

Ratio Test for Changes Heavy Index Based on M-Estimation under the Background of Big Data

Meiting Liu*, Hao Jin, Yangyi Cheng, Zini Wang

Department of Sciences, Xi'an University of Science and Technology, Xi'an, Shaanxi, China

Abstract

With the development of the big data age, the demand of modern emerging disciplines for the fine quantitative tail characteristics of financial data analysis is constantly developing, so the research method of tail index change point in the thick tail sequence is particularly important. In this paper, we extend the Ratio statistic proposed by Kim to the test of change point of tail index in infinite variance observations. The null distribution of the statistic and its consistency under alternative hypothesis were obtained. Prior with the least squares estimation method used in previous articles, we used robust M estimation to estimate unknown parameter, making parameter estimation in the model more accurate. The Monte Carlo numerical simulation shows that our test works well.

Keywords

Heavy-Tailed, Tail Index, Ratio Test, M-Estimation, Big Data

1. Introduction

In the age of big data, pre-detecting structural changes allows us to better interpret data, predict data more accurately and avoid risk. Therefore, the study of structural change points has attracted wide attention from many scholars. Tail index change point is one of the core contents of change point research, which is widely used in practical life. Such as finance, hydrology, communication engineering and other fields. Several tail index change-point tests have been proposed in the past decade. Quintos, Fan, and Phillips [1] employed three tests to verify tail index change-points in independent samples and ARCH models. And later Kim and Lee [4] studied the tail index change-point test based on autoregressive residuals.

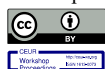
Reasonable estimates of the parameters in the models are also momentous. The least squares estimation is the most commonly used point estimation method in parameter estimation, but data with anomalies or strong influence points are often encountered in practical problems, while the least-squares estimator is sensitive to the emergence of outliers in the data. Therefore, in order to eliminate or reduce the effect of outliers, the estimator is required to have a robustness. While M estimation is a commonly used class of estimates in current robust estimation methods, and there have been a series of extensive studies on M estimation. Richard, Keight and Liu [10] used M estimation to study the estimates of autoregressive parameters in heavy-tailed sequences, and later Keith Knight proposed the limit theory of M estimation in heavy-tailed sequences.

In this paper, We employ robust M estimation to estimate unknown parameters in the model and obtain the asymptotic behaviour of the estimator. To define the test statistic, we cite the test statistic based on detecting change point in the persistence by Kim[8] to test the change point of heavy index.

This paper is arranged as follows. Section 2 we show the model and assumptions. Section 3 represent test statistic and asymptotic distribution. Section 4 the Monte Carlo simulation is made. Section 5 summarize this paper.

ICBASE2022@3rd International Conference on Big Data & Artificial Intelligence & Software Engineering, October 21-23, 2022, Guangzhou, China

*Corresponding author: Lmt137341@126.com (Meiting Liu)



© 2022 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

2. Model and Assumption

Our model is presented as follows:

$$y_t = u + \varepsilon_t, t = 1, 2, \dots, T \quad (1)$$

where μ is unknown parameter and we employ the M estimation to estimate μ . This paper is based on the following assumption:

(A₁) $\{\varepsilon_t\}$ is an i.i.d. sequence of r.v.'s with $E(\varepsilon_t^2) = \infty$ which are in the domain of attraction of a stable law of order $\alpha \in (0, 2]$. The normal distribution corresponds to $\alpha = 2$.

(A₂) $\rho' = \psi(\cdot); E(\psi(\eta_t)) = 0; E(|\psi(\eta_t)|^{2+\gamma}) < \infty$, for some $\gamma > 1$.

(A₃) $\psi'(\cdot)$ is Lipschitz continuous; $0 < |E(\psi'(\eta_t))| < \infty$.

Lemma.1. if the assumption (A₁) holds, then

$$\left(a_T^{-1} \sum_{t=1}^{[T\tau]} \varepsilon_t, a_T^{-2} \sum_{t=1}^{[T\tau]} \varepsilon_t^2 \right) \Rightarrow (U_k(r), \int_0^r (dU_k)^2) \quad (2)$$

Where $a_T = T^{1/k}L(T)$, for slowly varying function, the $U_k(r)$ is a standard stable process with index k and the characteristic function of $U_k(1)$ has the form $e^{-c|s|^k}$ where

$$c = \begin{cases} \Gamma(1-k) \cos(\pi k/2), & k \neq 1 \\ \pi/2, & k = 1 \end{cases} \quad (3)$$

Moreover, $U_k(r) = r^{1/k}U_k(1)$.

Remark 1. Specific reference Phillips [13].

Lemma.2. If assumption (A₁) – (A₃) holds, then

$$T^{1/2}(\hat{\mu}_1 - \mu) \rightarrow \frac{B(\tau)}{\tau \cdot E(\psi'(\varepsilon_t))}, t = 1, 2, \dots, [T\tau] \quad (4)$$

$$T^{1/2}(\hat{\mu}_2 - \mu) \rightarrow \frac{B(1)-B(\tau)}{(1-\tau) \cdot E(\psi'(\varepsilon_t))}, t = [T\tau] + 1, [T\tau] + 2, \dots, T \quad (5)$$

where $\hat{\mu}_1, \hat{\mu}_2$ represents M estimation based on respectively $y_1, y_2, \dots, y_{[T\tau]}$ and $y_{[T\tau]+1}, \dots, y_T$. $B(\tau)$ is Brownian motion.

Proof of Lemma.2.

We consider estimates $\hat{\mu}_1$ defined by minimizing $\sum_{t=2}^n \rho(y_t - \mu)$. Define the process

$$Z_n(v) = \sum_{t=1}^{[T\tau]} \left(\rho\left(\varepsilon_t + vT^{-\frac{1}{2}}\right) - \rho(\varepsilon_t) \right),$$

and note that \hat{v} which minimizes Z_n is simply $T^{\frac{1}{2}}(\hat{\mu}_1 - \mu)$. Then we using the Taylor series expansion of each summand of Z_n around $v = 0$, we get

$$Z_n(v) = T^{-\frac{1}{2}}v \sum_{t=1}^{[T\tau]} \psi(\varepsilon_t) + \frac{1}{2}T^{-1}v^2 \sum_{t=1}^{[T\tau]} \psi'(\varepsilon_t^*),$$

where ε_t^* lies between ε_t and $\varepsilon_t + T^{-\frac{1}{2}}v$. Using the fact that ψ' is Lipschitz-continuous $|\psi'(\varepsilon_t) - \psi'(\varepsilon_t^*)| \leq K \left| T^{-\frac{1}{2}}v \right|$, and reference Keith Knight[11], we can get

$$T^{-\frac{1}{2}} \sum_{i=1}^{[T\tau]} \psi(\eta_i) \rightarrow B(\tau), T^{-\frac{1}{2}} \sum_{i=[T\tau]+1}^T \psi(\eta_i) \rightarrow (B(1) - B(\tau)).$$

$$T^{-1}v^2 \sum_{t=1}^{[T\tau]} |\psi'(\varepsilon_t) - \psi'(\varepsilon_t^*)| \leq v^2 T^{-1} \cdot [T\tau] \cdot k \left| T^{-\frac{1}{2}}v \right| \xrightarrow{p} 0$$

To see this, note that $\sup_{0 \leq \tau \leq 1} \left| T^{-1} \sum_{t=1}^{[T\tau]} (\psi'(\varepsilon_t) - E(\psi'(\varepsilon_t))) \right| \xrightarrow{p} 0$,

Thus, the finite dimensional distributions of $Z_n(\cdot)$ converge weakly to those of $Z(\cdot)$,

where $Z(v) = vB(\tau) + \frac{1}{2}v^2 \tau E(\psi'(\varepsilon_t))$, let $Z'(v) = 0$, we get $v = T^{1/2}(\hat{\mu}_1 - \mu) \rightarrow \frac{B(\tau)}{\tau \cdot E(\psi'(\varepsilon_t))}$,

In the same way, $t = T\tau + 1, T\tau + 2, \dots, T, v = T^{1/2}(\hat{\mu}_2 - \mu) \rightarrow \frac{B(1)-B(\tau)}{(1-\tau) \cdot E(\psi'(\varepsilon_t))}$.

specific reference Keith Knight[10].

Now, we take into account the following null hypothesis and alternative hypothesis:

$H_0:k$ is constant
 $H_1:k = k_1 I_{\{t \leq k^*\}} + k_2 I_{\{t > k^*\}}$

3. Test Statistic and Asymptotic Distribution

We consider the Ratio statistic presented by Kim:

$$\Xi_T(\tau) = \frac{([1-\tau]T)^{-1} \sum_{t=[\tau T]+1}^T (\sum_{i=[\tau T]+1}^t \hat{\epsilon}_{1,i})^2}{[\tau T]^{-1} \sum_{t=1}^{[\tau T]} (\sum_{i=1}^t \hat{\epsilon}_{0,i})^2} \quad (6)$$

where $\hat{\epsilon}_{0,i}$ and $\hat{\epsilon}_{1,i}$ represents M estimation residuals based on respectively $y_1, y_2, \dots, y_{[\tau T]}$ and $y_{[\tau T]+1}, \dots, y_T$. According to ratio statistic, a Maximum Chow-type test is proposed

$$\Xi_T = \max_{\tau \in T} \Xi_T(\tau), T \in (0,1) \quad (7)$$

Theorem 3.1 Suppose the assumption $A_1 - A_3$ holds, then under the null hypothesis,

$$\Xi_T(\tau) \Rightarrow \frac{(1-\tau)^{-1} \int_{\tau}^1 G_1^2(r-\tau) dr}{\tau^{-1} \int_0^{\tau} G_1^2(r) dr} \quad (8)$$

And $\Xi_T \Rightarrow \Xi_T(\tau), G_1(r) = U_k(r), r \in (0, \tau); G_1(r-\tau) = U_k(r) - U_k(\tau), r \in (\tau, 1)$

Proof of Theorem 3.1

According to Lemma.1, Lemma.2, let $t = Tr$, we have

$a_T^{-1} \sum_{i=1}^{Tr} \hat{\epsilon}_{0,i} = a_T^{-1} \sum_{i=1}^{Tr} (\epsilon_i - (\hat{u}_1 - u)) \Rightarrow U_k(r) = G_1(r)$, so

$[\tau T]^{-1} \sum_{t=1}^{[\tau T]} (\sum_{i=1}^t \hat{\epsilon}_{0,i})^2 \Rightarrow \tau^{-1} \int_0^{\tau} G_1^2(r) dr$.

Similarly, $a_T^{-1} \sum_{i=T\tau+1}^T \hat{\epsilon}_{1,i} = a_T^{-1} \sum_{i=T\tau+1}^T (\epsilon_i - (\hat{u}_2 - u)) \Rightarrow U_k(r-\tau) = G_1(r-\tau)$.

Finally, we get the limit distribution of the statistics under the null hypothesis

$$\Xi_T(\tau) \Rightarrow \frac{(1-\tau)^{-1} \int_{\tau}^1 G_1^2(r-\tau) dr}{\tau^{-1} \int_0^{\tau} G_1^2(r) dr}$$

Theorem 3.2 If assumption $A_1 - A_3$ holds, and τ^* is the break time, then under the alternative hypothesis, $(1)k_1 > k_2$ and $0 < \tau \leq \tau^*$, we have

$$\Xi_T = \max_{\tau \in T} \Xi_T(\tau) = O_p\left(\frac{a_{T_2}^2}{a_{T_1}^2}\right), \text{ As } T \rightarrow \infty, \Xi_T \rightarrow \infty \quad (9)$$

For $\tau^* < \tau < 1$, we have $\Xi_T = \max_{\tau \in T} \Xi_T(\tau) = O_p(1)$.

(2) $k_1 < k_2$ and $0 < \tau \leq \tau^*$, we have $\Xi_T^{-1} = \max_{\tau \in T} \Xi_T^{-1}(\tau) = O_p(1)$,

For $\tau^* < \tau < 1$, we have

$$\Xi_T^{-1} = \max_{\tau \in T} \Xi_T^{-1}(\tau) = O_p\left(\frac{a_{T_1}^2}{a_{T_2}^2}\right) \rightarrow \infty, \text{ As } T \rightarrow \infty, \Xi_T^{-1} \rightarrow \infty \quad (10)$$

$a_{T_1} = T^{1/k_1} L(T), a_{T_2} = T^{1/k_2} L(T), L$ is slowly function.

Proof of Theorem 3.2

We first consider the case (i) $k_1 > k_2, 0 < \tau \leq \tau^*$, similar to Proof of Theorem 3.1, the denominator is $[\tau T]^{-1} \sum_{t=1}^{[\tau T]} (\sum_{i=1}^t \hat{\epsilon}_{0,i})^2 = O_p(a_{T_1}^2)$. The numerator is

$$([1-\tau]T)^{-1} \left\{ \sum_{t=[\tau T]+1}^{[\tau^* T]} (\sum_{i=[\tau T]+1}^t \hat{\epsilon}_{0,i})^2 + \sum_{t=[\tau^* T]+1}^T (\sum_{i=[\tau T]+1}^t \hat{\epsilon}_{1,i})^2 \right\} = O_p(a_{T_1}^2) + O_p(a_{T_2}^2).$$

When $k_1 > k_2, \Xi_T(\tau) = O_p\left(\frac{a_{T_2}^2}{a_{T_1}^2}\right) \rightarrow \infty$. The case $\tau^* < \tau < 1$,

the denominator is $[\tau T]^{-1} \left\{ \sum_{t=1}^{[\tau^* T]} (\sum_{i=1}^t \hat{\epsilon}_{0,i})^2 + \sum_{t=[\tau^* T]+1}^{[\tau T]} (\sum_{i=1}^t \hat{\epsilon}_{1,i})^2 \right\} = O_p(a_{T_1}^2) + O_p(a_{T_2}^2)$,

and the numerator is $([1-\tau]T)^{-1} \sum_{t=[\tau T]+1}^T (\sum_{i=[\tau T]+1}^t \hat{\epsilon}_{1,i})^2 = O_p(a_{T_2}^2)$.

Above all, $\Xi_T = \max_{\tau \in T} \Xi_T(\tau) = O_p\left(\frac{a_{T_2}^2}{a_{T_1}^2}\right)$. As $T \rightarrow \infty$, $\Xi_T \rightarrow \infty$.

Then the case (ii) $k_1 < k_2$, similar to the case (i), The proof is omitted .

Theorem 3.3 If Assumption $A_1 - A_3$ holds, then under the alternative hypothesis, if $k_1 \neq k_2$, we have

$$\Xi_T^* = \max\{\Xi_T, \Xi_T^{-1}\}, \text{ As } T \rightarrow \infty, \Xi_T^* \rightarrow \infty. \quad (11)$$

4. Monte Carlo simulation

This section we use the Monte Carlo numerical simulation method to verify the effectiveness of our Ratio test. We obtain the empirical level value and empirical potential function value . Consider the following data generation process: $y_t = u + \varepsilon_t$, $t = 1, 2, \dots, T$. The sample size is $T = 200, 500, 1000$. heavy-tailed index $\{\kappa = 0.6, 0.8, 0.9, 1.6, 1.8, 1.9\}$. The test was repeated 2000 times and significance level $\alpha = 0.05$. The original statistic and the inverted statistic are identically distributed, so we only give the critical value of the original statistic.

Table 1

Critical values of Maximum-Chow

k		0.6	0.8	0.9	1.6	1.8	1.9
T=200	$\Xi_T(\tau)$	70.6102	64.2800	55.7735	18.3333	16.1118	14.4291
T=500	$\Xi_T(\tau)$	70.3438	60.0030	59.5982	14.8721	18.5255	13.9622
T=1000	$\Xi_T(\tau)$	71.2121	68.7985	52.4784	17.0133	18.4640	13.9873

Table 2

Empirical size of Maximum-Chow

k		0.6	0.8	0.9	1.6	1.8	1.9
T=200	$\Xi_T(\tau)$	0.0470	0.0500	0.0510	0.0470	0.0500	0.0500
T=500	$\Xi_T(\tau)$	0.0490	0.0510	0.0500	0.0460	0.0510	0.0520
T=1000	$\Xi_T(\tau)$	0.0498	0.0499	0.0510	0.0510	0.0490	0.0486

Table 3

power experience of Maximum-Chow

T	$k_1 \rightarrow k_2$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$k_1 \rightarrow k_2$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$
200	1.9 \rightarrow 1.8	0.2260	0.2050	0.2720	0.9 \rightarrow 0.8	0.2230	0.2040	0.2300
	1.9 \rightarrow 1.6	0.2460	0.2640	0.3660	0.9 \rightarrow 0.6	0.2240	0.2620	0.2940
	1.8 \rightarrow 1.6	0.2090	0.2130	0.2330	0.8 \rightarrow 0.6	0.2050	0.2110	0.2290
	1.8 \rightarrow 1.5	0.2150	0.2220	0.2450	0.8 \rightarrow 0.5	0.2120	0.2260	0.2330
500	1.9 \rightarrow 1.8	0.2320	0.2480	0.2820	0.9 \rightarrow 0.8	0.2130	0.2210	0.2200
	1.9 \rightarrow 1.6	0.2560	0.2670	0.3690	0.9 \rightarrow 0.6	0.2330	0.2580	0.2850
	1.8 \rightarrow 1.6	0.2360	0.2540	0.2940	0.8 \rightarrow 0.6	0.2540	0.2590	0.2860
	1.8 \rightarrow 1.5	0.2680	0.2720	0.2950	0.8 \rightarrow 0.5	0.2580	0.2620	0.2880
1000	1.9 \rightarrow 1.8	0.2350	0.2890	0.2990	0.9 \rightarrow 0.8	0.2230	0.2520	0.2610
	1.9 \rightarrow 1.6	0.2590	0.2690	0.3740	0.9 \rightarrow 0.6	0.2600	0.2640	0.2680
	1.8 \rightarrow 1.6	0.2660	0.2670	0.2950	0.8 \rightarrow 0.6	0.2490	0.2540	0.2640
	1.8 \rightarrow 1.5	0.2690	0.2900	0.2990	0.8 \rightarrow 0.5	0.2500	0.2620	0.2980

Table 1 is the critical value of the statistic, we can see that with the increase of the sample size, the critical value gradually stabilizes, Table 2 is the rejection rate of the statistic under the null hypothesis, and the empirical size is close to the significance level 5%.

Table 3 is the empirical powers under the alternative hypothesis, we consider $\tau = 0.3, 0.5, 0.7$, from the table we can get that :power increases as the jump amplitude increases; when the jump amplitude is constant, power decreases as τ increases and as the sample size increases, because the larger the sample size T , the more dispersed the statistics are. In general, the larger the sample size, the greater the jump amplitude, and the smaller the τ , the better the statistical test effect. Furthermore $k_1 < k_2$, We can get the same results. For reasons of space, I will not repeat them here.

5. Conclusions

In this paper, we investigate the heavy index change point in thick tail sequence, and this theme is closely related to the big data setting and financial market. we adopt M-estimate to estimate unknown parameter, and absorb to the statistic proposed by Kim, obtained the limit distribution of the statistics under the null hypothesis and its consistency under alternative hypothesis. Numerical simulations verify that our statistics work very well.

6. References

- [1] Quintos, C., Fan, Z., & Phillips, P. C. B. (2001). Structural change tests in tail behaviour and the Asian crisis. *Review of Economic Studies*, 68, 633–663.
- [2] Relevant parameter changes in structural break models. *Journal of Econometrics*, 2020, 217(1): 46-78.
- [3] Peluso, S., S. Chib, and A. Mira. 2019. “Semiparametric Multivariate and Multiple Change-point Modelling.” *Bayesian Analysis* 14 (3): 727–751.
- [4] Kim, M. & S. Lee (2012) Change point test of tail index for autoregressive processes. *Journal of the Korean Statistical Society* 41, 305–312.
- [5] Page, E.S. Continue inspection schemes, *Biometrika*, 1954, 42(1): 100-114.
- [6] Chen J, A. K. Gupta. *Parametric Statistical Change Point Analysis*. Boston: Birkha-user, 2000.
- [7] Kim, J.Y., Detection of change in persistence of a linear time series. *Journal of Econometrics*, 2000, 95:97-116.
- [8] Kim, J.Y., Corrigendum to “Detection of change in persistence of a linear time series”. *Journal of Econometrics*, 2002, 109: 389-392.
- [9] GAMAGE, RAMADHA D. PIYADI, NING, WEI. Empirical likelihood for change point detection in autoregressive models. 2021(1).
- [10] Davis Richard A., Knight Keith, Liu Jian. M-estimation for autoregressions with infinite variance. *Stochastic Processes and their Applications*, 1992, 40(1):
- [11] Keith Knight. Limit Theory for M-Estimates in an Integrated Infinite Variance Process. *Econometric Theory*, 1991, 7(2):
- [12] Maryam Sohrabi, Mahmoud Zarepour. Asymptotic theory for M-estimates in unstable AR(p) processes with infinite variance innovations. *Journal of Statistical Planning and Inference*, 2019, 198:
- [13] Phillips P C B 1990 *Econometric Theory* 6 44-62.