	0.958	0.971
		500	0.673	0.912	0.968	0.984	0.990	0.993	0.995	0.999
	10	200	0.280	0.432	0.529	0.608	0.653	0.697	0.785	0.835
		300	0.368	0.592	0.736	0.824	0.877	0.909	0.957	0.964
		500	0.559	0.828	0.934	0.964	0.979	0.984	0.996	0.997
0.8	0	200	0.263	0.315	0.334	0.338	0.343	0.344	0.350	0.344
		300	0.316	0.387	0.413	0.425	0.436	0.441	0.452	0.438
		500	0.460	0.545	0.583	0.608	0.619	0.623	0.630	0.621
	5	200	0.255	0.303	0.321	0.335	0.342	0.341	0.342	0.347
		300	0.311	0.381	0.408	0.422	0.433	0.438	0.447	0.448
		500	0.492	0.568	0.604	0.613	0.626	0.632	0.644	0.642
	10	200	0.215	0.268	0.287	0.302	0.309	0.320	0.331	0.360
		300	0.294	0.347	0.384	0.409	0.432	0.435	0.455	0.477
		500	0.394	0.516	0.575	0.605	0.609	0.613	0.626	0.634

(a) *Θ is compact,*

(b) *$A(x,\theta)$ is continuous in θ for all x,*

(c) *There exists a function $B(x)$ such that $EB(X) < \infty$ and $|A(x,\theta)| \le B(x)$ for all x and θ,*

then

$$P\left\{\lim_{n\to\infty}\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}A(X_t,\theta)-a(\theta)\right|=0\right\}=1, \tag{4.1}$$

where $a(\theta) = EA(X,\theta)$.

In addition, if there exists $\theta_0 = \arg\min_{\theta\in\Theta} a(\theta)$ and it is unique, then

$$P\{\hat\theta_n \to \theta_0, n\to\infty\} = 1, \tag{4.2}$$

where $\hat\theta_n = \arg\min_{\theta\in\Theta} n^{-1}\sum_{t=1}^{n} A(X_t,\theta)$.

PROOF. The proof of (4.1) is essentially the same as in Ferguson ((1996), pp. 107–111), and is omitted for brevity. Now we prove (4.2). For any $\delta > 0$, $\|\hat{\theta}_n - \theta_0\| > \delta$ implies that

$$\min_{\theta\in\Theta} \frac{1}{n}\sum_{t=1}^{n} A(X_t,\theta) = \frac{1}{n}\sum_{t=1}^{n} A(X_t,\hat{\theta}_n) = \min_{\theta\in S} \frac{1}{n}\sum_{t=1}^{n} A(X_t,\theta),$$

where $S = \{\theta \in \Theta : \|\theta - \theta_0\| \geq \delta\}$ is a compact subset of Θ. Therefore,

$$\begin{aligned}\left|a(\theta_0) - \min_{\theta\in S} a(\theta)\right| &\leq \left|\min_{\theta\in S} \frac{1}{n}\sum_{t=1}^{n} A(X_t,\theta) - \min_{\theta\in S} a(\theta)\right| + \left|\min_{\theta\in\Theta} \frac{1}{n}\sum_{t=1}^{n} A(X_t,\theta) - \min_{\theta\in\Theta} a(\theta)\right| \\ &\leq 2\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n} A(X_t,\theta) - a(\theta)\right|.\end{aligned}$$

Note that due to conditions (b), (c) and the dominated convergence theorem, we have

$$a(\theta') = EA(X,\theta') \to EA(X,\theta) = a(\theta), \qquad \text{as} \quad \theta' \to \theta,$$

and consequently, $a(\theta)$ is continuous in θ. Hence, from the uniqueness of θ_0, it follows that

$$\left|a(\theta_0) - \min_{\theta\in S} a(\theta)\right| = \min_{\theta\in S} a(\theta) - a(\theta_0) := \epsilon_\delta > 0.$$

Therefore, by (4.1) we have

$$\begin{aligned} 0 &\leq P\left\{\bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty}(\|\hat{\theta}_n - \theta_0\| > \delta)\right\} \\ &\leq P\left\{\bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty}\left(\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n} A(X_t,\theta) - a(\theta)\right| \geq \frac{\epsilon_\delta}{2} > 0\right)\right\} = 0,\end{aligned}$$

which completes the proof. □

For notational convenience, without any confusion, we drop the subscript α from the quantities $\hat{\theta}_n$, H_n, V, U_n, and etc.

PROOF OF THEOREM 2.1. Since ϑ is an open set with $\theta_\alpha \in \vartheta$, there exists a positive δ such that

$$B_\delta := \{\theta : \|\theta - \theta_\alpha\| \leq \delta\} \subset \vartheta.$$

Note that from A2 and A3, $V(\theta; x)$ is r-times differentiable with respect to θ on ϑ, and consequently, by Taylor's series expansion we can write

$$\begin{aligned} V(\theta;x) &= V(\theta_\alpha;x) + \sum_{p=1}^{r-1}\frac{1}{p!}\sum_{i_1,\ldots,i_p=1}^{m}\frac{\partial^p V(\theta_\alpha;x)}{\partial\theta_{i_1}\cdots\partial\theta_{i_p}}\prod_{q=1}^{p}(\theta_{i_q} - \theta_{\alpha,i_q}) \\ &\quad + \frac{1}{(r-1)!}\int_0^1\sum_{i_1,\ldots,i_r=1}^{m}\frac{\partial^p V(\theta_\alpha + u(\theta-\theta_\alpha);x)}{\partial\theta_{i_1}\cdots\partial\theta_{i_r}}\prod_{q=1}^{r}(\theta_{i_q} - \theta_{\alpha,i_q})(1-u)^{r-1}du,\end{aligned}$$

for all $\theta \in B_\delta$ and almost all $x \in \mathcal{X}$. Thus, for almost all $x \in \mathcal{X}$,

$$\begin{aligned}(4.3)\qquad \sup_{\theta\in B_\delta} |V(\theta;x)| \le |V(\theta_\alpha;x)| &+ \sum_{p=1}^{r-1}\frac{\delta^p}{p!}\sum_{i_1,\ldots,i_p=1}^{m}\left|\frac{\partial^p V(\theta_\alpha;x)}{\partial\theta_{i_1}\cdots\partial\theta_{i_p}}\right| \\ &+ \frac{\delta^r}{r!}\sum_{i_1,\ldots,i_r=1}^{m} M_{i_1\cdots i_r}(x)\end{aligned}$$

and the expected value of the term of the right hand side of (4.3) is finite by A4. Therefore, the conditions in Lemma 4.1 for compact $B_\delta \subset \vartheta$ are satisfied, and consequently, if we set $\hat\theta_n(\delta) := \arg\min_{\theta\in B_\delta} H_n(\theta)$, we get

$$(4.4)\qquad P\{\hat\theta_n(\delta) \to \theta_\alpha,\ \text{as } n\to\infty\} = 1$$

and

$$(4.5)\qquad P\left\{\sup_{\theta\in B_\delta}|H_n(\theta)-\nu(\theta)| \to 0, n\to\infty\right\} = 1,$$

where $\nu(\theta) = EV(\theta;X)$, $X \sim G$.

In order to obtain $\hat\theta_n$, which is independent of δ, we define $\hat\theta_n$ as the closest $\hat\theta_n(\delta)$ to θ_α, i.e.,

$$(4.6)\qquad \|\hat\theta_n - \theta_\alpha\| = \inf\{\|\hat\theta_n(\delta)-\theta_\alpha\|; \hat\theta_n(\delta) \text{ satisfies } (4.4)\}.$$

Then, with this $\hat\theta_n$, (2.2) holds obviously.

Next, note that $\nu(\theta) = d_\alpha(g, f_\theta) - \alpha^{-1}\int g^{1+\alpha}(z)dz$, and so

$$\theta_\alpha = \arg\min_\theta d_\alpha(g, f_\theta) = \arg\min_\theta \nu(\theta).$$

In addition, since $\nu(\theta)$ is continuous in θ and θ_α is unique,

$$\epsilon(\delta) := -\sup_{\|\theta-\theta_\alpha\|=\delta}\left(-\frac{1}{2}(\nu(\theta)-\nu(\theta_\alpha))\right) = \frac{1}{2}\inf_{\|\theta-\theta_\alpha\|=\delta}(\nu(\theta)-\nu(\theta_\alpha)) > 0.$$

Then using these facts, we have

$$\begin{aligned}B_n &:= \left\{\sup_{\theta\in B_\delta}|H_n(\theta)-\nu(\theta)| < 2\epsilon(\delta)\right\} \\ &\subset \left\{\sup_{\theta\in B_\delta}|H_n(\theta)-\nu(\theta)-H_n(\theta_\alpha)+\nu(\theta_\alpha)| < \epsilon(\delta)\right\} \\ &\subset \{H_n(\theta)-\nu(\theta)-H_n(\theta_\alpha)+\nu(\theta_\alpha) > -\epsilon(\delta), \forall\theta\in B_\delta\} \\ &\subset \Bigg\{H_n(\theta)-H_n(\theta_\alpha)-\nu(\theta)+\nu(\theta_\alpha) > \sup_{\|\theta-\theta_\alpha\|=\delta}\left(-\frac{1}{2}(\nu(\theta)-\nu(\theta_\alpha))\right), \\ &\qquad\qquad \forall\theta : \|\theta-\theta_\alpha\| = \delta\Bigg\} \\ &\subset \left\{H_n(\theta)-H_n(\theta_\alpha)-\nu(\theta)+\nu(\theta_\alpha) > -\frac{1}{2}(\nu(\theta)-\nu(\theta_\alpha)), \forall\theta : \|\theta-\theta_\alpha\| = \delta\right\}\end{aligned}$$

$$
\begin{aligned}
&= \left\{ H_n(\theta) - H_n(\theta_\alpha) > \frac{1}{2}(\nu(\theta) - \nu(\theta_\alpha)), \forall\theta : \|\theta - \theta_\alpha\| = \delta \right\} \\
&\subset \left\{ H_n(\theta) - H_n(\theta_\alpha) > \frac{1}{2} \inf_{\|\theta-\theta_\alpha\|=\delta} (\nu(\theta) - \nu(\theta_\alpha)) = \epsilon(\delta) > 0, \forall\theta : \|\theta - \theta_\alpha\| = \delta \right\} \\
&\subset \{H_n(\theta) > H_n(\theta_\alpha), \forall\theta : \|\theta - \theta_\alpha\| = \delta\} := A_n.
\end{aligned}
$$

Therefore, due to (4.5), we have

$$
1 \geq P\left\{ \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} A_n \right\} \geq P\left\{ \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} B_n \right\} = 1.
$$

Thus, a local minimum is attained in the interior of B_δ, and therefore,

$$
U_n(\hat{\theta}_n(\delta)) = 0, \tag{4.7}
$$

if $V(\theta; x)$ is differentiable with respect to θ. Hence, for $\hat{\theta}_n$ satisfying (4.6), (2.3) holds. This completes the proof. □

Now we prove Theorem 2.2.

LEMMA 4.2. *Under the assumptions of Theorem 2.2,*

$$
\frac{[ns]}{\sqrt{n}} U_{[ns]}(\theta_\alpha) \Rightarrow K^{1/2} W_m(s),
$$

in the $D^m[0,1]$ space.

PROOF. We prove the theorem using the Cramér-Wold device. For any $\lambda = (\lambda_1, \ldots, \lambda_m)' \in R^m$ with $\|\lambda\| = 1$, let

$$
Y_t^\lambda = \frac{1}{1+\alpha} \sum_{i=1}^{m} \lambda_i \cdot \frac{\partial V(\theta_\alpha; X_t)}{\partial \theta_i}.
$$

Note that due to conditions 1–3, $\{Y_t^\lambda\}$ is a strictly stationary real-valued process, such that

(a) $\{Y_t^\lambda\}$ is strong mixing of size $-\gamma/\gamma - 2$,
(b) $EY_t^\lambda = 0$ and $E|Y_t^\lambda|^\gamma < \infty$,
(c) $n^{-1}\operatorname{Var}(\sum_{t=1}^n Y_t^\lambda) = \lambda'(nK_n)\lambda \to \lambda' K \lambda > 0$.

Therefore, we have

$$
\lambda' \frac{[ns]}{\sqrt{n}} U_{[ns]}(\theta_\alpha) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} Y_t^\lambda \Rightarrow (\lambda' K \lambda)^{1/2} W_1(s)
$$

in the $D[0,1]$ space (cf. Peligrad (1986), p. 202). Hence, the lemma follows. □

LEMMA 4.3. *Under the assumptions of Theorem 2.2, R_n converges a.s. to J.*

PROOF. From condition A4 with $r = 3$ and (2.2), the second term of the right hand side of (2.4) converges to 0 almost surely since

$$0 \leq \left| \sum_{k=1}^{m} \frac{\partial^3 H_n(\theta_n^*)}{\partial\theta_i \partial\theta_j \partial\theta_k} (\hat{\theta}_{n,k} - \theta_{\alpha,k}) \right| \leq \frac{1}{n} \sum_{t=1}^{n} \sum_{k=1}^{m} M_{ijk}(X_t) \cdot \|\hat{\theta}_n - \theta_\alpha\|$$

and $n^{-1} \sum_{t=1}^{n} \sum_{k=1}^{m} M_{ijk}(X_t) \to \sum_{k=1}^{m} EM_{ijk}(X_1)$ a.s. by the strong law of large numbers.

Meanwhile, since $\{X_t\}$ is strictly stationary and ergodic, $\{\partial^2 V(\theta; X_t)/\partial\theta_i \partial\theta_j\}$ is also strictly stationary and ergodic for all $1 \leq i, j \leq m$. Hence,

$$\frac{\partial^2 H_n(\theta_\alpha)}{\partial\theta_i \partial\theta_j} = \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 V(\theta_\alpha; X_t)}{\partial\theta_i \partial\theta_j} \overset{\text{a.s.}}{\longrightarrow} E \frac{\partial^2 V(\theta_\alpha; X_1)}{\partial\theta_i \partial\theta_j} = -(1+\alpha) J^{ij},$$

that is, the first term of the right hand side of (2.4) converges almost surely to the (i, j)-th component of J. Therefore, $R_n \to J$ a.s. □

The following lemma is concerned with the negligibility of Δ_k.

LEMMA 4.4. *Under the assumptions of Theorem 2.2,*

$$\max_k \left(\frac{k}{\sqrt{n}} \|\Delta_k\| \right) = o_P(1). \tag{4.8}$$

PROOF. Since $\det(R_n) \overset{\text{a.s.}}{\to} \det(J)$ in view of Lemma 4.3, it follows from Egoroff's theorem that given $\epsilon > 0$, there exists an event A with $P(A) < \epsilon$ an integer N_0 such that on A^c and for all $n > N_0$,

$$\big|\, |\det(R_n)| - |\det(J)| \,\big| \leq |\det(R_n) - \det(J)| < 2^{-1} |\det(J)| \neq 0,$$

and consequently,

$$|\det(R_n)| > 2^{-1} |\det(J)| \neq 0.$$

Therefore, on A^c and for all $n > N_0$, there exists an inverse matrix of R_n.

Note that for any $\delta > 0$,

$$\begin{aligned} P\left\{ \max_{1 \leq k \leq n} \frac{k}{\sqrt{n}} \|\Delta_k\| \geq 3\delta \right\} \leq\ & P\left\{ \max_{1 \leq k \leq N_0} \frac{k}{\sqrt{n}} \|\Delta_k\| \geq \delta \right\} \\ & + P\left\{ \max_{N_0 < k \leq n^\nu} \frac{k}{\sqrt{n}} \|\Delta_k\| \geq \delta, A^c \right\} \\ & + P\left\{ \max_{n^\nu < k \leq n} \frac{k}{\sqrt{n}} \|\Delta_k\| \geq \delta, A^c \right\} \\ & + P(A). \end{aligned} \tag{4.9}$$

First, notice that there exists $N_1 \geq 1$, such that for all $n > N_1$, the first term of the right hand side of (4.9) is less than ϵ, since

$$\max_{1 \leq k \leq N_0} \frac{k}{\sqrt{n}} \|\Delta_k\| \leq \frac{1}{\sqrt{n}} \max_{1 \leq k \leq N_0} k \|J^{-1}(J - R_k)\| \|\hat{\theta}_k - \theta_\alpha\| = o_P(1) \quad \text{as} \quad n \to \infty.$$

Second, since on A^c, the inverse matrix of R_k exists for all $k \geq N_0$, we have

$$\Delta_k = J^{-1}(J - R_k)(\hat{\theta}_k - \theta_\alpha) = J^{-1}(J - R_k)R_k^{-1}U_k(\theta_\alpha),$$

so that

$$\max_{N_0 \leq k \leq n^\nu} \frac{k}{\sqrt{n}} \|\Delta_k\| \leq \max_{N_0 \leq k \leq n^\nu} \|R_k^{-1} - J^{-1}\| \cdot \frac{1}{\sqrt{n}} \sum_{t=1}^{n^\nu} \|U(\theta_\alpha; X_t)\|. \tag{4.10}$$

Note that

$$E \frac{1}{\sqrt{n}} \sum_{t=1}^{n^\nu} \|U(\theta_\alpha; X_t)\| = n^{\nu - 1/2} E\|U(\theta_\alpha; X_1)\| = O(n^{\nu - 1/2}),$$

since $\{X_t\}$ is strictly stationary and $E\|U(\theta_\alpha; X_1)\| \leq \sum_{i=1}^m E|U_i(\theta_\alpha; X_1)| < \infty$. Therefore, for some $0 < \nu < 1/2$, we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n^\nu} \|U(\theta_\alpha; X_t)\| = o_P(1). \tag{4.11}$$

On the other hand, $R_n^{-1} \to J^{-1}$ on A^c and therefore

$$\max_{N_0 \leq k \leq n^\nu} \|R_k^{-1} - J^{-1}\| = O_P(1). \tag{4.12}$$

By (4.11) and (4.12), the right hand side of (4.10) becomes $o_P(1)$, and thus there exists an integer N_2, such that for all $n > N_2$, the second term of the right hand side of (4.9) is less than ϵ.

Finally, we have

$$\max_{n^\nu < k \leq n} \frac{k}{\sqrt{n}} \|\Delta_k\| \leq \max_{n^\nu < k \leq n} \|R_k^{-1} - J^{-1}\| \cdot \max_{1 \leq k \leq n} \left\| \frac{k}{\sqrt{n}} U_k(\theta_\alpha) \right\| = o_P(1),$$

since

$$\max_{n^\nu < k \leq n} \|R_k^{-1} - J^{-1}\| \xrightarrow{\text{a.s.}} 0$$

and

$$\max_{1 \leq k \leq n} \left\| \frac{k}{\sqrt{n}} U_k(\theta_\alpha) \right\| \xrightarrow{d} \sup_{0 \leq s \leq 1} \|K^{1/2} W(s)\|.$$

Therefore, there exists $N_3 \geq 1$, such that for $n > N_3$, the third term of the right hand side of (4.9) is less than ϵ. Now, if we put $N = \max_{0 \leq j \leq 3}\{N_j\}$, the right hand side of (4.9) is less than 5ϵ for all $n > N$. This establishes the lemma. □

LEMMA 4.5. *Under the assumptions of Theorem* 2.2, $\hat{J}$ *converges a.s. to* J.

PROOF. Define

$$J_1(\theta) = \int \{(1 + \alpha) u_\theta(z) u_\theta(z)' - i_\theta(z)\} f_\theta^{1+\alpha}(z) dz.$$

Note that

$$J_1(\theta) = -(1+\alpha)^{-1} \cdot \frac{\partial^2}{\partial \theta^2} \int f_\theta^{1+\alpha}(z) dz,$$

and thus, $J_1(\theta)$ is continuous in θ by condition A3. Therefore, since $\hat{\theta}_n \overset{\text{a.s.}}{\to} \theta_\alpha$,

$$J_1(\hat{\theta}_n) \overset{\text{a.s.}}{\to} J_1(\theta_\alpha). \tag{4.13}$$

Let

$$\eta(x,\theta) = \{i_\theta(x) - \alpha u_\theta(x) u_\theta(x)'\} f_\theta^\alpha(x)$$

and let $J_2(\theta) = E\eta(X,\theta)$, where $X \sim G$. Note that for any $\epsilon > 0$ and $\delta > 0$,

$$\begin{aligned}
A_n &:= \left\{ \left\| \frac{1}{n} \sum_{t=1}^n \eta(X_t, \hat{\theta}_n) - J_2(\hat{\theta}_n) \right\| > \epsilon \right\} \\
&\subset \left\{ \left\| \frac{1}{n} \sum_{t=1}^n \eta(X_t, \hat{\theta}_n) - J_2(\hat{\theta}_n) \right\| > \epsilon, \|\hat{\theta}_n - \theta_\alpha\| \le \delta \right\} \cup \{\|\hat{\theta}_n - \theta_\alpha\| > \delta\} \\
&\subset \left\{ \sup_{\|\theta - \theta_\alpha\| \le \delta} \left\| \frac{1}{n} \sum_{t=1}^n \eta(X_t, \theta) - J_2(\theta) \right\| > \epsilon, \|\hat{\theta}_n - \theta_\alpha\| \le \delta \right\} \cup \{\|\hat{\theta}_n - \theta_\alpha\| > \delta\} \\
&\subset \left\{ \sup_{\|\theta - \theta_\alpha\| \le \delta} \left\| \frac{1}{n} \sum_{t=1}^n \eta(X_t, \theta) - J_2(\theta) \right\| > \epsilon \right\} \cup \{\|\hat{\theta}_n - \theta_\alpha\| > \delta\} := B_n \cup C_n,
\end{aligned}$$

where $B_n := \{\sup_{\|\theta-\theta_\alpha\| \le \delta} \|n^{-1} \sum_{t=1}^n \eta(X_t,\theta) - J_2(\theta)\| > \epsilon\}$ and $C_n := \{\|\hat{\theta}_n - \theta\| > \delta\}$. Moreover, since the conditions in Lemma 4.1 are satisfied,

$$\sup_{\|\theta - \theta_\alpha\| \le \delta} \left\| \frac{1}{n} \sum_{t=1}^n \eta(X_t, \theta) - J_2(\theta) \right\| \overset{\text{a.s.}}{\to} 0.$$

Therefore, we have

$$\begin{aligned}
P\left\{ \bigcap_{N=1}^\infty \bigcup_{n=N}^\infty A_n \right\} &\le P\left\{ \bigcap_{N=1}^\infty \bigcup_{n=N}^\infty (B_n \cup C_n) \right\} \\
&\le P\left\{ \bigcap_{N=1}^\infty \bigcup_{n=N}^\infty B_n \right\} + P\left\{ \bigcap_{N=1}^\infty \bigcup_{n=N}^\infty C_n \right\} = 0.
\end{aligned}$$

That is,

$$\left\| \frac{1}{n} \sum_{t=1}^n \eta(X_t, \hat{\theta}_n) - J_2(\hat{\theta}_n) \right\| \overset{\text{a.s.}}{\to} 0. \tag{4.14}$$

Then from(4.13), (4.14), and the identities

$$J = J_1(\theta_\alpha) + J_2(\theta_\alpha), \qquad \hat{J} = J_1(\hat{\theta}_n) + \frac{1}{n} \sum_{t=1}^n \eta(X_t, \hat{\theta}_n),$$

we have

$$\begin{aligned}\|\hat{J} - J\| &\le \|J_1(\hat{\theta}_n) - J_1(\theta_\alpha)\| + \left\|\frac{1}{n}\sum_{t=1}^{n}\eta(X_t,\hat{\theta}_n) - J_2(\theta_\alpha)\right\| \\ &\le \|J_1(\hat{\theta}_n) - J_1(\theta_\alpha)\| + \left\|\frac{1}{n}\sum_{i=1}^{n}\eta(X_i,\hat{\theta}_n) - J_2(\hat{\theta}_n)\right\| + \|J_2(\hat{\theta}_n) - J_2(\theta_\alpha)\| \xrightarrow{\text{a.s.}} 0,\end{aligned}$$

where we have used the fact $\hat{\theta}_n \to \theta$ a.s., the continuity of $J_2(t)$, and Slutsky's theorem. This completes the proof. □

To prove Lemma 4.8, we need the following lemmas.

LEMMA 4.6. (1-dimensional case) *Let $\{X_t\}$ be a zero-mean real-valued stationary process such that*

$$\sup_t E|X_t|^{2(q\vee r)} < \infty \tag{4.15}$$

and

$$\sum_{k=1}^{\infty}\beta(k)^{1/p} < \infty, \tag{4.16}$$

where $q > 1$, $r > 1$ and $1/q + 1/r = 1 - 1/p$. Then the following hold:

(i) *The series $\sum_k \mathrm{Cov}(X_0, X_k)$ is absolutely convergent and has a nonnegative sum σ^2.*

(ii) *If $\{h_n\}$ is a sequence of positive integers satisfying the property in* (2.9), *then*

$$\sum_{k=-h_n}^{h_n}\hat{\gamma}_n(k) \xrightarrow{P} \sigma^2 \tag{4.17}$$

where $\hat{\gamma}_n(k) = n^{-1}\sum_{t=1}^{n-k} X_t X_{t+k}$.

PROOF. First, note that

$$\frac{1}{p} = 1 - \frac{1}{q} - \frac{1}{r} \le 1 - \frac{2}{q\vee r} \le 1 - \frac{2}{2(q\vee r)},$$

and consequently, (4.16) implies that $\sum_{k=1}^{\infty}\beta(k)^{1-2/2(q\vee r)} < \infty$. Therefore, from (4.15), (i) holds (cf. Bosq (1996), p. 32).

Let $\gamma(k) = \mathrm{Cov}(X_0, X_k)$ and note that

$$\left|\sum_{k=-h_n}^{h_n}\hat{\gamma}_n(k) - \sigma^2\right| \le \sum_{k=-h_n}^{h_n}|\hat{\gamma}_n(k) - \gamma(k)| + \sum_{|k|>h_n}|\gamma(k)|. \tag{4.18}$$

Since $\sum_k |\gamma(k)| < \infty$ and $h_n \to \infty$, the second term of the right hand side of (4.18) converges to 0 as $n \to \infty$. Let $\gamma_n^*(k) = n^{-1}\sum_{t=1}^{n} X_t X_{t+k}$. Then from the fact that $\hat{\gamma}_n(k) = (1 - k/n)\gamma_{n-k}^*(k)$, we have

$$\sum_{k=-h_n}^{h_n}|\hat{\gamma}_n(k) - \gamma(k)| \le \sum_{k=-h_n}^{h_n}|\gamma_{n-k}^*(k) - \gamma(k)| + \frac{\gamma(0)}{n}\sum_{k=-h_n}^{h_n}|k|. \tag{4.19}$$

Since $h_n/\sqrt{n} \to 0$, the second term of the right hand side of (4.19) is $o(1)$ as $n \to \infty$. Therefore, if the first term of the right hand side of (4.19) is $o_P(1)$, then (4.17) holds. Since

$$\left\| \sum_{k=-h_n}^{h_n} |\gamma^*_{n-k}(k) - \gamma(k)| \right\|_2 \le \sum_{k=-h_n}^{h_n} \|\gamma^*_{n-k}(k) - \gamma(k)\|_2$$
$$\le (2h_n + 1) \max_{-h_n \le k \le h_n} \|\gamma^*_{n-k}(k) - \gamma(k)\|_2,$$

it is enough to show that

$$\sup_k \|\gamma^*_n(k) - \gamma(k)\|_2 = O(1/\sqrt{n}). \tag{4.20}$$

However, by Davydov's inequality, we have that

$$\begin{aligned}
\|\gamma^*_n(k) - \gamma(k)\|_2^2 &= \frac{1}{n^2} E\left\{ \sum_{t=1}^{n} (X_t X_{t+k} - EX_t X_{t+k}) \right\}^2 \\
&\le \frac{1}{n^2} \sum_{t,s=1}^{n} |\operatorname{Cov}(X_t X_{t+k}, X_s X_{s+k})| \\
&\le \frac{1}{n^2} \sum_{t,s=1}^{n} 2p(2\beta(|t-s|))^{1/p} \|X_t X_{t+k}\|_q \|X_s X_{s+k}\|_r \\
&\le \frac{1}{n} \cdot 8p \cdot \|X_0\|_{2q}^2 \|X_0\|_{2r}^2 \cdot \sum_{k=1}^{\infty} \beta(k)^{1/p}
\end{aligned}$$

for all k. Then (4.20) is immediate in view of (4.15) and (4.16). □

LEMMA 4.7. *Let $\{X_t\}$ be a strictly stationary process and let $A(x, \theta)$ be a measurable function of x for all θ in some parameter space $\Theta \subset R^m$. Let $\hat{\theta}_n$ be a consistent estimator of $\theta_0 \in \Theta$, based on $X_1, \ldots, X_n$. Assume that*

(1) $EA(X, \theta_0)^2 < \infty$,

(2) $A(x, \theta)$ is differentiable on Θ for all x,

(3) There exists a function $B(x)$ such that $EB(X)^2 < \infty$ and $\sum_{i=1}^m |\partial A(x,\theta)/\partial\theta_i| \le B(x)$ for all x and θ,

(4) $\hat{\theta}_n - \theta_0 = O_P(1/\sqrt{n})$ as $n \to \infty$.

Then for a sequence $\{h_n\}$ satisfying the property in (2.9), as $n \to \infty$,

$$\frac{1}{n} \sum_{k=-h_n}^{h_n} \sum_{t=1}^{n-k} \{A(X_t, \hat{\theta}_n) A(X_{t+k}, \hat{\theta}_n) - A(X_t, \theta_0) A(X_{t+k}, \theta_0)\} = o_P(1).$$

PROOF. By the mean value theorem, we have

$$A(x, \theta) - A(x, \theta_0) = \partial A(x, \theta_0 + A(\theta - \theta_0))/\partial\theta \cdot (\theta - \theta_0)$$

for some $0 < u < 1$, and consequently,

$$|A(x, \theta) - A(x, \theta_0)| \le \|\partial A(x, \theta_0 + A(\theta - \theta_0))/\partial\theta\| \cdot \|\theta - \theta_0\| \le B(x) \cdot \|\theta - \theta_0\|$$

for all θ and x. Therefore, we have

$$\begin{aligned}
&|A(X_t,\hat\theta_n)A(X_{t+k},\hat\theta_n) - A(X_t,\theta_0)A(X_{t+k},\theta_0)| \\
&\quad \le |A(X_t,\hat\theta_n)||A(X_{t+k},\hat\theta_n) - A(X_{t+k},\theta_0)| + |A(X_t,\hat\theta_n) - A(X_t,\theta_0)||A(X_{t+k},\theta_0)| \\
&\quad \le \|\hat\theta_n - \theta_0\|(|A(X_t,\theta_0)|B(X_{t+k}) + B(X_t)|A(X_{t+k},\theta_0)|) \\
&\qquad + \|\hat\theta_n - \theta_0\|^2 B(X_t)B(X_{t+k}) \\
&\quad \le \frac{1}{2}\|\hat\theta_n - \theta_0\|(A(X_t,\theta_0)^2 + A(X_{t+k},\theta_0)^2 + B(X_t)^2 + B(X_{t+k})^2) \\
&\qquad + \frac{1}{2}\|\hat\theta_n - \theta_0\|^2(B(X_t)^2 + B(X_{t+k})^2),
\end{aligned}$$

and consequently,

$$\begin{aligned}
(4.21)\qquad &\left|\frac{1}{n}\sum_{k=-h_n}^{h_n}\sum_{t=1}^{n-k}\{A(X_t,\hat\theta_n)A(X_{t+k},\hat\theta_n) - A(X_t,\theta_0)A(X_{t+k},\theta_0)\}\right| \\
&\quad \le \|\hat\theta_n - \theta_0\|\cdot\frac{1}{2n}\sum_{k=-h_n}^{h_n}\sum_{t=1}^{n-k}(A(X_t,\theta_0)^2 + A(X_{t+k},\theta_0)^2 \\
&\qquad\qquad + B(X_t)^2 + B(X_{t+k})^2) \\
&\qquad + \|\hat\theta_n - \theta_0\|^2\cdot\frac{1}{2n}\sum_{k=-h_n}^{h_n}\sum_{t=1}^{n-k}(B(X_t)^2 + B(X_{t+k})^2).
\end{aligned}$$

From conditions (1), (3) and the strict stationarity of $\{X_t\}$, it follows that as $n \to \infty$,

$$\begin{aligned}
&E\left\{\frac{1}{2n\sqrt{n}}\sum_{k=-h_n}^{h_n}\sum_{t=1}^{n-k}(B(X_t)^2 + B(X_{t+k})^2)\right\} \\
&\quad \le E\left\{\frac{1}{2n\sqrt{n}}\sum_{k=-h_n}^{h_n}\sum_{t=1}^{n-k}(A(X_t,\theta_0)^2 + A(X_{t+k},\theta_0)^2 + B(X_t)^2 + B(X_{t+k})^2)\right\} \\
&\quad = \frac{1}{2n\sqrt{n}}\sum_{k=-h_n}^{h_n}\sum_{t=1}^{n-k}(EA(X_t,\theta_0)^2 + EA(X_{t+k},\theta_0)^2 + EB(X_t)^2 + EB(X_{t+k})^2) \\
&\quad \le \frac{2h_n+1}{\sqrt{n}}(EA(X,\theta_0)^2 + EB(X)^2) \to 0.
\end{aligned}$$

Since $\|\hat\theta_n - \theta_0\| = O_P(1/\sqrt{n})$, the term of the right hand side of (4.21) is $o_P(1)$. This completes the proof. □

LEMMA 4.8. *Under the assumptions of Theorem 2.4, $\hat K \to K$ in probability.*

PROOF. Let

$$\hat\gamma_n(\theta;k) = \frac{1}{n(1+\alpha)^2}\sum_{t=1}^{n-k}\frac{\partial V(\theta;X_t)}{\partial\theta}\cdot\frac{\partial V(\theta;X_{t+k})'}{\partial\theta}.$$

Note that for any $\lambda \in R^m$ with $\|\lambda\| = 1$, $\{\lambda' \partial V(\theta_\alpha; X_t)/\partial\theta\}$ is a zero mean real-valued stationary process and that conditions K1 and K2 imply conditions (4.15) and (4.16) for $p = q = r = 3$. Therefore, the conditions in Lemma 4.6 for $\{\lambda' \partial V(\theta_\alpha; X_t)/\partial\theta\}$ are fulfilled, and thus

$$\sum_{k=-h_n}^{h_n} \hat{\gamma}_n(\theta_\alpha; k) \xrightarrow{P} \lambda' K \lambda.$$

Since conditions A1–A5 with $r = 3$ and K1–K3 imply the conditions in Lemma 4.7, we have

$$\sum_{k=-h_n}^{h_n} \hat{\gamma}_n(\hat{\theta}_n; k) - \sum_{k=-h_n}^{h_n} \hat{\gamma}_n(\theta_\alpha; k) = o_P(1).$$

Therefore, by the Slutsky's theorem we have

$$\lambda' \hat{K} \lambda = \sum_{k=-h_n}^{h_n} \hat{\gamma}_n(\hat{\theta}_n; k) \xrightarrow{P} \lambda' K \lambda,$$

which establishes the lemma. □

5. Concluding remarks

In this paper, we constructed the cusum test statistic as proposed in Lee *et al.* (2003) based on the estimator minimizing density-based divergence measures. It was shown that the test statistic converges to the sup of a Brownian bridge under regularity conditions. For this, we verified the strong convergence result of the estimator. Through a simulation study, we have seen that the cusum test $T_{\alpha,n}$ with $\alpha > 0$ constitutes a robust test against outliers, while it still keeps the same efficiency as the MLE based cusum test when there are no outliers. The simulation result strongly supports the validity of the test. In actual practice, it may be an important issue to select an optimal α. One possible way is, as in the trimmed mean context, to choose an α to produce the smallest asymptotic variance of the minimum density power divergence estimator. See, for instance, Hong and Kim (2001). They only handled one dimensional parameter space, but the idea can be extended to multi-dimensional cases adopting the largest eigenvalue as a measure of the magnitude of asymptotic covariance matrices. However, taking account of possible parameter changes, it is not clear whether one can select an optimal α in all situations. Actually, a local power study for a robust cusum test has been conducted in Lee and Park (2001). According to their result, the power of the cusum test based on trimmed r.v.'s depends upon trimming portions and unknown density of a given data. Thus, a serious difficulty arises in choosing optimal trimming portions. From this example, one can also reason that obtaining an optimal α is not an easy task in our case. We recommend to use $\alpha = 0.1 \sim 0.2$ since in our simulation study, it keeps the efficient of the test when there are no outliers, while it makes a robust test against outliers. Here, we restricted ourselves to the case of the strong mixing process, but the robust cusum test has a great potential to be applied to various statistical models. We leave the task of extension to other models as a future study.

Acknowledgements

We like to thank the referee for the valuable comments. This research was supported (in part) by KOSEF through Statistical Research Center for Complex Systems at Seoul National University.

References

Basu, A. and Lindsay, B. G. (1994). Minimum disparity estimation for continuous models: Efficiency, distributions and robustness, *Annals of the Institute of Statistical Mathematics*, **46**, 683–705.

Basu, A., Harris, I. R., Hjort, N. L. and Jones, M. C. (1998). Robust and efficient estimation by minimizing a density power divergence, *Biometrika*, **85**, 549–559.

Beran, R. (1977). Minimum hellinger distance estimates for parametric models, *Annals of Statistics*, **5**, 445–463.

Bosq, D. (1996). *Nonparametric Statistics for Stochastic Processes; Estimation and Prediction*, Lecture Notes in Statistics, No. 10, Springer, New York.

Bustos, O. H. (1982). General M-estimates for contaminated pth-order autoregressive processes: Consistency and asymptotic normality, *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete*, **59**, 491–504.

Cao, R., Cuevas, A. and Fraiman, R. (1995). Minimum distance density-based estimation, *Computational Statistics and Data Analysis*, **20**, 611–631.

Chan, J. and Gupta, A. K. (2000). *Parametric Statistical Change Point Analysis*, Birkhäuser, Boston.

Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-point Analysis*, Wiley, New York.

Denby, L. and Martin, D. (1979). Robust estimation of the first-order autoregressive parameter, *Journal of the American Statistical Association*, **74**, 140–146.

Ferguson, T. S. (1996). *A Course in Large Sample Theory*, Chapman & Hall, London.

Fox, A. J. (1972). Outliers in time series, *Journal of the Royal Statistical Society. Series B*, **34**, 350–363.

Hong, C. and Kim, Y. (2001). Automatic selection of the tuning parameter in the minimum density power divergence estimation, *Journal of the Korean Statistical Society*, **30**, 453–465.

Inclán, C. and Tiao, G. C. (1994). Use of cumulative sums of squares for retrospective detection of changes of variances, *Journal of the American Statistical Association*, **89**, 913–923.

Lee, S. and Park, S. (2001). The cusum of squares test for scale changes in infinite order moving average processes, *Scandinavian Journal of Statistics*, **28**, 625–644.

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

Page, E. S. (1955). A test for change in a parameter occurring at an unknown point, *Biometrika*, **42**, 523–527.

Peligrad, M. (1986). Recent advances in the central limit theorem and its weak invariance principle for mixing sequences of random variables (a survey), *Dependence in Probability and Statistics* (eds. E. Eberlein and M. S. Taqqu), 193–223, Birkhäuser, Boston.

Simpson, D. G. (1987). Minimum hellinger distance estimation for the analysis of count data, *Journal of the American Statistical Association*, **82**, 802–807.

Tamura, R. N. and Boos, D. D. (1986). Minimum hellinger distance estimation for multivariate location and covariance, *Journal of the American Statistical Association*, **81**, 223–239.