View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by International Institute for Science, Technology and Education (IISTE): E-Journals

ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.3, No.1, 2013

Multiple Change Points Detection in the Mean of Observations with Time Varying Variances: An MCMC Approach

Reza Habibi
Department of Statistics, Central Bank of Iran, Tehran, Iran

Abstract. This paper is concerned with multiple change points detection in the mean of observations with time varying variances. The posterior distribution is estimated using a MCMC method and the hyper-parameters are estimated via SAEM algorithm. The simulation results are also given.

Keywords: Change point; GARCH models; MCMC; SAEM algorithm; Weighted energy function

1 Introduction. The stability of parameters of a statistical model over time is a necessary condition for making inference. The validity of predictions and interpretations depends on the model stability. In the quality control setting, when the stability of model assumption is violated the underlying process is out of control. Therefore, it is very important for a researcher to know if the parameters of the statistical model are constant at least within the given sample. Because of this need, the change point analysis should be performed in a specified inferential problem. Page (1954) first proposed change point problem in the context of quality control. So far, this problem has been received considerable attentions in statistical literatures. Some excellent references in this field are Basseville and Nikiforov (1993), Brodsky and Darkhovsky (1993), Csorgo and Horvath (1997) and Chen and Gupta (2001).

The change point may occur in time series models. For example, Hansen (1992) examined whether an AR(1) model for annual U.S. output growth rates has remained constant over 1889-1987. For a review, we refer to Ray and Tsay (2001), Lavielle and Teyssieve (2006) and Kawano et al. (2008) and references therein. Change point phenomenon also happens in financial time series models. For example, Lee et al. (2004) detected multiple change points in the return of yen/dollar exchange rate data. Three stylized facts of a financial time series are volatility clustering, fat tail, and volatility mean reversion. These properties suggest GARCH modeling in these cases (see Zivot and Wang (2005)). Therefore, performing the change point analysis in a GARCH time series is too important. For a general review in this area, see Kim et al. (2000), Kokoszka and Leipus (2000), Lee *et al.* (2004), Wang and Wang (2006) and Zhao *et al.* (2010).

Lavielle and Lebarbier (2001) (hereafter LL) detected multiple change points in the mean of independent normal observations. They adopted a Bayesian approach. They estimated the posterior distribution by MCMC algorithm. A crucial assumption for LLs paper is that the variance of normal data remains fixed over the time. In this note, we consider the observations with multiple change points in their means and time varying variances. Author believes that the problem in this case is more difficult than the constant variance case. LL (2001) showed that the posterior distribution is a map of a energy function. Here, we show that the posterior, in our case, is function of weighted energy function. This paper is organized as follows. In Section 2, we propose the models and prior assumptions. The MCMC method for estimating the posterior and SAEM algorithm to estimate the hyper-parameters are also given in this Section. The simulation results about time varying independent Normal observations and GARCH(1,1) time series are given in Section 3.

2 Time varying variance case. Following LL (2001), suppose that the underlying process is $y = \{y_t\}_{t\geq 1}$ at which

$$y_t = s_t + \varepsilon_t,$$

for $t \geq 1$. The error process ε_t are independent zero mean random variables such that $var(\varepsilon_t) = \sigma^2 h_t$, and let $h = \{h_t\}_{t\geq 1}$. The mean function s_t is piecewise constant, i.e., $s_t = m_k$ for $\tau_{k-1}+1 \leq t \leq \tau_k$, $k = 1, ..., R$ and $m = \{m_t\}_{t\geq 1}$. The parameter R is unknown number of change points. Following LL, the change point process$m = \{m_t\}_{t\geq 1}$ defined by

$$r_t = \begin{cases} 1 \text{ if there exists } k \text{ s.t } t = \tau_k, \\ 0 \text{ otherwise,} \end{cases}$$

$t = 1, ..., n$ are iid Bernoulli random variables with parameter λ. Suppose that $s_1, ..., s_n$ are independent random variables with mean μ and variance $v^2 d_i$, $i = 1, ..., n$. It is easy to see that m_k's are independent and Gaussian with mean μ and variance $v^2/\sum_{i=\tau_{k-1}+1}^{\tau_k} d_i^{-1}$. The conditional distribution of the observations is given by

$$\begin{aligned} h(y|r, m, h, \sigma^2) &= (2\pi\sigma^2)^{\frac{-n}{2}}(\prod_{i-1}^{n} h_i^{-1/2}) \\ &\quad \times \exp\{\frac{-1}{2\sigma^2}\sum_{k=1}^{R}\sum_{i=\tau_{k-1}+1}^{\tau_k}(y_i - m_k)^2/h_i\}. \end{aligned}$$

It is seen that given observations y, the m_k's remain independent and each m_k is distributed as $N(\alpha_k, \sigma_k^2)$ where

$$\alpha_k = \frac{\sigma^2 v^2}{v^2 + l_k\sigma^2}\{\frac{\overline{y}_k^w}{\sigma^2} + \frac{l_k}{v^2}\mu_k\},$$

and

$$\sigma_k^2 = \frac{\sigma^2 v^2}{v^2 + l_k\sigma^2}\{\sum_{i=\tau_{k-1}+1}^{\tau_k} h_i^{-1}\},$$

at which

$$l_k = \frac{\sum_{i=\tau_{k-1}+1}^{\tau_k} d_i^{-1}}{\sum_{i=\tau_{k-1}+1}^{\tau_k} h_i^{-1}} \text{ and } \overline{y}_k^w = \frac{\sum_{i=\tau_{k-1}+1}^{\tau_k} h_i^{-1} y_i}{\sum_{i=\tau_{k-1}+1}^{\tau_k} h_i^{-1}},$$

is the weighted mean of $k-th$ segment. Note that by letting $h_i = d_i = 1$, then the LL formula is derived. Since there is no information about weights d_i's, and to make consistency between h_i and d_i, we let $h_i = d_i$. Denote hyper-parameters by $\theta = (\mu, \sigma^2, v^2, \lambda)$. Then, we can show that

$$P(r|y,\theta) = \exp\{-\Phi S_r^w - \gamma R\},$$

where

$$S_r^w = \sum_{k=1}^{R} \sum_{t=\tau_{k-1}+1}^{\tau_k} h_t^{-1}(y_t - \overline{y}_k^w)^2.$$

Here, $\Phi = \frac{v^2}{2\sigma^2(v^2+\sigma^2)}$ and $\gamma = (1/2)\log(\frac{v^2+\sigma^2}{\sigma^2}) + \log(\frac{1-\lambda}{\lambda})$.

Remark 1. The posterior, in our case, is function of weighted energy function $U_\theta^w(y,r) = \Phi S_r^w + \gamma R$. We call U_θ^w since sum of square error S_r^w is weighted. Following LL (2001), the most probable configuration under posterior distribution is a MAP estimator. As it is stated by LL (2001), a low temperature version of posterior can be obtained by involving a temperature parameter T to the posterior distribution, i.e.,

$$P_T(r|y;\theta) = C_T(y,\theta)\exp\{U_\theta^w(y,r)/T\}.$$

They described that T plays an important role to discriminate the global and local maxima of the posterior distribution.

Remark 2. The probability $P(r_t = 1|y;\theta)$ of instant t to be a change point as well as the probability of having exact k change point between instants a, b, i.e., $P(\sum_{t=a}^{b} r_t = k|y,\theta)$ are replaced by their MCMC estimators, in practical cases. For example, the first probability is estimated by $(1/N)\sum_{i=1}^{N} r_t^{(i)}$, where $\{r^{(i)}, i \geq 1\}$ are ergodic Markov chains generated by MCMC method. For another example, $\#(R(i) = k)/N$ converges $a.s.$ to $P(R = k|y;\theta)$. Since the result of estimation by MCMC method depends weakly on the initial guess some initial burn-in period is considered before collecting samples. Following LL (2001), three types of $\alpha(r,\widetilde{r})$ probability of accepting $\widetilde{r}$ as a new sample in MCMC methods are considered. They are given as follows.

$$\alpha_1(r,\widetilde{r}) = \min\{1, \exp\{-\Phi(S_{\widetilde{r}}^w - S_r^w) - \beta(R_{\widetilde{r}} - R_r)\}\},$$

where $\beta_r = (1/2)\log(\frac{v^2+\sigma^2}{\sigma^2})$. Hereafter, we use the notation R_r to show that R_r is the number of change points related to configuration r. The other two types of $\alpha(r,\widetilde{r})$ are

$$\begin{aligned} \alpha_2(r,\widetilde{r}) &= \min\{1, \exp\{-\Phi(S_{\widetilde{r}}^w - S_r^w) \pm \gamma\}\}, \\ \alpha_3(r,\widetilde{r}) &= \min\{1, \exp\{-\Phi(S_{\widetilde{r}}^w - S_r^w)\}\}. \end{aligned}$$

Remark 3. In above, we assumed that the hyper-parameters $\theta = (\mu, \sigma^2, v^2, \lambda)$ are known. This is not the usual case, in practice. Here, following LL (2001), we also advise to estimate θ. Since the closed forms of ML estimators dont exist, we use the Stochastic Approximation Expectation Maximization (SAEM) algorithm, for more details see Givens and Hoeting (2005). The hyper-parameters are estimated by minimizing the likelihood based on complete data (y, r); i.e.,

$$L(\theta) = f(y, r; \theta).$$

One can see that

$$\widehat{\mu} = \overline{y}^w = \frac{\sum_{i=1}^n h_i^{-1} y_i}{\sum_{i=1}^n h_i^{-1}},\ \widehat{\sigma}^2 = S_r^w/(n - R_r)$$

and

$$\widehat{v}^2 = \frac{\sum_{t=1}^n h_t^{-1}(y_t - \overline{y}^w)^2 - S_r^w}{R_r} - \widehat{\sigma}^2 \text{ and } \widehat{\lambda} = \frac{R_r - 1}{n - 1}.$$

Remark 4. To perform SAEM and MCMC algorithms, we start by initial configuration $r^{(0)}$ and initial guess $\theta^{(0)}$. Using these values as well as using M iterations of MCMC algorithm, a new configuration r(1) is generated. Then, by SAEM algorithm $\theta^{(0)}$ is updated to $\theta^{(1)}$. Again, using $r^{(1)}$, $\theta^{(1)}$ and MCMC a new $r^{(2)}$ is constructed. This iterative scheme is continued, for more details see LL (2001).

3 Examples. In this Section, we apply our method to two known situations. The first one relates to Normal observations with time varying variances which appears frequently in regression models and the second one is GARCH time series models.

Example 1 Normal observations: time varying variances. In this subsection, we assume that ε_t are independent and distributed as $N(0, \sigma^2 h_t)$.Here, we let $n = 700$. There are five change points at $\tau_1 = 100$, $\tau_2 = 250$, $\tau_3 = 425$, $\tau_4 = 550$, $\tau_5 = 675$. The common part of variance, $\sigma^2 = 0.1$ and we let $h_t = 1 + 0.01t$. We let the vector of mean be $m = (0.1, 0.45, 0.5, 0.45, 0.2, 0.1)$. The the estimated hyper parameter are $\widehat{\theta} = (\widehat{\mu}, \widehat{\sigma}^2, \widehat{v}^2, \widehat{\lambda}) = (0.38, 0.11, 2.8, 0.015)$. It is seen that MCMC method with 400 burn-in iterations converges after 20000 iterations. The MAP estimators of R and τ_i's are $\widehat{R} = 6$ and $\widehat{\tau}_1 = 96$, $\widehat{\tau}_2 = 251$, $\widehat{\tau}_3 = 415$, $\widehat{\tau}_4 = 553$, $\widehat{\tau}_5 = 670$. The estimated mean vector is $\widehat{m} = (0.15, 0.48, 0.55, 0.43, 0.21, 0.13)$. This example shows that our method-works well.

Example 2 GARCH time series. In the previous section we assumed that weights h_t are known, the usual assumption in the Bayesian setting. Here,we let ε_t be a GARCH(1,1) time series, that is h_t is a linear combination of h_{t-1} and ε_{t-1}^2. Here, we let

$$ht = 0.14 + 0.175\varepsilon_{t-1}^2 + 0.686h_{t-1}.$$

Following the previous example, we let $\sigma^2 = 0.1$ and $\tau_1 = 100$, $\tau_2 = 250$, $\tau_3 = 425$, $\tau_4 = 550$, $\tau_5 = 675$. Here, again $\widehat{m} = (0.1, 0.45, 0.5, 0.45, 0.2, 0.1)$. Then, $\widehat{m} = (0.17, 0.43, 0.51, 0.47, 0.25, 0.11)$. It is seen that $\widehat{R} = 6$ and $\widehat{\tau}_1 = 9$,

$\widehat{\tau}_2 = 245, \widehat{\tau}_3 = 420, \widehat{\tau}_4 = 548, \widehat{\tau}_5 = 666$. The hyper parameter is estimated as (0.55, 0.21, 3.2, 0.019). It is seen that our method again works well in this case.

References

[1] Basseville, M. and Nikiforov, L. V. (1993). *Detection of abrupt changes-theory and application*. Englewood Cliffs.

[2] Brodsky, B. E. and Darkhovsky, B. S. (1993). *Nonparametric methods in change-point problems*. Kluwer Academic Publishers.

[3] Chen, J., Gupta, A.K. (2000). *Parametric statistical change point analysis*. Birkhauser.

[4] Csorgo, M. and Horvath, L. (1997). *Limit theorems in change point analysis*. Wiley.

[5] Givens, G. H. and Hoeting, J. A. (2005). *Computational Statistics*. Wiley.

[6] Hansen, B. E. (1992). Testing for parameter instability in linear models. *Journal of Policy modeling* **14**, 517-533.

[7] Kawano, H., Hattori, T., and Nishimatsu, K. (2008). Structural change point detection method of time series using sequential probability ratio test. *Tech. Report*. NTT Lab. Japan.

[8] Kim, S., Cho, S. and Lee, S. (2000). On the cusum test for parameter changes in GARCH(1,1) models. *Communications in Statistics. Theory & Methods* **29**, 445-462.

[9] Kokoszka, P. and Leipus, R. (2000). Change-point estimation in ARCH models. *Bernoulli* **6**, 513-539.

[10] Lavielle, M. and Lebarbier, E. (2001). An application of MCMC methods for multiple change point problem. *Signal Processing* **81**, 39-53.

[11] Lavielle, M. and Teyssieve, G. (2006). Detection of multiple change points in multivariate time series. Tech. Report. Universite Rene Descartes and Universit´e ParisSud.

[12] Lee, S., Tokutsu, Y., and Maekawa, K. (2004). The cusum test for parameter change in regression models with ARCH errors. *Journal of Japan Statistical Society* **2**, 173-188.

[13] Page, E. S. (1954). Continuous inspection schemes. *Biometrika* **41**, 100-115.

[14] Ray, B. K. and Tsay, R. S. (2001). Bayesian methods for change-point detection in long-range dependent processes. *Tech. Report*. Graduate School of Business. University of Chicago

[15] Wang, L. and Wang, J. (2006). Change-of-variance problem for linear processes with long memory. *Statistical Papers* **47**, 279-298.

[16] Zhao, W., Tian, Z., and Xia, Z. (2010). Ratio test for variance change point in linear process with long memory. *Statistical Papers* **51**, 397-407.

[17] Zivot, E. and Wang, J. (2005). *Modeling financial time series with S-PLUS*. Springer.