Bayesian Analysis of Structural Changes in ARFIMA Models with an Application to Realized Volatility*

Toshiaki Watanabe†

October 23, 2008

Abstract

Many researchers have documented that realized volatility, which is the sum of squared intraday returns, may follow a long-memory process, but the long-memory property may be caused by structural changes. This article develops a new method for the analysis of long memory and structural changes jointly by combining the irreversible Markov switching model proposed by Chib (1998) with the ARFIMA model. Since the resulting model cannot be estimated using the maximum likelihood method, a Bayesian method is developed using MCMC techniques. Specifically, the parameters and a state variable representing the number of change points up to each period are sampled from their joint posterior distribution by sampling them from their full conditional distributions iteratively. The state variable is divided into several blocks and each block is sampled at a time from its full conditional distribution given the parameters and the other blocks, which mitigates the computational burden and makes the method implementable. Our method is illustrated by applying to the realized volatility (RV) of the TOPIX between 1/7/1997 and 12/28/2006. Using the marginal likelihood, it is found that there are two change points around 6/22/2004 and 8/4/2005. It is also found that the mean of RV is low during these two change points. The long-memory is still detected even if structural changes are allowed, providing evidence that the long-memory property of the RV of the TOPIX cannot fully be attributed to structural changes.

Key words: ARFIMA, Irreversible Markov switching, MCMC, Realized volatility, Structural change

*This research is partially supported by Grant-in-Aid for Scientific Research 18203901 from the Japanese Ministry of Education, Science, Sports, Culture and Technology. The author thanks the participants in JEuBES 2007 in the National Bank of Hungary, the 2007 Japanese joint statistical meeting in Kobe University and the 2007 Economic Statistics Workshop in Hitotsubashi University for useful comments. All remaining errors are clearly my own.

†Institute of Economic Research, Hitotsubashi University, 2-1 Naka, Kunitachi City, Tokyo 186-8603, JAPAN. E-mail: watanabe@ier.hit-u.ac.jp

1. Introduction

The long-memory property has been detected in some economic time series data and the ARFIMA (autoregressive fractionally integrated moving average) model has been applied to such data. One representative variable is realized volatility (RV), which is the sum of squared intraday returns in financial markets and has recently attracted the attentions of financial econometricians as an accurate estimator of volatility (Andersen et al. (2001, 2003)). The long-memory property may, however, be spurious and caused by structural changes. Thus, it is important to analyze long-memory and structural changes jointly.

Chib (1998) proposes a model for the analysis of multiple structural changes by modifying the Markov switching model proposed by Hamilton (1989). He introduces a state variable D_t representing the number of change points up to period t and assumes that D_t follows a Markov process. Let T and N denote the sample size and the number of change points during the sample period respectively, and D_1 and D_T fixed as 0 and N. Suppose that $D_{t-1} = n$. Then, D_t takes n or $n+1$ when $n \leq N-1$ and only N when $n = N$. Since D_t never decreases, we call this model as an irreversible Markov switching model. Chib (1998) develops a Bayesian method for the analysis of this model using Markov chain Monte Carlo (MCMC) techniques and proposes to select the number of change points N using the marginal likelihood.

We combine this model with the ARFIMA model. Hamilton (1989) combines the Markov switching model with the AR model and proposes the filter to calculate the likelihood of this model. This filter requires us to evaluate probabilities for all possible combinations of $(D_t, D_{t-1}, \ldots, D_{t-r})$ for each period $t = r+1, \ldots, T$ where r is the lag-length of AR model, and hence the computational burden will increase with the lag-length. Since the ARFIMA model is represented by the AR model with an infinite lag-length, the likelihood of our model cannot be calculated using the Hamilton (1989) filter. We may approximate the ARFIMA model by an AR model with a finite lag-length, but the Hamilton (1989) filter is still not applicable if a sufficiently large lag-length is chosen such that the approximation error may be negligible. Since the likelihood of our model cannot be evaluated and the parameters of our model cannot be estimated using the maximum likelihood method, we develop a Bayesian method for the analysis of our model. We sample the parameters and $(D_2, \ldots D_{T-1})$ from their joint posterior distribution and use the obtained samples for the

Bayesian posterior analysis. This sampling can be done using Gibbs sampler, which is one of the MCMC techniques, i.e., sampling the parameters and $(D_2, \dots D_{T-1})$ iteratively from their full conditional distributions.

The problem is how to sample $(D_2, \dots D_{T-1})$ from their full conditional distribution. Two methods are available for this sampling (Kim and Nelson (1999b) Section 9.1.1). One is the single-move sampler where a single D_t $(t = 2, \dots, T-1)$ is sampled at a time from its full conditional distribution with $(D_1, \dots, D_{t-1}, D_{t+1}, \dots, D_T)$ and the parameters given. This method is applicable to our model but inefficient in the sense that it produces highly autocorrelated samples and the speed of convergence to the posterior distribution is slow. The other is the multi-move sampler where the vector $(D_2, \dots D_{T-1})$ is sampled at a time from the full conditional distribution with the parameters given. This sampler uses the Hamilton (1989) which is not applicable to our model as mentioned above. In this article, we propose to divide $(D_2, \dots, D_{T-1})$ into several blocks and sample each block at a time from the full conditional distribution with the other blocks and the parameters given. This sampling can be done by applying the multi-move sampler to each block. This sampling makes our method feasible by mitigating the computational burden for the following two reasons. First, we need not sample the blocks unless D_t at the beginning of the next block is larger than that at the end of the previous block. Second, the computational burden depends on the number of periods within each block instead of the lag-length because the values of D_t outside the block are given. If the blocks are fixed over the MCMC iterations, there are some blocks which are not sampled as long as D_t at the beginning of the next block is the same as that at the end of the previous block, which may make this method inefficient. We select the knots of blocks randomly to accelerate the convergence to the posterior distribution.

Our method is illustrated by applying to the daily RV of TOPIX (Tokyo Stock Price Index) between 1/7/1997 and 12/28/2006. Using the marginal likelihood, we find the evidence that there are two change points around 6/22/2004 and 8/4/2005. We also find that the mean of RV is low during these two change points. The long-memory is still detected even if we take account of structural changes, providing evidence that the long-memory property of the RV of the TOPIX cannot fully be explained by structural changes.

The article proceeds as follows. Section 2 explains the irreversible Markov switching ARFIMA model for the joint analysis of long-memory and structural changes. Section 3 explains our Bayesian method for the analysis of this model. Section 4 reviews RV and explains how we calculate the RV of the TOPIX. Section 5 applies our method to the RV of the TOPIX and summarizes the results. Section 6 concludes.

2. Irreversible Markov Switching ARFIMA Model

We start with a definition of long-memory. Let $\rho(h)$ denote the h-th order autocorrelation coefficient of variable y_t. Then, y_t follows a short-memory process if $\sum_{h=0}^{\infty} |\rho(h)| < \infty$ and a long-memory process if $\sum_{h=0}^{\infty} |\rho(h)| = \infty$. A stationary ARMA model is a short-memory process. As h increases, the autocorrelation coefficient $\rho(h)$ of the long-memory process decays to zero more slowly than that of the short-memory process. More specifically, the former decays hyperbolically and the latter decays geometrically.

The most widely used for a long-memory process is ARFIMA(p, d, q) model[1]:

$$\Phi(L)(1-L)^d(y_t - \mu) = \Theta(L)u_t, \quad u_t \sim WN(0, \sigma^2), \tag{1}$$

where μ is the mean of y_t and $WN(0, \sigma^2)$ represents a white noise with mean 0 and variance σ^2. L denotes the lag operator and $\Phi(L) = 1-\phi_1 L - \cdots - \phi_p L^p$ and $\Theta(L) = 1-\theta_1 L - \cdots - \theta_q L^q$ are the p-th and q-th order lag-polynomials. For these lag-polynomials, we assume the stationarity and reversibility, i.e., all roots of $\Phi(\lambda) = 0$ and $\Theta(\lambda) = 0$ lie outside the unit circle.

The order of integration d is allowed to take non-integer values. If $d = 0$, ARFIMA(p, d, q) model collapses to stationary ARMA(p, q) model and if $d = 1$, it becomes ARIMA$(p, 1, q)$ model. If $0 < d < 0.5$, y_t follows a stationary long-memory process. If $0.5 \le d < 1$, y_t follows a nonstationary long-memory process[2]. $(1-L)^d$ may be written as follows.

$$(1-L)^d = 1 + \sum_{k=1}^{\infty} \frac{d(d-1)\cdots(d-k+1)}{k!}(-L)^k \tag{2}$$

It follows that ARFIMA model can be represented by ARMA(∞, q) model.

To analyze long-memory and structural changes jointly, we combine the ARFIMA model with the irreversible Markov switching model proposed by Chib (1998). Let T denote the sample size and N and $(\tau^{(1)}, \ldots, \tau^{(N)})$ denote the number and the points of structural changes

[1] See Beran (1994) for the details of long-memory and ARFIMA model.

[2] In this case, the above definition of long-memory is not applicable because the autocorrelation coefficients $\rho(h)$ cannot be defined.

respectively, where $1 < \tau^{(1)} < \tau^{(2)} < \cdots < \tau^{(N)} \leq T$. For simplicity, we assume that the only mean μ may be subject to structural changes as follows.

$$\mu_t = \begin{cases} m^{(0)}, & 1 \leq t < \tau^{(1)} \\ m^{(1)}, & \tau^{(1)} \leq t < \tau^{(2)} \\ \vdots & \\ m^{(N-1)}, & \tau^{(N-1)} \leq t < \tau^{(N)} \\ m^{(N)}, & \tau^{(N)} \leq t \leq T \end{cases} \tag{3}$$

It is straightforward to extend such that the other parameters may also change. We also assume that u_t follows a normal distribution. Then, our model is represented by

$$\Phi(L)(1-L)^d(y_t - \mu_t) = \Theta(L)u_t,\ u_t \sim NID(0, \sigma^2) \tag{4}$$

Let D_t denote the number of change points up to period t, i.e.,

$$D_t = \begin{cases} 0, & 0 \leq t < \tau^{(1)} \\ 1, & \tau^{(1)} \leq t < \tau^{(2)} \\ \vdots & \\ N-1, & \tau^{(N-1)} \leq t < \tau^{(N)} \\ N, & \tau^{(N)} \leq t \leq T \end{cases} \tag{5}$$

where D_1 and D_T are fixed as 0 and N.

Following Chib (1998), we assume that D_t follows an irreversible Markov process with transition probabilities:

$$\begin{aligned} &P(D_t = 0|D_{t-1} = 0) = \pi_{00}, \quad P(D_t = 1|D_{t-1} = 0) = 1 - \pi_{00}, \\ &P(D_t = 1|D_{t-1} = 1) = \pi_{11}, \quad P(D_t = 2|D_{t-1} = 1) = 1 - \pi_{11}, \\ &\qquad\qquad\vdots \\ &P(D_t = N-1|D_{t-1} = N-1) = \pi_{N-1,N-1}, \quad P(D_t = N|D_{t-1} = N-1) = 1 - \pi_{N-1,N-1}, \\ &P(D_t = N|D_{t-1} = N) = 1. \end{aligned} \tag{6}$$

These transition probabilities mean that D_t takes n or $n+1$ if $D_{t-1} = n \leq N-1$ and takes only N if $D_{t-1} = N$, which is a natural assumption for structural changes. Since D_t never decreases, we call this process "irreversible" Markov process.

3. MCMC Bayesian Analysis

Hamilton (1989) considers a Markov switching AR(p) model:

$$y_t = \mu_t + \phi_1(y_{t-1} - \mu_{t-1}) + \cdots + \phi_p(y_{t-p} - \mu_{t-p}) + u_t, \quad u_t \sim NID(0, \sigma^2) \tag{7}$$

If the lag-length p of this model is not large, it is straightforward to calculate the likelihood by using the filter proposed by Hamilton (1989) and estimate the parameters by the maximum likelihood method. This filter requires us to calculate probabilities for all combinations of $(D_t, D_{t-1}, \ldots, D_{t-p})$ for $t = 1, \ldots, N$. If p is large, it is difficult to calculate the likelihood because of the computational burden. It is impossible to apply the Hamilton (1989) filter directly to the irreversible Markov switching ARFIMA model because the ARFIMA model may be represented by an AR model but the lag-length is infinite unless $d = q = 0$. To overcome this difficulty, we resort to a Bayesian method using MCMC techniques.

In this method, we first sample the parameters and $(D_2, \ldots, D_{T-1})$ from their joint posterior distribution

$$f(d, \theta, \sigma^2, m^{(0)}, \ldots, m^{(N)}, \pi_{00}, \ldots, \pi_{N-1,N-1}, D_2, \ldots, D_{T-1} | \tilde{y}_T, D_1, D_T), \tag{8}$$

where $\tilde{y}_T = (y_1, \ldots, y_T)$.
Then, we use the obtained samples for the Bayesian posterior analysis.

Sampling from (8) can be done by sampling iteratively from the following full conditional distributions.

$$f(d|\theta, \sigma^2, m^{(0)}, \ldots, m^{(N)}, \pi_{00}, \ldots, \pi_{N-1,N-1}, \tilde{D}_T, \tilde{y}_T), \tag{9}$$

$$f(\theta|d, \sigma^2, m^{(0)}, \ldots, m^{(N)}, \pi_{00}, \ldots, \pi_{N-1,N-1}, \tilde{D}_T, \tilde{y}_T), \tag{10}$$

$$f(\sigma^2|d, \theta, m^{(0)}, \ldots, m^{(N)}, \pi_{00}, \ldots, \pi_{N-1,N-1}, \tilde{D}_T, \tilde{y}_T), \tag{11}$$

$$f(m^{(0)}, \ldots, m^{(N)}|d, \theta, \sigma^2, \pi_{00}, \ldots, \pi_{N-1,N-1}, \tilde{D}_T, \tilde{y}_T), \tag{12}$$

$$f(\pi_{00}, \ldots, \pi_{N-1,N-1}|d, \theta, \sigma^2, m^{(0)}, \ldots, m^{(N)}, \tilde{D}_T, \tilde{y}_T), \tag{13}$$

$$f(D_2, \ldots, D_{T-1}|d, \theta, \sigma^2, m^{(0)}, \ldots, m^{(N)}, \pi_{00}, \ldots, \pi_{N-1,N-1}, D_1, D_T, \tilde{y}_T), \tag{14}$$

where $\tilde{D}_T = (D_1, \ldots, D_T)$.

If we use the normal prior for $(m^{(0)}, \cdots, m^{(N)})$, the inverse gamma prior for σ^2 and the beta prior for π_{ii} $(i = 0, \ldots, N-1)$, (10) and (12) are normal and (11) and (13) are inverse gamma and beta respectively. It is straightforward to sample from these distributions.

It is not true for (9), (10) and (14). To sample from (9) and (10), we use the ARMH (Accept-Reject Metropolis-Hastings) algorithm proposed by Tierney (1994) (see also Chib and Greenberg (1995) and Watanabe (2001)) with the normal proposal density obtained by applying the second-order Taylor expansion to the true log-density around its mode (see Appendix A).

If the model is Markov switching AR model (7) and the lag-length p is not large, we can sample from (14) efficiently using the multi-move sampler proposed by Kim and Nelson (1998, 1999a,b). If p is large, this sampler is not applicable because it uses the Hamilton (1989) filter. The irreversible Markov switchig ARFIMA model (4) is represented by the AR model with infinite lag-length:

$$y_t = \mu_t + \sum_{i=0}^{\infty} \beta_i (y_{t-i} - \mu_{t-i}) + u_t, \tag{15}$$

where β_i $(i = 1, 2, \ldots)$ can be calculated depending on the lag-length p and q. For example, if $p = 0$ and $q = 1$, they can be calculated sequentially as follows.

$$\alpha_1 = d,\ \beta_1 = d - \theta,\ \alpha_i = \frac{i-d-1}{i}\alpha_{i-1},\ \beta_i = \alpha_i + \theta\phi_{i-1}\ (i \geq 2) \tag{16}$$

We assume that $y_0 = \mu_0, \ldots y_{-r} = \mu_{-r}$. Then, the lag-length of (15) will be $t-1$ as follows.

$$y_t = \mu_t + \sum_{i=0}^{t-1} \beta_i (y_{t-i} - \mu_{t-i}) + u_t \tag{17}$$

If the sample size is large, the lag-length $t-1$ will also be large and it is still difficult to sample from (14) using the multi-move sampler.

Another method is the single-move sampler where a single D_t $(t = 2, \ldots, T-1)$ is sampled from

$$f(D_t | D_1, \ldots, D_{t-1}, D_{t+1}, \ldots, D_T, d, \theta, \sigma^2, m^{(0)}, \ldots, m^{(N)}, \pi_{00}, \ldots, \pi_{N-1,N-1}, \tilde{y}_T) \tag{18}$$

This method is applicable to our model but inefficient in the sense that it produces highly autocorrelated samples and the speed of convergence to the posterior distribution is slow.

Thus, we divide $(D_2, \ldots, D_{T-1})$ into $K+1$ blocks $(D_{k_{i-1}+1}, \ldots, D_{k_i})$ by selecting $(k_1, \ldots, k_K)$ with $k_0 = 1$ and $k_{K+1} = T-1$ where $k_i - k_{i-1} \geq 2$ for $i = 1, \ldots, K+1$. Then, we sample each block $(D_{k_{i-1}+1}, \ldots, D_{k_i})$ at a time from the full conditional distribution:

$$f(D_{k_{i-1}+1}, \ldots, D_{k_i} | \tilde{D}_{k_{i-1}}, D_{k_i+1}, \ldots, D_T, d, \theta, \sigma^2, m^{(0)}, \ldots, m^{(N)}, \pi_{00}, \ldots, \pi_{N-1,N-1}, \tilde{y}_T) \tag{19}$$

This sampling can be done by applying the multi-move sampler to each block (see Appendix B). This method mitigates the computational burden for the following reasons. First, we have only to sample the block where $D_{k_{i-1}} < D_{k_i+1}$ assuming that there are $D_{k_i+1} - D_{k_{i-1}}$ change points in this block. Second, this method makes the D_t outside the block as given. Thus, we need not calculate the probabilities for all possible combinations of $(D_2, \ldots, D_{T-1})$. All we have to calculate is the probabilities for all possible combinations of $(D_{k_{i-1}+1}, \ldots, D_{k_i})$.

We select the knots $(k_1, \ldots, k_K)$ randomly as follows.

$$k_i = \text{int}\left[(T-1) \times (i + U_i)/(K+2)\right], \quad i = 2, \ldots, K. \tag{20}$$

where U_i's are independent uniform random variables on $(0, 1)$ and $\text{int}[x]$ denotes the integer part of x.

These stochastic knots may accelerate the convergence to the posterior distribution by preventing the sampled change points from staying in the same blocks over the MCMC iterations (see Shephard and Pitt (1997) and Watanabe and Omori (2004)).

4. Realized Volatility

Since many researchers have documented that RV may follow a long-memory process (Andersen et al. (2001, 2003)), we illustrate our method by applying to the RV of the TOPIX. In this section, we explain RV and how we calculate the RV of the TOPIX.

Suppose that the log-price $p(s)$ follows the simple diffusion process:

$$dp(s) = \mu(s)dt + \sigma(s)dW(s), \tag{21}$$

where $W(s)$ is a standard Brownian process and $\mu(s)$ and $\sigma(s)$ are the mean and the standard deviation of $dp(s)$ respectively, which may be time-varying but are assumed to be independent of $dW(s)$. In this article, we call $\sigma^2(s)$ volatility although $\sigma(s)$ is usually called volatility in the finance literature.

Then, the true volatility for day t is defined as the integral of $\sigma^2(s)$ over the interval $(t-1, t)$, i.e.,

$$IV_t = \int_{t-1}^{t} \sigma^2(s)ds, \tag{22}$$

which is called integrated volatility. In this article, we define $t-1$ and t as the market closing time on days $t-1$ and t respectively.

The integrated volatility is unobservable, but if we have the intraday return data $(r_{t-1+1/n}, r_{t-1+2/n}, \ldots, r_t)$, we can estimate it as the sum of their squares

$$RV_t = \sum_{i=1}^{n} r_{t-1+i/n}^2, \tag{23}$$

which is called realized volatility (RV). RV_t will provide a consistent estimate of IV_t, i.e.,

$$\operatorname*{plim}_{n\to\infty} RV_t = IV_t. \tag{24}$$

We calculate the RV of TOPIX from 1/7/1997 to 12/28/2006. The sample size is 2440. There are two problems in calculating RV. One problem is the presence of microstructure noise (see Campbell et al. (1997) Chapter 3). If there presents microstructure noise, equation (24) may not be true. As the time interval of intraday returns used for calculating RV becomes smaller, the influence of microstructure noise on realized volatility will increase. While some methods for mitigating the effect of microstructure noise have been proposed[3], we simply use 5-minute returns to calculate RV although our dataset contains 1-minute prices.

The other problem is the presence of non-trading hours. Tokyo Stock Exchange is open only for 9:00–11:00 (morning session) and 12:30–15:00 (afternoon session) except for the first and last trading days in every year, when it is open only for 9:00-11:00. It is impossible to obtain high frequency returns for 15:00–9:00 (overnight) and 11:00–12:30 (lunch-time). Using the overnight and lunch-time returns may cause a severe discretization error because their time intervals are long. Hansen and Lunde (2005a) propose to calculate RV only when the market is open $RV_t^{(o)}$ and multiply a constant c such that the sample mean of realized volatility to the sample variance of daily returns, i.e.,

$$RV_t = cRV_t^{(o)}, \quad c = \frac{\sum_{t=1}^{T}(R_t - \overline{R})^2}{\sum_{t=1}^{T} RV_t^{(o)}}, \tag{25}$$

where $\overline{R}$ is the sample mean of daily returns[4]. We use this method.

Figure 1 plots the RV of the TOPIX calculated using the above method. This figure shows that there is a low volatility period from the middle of 2004 to the middle of 2005. What caused the low volatility in this period has not yet reached consensus although some possible reasons are proposed such as the improvement of economic indicators and the trading of Nikkei-linked bond. Table 1 summarizes the descriptive statistics of RV and log-RV. According to the skewness and kurtosis in Table 1, the distribution of log-RV is much closer to normal distribution than that of RV. Thus, we use log-RV as y_t in equation (4). LB(10), which

[3]Methods for mitigating the effect of microstructure noise are the selection of optimal time interval of intraday returns (Aït-Sahalia et al. (2005) and Bandi and Russell (2006, 2008)) and two (multi) scale estimator (Zhang et al. (2005) and Zhang (2006)) and the kernel estimation (Hansen and Lunde (2006) and Barndorff-Nielsen et al. (2008)).

[4]Hansen and Lunde (2005b) propose a more elaborate method.

represents the Ljung and Box statistic to test the null hypothesis of no autocorrelations up to the 10th lag, both for RV and log-RV are large enough to reject the null hypothesis. Figure 2 plots the autocorrelation coefficients in RV and log-RV. The autocorrelations decreases with the lag-length very slowly, providing evidence that RV and log-RV may follow a long-memory process.

5. Estimation Results

We estimate the irreversible Markov switching model (4) setting $N = 0, 1, 2, 3$. Since the lag length of p and q in AR and MA parts are selected as 0 and 1 by calculating the SBIC of the plain ARFIMA model, we concentrate on the irreversible Markov switching ARFIMA$(0, d, 1)$ model:

$$(1 - L)^d (y_t - \mu_t) = u_t - \theta u_{t-1} \quad u_t \sim NID(0, \sigma^2) \tag{26}$$

The priors are set as follows.

$$(m^{(0)}, \dots, m^{(N)})' \sim N\left(0, 10 \times I_{N+1}\right) \tag{27}$$

$$d \sim U(0, 1),\ \theta \sim U(-1, 1) \tag{28}$$

$$\sigma^2 \sim IG(3, 0.6) \tag{29}$$

$$\pi_{i,i} \sim Beta(20, 0.1),\ (i = 1, \dots, N - 1) \tag{30}$$

where I_{N+1} represents the $N + 1 \times N + 1$ identity matrix. $U(a, b)$ is uniform distribution in (a, b) and $IG(\cdot, \cdot)$ and $Beta(\cdot, \cdot)$ are inverse gammma and beta distributions respectively.

Following Kim, Morley and Nelson (2005), I impose the assumption that the periods between the two consecutive structural change points are more than 65 days (roughly 3 months). As explained in Section 3, we divide $(D_2, \dots, D_{T-1})$ into several blocks and sample each block at a time. We divide $(D_2, \dots, D_{T-1})$ into 25 blocks.

We sample from the joint posterior distribution (8) by sampling iteratively from the full conditional distributions (9)–(14) and (19). The first 5000 draws are discarded and then the next 10000 are used for the Bayesian posterior analysis.

Table 2 shows the posterior means, the standard errors of the posterior means, the 95% intervals and the convergence diagnostic (CD) statistics proposed by Geweke (1992) for each parameters and change points. The posterior means are computed by averaging the simulated draws. The standard errors of the posterior means are computed using a Parzen window with a bandwidth of 1000 (Shephard and Pitt (1997)). The 95% intervals are calculated using the 2.5th and 97.5th percentiles of the simulated draws. Geweke (1992) suggests assessing the convergence of MCMC by comparing values early in the sequence with those late in the sequence. Let $x^{(i)}$ be the ith draw of a parameter in the recorded 10000 draws, and let $\overline{x}_A = (1/n_A)\sum_{i=1}^{n_A} x^{(i)}$ and $\overline{x}_B = (1/n_A)\sum_{i=10001-n_B}^{n_B} x^{(i)}$. Using these values, Geweke (1989) proposes the following CD statistics.

$$\text{CD} = \frac{\overline{x}_A - \overline{x}_B}{\sqrt{\hat{\sigma}_A^2/n_A + \hat{\sigma}_B^2/n_B}}, \tag{31}$$

where $\sqrt{\hat{\sigma}_A^2/n_A}$ and $\sqrt{\hat{\sigma}_B^2/n_B}$ are the standard errors of $\overline{x}_A$ and $\overline{x}_B$. If the sequence of $x^{(i)}$ is stationary, it converges in distribution to the standard normal. We set $n_A = 1000$ and $n_B = 5000$ and compute $\hat{\sigma}_A^2$ and $\hat{\sigma}_B^2$ using Parzen window with bandwidth of 100 and 500 respectively.

According to the CD values, the null hypothesis of the convergence to the posterior distribution is accepted for all parameters and change points at the 1% significance level.

We select the number of change points using the posterior odds ratio. The posterior odds ratio in favor of model i, M_i, to model j, M_j, is given by

$$\text{POR} = \frac{f(M_i|\tilde{y}_T)}{f(M_i|\tilde{y}_T)} = \frac{f(\tilde{y}_T|M_i)}{f(\tilde{y}_T|M_j)}\frac{f(M_i)}{f(M_j)}, \tag{32}$$

where $f(\tilde{y}_T|M_i)/f(\tilde{y}_T|M_j)$ and $f(M_i)/f(M_j)$ are called Bayes factor and prior odds ratio respectively.

If this posterior odds ratio is greater (less) than one, M_i (M_j) will be selected. As is the usual practice, we set the prior odds ratio to be one, so that the posterior odds ratio is equal to the Bayes factor. The Bayes factor is the ratio of marginal likelihoods $f(\tilde{y}_T|M_i)$ and $f(\tilde{y}_T|M_j)$, so that the model with a higher marginal likelihood will be selected. We calculate the marginal likelihood using the harmonic mean method proposed by Newton and Raftery (1994).

Table 2 also shows the log-marginal likelihood to select the number of change points. According to the log marginal likelihood, we may choose the number of change points as two. Figure 2 plots the histogram of change points (τ_1, τ_2) sampled from their posterior distribution when the number of change points is two. Judging from their modes, we may conclude that the change points are around 6/22/2004 and 8/4/2005. Posterior means of m_0, m_1 and m_2 when the number of change points is two are 0.4706, 0.0593 and 0.1910, indicating that the mean of RV is low during these two change points.

The posterior mean of d decreases slightly as the number of change points increases, but the differences are small. We may conclude that RV follows a long memory process even if we take account of structural changes.

6. Conclusions

This article presents a Bayesian method for the joint analysis of long-memory and structural changes using the irreversible Markov switching ARFIMA model. Our method is illustrated by applying to the RV of the TOPIX from 1/7/1997 to 12/28/2006. We detect two change points around 6/22/2004 and 8/4/2005 and find that the mean of RV is low during these two change points. We also find that the value of memory parameter d does not decrease so much even if we consider structural changes.

This article is just a starting point and several extensions are possible. First, we detect long-memory even if we consider structural changes. To examine whether this result might be specific to our data, we should also apply our method to different RV data. Second, we concentrate on the structural changes in the mean of RV. It is important to extend our analysis such that other parameters such as d, θ and σ^2 may also change. Third, it is also important to examine whether the result may change if we calculate RV taking account of jumps in prices (Barndorff-Nielsen and Shephard (2004)). Fourth, we only used the ARFIMA model for RV. It is worthwhile using the other models such as HAR (heterogeneous interval autoregressive) model (Corsi (2004)) and UC (unobserved components) model (Barndorff-Nielsen and Shephard (2001, 2002)).

Appendix A: Sampling from (9) and (10)

We sample from (9) and (10) using the ARMH (Accept-Reject Metropolis-Hastings) algorithm proposed by Tierney (1994) with the normal proposal density obtained by applying the second order Taylor expansion to the true log-density around its mode. In this Appendix, we explain this method.

Let $f(x)$ denote the true density from which we like to sample and $l(x) = \log f(x)$. We approximte $l(x)$ using the second-order Taylor expansion around x^*.

$$\begin{aligned} l(x) &\approx l(x^*) + l'(x^*)(x - x^*) + l''(x^*)(x - x^*)^2 \\ &\equiv \log h(x) \end{aligned} \tag{A1}$$

The normalized version of $h(x)$ is a normal density with mean $x^* - l'(x^*)/l''(x^*)$ and variance $-1/l''(x^*)$. Using this normal density as the proposal density, we apply the ARMH algorithm.

Denote the $(i-1)$th sampled value of x by x_{i-1} and consider the ith sampling. Then, the ARMH algorithm proceeds as follows.

1. Sample a proposal x from $h(x)$ and a value u from the uniform distribution on $(0, 1)$.

2. If $u \leq f(x)/h(x)$, go to 3. Else, return to 1.

3. If $f(x_{i-1}) < h(x_{i-1})$, then let $q = 1$.
   If $f(x_{i-1}) \geq h(x_{i-1})$ and $f(x) < h(x)$, then let $q = h(x_{i-1})/f(x_{i-1})$.
   If $f(x_{i-1}) \geq h(x_{i-1})$ and $f(x) \geq h(x)$, then let
   $$q = \min \left[\frac{h(x)h(_{i-1})}{f(x_{i-1})h(x)}, 1 \right].$$

4. Sample a value u from the uniform distribution $(0, 1)$.

5. If $u \leq q$, return $x_i = x$. Else, return $x_i = x_{i-1}$.

The closer to one is $f(x)/h(x)$ around the mode of $f(x)$, the more efficient is the draw in the AR part 1-2, in the sense that it accepts the draw with higher probability. Hence, we select x^* as the mode of $f(x)$, which we find by applying the Newton method to $f(x)$ several times.

Appendix B: Sampling from (19)

(19) can be represented as follows.

$$\begin{aligned} & P(D_{k_{i-1}+1}, \ldots, D_{k_i} | \tilde{D}_{k_{i-1}}, D_{k_i+1}, \ldots, D_T, \tilde{y}_{k_i}) \\ &= P(D_{k_i} | \tilde{D}_{k_{i-1}}, D_{k_i+1}, \tilde{y}_{k_i}) \prod_{t=k_{i-1}+1}^{k_i - 1} P(D_t | \tilde{D}_{k_{i-1}}, D_{t+1}, \tilde{y}_t), \end{aligned} \tag{B1}$$

where we omit the parameters $(\theta, \sigma^2, m^{(0)}, \ldots, m^{(N)}, \pi_{00}, \ldots, \pi_{N-1,N-1})$ in the condition in this Appendix.

Thus, we first sample D_{k_i} from $P(D_{k_i}|\tilde{D}_{k_{i-1}}, D_{k_i+1}, \tilde{y}_{k_i})$. Using the sampled D_{k_i}, we next sample D_{k_i-1} from $P(D_{k_i-1}|\tilde{D}_{k_{i-1}}, D_{k_i}, \tilde{y}_{k_i-1})$. Using the sampled D_{k_i-1}, we next sample D_{k_i-2} from $P(D_{k_i-2}|\tilde{D}_{k_{i-1}}, D_{k_i-1}, \tilde{y}_{k_i-2})$. By repeating this procedure, we can sample from (19).

$P(D_t|\tilde{D}_{k_{i-1}}, D_{t+1}, \tilde{y}_t)$ in (B1) is represented as

$$P(D_t|\tilde{D}_{k_{i-1}}, D_{t+1}, \tilde{y}_T) \propto P(D_{t+1}|D_t)P(D_t|\tilde{D}_{k_{i-1}}, \tilde{y}_t), \tag{B2}$$

where $P(D_{t+1}|D_t)$ is the transition probability defined by (6).

To calculate $P(D_t|\tilde{D}_{k_{i-1}}, \tilde{y}_t)$ in (B2), we must calculate $P(D_{k_{i-1}+1}, \ldots, D_t|\tilde{D}_{k_{i-1}}, \tilde{y}_t)$ $(t = k_{i-1}+1, \ldots, k_i)$ sequentially starting from $D_{k_{i-1}}$. Suppose that $P(D_{k_{i-1}+1}, \ldots, D_{t-1}|\ \tilde{D}_{k_{i-1}}, \tilde{y}_{t-1})$ is given. Then, we can update it to $P(D_{k_{i-1}+1}, \ldots, D_t|\tilde{D}_{k_{i-1}}, \tilde{y}_t)$ using the following Hamilton (1989) filter.

Step 1:

$$P(D_{k_{i-1}+1}, \ldots, D_t|\tilde{D}_{k_{i-1}}, \tilde{y}_{t-1}) = P(D_t|D_{t-1})P(D_{k_{i-1}+1}, \ldots, D_{t-1}|\tilde{D}_{k_{i-1}}, \tilde{y}_{t-1}) \tag{B3}$$

Step 2:

$$f(y_t, D_{k_{i-1}+1}, \ldots, D_t|\tilde{D}_{k_{i-1}}, \tilde{y}_{t-1}) = f(y_t|\tilde{D}_t, \tilde{y}_{t-1})P(D_{k_{i-1}+1}, \ldots, D_t|\tilde{D}_{k_{i-1}}, \tilde{y}_{t-1}), \tag{B4}$$

where

$$f(y_t|\tilde{D}_t, \tilde{y}_{t-1}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left\{(y_t - \mu_t) - \sum_{i=1}^{t-1} \phi_i(y_{t-i} - \mu_{t-i})\right\}^2\right] \tag{B5}$$

Step 3:

$$f(y_t|\tilde{D}_{k_{i-1}}, \tilde{y}_{t-1}) = \sum_{D_{k_{i-1}+1}=0}^{N} \cdots \sum_{D_t=0}^{N} f(y_t, D_{k_{i-1}+1}, \ldots, D_t|\tilde{D}_{k_{i-1}}, \tilde{y}_{t-1}) \tag{B6}$$

Step 4:

$$P(D_{k_{i-1}+1}, \ldots, D_t|\tilde{D}_{k_{i-1}}, \tilde{y}_t) = \frac{f(y_t, D_{k_{i-1}+1}, \ldots, D_t|\tilde{D}_{k_{i-1}}, \tilde{y}_{t-1})}{f(y_t|\tilde{D}_{k_{i-1}}, \tilde{y}_{t-1})} \tag{B7}$$

$P(D_t|\tilde{D}_{k_{i-1}}, \tilde{y}_t)$ $(t = k_{i-1}+1, \ldots, k_i)$ are calculated as

$$P(D_t|\tilde{D}_{k_{i-1}}, \tilde{y}_t) = \sum_{D_{k_{i-1}+1}=0}^{N} \cdots \sum_{D_{t-1}=0}^{N} P(D_{k_{i-1}+1}, \ldots, D_t|\tilde{D}_{k_{i-1}}, \tilde{y}_t), \tag{B8}$$

where $P(D_{k_{i-1}+1}, \ldots, D_t|\tilde{D}_{k_{i-1}}, \tilde{y}_t)$ is given by (B7).

References

[1] Aït-Sahalia, Y., Mykland, P. A. and Zhang, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies*,18: 351–416.

[2] Andersen, T. G., Bollerslev, T., Diebold, F. X. and Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, 61: 43–76.

[3] Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71: 579–625.

[4] Bandi, F. M. and Russell, J. R. (2006). Separating microstructure noise from volatility. *Journal of Financial Economics*, 79: 655–692.

[5] Bandi, F. M. and Russell, J. R. (2008). Microstructure noise, realized variance, and optimal sampling. *Review of Economic Studies*, 75: 339–369.

[6] Barndorff-Nielsen, O.E., Shephard, N. (2001). Non-Gaussian OU based models and some of their uses in financial economics (with discussion). Journal of the Royal Statistical Society. Series B 63, 167–241.

[7] Barndorff-Nielsen, O. E. and Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B*, 64: 253–280.

[8] Barndorff-Nielsen, O. E. and Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps (with discussion). *Journal of Financial Econometrics*, 2: 1–37.

[9] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2008). Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica*, forthcoming.

[10] Beran, J. (1994). *Statistics for long-memory processes*, Chapman & Hall.

[11] Campbell, J. Y., Lo, A. W. and MacKinlay, A. C. (1997). *The econometrics of financial markets*, Princeton University Press.

[12] Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of Econometrics*, 86: 221-241.

[13] Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *American Statistician*, 49: 327–335.

[14] Corsi, F. (2004). A simple long memory model of realized volatility. Available at SSRN: http://ssrn.com/abstract=626064.

[15] Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (eds), *Bayesian Statistics 4*, Oxford University Press.

[16] Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and business cycle," *Econometrica*, 57: 357-384.

[17] Hansen, P. R. and Lunde, A. (2005a). A forecast comparison of volatility models: does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, 20: 873–889.

[18] Hansen, P. R. and Lunde, A. (2005b). A realized variance for the whole day based on intermittent high-frequency data. *Journal of Financial Econometrics*, 3: 525–554.

[19] Hansen, P. R. and Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24: 127–161.

[20] Kim, C. J., Morley, J. C. and Nelson, C. (2005). The structural break in the equity premium. *Journal of Business and Economic Statistics*, 23: 181–191.

[21] Kim, C.-J. and Nelson, C. R. (1998). Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime-switching. *Review of Economics and Statistics*, 80: 188–201.

[22] Kim, C.-J. and Nelson, C. R. (1999a). Has the U.S. economy become more stable? A Bayesian approach based on a Markov-Switching model of the business cycle. *Review of Economics and Statistics*, 81: 608–616.

[23] Kim, C.-J. and Nelson, C. R. (1999b). *State-Space Models with Regime Switching*, MIT Press.

[24] Newton, M. A. and Raftery, A. E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society, Series B*, 56: 3–26.

[25] Shephard, N. and Pitt, M. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, 84: 653–667.

[26] Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics*, 21: 1701–1762.

[27] Watanabe, T. (2001). On sampling the degree-of-freedom of student-t disturbances. *Statistics and Probability Letters*, 52: 177–181.

[28] Watanabe, T. and Omori, Y. (2004). A multi-move sampler for estimating non-Gaussian time series models: comments on Shephard and Pitt (1997). *Biometrika*, 91: 246–248.

[29] Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. *Bernoulli*, 12: 1019–1043.

[30] Zhang, L., Mykland, P.A. and Aït-Sahalia, Y. (2005). A tale of two time scales: determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100: 1394–1411.

Table 1: Descriptive Statistics of the Realized Volatility

	RV	log-RV
Mean	1.5453	0.1630
	(0.0260)	(0.0139)
Standard Deviation	1.3819	0.7366
Min	0.0986	−2.3166
Max	21.2073	3.0543
Skewness	4.1628	−0.0536
	(0.0461)	(0.0461)
Kurtosis	36.5959	3.3243
	(0.0922)	(0.0922)
LB(10)	3518.60	8848.73

Note: Numbers in parentheses are standard errors. LB(10) represents the Ljung-Box statistics to test the null hypothesis of no autocorrelations up to the 10th lag.

Table 2: Estimation Results

Parameter	Mean	Standard Error	95% Interval	CD
Number of Change Points = 0 (Log Marginal Likelihood = -1733.59)				
d	0.5011	0.0007	$[0.4560, 0.5487]$	0.97
m	0.4501	0.0030	$[0.0779, 0.8426]$	0.82
θ	0.1337	0.0010	$[0.0735, 0.1952]$	1.16
σ^2	0.2419	0.0001	$[0.2306, 0.2537]$	0.20
Number of Change Points = 1 (Log Marginal Likelihood = -1732.67)				
d	0.4939	0.0012	$[0.4469, 0.5431]$	2.44
$m^{(0)}$	0.4676	0.0047	$[0.0866, 0.8453]$	−0.98
$m^{(1)}$	0.1986	0.0269	$[-0.4557, 0.9948]$	1.79
θ	0.1256	0.0014	$[0.0654, 0.1892]$	1.99
σ^2	0.2417	0.0001	$[0.2305, 0.2532]$	0.69
π_{00}	0.9994	0.0000	$[0.9982, 1.0000]$	2.11
$\tau^{(1)}$	1881.01	51.87	$[1431, 2387]$	1.74

Note: The first 5000 draws are discarded and then the next 10000 draws are used for calculating the posterior means, the standard errors of the posterior means, the 95% intervals and the convergence diagnostic (CD) statistics proposed by Geweke (1992) for each parameters and change points. The posterior means are computed by averaging the simulated draws. The standard errors of the posterior means are computed using a Parzen window with a bandwidth of 1000. The 95% intervals are calculated using the 2.5th and 97.5th percentiles of the simulated draws. The CD is computed using equation (31) with $n_A = 1000$ and $n_B = 5000$, where we calculate $\hat{\sigma}_A^2$ and $\hat{\sigma}_B^2$ using Parzen window with bandwidth of 100 and 500 respectively.

Table 2: Estimation Results (Continued)

Parameter	Mean	Standard Error	95% Interval	CD
Number of Change Points = 2 (Log Marginal Likelihood = -1731.79)				
d	0.4875	0.0010	$[0.4378, 0.5387]$	-0.94
$m^{(0)}$	0.4706	0.0031	$[0.0971, 0.8439]$	0.14
$m^{(1)}$	0.0593	0.0182	$[-0.5915, 0.7805]$	-0.66
$m^{(2)}$	0.1910	0.0172	$[-0.5941, 0.9918]$	-0.76
θ	0.1217	0.0011	$[0.0598, 0.1876]$	-1.28
σ^2	0.2415	0.0001	$[0.2303, 0.2532]$	2.22
π_{00}	0.9994	0.0000	$[0.9982, 1.0000]$	1.33
π_{11}	0.9995	0.0000	$[0.9985, 1.0000]$	0.39
$\tau^{(1)}$	1787.90	16.76	$[1461, 2058]$	0.76
$\tau^{(2)}$	2104.30	17.65	$[1810, 2416]$	0.52
Number of Change Points = 3 (Log Marginal Likelihood = -1733.51)				
d	0.4747	0.0010	$[0.4223, 0.5303]$	0.35
$m^{(0)}$	0.4511	0.0038	$[0.1077, 0.8119]$	1.86
$m^{(1)}$	-0.1101	0.0225	$[-0.7067, 0.5715]$	1.18
$m^{(2)}$	0.2505	0.0791	$[-0.7654, 1.0947]$	-2.01
$m^{(3)}$	0.2104	0.0158	$[-0.5944, 1.0010]$	2.04
θ	0.1101	0.0010	$[0.0456, 0.1776]$	0.62
σ_2	0.2410	0.0001	$[0.2300, 0.2528]$	-1.38
π_{00}	0.9994	0.0000	$[0.9983, 1.0000]$	-0.08
π_{11}	0.9995	0.0000	$[0.9985, 1.0000]$	-2.40
π_{22}	0.9995	0.0000	$[0.9986, 1.0000]$	-1.40
$\tau^{(1)}$	1872.68	10.46	$[1761, 2025]$	-2.44
$\tau^{(2)}$	2084.40	16.15	$[1913, 2203]$	-2.19
$\tau^{(3)}$	2311.38	23.83	$[2096, 2435]$	-2.12

Note: The first 5000 draws are discarded and then the next 10000 draws are used for calculating the posterior means, the standard errors of the posterior means, the 95% intervals and the convergence diagnostic (CD) statistics proposed by Geweke (1992) for each parameters and change points. The posterior means are computed by averaging the simulated draws. The standard errors of the posterior means are computed using a Parzen window with a bandwidth of 1000. The 95% intervals are calculated using the 2.5th and 97.5th percentiles of the simulated draws. The CD is computed using equation (31) with $n_A = 1000$ and $n_B = 5000$, where we calculate $\hat{\sigma}_A^2$ and $\hat{\sigma}_B^2$ using Parzen window with bandwidth of 100 and 500 respectively.

Figure 1: RV of TOPIX

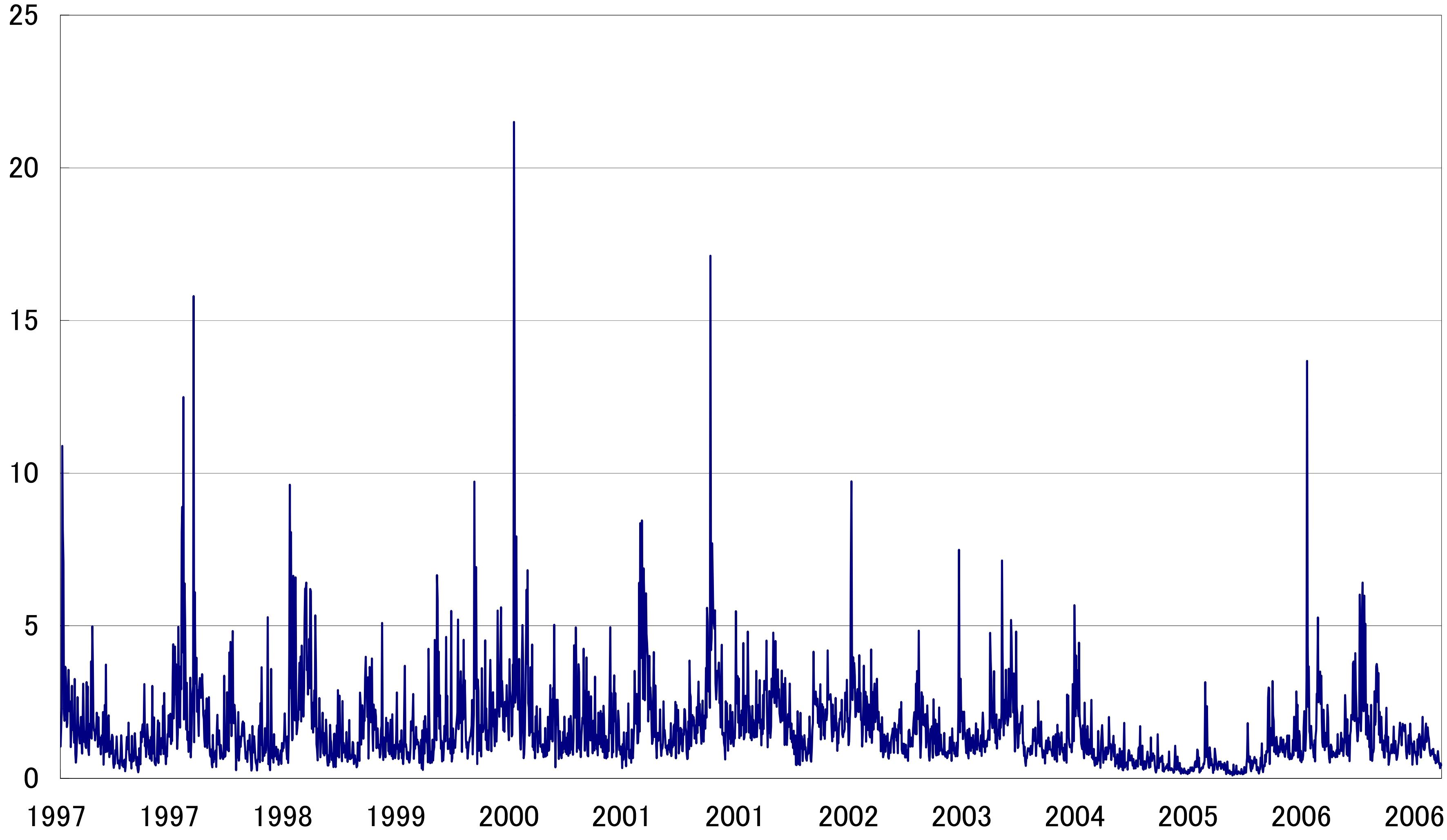

Figure 2: Sample Autocorrelation Coefficients in RV and log−RV

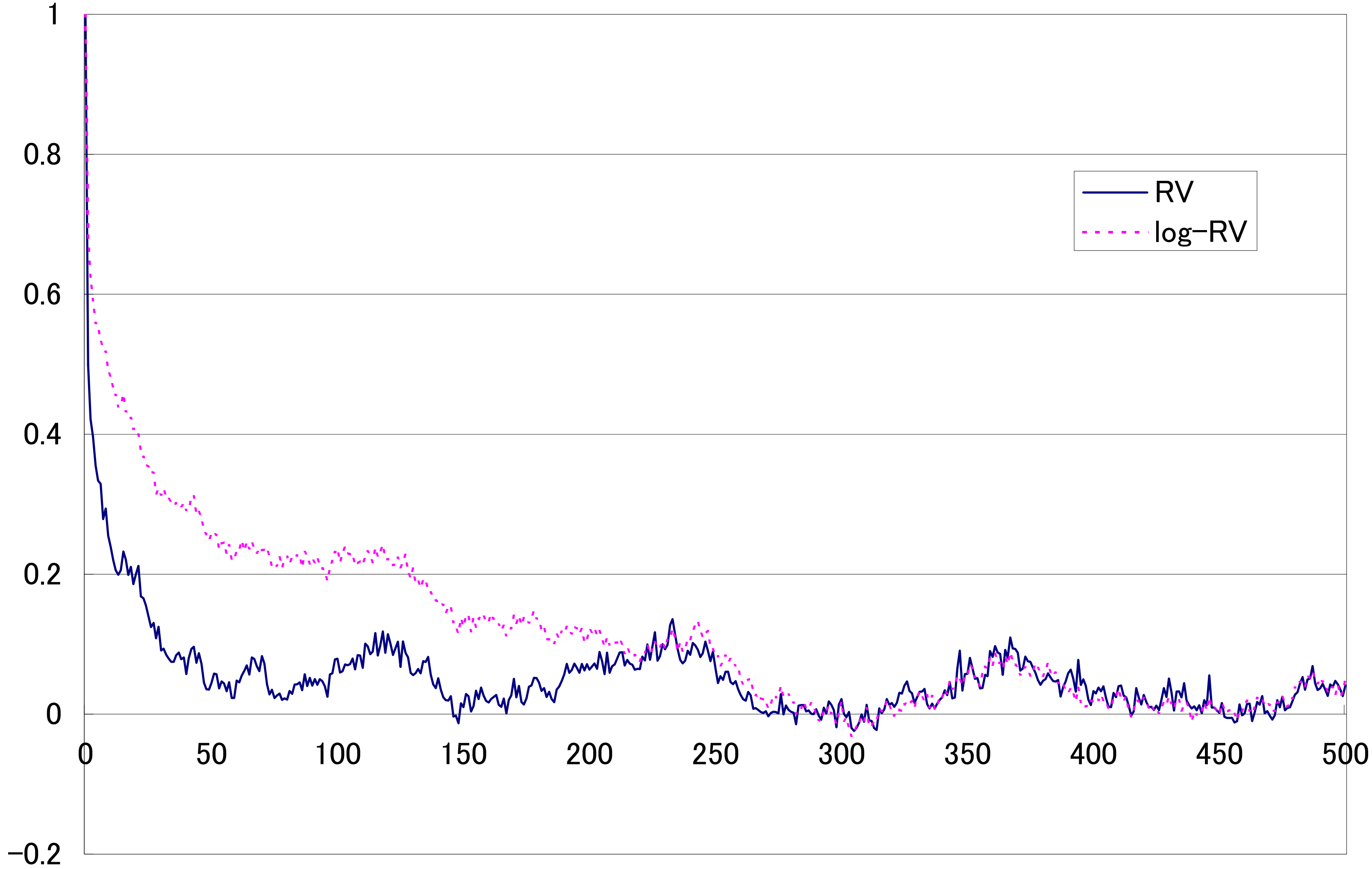

Figure 3: Histogram of Change Points