Test for Parameter Change based on the Estimator Minimizing Density-based Divergence Measures

Okyoung Na [1], Sangyeol Lee[2] and Siyun Park[3]

ABSTRACT

In this paper we consider the problem of parameter change based on the cusum test proposed by Lee et al. (2003). The cusum test statistic is constructed utilizing 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 there are outliers.

Keywords. Cusum test, density-based divergence measures, robust property, weak convergence, Brownian bridge.
Running title. Test for parameter change.

1. Introduction

The problem of parameter change 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 suitable in the situation that no assumptions are imposed on the underlying distribution of observations. For instance, any parametric approach is not directly applicable to the test for a change in the autocorrelations of stationary time series. To overcome such a problem, Lee et al. devised a cusum test adopting an idea 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 ARCH parameters. A merit of the cusum test is that it can test the existence of a change point and at the same time detect the locations of change points. But the most important advantage is that any estimators can be employed to construct the cusum test as far as

[1]Ph.D. Student, Department of Statistics, Seoul National University.
[2]Associate Professor, Department of Statistics, Seoul National University.
[3]Ph.D, Marketing Consulting Team, SK Corporation.

they satisfy regularity conditions. For instance, when there is a concern about outliers, a robust estimator can utilized. This is not a task that can be achieved immediately by the parametric approach.

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 divergence methods, such as Beran (1977), Tamura and Boos (1986) and Simpson (1987), which use the Hellinger distance, and Basu and Lindsey (1994) and Cao, Cuevas, and Fairman (1994), the new method has a merit of not requiring any smoothing method. In this case, one can avoid 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) under model conditions. Therefore, their estimator can be viewed as a good alternative to the MLE in terms of efficiency and robustness. Seemingly, this result can be reflected in constructing 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: the same phenomenon is anticipated to occur in other situations, we are led to consider a robust cusum test for the general parameter case. Here, 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 dependent observations. Thus in our set-up, the observations are assumed to be dependent and to satisfy the 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 a maximum of 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 MLE. In Section 4, we provide the proofs. Finally in Section 5, we provide concluding remarks.

2. Main result

BHHJ introduced a family of density power divergences d_α, $\alpha \geq 0$;

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

where g and f are density functions.

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. For any given α, they defined the minimum density power divergence functional $T_\alpha(\cdot)$ by the requirement that for every G in $\mathcal{G}$,

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

where g is the density of G.

Let $\hat{\theta}_{\alpha,n}$ be the minimum density power divergence estimator based on $X_1, \ldots, X_n$;

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

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

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

When $X_1, \ldots, X_n$ are iid with distribution G with corresponding density g, BHHJ showed that under the below conditions with $r = 3$, $\hat{\theta}_{\alpha,n}$ 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.

Conditions

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

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

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

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

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

A5. There exists a nonsingular matrix J, defined by

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

Now, let us consider the ergodic and strictly stationary process $\{X_t; t = 1, 2, \ldots\}$. We assume here that $\theta_\alpha = T_\alpha(G)$ exists and is unique. Define the estimator $\hat{\theta}_{\alpha,n}$ of θ_α as the minimizer of $H_{\alpha,n}(\theta)$ like (2.1). In fact, the estimator is obtained by solving the estimating equations

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

To prove Theorem 1.1, we need the following lemma.

Lemma 2.1. *Let $X_1, X_2, \ldots$ be strictly stationary and ergodic. If*

1. *Θ is compact,*
2. *$A(x, \theta)$ is continuous in θ for all x,*
3. *There exists a function $B(x)$ such that $EB(X) < \infty$ and $|A(x,\theta)| \le B(x)$ for all x and θ,*

then

$$P\left\{ \lim_{n\to\infty} \sup_{\theta\in\Theta} \left| \frac{1}{n} \sum_{t=1}^{n} A(X_t, \theta) - a(\theta) \right| = 0 \right\} = 1, \tag{2.2}$$

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

$$P\left\{ \hat{\theta}_n \to \theta^0, n \to \infty \right\} = 1 \tag{2.3}$$

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

Theorem 2.1. ***(Strong consistency)***

Assume that Conditions A1 - A4 hold with $r = 1$. Then there exists a sequence $\{\hat{\theta}_{\alpha,n}\}$ such that

$$(i) \quad U_{\alpha,n}(\hat{\theta}_{\alpha,n}) = 0 \text{ for sufficiently large } n, \tag{2.4}$$

$$(ii) \quad P\left\{ \hat{\theta}_{\alpha,n} \to \theta_\alpha, \text{ as } n \to \infty \right\} = 1. \tag{2.5}$$

Next consider the limit distribution of $\hat{\theta}_{\alpha,n}$.

Assume that Conditions A1-A5 hold. Since (2.4) 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

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

where R_n is the $m \times m$ matrix whose (i,j)-th component is

$$R_n^{ij} := -\frac{1}{1+\alpha}\left\{ \frac{\partial^2 H_{\alpha,n}(\theta_\alpha)}{\partial\theta_i \partial\theta_j} + \frac{1}{2}\sum_{k=1}^{m} \frac{\partial^3 H_{\alpha,n}(\theta^*_{\alpha,n})}{\partial\theta_i\partial\theta_j\partial\theta_k}\left(\hat{\theta}^k_{\alpha,n} - \theta^k_\alpha\right)\right\} \tag{2.6}$$

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

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

where $\Delta_n = J^{-1}(J - R_n)\left(\hat{\theta}_{\alpha,n} - \theta_\alpha\right)$, and consequently

$$\frac{[ns]}{\sqrt{n}}\left(\hat{\theta}_{\alpha,[ns]} - \theta_\alpha\right) = J^{-1} \cdot \frac{[ns]}{\sqrt{n}} U_{\alpha,[ns]}(\theta_\alpha) + \frac{[ns]}{\sqrt{n}} \Delta_{[ns]}. \tag{2.7}$$

Suppose that there exists a positive definite and symmetric matrix K such that

$$\frac{[ns]}{\sqrt{n}} U_{\alpha,[ns]}(\theta_\alpha) \Rightarrow K^{1/2} W(s) \tag{2.8}$$

in the $D^m[0,1]$ space, where W denotes a m-dimensional standard Brownian motion. In view of (2.7) and (2.8), if it holds that

$$\max_{1 \le k \le n} \frac{k}{\sqrt{n}} \Delta_k = o_P(1), \tag{2.9}$$

we obtain the convergence result as in Theorem 2.2.. The following is the main result of this section.

Theorem 2.2. *(The functional central limit theorem)*
Assume that Conditions A1-A5 hold with r=3. Also, suppose that

1. *$\{X_t\}$ is α-mixing of size $-\gamma/(\gamma-2)$ for $\gamma > 2$, i.e., $\sum_{n=1}^{\infty} \alpha(n)^{1-2/\gamma} < \infty$.*
2. *$E|\partial V_\alpha(\theta_\alpha; X)/\partial\theta_i|^\gamma < \infty$ for $i = 1, \ldots, m$.*
3. *$nK_n \to K$ for some positive definite and symmetric matrix K, where K_n is the covariance matrix of $U_{\alpha,n}(\theta_\alpha)$.*

Then we have

$$\frac{[ns]}{\sqrt{n}}\left(\hat{\theta}_{\alpha,[ns]} - \theta_\alpha\right) \Rightarrow J^{-1}K^{1/2}W(s).$$

The following lemma is concerned with the negligibility of Δ_k.

Lemma 2.2. *Under the assumptions of Theorem 2.2.,*

$$\max_k \left(n^{-1/2} k \,\|\Delta_k\|\right) = o_P(1).$$

Finally, we consider the problem of testing

H_0 : θ_α does not change over $X_1, \ldots, X_n$. vs.
H_1 : not H_0.

As a consequence of Theorem 2.2., one can readily construct the tests for the null hypothesis. The following is the result.

Theorem 2.3. *Define the test statistic* $T^0_{\alpha,n}$ *by*

$$T^0_{\alpha,n} := \max_{m\le k\le n} \frac{k^2}{n}\left(\hat{\theta}_{\alpha,k}-\hat{\theta}_{\alpha,n}\right)' JK^{-1}J\left(\hat{\theta}_{\alpha,k}-\hat{\theta}_{\alpha,n}\right). \tag{2.10}$$

Suppose that Assumptions of Theorem 2.2. hold. Then, under H_0,

$$T^0_{\alpha,n} \Rightarrow \sup_{0\le s\le 1} \|W^o(s)\|^2.$$

We reject H_0 *if* $T^0_{\alpha,n}$ *is large.*

Since J and K are unknown, we should replace them by consistent estimators $\hat{J}$ and $\hat{K}$. First, note that

$$J = \int u_{\theta_\alpha}(z)u_{\theta_\alpha}(z)' f^{1+\alpha}_{\theta_\alpha}(z)dz + \int \left(i_{\theta_\alpha}(z) - \alpha u_{\theta_\alpha}(z)u_{\theta_\alpha}(z)'\right)\left(g(z)-f_{\theta_\alpha}(z)\right) f^{\alpha}_{\theta_\alpha}(z)dz,$$

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

$$\begin{aligned}\hat{J} &= \int \left\{(1+\alpha)u_{\hat{\theta}_{\alpha,n}}(z)u_{\hat{\theta}_{\alpha,n}}(z)' - I_{\hat{\theta}_{\alpha,n}}(z)\right\} f^{1+\alpha}_{\hat{\theta}_{\alpha,n}}(z)dz \\ &+ \frac{1}{n}\sum_{t=1}^{n}\left\{I_{\hat{\theta}_{\alpha,n}}(X_t) - \alpha u_{\hat{\theta}_{\alpha,n}}(X_t)u_{\hat{\theta}_{\alpha,n}}(X_t)'\right\} f^{\alpha}_{\hat{\theta}_{\alpha,n}}(X_t),\end{aligned}$$

then $\hat{J}$ converges to J almost surely.

Let $DV_\alpha(\theta;x)$ be the $m\times 1$ vector of partial derivatives of $V_\alpha(\theta;x)$ with respect to θ. Under the assumptions of Theorem 2.2., note that

$$K = \sum_{k=-\infty}^{\infty} Cov\left(\frac{DV_\alpha(\theta_\alpha, X_0)}{1+\alpha}, \frac{DV_\alpha(\theta_\alpha, X_k)}{1+\alpha}\right),$$

due to theorem 1.5 in Bosq(1996, page, 32). Assume that

- **K1.** $E\|DV_\alpha(\theta_\alpha;X)\|^6 < \infty$,
- **K2.** $\sum_{n=1}^{\infty}\alpha(n)^{1/3} < \infty$,
- **K3.** there exists a function $M(x)$ with $EM(X)^2<\infty$ such that $\|\partial^2 V_\alpha(\theta;x)/\partial\theta^2\| \le M(x)$ for all $\theta\in\Theta$ and $x\in\mathcal{X}$.

Then $\hat{K}\to K$ in probability, where

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

and $\{h_n\}$ is a sequence of positive integers such that $h_n\to\infty$ and $h_n/\sqrt{n}\to 0$.

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

$$T_{\alpha,n} := \max_{m\le k\le n} \frac{k^2}{n}\left(\hat{\theta}_{\alpha,k}-\hat{\theta}_{\alpha,n}\right)' \hat{J}\hat{K}^{-1}\hat{J}\left(\hat{\theta}_{\alpha,k}-\hat{\theta}_{\alpha,n}\right). \tag{2.11}$$

Suppose that Assumptions of Theorem 2.2. and K1-K3 hold. Then, under H_0,

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

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

REFERENCES

Base, A. & Lindsay, B. G. (1994). Minimum disparity estimation for continuous models : efficiency, distributions and robustness. *Ann. Inst. Statist. Math.* **48**, 6833-705.

Basu, A., Harris, I. R., Hjort, N. L. & 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. *Ann. Statist.* **5**, 445-463.

Bosq, D. (1996). *Nonparametric Statistics for Stochastic Processes; Estimation and Prediction.* Lecture Notes in Statistics. Springer, New York.

Cao, R., Cuevas, A. & Fraiman, R. (1995). Minimum distance density-based estimation. *Comput. Statist. Data Anal.* **20**, 611-631.

Csörgő, M. & Horváth, L. (1997). *Limit theorems in change-point analysis.* Wiley, New York.

Ferguson, T. S. (1996). *A course in large sample theory.* Chapman & Hall.

Inclán, C. & Tiao, G. C. (1994). Use of cumulative sums of squares for retrospective detection of changes of variances. *J. Amer. Statist. Assoc.* **89**, 913-923.

Lee, S. & Park, S. (2001). The cusum of squares test for scale changes in infinite order moving average processes. *Scand. J. Statist.* **28**, 625-644.

Lee, S., Ha, J., Na, O. & Na, S. (2003) The Cusum Test for Parameter Change in Time Series Models. *To appear in Scand. J. Statist.*

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

Simpson, D. G. (1987). Minimum hellinger distance estimation for the analysis of count data. *J. Amer. Statist. Assoc.* **82**, 802-807.

Tamura, R. N. & Boos, D. D. (1986). Minimum hellinger distance estimation for multivariate location and covariance. *J. Amer. Statist. Assoc.* **81**, 223-239.