Bayesian Change Point Analysis of ARFIMA model for Realized Volatility

Naonori Kurata* Haruhisa Nishino†

*Graduate School of Humanities and Social Sciences, Chiba University

†Faculty of Law and Economics, Chiba University

November 17, 2012

Outline

Introduction (Realized Volatility)

- ARFIMA model has often used to estimate data that has a long memory property, for example financial data.

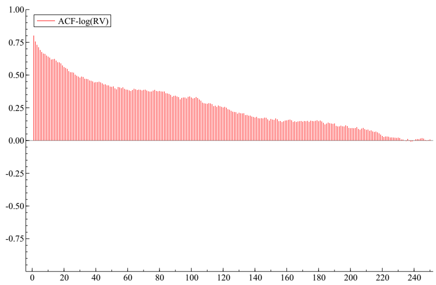

Figure: The sample ACF of the log realized volatility of Nikkei 225, 2001.7.2-2010.6.30

Introduction (Realized Volatility)

- This figure shows that the log realized volatility has a long memory property.
- It has been shown that the log realized volatility had a long memory property in previous studies, for example Watanabe (2007), Watanabe (2010), and Nishino (2010) etc.

Introduction (Estimation methods of ARFIMA model)

- Some methods are surveyed.

Estimation methods of ARFIMA model

- Beran (1995) proposes estimation model using an approximated AR(M) model.
- Chan and Palma (1998) proposes estimation model using an approximated MA(M) model.
- Robinson (2006) proposes the method using Conditional-sum-of-squares estimation (CSS) method.

- We want to estimate an ARFIMA model with change points of fractional difference or mean.

Introduction (Detecting the change points)

Some methods that can estimate the ARFIMA model with change points have been proposed in Bayesian framework.

- Ray and Tsay (2002) detects change points using the approximated MA(M) model.

Detect change points of μ

$$\mu_t = \mu_0 + \sum_{j=1}^{t} \delta_j \beta_j = \mu_{t-1} + \delta_t \beta_t. \quad (1)$$

If there is a change point, we set $\delta_t = 1$ otherwise $\delta_t = 0$.
β_t is a scale value from μ_{t-1} to μ_t when $\delta_t = 1$.

- Watanabe (2010) used an approximated AR(M) model which introduced by Beran (1995) and a hidden Markov model to detect change points.

  An approximated AR(M) model and a hidden Markov model are described later.

Introduction

- These methods in previous studies need the decision of the order of the approximated MA or AR model before estimation.
- Ray and Tsay (2002)'s method needs much time until finishing the calculation.

Our goal

We propose the method that can estimate ARFIMA model with change points using the Markov chain Monte Carlo (MCMC) method.

- The proposed method also uses an approximated AR model and a hidden Markov model.
- Conditional-sum-of-squares estimation (CSS) method with the approximated AR model is introduced by Robinson (2006).
- CSS method uses all observed residuals, so we need not decide the order of the approximated AR model.
- The hidden Markov model is used to detect multiple change points.
- The proposed method needs less calculation time than the method by Ray and Tsay (2002).
- We apply the proposed method to the simulation data, to the yearly minima of Nile river, and to the log realized volatility of Nikkei 225.

ARFIMA model

Introduce an ARFIMA(p, d, q) model estimates data with a long memory process.

Let $\{y_t\}$ is a long memory process.

- ARFIMA(p, d, q) model

$$\phi(L)(1-L)^d(y_t-\mu) = \theta(L)\varepsilon_t, \quad t = 1, 2, \dots, T. \qquad (2)$$

  - $\{\varepsilon_t\} \overset{i.i.d.}{\sim} WN(0, \sigma_\varepsilon^2)$, we use a Gaussian white noise.
  - d is a fractional difference and $0 < d < \frac{1}{2}$.
  - μ is mean.
  - L is the lag operator, $Ly_t = y_{t-1}$.
  - If the roots of $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p = 0$ and $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q = 0$ lie outside of the unit circle, the process has stationary and invertible.
  - And the roots have no common root.

Representation of likelihood function

To use the MCMC method, we need a likelihood function.

Beran (1994), Robinson (2003), and Palma (2007) give the survey of the estimation methods.

Beran (1995), Chan and Palma (1998), and Robinson (2006) propose the following estimation methods.

Various approximated likelihood functions

- Beran (1995) proposes the AR approximation method.
- Chan and Palma (1998) proposes the MA approximation method.
- Robinson (2006) proposes the conditional-sum-of-squares estimation (CSS) method.

Beran's AR approximation method

The difference between Beran's method and CSS method is whether uses M or not in the residuals.

The Beran's AR approximation method is

- The likelihood function is

$$L(Y_T|d, \mu, \sigma_\varepsilon^2, \Phi, \Theta) \propto \left(\frac{1}{\sigma_\varepsilon^2}\right)^{\frac{T}{2}} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\sum_{t=1}^{T} e_t^2\right\}, \qquad (3)$$

$$e_t = \sum_{j=0}^{\min(t-1, M)} \pi_j(d, \Phi, \Theta)(y_{t-j}-\mu). \qquad (4)$$

  - M is the order of the approximated AR model.
  - $Y_T = (y_1, y_2, \dots, y_T)'$.

Conditional-sum-of-squares estimation method

- The conditional-sum-of-squares estimation method
  - The likelihood function is represented as an AR approximation.
  - The CSS method needs less calculation time than the MA approximation method.
  - The CSS method needs not to decide the order of an approximated AR model.
  - The Beran's method can be seen as a special case of the CSS method.
- The likelihood function is

$$L(Y_T|d, \mu, \sigma_\varepsilon^2, \Phi, \Theta) \propto \left(\frac{1}{\sigma_\varepsilon^2}\right)^{\frac{T}{2}} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\sum_{t=1}^{T} e_t^2\right\}, \qquad (5)$$

$$e_t = \sum_{j=0}^{t-1} \pi_j(d, \Phi, \Theta)(y_{t-j}-\mu). \qquad (6)$$

Hidden Markov model

To estimate multiple change points of time series data, we use the irreversible hidden Markov model.

If the state shifts from state i to state j, the state can never shift to state i.

From Chib (1998), the hidden Markov model is

- Variable parameters are

$$\theta_k = \begin{cases} \theta_0, & 0 < t \leq t_1, \\ \theta_1, & t_1 < t \leq t_2, \\ \vdots & \\ \theta_m, & t_m < t \leq T. \end{cases} \tag{7}$$

- The hidden values are

$$S_T = (s_1, s_2, \ldots, s_T)', \tag{8}$$
$$s_t \in \{0, 1, \ldots, m\}. \tag{9}$$

Hidden Markov model

- The transition probability matrix is

$$P = \begin{pmatrix} p_{00} & p_{01} & 0 & \cdots & 0 & 0 \\ 0 & p_{11} & p_{12} & \cdots & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & & \cdots & p_{m-1,m-1} & p_{m-1,m} \\ 0 & 0 & & \cdots & 0 & 1 \end{pmatrix}. \tag{10}$$

  - The element $p_{ij} = P(s_t = j | s_{t-1} = i)$ on the matrix is the probability of state i to state j.
  - When we use this model, we decide the number of change points before estimation.
  - The number of change points should be compared with the log marginal likelihood.

Sampling $\{s_t\}$

When using the MCMC method, the hidden values $S_T = (s_1, \ldots, s_T)'$ and the elements of P are estimated.

And the $\{\theta_k\}$'s sampling method is described later.

- The conditional posterior distribution of S_T, $\pi(S_T|Y_T)$ by Chib (1996)

$$\begin{aligned} \pi(S_T|Y_T) &= \pi(s_{T-1}|Y_T, s_T, \boldsymbol{\theta}, P) \times \cdots \times \pi(s_t|Y_T, S^{t+1}, \boldsymbol{\theta}, P) \times \cdots \\ &\times \pi(s_1|Y_T, S^2, \boldsymbol{\theta}, P). \end{aligned} \tag{11}$$

From this conditional distribution, $\{s_t\}$ can be sampled from s_{T-1} to s_1.

Sampling $\{s_t\}$

$\{s_t\}$ is sampled by the following the probability function.

- Sampling $\{s_t\}$
  - s_t is sampled by $p(s_t|Y_T, S^{t+1}, \boldsymbol{\theta}, P)$

$$p(s_t|Y_T, S^{t+1}, \boldsymbol{\theta}, P) \propto p(s_t|Y_T, \boldsymbol{\theta}, P)p(s_{t+1}|s_t, P). \tag{12}$$

  - $\boldsymbol{\theta}$ is the parameter vector that consists of the variable parameters and the another parameters.
  - $S^{t+1} = (s_{t+1}, \ldots, s_T)'$, $S_t = (s_1, \ldots, s_t)'$.
  - $p(s_t|Y_T, \boldsymbol{\theta}, P)$ is the mass function.
  - $p(s_{t+1}|s_t, P)$ is the element of the transition probability matrix.

Sampling $\{s_t\}$

To sample $\{s_t\}$, we have to calculate the mass function.

- The mass function

$$p(s_t = k|Y_t, \boldsymbol{\theta}, P) = \frac{p(s_t = k|Y_{t-1}, \boldsymbol{\theta}, P) \times f(y_t|Y_{t-1}, \theta_k)}{\sum_{l=k-1}^{k} p(s_{t-1} = l|Y_{t-1}, \boldsymbol{\theta}, P) \times f(y_t|Y_{t-1}, \theta_l)}. \tag{13}$$

- The part of the numerator

$$p(s_t = k|Y_{t-1}, \boldsymbol{\theta}, P) = \sum_{l=k-1}^{k} p_{lk} \times p(s_{t-1} = l|Y_{t-1}, \boldsymbol{\theta}, P). \tag{14}$$

- The initial value

$$p(s_1 = 0|Y_0, \boldsymbol{\theta}) = 1. \tag{15}$$

where $f(y_t|Y_{t-1}, \theta_l)$ is the conditional distribution.

Sampling $\{p_{ii}\}$

$\{p_{ii}\}$ are sampled by the Gibbs sampler.

- Sampling $\{p_{ii}\}$
  - The prior distribution

$$p_{ii} \sim Beta(\gamma_1, \gamma_2). \tag{16}$$

  - The conditional posterior distribution

$$p_{ii}|\boldsymbol{\theta}, S_T \sim Beta(\gamma_1 + n_{ii}, \gamma_2 + 1). \tag{17}$$

    - n_{ii} is the number of one-step transitions from state i to state i.

Sampling step of the hidden Markov model

- Sampling step
  1. Sampling $\boldsymbol{\theta}$ and P
  2. Calculate the mass function $p(s_t|Y_T, S^{t+1}, \boldsymbol{\theta}, P)$
  3. Sampling $\{s_t\}$
     - s_{T-1} is sampled from $p(s_{T-1}|Y_T, s_T = m, \boldsymbol{\theta}, P)$,
     - s_{T-2} is sampled from $p(s_{T-2}|Y_T, S^{T-1}, \boldsymbol{\theta}, P)$,
     - $\vdots$
     - s_1 is sampled from $p(s_1|Y_T, s^2, \boldsymbol{\theta}, P)$.

  where we set $s_T = m$.
- Detect the time having a change point from $P(s_t|Y_T)$
  - We can get $P(s_t|Y_T)$ taking average of $p(s_t|Y_{t-1}, \boldsymbol{\theta}, P)$ over the MCMC iteration.

Previous studies

Ray and Tsay (2002) proposes a random persistence-shift (RPS) ARFIMA model and a random level-shift ARFIMA (RLS) model.

RPS and RLS-ARFIMA model (Ray and Tsay (2002))

- Estimate multiple change points of fractional difference (RPS).
- Estimate multiple change points of mean (RLS).
- Use the MA approximation method by Chan and Palma (1998).
- The hidden Markov model isn't used.

ARFIMA model with change points (Watanabe (2010))

- Estimate multiple change points of fractional difference, mean, and variance.
- Use the AR approximation method by Beran (1995).
- The hidden Markov model by Chib (1998) is used.

Proposed method

The proposed model

- Estimate multiple change points of fractional difference (RPS).
- Estimate multiple change points of mean (RLS).
- Use the CSS method by Robinson (2006).
- The hidden Markov model by Chib (1998) is used.

The proposed method follows the method by Watanabe (2010).

RPS-ARFIMA+CSS+HMM

Propose the another estimation method for RPS-ARFIMA model.

- RPS-ARFIMA+CSS+HMM

$$\phi(L)(1-L)^{d_{s_t}}(y_t - \mu) = \theta(L)\varepsilon_t, \tag{18}$$

$$d_k = \begin{cases} d_0, & 0 < t \le t_1, \\ d_1, & t_1 < t \le t_2, \\ \vdots & \\ d_m, & t_m < t \le T. \end{cases} \tag{19}$$

$$S_T = (s_1, s_2, \ldots, s_T)', \tag{20}$$

$$s_t \in \{0, 1, \ldots, m\}. \tag{21}$$

RPS-ARFIMA+CSS+HMM

- RPS-ARFIMA+CSS+HMM

$$P = \begin{pmatrix} p_{00} & p_{01} & 0 & \cdots & 0 & 0 \\ 0 & p_{11} & p_{12} & \cdots & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & & \cdots & p_{m-1,m-1} & p_{m-1,m} \\ 0 & 0 & & \cdots & 0 & 1 \end{pmatrix}. \tag{22}$$

- The likelihood function with CSS method

$$L(Y_T|d_0, \ldots, d_m, \mu, \sigma_\varepsilon^2, P, S_T) \propto \left(\frac{1}{\sigma_\varepsilon^2}\right)^{\frac{T}{2}} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\sum_{t=1}^{T} e_t^2\right\}, \tag{23}$$

$$e_t = \sum_{j=0}^{t-1} \pi_j(d_{s_t}, \Phi, \Theta)(y_{t-j} - \mu). \tag{24}$$

where we call this model as RPS-ARFIMA+CSS+HMM.

The prior distribution (RPS)

We describe RPS-ARFIMA$(0, d, 0)$+CSS+HMM to use the MCMC method.

- The prior distributions

$$d_k \sim \mathcal{U}(0, 0.5), \tag{25}$$

$$\mu \sim \mathcal{N}(\mu_0, \sigma_0^2), \tag{26}$$

$$\sigma_\varepsilon^2 \sim \mathcal{IG}\left(\frac{\nu_0}{2}, \frac{\lambda_0}{2}\right), \tag{27}$$

$$p_{ii} \sim Beta(a, b). \tag{28}$$

- The posterior distributions

$$\pi(d_0, \ldots, d_m, \mu, \sigma_\varepsilon^2, P, S_T) \propto L(\boldsymbol{\theta})\pi(d_0)\cdots\pi(d_m)\pi(\mu)\pi(\sigma_\varepsilon^2)\pi(p_{00})\cdots\pi(p_{mm}). \tag{29}$$

The conditional posterior distribution (RPS)

$\{d_k\}$ is estimated by an acceptance-reject (AR) MH algorithm.

The another parameters are estimated by the Gibbs sampler.

First, we show about the another parameters.

- The conditional posterior distributions

$$\sigma_\varepsilon^2 | d_0, \ldots, d_m, \mu, P, S_T, Y_T \sim \mathcal{IG}\left(\frac{\nu_0 + T}{2}, \frac{\lambda_0}{2} + \frac{1}{2}\sum_{t=1}^{T} e_t^2\right), \quad (30)$$

$$p_{ii} | d_0, \ldots, d_m, \mu, \sigma_\varepsilon^2, S_T, Y_T \sim Beta(a + n_{ii}, b + 1), \quad (31)$$

Conditional posterior distribution (RPS)

- The conditional posterior distributions

$$\mu | d_0, \ldots, d_m, \sigma_\varepsilon^2, P, S_T, Y_T \sim N(\mu^*, \sigma^{*2}), \quad (32)$$

$$\mu^* = \frac{\sigma_0^2 \sum_{t=1}^{T} c_t a_t + \sigma_\varepsilon^2 \mu_0}{\sigma_0^2 \sum_{t=1}^{T} c_t^2 + \sigma_\varepsilon^2}, \quad (33)$$

$$\sigma^{*2} = \frac{\sigma_0^2 \sigma_\varepsilon^2}{\sigma_0^2 \sum_{t=1}^{T} c_t^2 + \sigma_\varepsilon^2}, \quad (34)$$

$$c_t = \sum_{j=0}^{t-1} \pi_j(d_{s_t}), \quad (35)$$

$$a_t = \sum_{j=0}^{t-1} \pi_j(d_{s_t}) y_{t-j}. \quad (36)$$

Sampling $\{d_k\}$

Sampling $\{d_k\}$, we use an acceptance-rejection (AR) MH algorithm.

- The log proposal distribution $\ln q(d_k)$ following by Chib and Greenberg (1995) and Watanabe (2001).

  The log proposal distribution is the second-order Taylor expansion of the likelihood function around d_k^*.

$$\ln L(d_k^* | \boldsymbol{\theta}) \approx \ln L(d_k^* | \boldsymbol{\theta}) + \frac{\partial \ln L(d_k^* | \boldsymbol{\theta})}{\partial d_k}(d_k - d_k^*) + \frac{1}{2}\frac{\partial^2 \ln L(d_k^* | \boldsymbol{\theta})}{\partial d_k^2}(d_k - d_k^*)^2 = \ln q(d_k). \quad (37)$$

- d_k^* is the posterior mode.
- $\ln L(d_k^* | \boldsymbol{\theta})$ is the log likelihood function.
- $\boldsymbol{\theta}$ exclude the parameter d_k from the parameters.
- Mean: $d_k^* - \left(\frac{\partial \ln L(d_k^* | \boldsymbol{\theta})}{\partial d_k}\right) / \left(\frac{\partial^2 \ln L(d_k^* | \boldsymbol{\theta})}{\partial^2 d_k}\right)$
- Variance: $-\left(\frac{\partial^2 \ln L(d_k^* | \boldsymbol{\theta})}{\partial d_k^2}\right)^{-1}$

The Conditional distribution (RPS)

When sampling S_T, we use the conditional distribution $f(y_t | Y_{t-1}, \boldsymbol{\theta})$.

- The conditional distribution

$$f(y_t | Y_{t-1}, \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi v_t}} \exp\left\{-\frac{1}{2v_t}\sum_{j=0}^{t-1} \pi_j(d_{s_t})(y_{t-j} - \mu)^2\right\}, \quad (38)$$

$$v_t = Var(y_t - \hat{y}_t) = \gamma_0(d_{s_0}) \times \prod_{j=1}^{t-1}(1 - \phi_{jj}^2(d_{s_j})), \quad (39)$$

$$\phi_{tj} = -\binom{t}{j}\frac{\Gamma(j-d)\Gamma(t-d-j+1)}{\Gamma(-d)\Gamma(t-d+1)}. \quad (40)$$

where $\Gamma(\cdot)$ is the gamma function.

Sampling step (RPS)

Sampling step (RPS)

Step 0 Set the hyperparameters of the prior distributions and the initial values of the parameters.

Step 1 For $i = 1, 2, \ldots$, we iterate the next step.

  a Sampling $\{d_k\}^{(i)}$, $\mu^{(i)}$, $\sigma_\varepsilon^{2(i)}$, $\{p_{ii}^{(i)}\}$.

  b Sampling $S_T^{(i)}$

Step 2 For a sufficient large number N, we save $\{d_k^{(i)}\}$, $\mu^{(i)}$, $\sigma_\varepsilon^{2(i)}$, $\{p_{ii}^{(i)}\}$, $S_T^{(i)}$, $i = N, N+1, \ldots$.

RLS-ARFIMA+CSS+HMM

Next, we propose the another estimation method for RLS-ARFIMA model.

- RLS-ARFIMA+CSS+HMM

$$\phi(L)(1-L)^d(y_t - \mu_{s_t}) = \theta(L)\varepsilon_t, \quad (41)$$

$$\varepsilon_t \overset{i.i.d.}{\sim} N(0, \sigma_\varepsilon^2), \quad (42)$$

$$\mu_k = \begin{cases} \mu_0, & 0 < t \le t_1, \\ \mu_1, & t_1 < t \le t_2, \\ \vdots & \\ \mu_m, & t_m < t \le T. \end{cases} \quad (43)$$

$$S_T = (s_1, s_2, \ldots, s_T)', \quad (44)$$

$$s_t \in \{0, 1, \ldots, m\}, \quad (45)$$

RLS-ARFIMA+CSS+HMM

$$P = \begin{pmatrix} p_{00} & p_{01} & 0 & \cdots & 0 & 0 \\ 0 & p_{11} & p_{12} & \cdots & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & & \cdots & p_{m-1,m-1} & p_{m-1,m} \\ 0 & 0 & & \cdots & 0 & 1 \end{pmatrix}. \quad (46)$$

The likelihood function of RLS-ARFIMA+CSS+HMM

$$L(Y_T|d,\boldsymbol{\mu},\sigma_\varepsilon^2,P,S_T) \propto \left(\frac{1}{\sigma_\varepsilon^2}\right)^{\frac{T}{2}} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\sum_{t=1}^{T} e_t^2\right\}, \quad (47)$$

$$e_t = \sum_{j=0}^{t-1} \pi_j (y_{t-j} - \mu_{s_{t-j}}). \quad (48)$$

where $\boldsymbol{\mu} = (\mu_0, \ldots, \mu_m)'$ and we call this model as RLS-ARFIMA+CSS+HMM.

The prior distribution (RLS)

We describe RLS-ARFIMA$(0, d, 0)$+CSS+HMM.

- The prior distributions

$$d \sim \mathcal{U}(0, 0.5), \quad (49)$$

$$\sigma_\varepsilon^2 \sim \mathcal{IG}\left(\frac{\nu_0}{2}, \frac{\lambda_0}{2}\right), \quad (50)$$

$$p_{ii} \sim Beta(a, b), \quad (51)$$

$$\mu_k \sim \mathcal{N}(\mu_0, \sigma_0^2). \quad (52)$$

The conditional posterior distribution (RLS)

d is estimated by the AR-MH algorithm.

Another parameters are estimated by the Gibbs sampler.

- The conditional posterior distributions

$$\sigma_\varepsilon^2|d,\boldsymbol{\mu},P,S_T,Y_T \sim \mathcal{IG}\left(\frac{\nu_0+T}{2}, \frac{\lambda_0}{2} + \frac{1}{2}\sum_{t=1}^{T} e_t^2\right), \quad (53)$$

$$p_{ii}|d,\boldsymbol{\mu},\sigma^2,S_T,Y_T \sim Beta(a+n_{ii}, b+1). \quad (54)$$

The conditional posterior distribution (RLS)

- The conditional posterior distributions

$$\mu_k|d,\boldsymbol{\mu}_{(-k)},\sigma_\varepsilon^2,P,S_T,Y_T \sim \mathcal{N}(\mu_0^*, \sigma_0^{*2}), \quad (55)$$

$$\mu_0^* = \begin{cases} \dfrac{\sigma_0^2 \sum_{j=1}^{T} c_j a_j + \sigma_\varepsilon^2 \mu_0}{\sigma_0^2 \sum_{j=1}^{T} c_j^2 + \sigma_\varepsilon^2}, & k = 0 \\ \dfrac{\sigma_0^2 \sum_{j=t_k+1}^{T} c_j a_j + \sigma_\varepsilon^2 \mu_0}{\sigma_0^2 \sum_{j=t_k+1}^{T} c_j^2 + \sigma_\varepsilon^2}, & 1 \le k \le m \end{cases} \quad (56)$$

  - $\boldsymbol{\mu}_{(-k)}$ exclude μ_k from $\boldsymbol{\mu}$.
  - a_t exclude the terms of μ_k from e_t.

The conditional posterior distribution (RLS)

- The conditional posterior distributions

$$c_t = \begin{cases} \sum_{j=0}^{t-1} \pi_j, & t \le t_1,\ k = 0 \\ \sum_{j=t-t_1}^{t-t_1} \pi_j, & t_1 < t,\ k = 0 \\ \sum_{j=0}^{t-t_k-1} \pi_j, & t \le t_{k+1},\ 1 \le k < m \\ \sum_{j=t-t_{k+1}}^{t-t_k-1} \pi_j, & t_{k+1} < t,\ 1 \le k < m \\ \sum_{j=0}^{t-t_m-1} \pi_j, & k = m, \end{cases} \quad (57)$$

The conditional posterior distribution (RLS)

- The conditional posterior distributions

$$\sigma_0^{*2} = \frac{\sigma_0^2 \sigma_\varepsilon^2}{G}, \quad (58)$$

$$G = \begin{cases} \sigma_0^2 \sum_{j=1}^{T} c_j^2 + \sigma_\varepsilon^2, & k = 0, \\ \sigma_0^2 \sum_{j=t_k+1}^{T} c_j^2 + \sigma_\varepsilon^2, & 1 \le k \le m. \end{cases} \quad (59)$$

The conditional distribution (RLS)

When sampling S_T, we use the conditional distribution $f(y_t|Y_{t-1},\boldsymbol{\theta})$.

- The conditional distribution

$$f(y_t|Y_{t-1},\theta_k) = \frac{1}{\sqrt{2\pi v_t}} \exp\left\{-\frac{1}{2v_t}\sum_{j=0}^{t-1}\pi_j(d)(y_{t-j}-\mu_{s_t})^2\right\}, \quad (60)$$

$$v_t = Var(y_t-\hat{y}_t) = \gamma_0(d)\times\prod_{j=1}^{t-1}(1-\phi_{jj}^2(d)). \quad (61)$$

When we calculate $p(s_t=k|Y_t,\boldsymbol{\theta},P)$, the conditional distribution is

$$f(y_t|Y_{t-1},\mu_k^{(m+1)}) = \frac{1}{\sqrt{2\pi v_t}} \exp\left\{-\frac{1}{2v_t}\left((y_t-\mu_k^{(m+1)})+\sum_{j=1}^{t-1}\pi_j(d)(y_{t-j}-\mu_{s_j}^{(m)})\right)^2\right\}. \quad (62)$$

where $\mu_{s_j}^{(m)}$ is drawn at the iteration of the MCMC method.

Sampling step of RLS-ARFIMA+CSS+HMM

Sampling step

Step 0 Set the hyperparameters of the prior distributions and the initial values of the parameters.

Step 1 For $i=1,2,\ldots$, we iterate the next step.
- a Sampling $d^{(i)}$, $\{\mu_k\}^{(i)}$, $\sigma_\varepsilon^{2(i)}$, $\{p_{ii}^{(i)}\}$.
- b Sampling $S_T^{(i)}$

Step 2 For a sufficient large number N, we save $d^{(i)}$, $\{\mu_k^{(i)}\}$, $\sigma_\varepsilon^{2(i)}$, $\{p_{ii}^{(i)}\}$, $i=N,N+1,\ldots$.

Model comparison

- Chib (1998) uses the log marginal likelihood to compare the number of change points.
- To calculate the log marginal likelihood, we use the modified harmonic mean estimator by Geweke (1999).

Simulation

In this section, we see whether the proposed model can detect multiple change points or not.

- We use the simulation data having change points of d or μ.
- We make a comparison of the calculation time between the proposed method and the method by Ray and Tsay (2002) in RLS model.
- Simulation
  - First, we estimate RPS-ARFIMA$(0,d,0)$+CSS+HMM for the simulation data having two change points of d.
  - Next, we estimate RLS-ARFIMA$(0,d,0)$+CSS+HMM for the simulation having two change points of μ.
- The computer spec
  OS: Mac OS X Lion 10.7.5, Processor: 2.5GHz Intel Core i7, Memory: 8GB, Software: Ox version 6.21.

Simulation data (RPS)

We explain the set up of the simulation for RPS-ARFIMA$(0,d,0)$+CSS+HMM.

In this simulation, the change points were detected by every 2 periods.

- Simulation data: Sample size $T=1200$, $\mu=1.0$, $\sigma_\varepsilon^2=1.0$ and

$$d_k = \begin{cases} d_0=0.15, & 0<t\leq 449, \\ d_1=0.45, & 449<t\leq 849, \\ d_2=0.10, & 849<t\leq 1200. \end{cases} \quad (63)$$

- The hyperparameters of the prior distributions

$$\mu_0=0.0,\ \sigma_0^2=5.0,\ \nu_0=4.0,\ \lambda_0=4.0,\ a=8.0,\ b=0.1. \quad (64)$$

and (burn-in,draw)=(15000,10000).

Simulation data (RPS)

- The simulation data

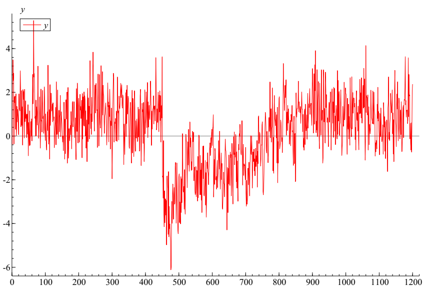

Figure: The simulation data having two change points of fractional difference

Log marignal likelihood (RPS)

The comparison of the proposed models with $\mathscr{M}_1$, $\mathscr{M}_2$, and $\mathscr{M}_3$.

Table: Log marginal likelihood of RPS-ARFIMA$(0, d, 0)$+CSS+HMM

	$\mathscr{M}_1$	$\mathscr{M}_2$	$\mathscr{M}_3$
Log marginal likelihood	-1749.1	-1738.5	-1741.5

Mean of the parameters with $\mathscr{M}_2$ (RPS)

Table: Esitmation result of RPS-ARFIMA$(0, d, 0)$+CSS+HMM with $\mathscr{M}_2$

	Mean	S.D.	2.5%	97.5%	CD	IF
μ	1.060	0.127	0.810	1.320	0.1	1.8
σ_ε^2	1.036	0.043	0.955	1.123	0.3	0.7
d_0	0.172	0.043	0.091	0.261	0.1	1.6
d_1	0.482	0.015	0.445	0.499	0.7	1.1
d_2	0.184	0.075	0.053	0.358	0.1	8.4
p_{00}	0.997	0.003	0.991	1.000	0.2	1.8
p_{11}	0.997	0.004	0.987	1.000	0.1	2.3

- The convergence diagnostic (CD) gives a criteria on whether a sample convergence or not, proposed by Geweke (1992).
- The inefficiency factor (IF) measures efficiency of sampling.

Posterior probability of $s_t = k$ with $\mathscr{M}_2$ (RPS)

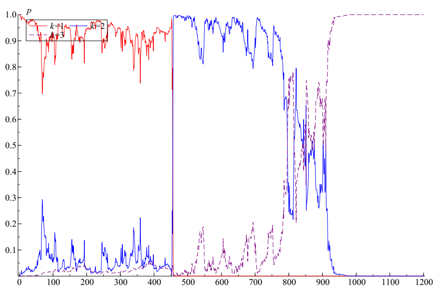

Figure: Posterior probability of $s_t = k$ with $\mathscr{M}_2$ given the data Y_T

Summary of RPS-ARFIMA$(0, d, 0)$+CSS+HMM

- From the log marginal likelihood, this model can estimate the true number of the change points
- From the posterior probability, RPS-ARFIMA$(0, d, 0)$+CSS+HMM can estimate the change points of the simulation data
- From the table, RPS-ARFIMA$(0, d, 0)$+CSS+HMM can also estimate change in parameters and another parameters

Simulation data (RLS)

We explain the set up of the simulation for RLS-ARFIMA$(0, d, 0)$+CSS+HMM.

In this simulation, the change points were detected by every 1 period.

And the calculation time were also compared between the proposed method and the method by Ray and Tsay (2002).

- The simulation data: Sample size $T = 1200$, $d = 0.4$, $\sigma_\varepsilon^2 = 1.0$ and

$$\mu_k = \begin{cases} \mu_0 = 0.0, & 0 < t < 350, \\ \mu_1 = 2.5, & 350 \leq t < 850, \\ \mu_2 = -1.0, & 850 \leq t \leq 1200. \end{cases} \quad (65)$$

- The hyperparameters

$$\mu_0 = 1.0,\ \sigma_0^2 = 5.0,\ \nu_0 = 4.0,\ \lambda_0 = 4.0,\ a = 8.0,\ b = 0.1. \quad (66)$$

Simulation data (RLS)

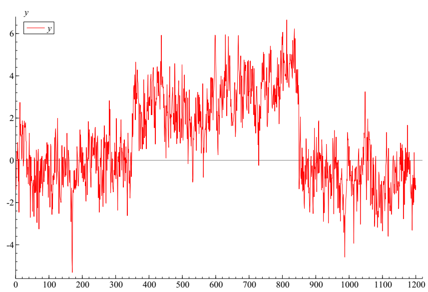

Figure: Simulation data having two change points of μ

Log marginal likelihood (RLS)

Table: Log marginal likelihood of RLS-ARFIMA$(0, d, 0)$+CSS+HMM

	$\mathscr{M}_1$	$\mathscr{M}_2$	$\mathscr{M}_3$
Log marginal likelihood	-1736.6	-1733.3	-1754.7

Estimation result of $\mathscr{M}_2$ (RLS)

Table: Estimation result of RLS-ARFIMA$(0, d, 0)$+CSS+HMM with $\mathscr{M}_2$

	Estimates	S.D.	2.5%	97.5%	CD	IF
d	0.396	0.023	0.350	0.443	0.9	1.7
σ_ε^2	0.997	0.041	0.920	1.081	0.5	0.9
μ_0	-0.065	0.380	-0.811	0.688	0.9	2.6
μ_1	2.790	0.426	1.935	3.618	0.9	4.1
μ_2	-1.135	0.496	-2.122	-0.171	0.5	3.3
p_{00}	0.997	0.003	0.989	1.000	0.6	1.0
p_{11}	0.998	0.002	0.992	1.000	0.1	1.0

Posterior probability of $s_t = k$ with $\mathscr{M}_2$ (RLS)

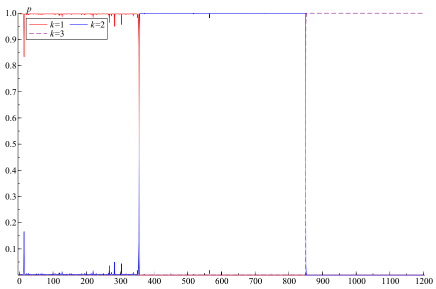

Figure: Posterior probability of $s_t = k$ with $\mathscr{M}_2$ given the data Y_T, Simulation data

Calculation time (RLS)

Table: Calculation time

	Ray and Tsay (2002)	$\mathscr{M}_1$	$\mathscr{M}_2$	$\mathscr{M}_3$
Calc.time	21:17'50.05	3:20'34.52	3:26'24.65	3:40'27.15

where the order is $M = 40$ and a change point was detected every 100 periods in Ray and Tsay (2002).

Summary of RLS-ARFIMA$(0, d, 0)$+CSS+HMM

- From the log marginal likelihood, we can select the true model $\mathscr{M}_2$ with two change points.
- From the table, this model can estimate the change points of this simulation data.
- From the table, this model can also estimate change in parameters and the another parameters.
- From the table, the proposed model needs less the calculation time than the method by Ray and Tsay (2002).

Summary of simulation result

- Using the hidden Markov model, we can estimate the ARFIMA model with multiple change points.
- The propose method can estimate the variable parameters.
- CSS method's calculation time is shorter than the time of the MA approximation method.
- The proposed method needn't to decide the order of an approximated AR model.

Applications

In this section, we estimate the data.

- The yearly minima of the Nile river
- The realized volatility of Nikkei 225

The yearly minima of the Nile river

It has been known that the yearly minima of the Nile river had a long memory property.

And this data that has one change point of d is shown by Beran and Terrin (1996).

Sample period is A.D.622-A.D.1284 and sample size is $T = 663$.

We use the models

- ARFIMA$(0, d, 0)$+CSS, $\mathscr{M}_0$
- RPS-ARFIMA$(0, d, 0)$+CSS+HMM with $\mathscr{M}_1$ and $\mathscr{M}_2$.

The yearly minima of the Nile river

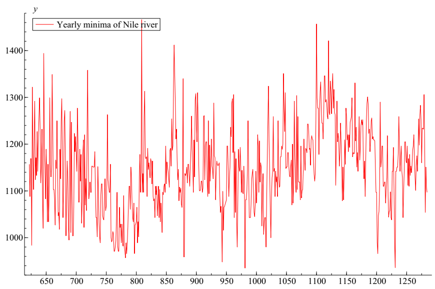

Figure: The yearly minima of the Nile River

Previous study (Nile river)

Beran and Terrin (1996) estimates the change points of d for the Nile river data.

Table: Estimation result of Beran and Terrin (1996)

	H	$d = H - \frac{1}{2}$
$t = 1, \ldots, 100$	0.5433	0.0433
$t = 101, \ldots, 200$	0.8531	0.3531
$t = 201, \ldots, 300$	0.8652	0.3652
$t = 301, \ldots, 400$	0.8281	0.3281
$t = 401, \ldots, 500$	0.8435	0.3435
$t = 501, \ldots, 600$	0.9354	0.4354

Beran and Terrin (1996) shows that d is different between for $t = 1, \ldots, 100$ and for $t = 101, \ldots$.

The prior distribution (Nile river)

We set the hyperparameters of the prior distributions.

- The hyperparameters of ARFIMA$(0, d, 0)$+CSS

$$\mu_0 = 1100.0,\ \sigma_0^2 = 200.0,\ \nu_0 = 4.0,\ \lambda_0 = 4.0. \quad (67)$$

- The hyperparameters of RPS-ARFIMA$(0, d, 0)$+CSS+HMM

$$\mu_0 = 1100.0,\ \sigma_0^2 = 200.0,\ \nu_0 = 4.0,\ \lambda_0 = 4.0,\ a = 8.0,\ b = 0.1. \quad (68)$$

- (burn-in,draw) are (10000, 10000).
- When we use RPS-ARFIMA+CSS+HMM, we detect a change point every 10 periods.

Log marginal likelihood (Nile river)

Table: Log marginal likelihood of the yearly minima of Nile River

	$\mathscr{M}_0$	$\mathscr{M}_1$	$\mathscr{M}_2$
Log marginal likelihood	-3778.0	-3777.5	-3778.9

Estimation result (Nile river)

Table: Estimation result of $\mathscr{M}_1$

	Estimates	S.D.	2.5%	97.5%	CD	IF
d_0	0.175	0.114	0.010	0.435	0.9	1.1
d_1	0.424	0.034	0.360	0.488	0.9	1.7
μ	1118.960	13.791	1091.838	1145.155	0.3	1.0
σ_ε^2	4856.900	274.044	4348.937	5425.553	0.4	1.0
p_{00}	0.956	0.048	0.827	0.999	0.5	0.9

Posterior probability of $s_t = k$ with $\mathscr{M}_1$ (Nile river)

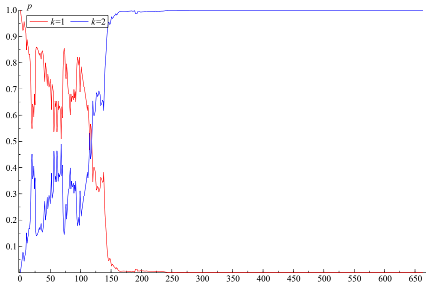

Figure: Posterior probability of $s_t = k$ with $\mathscr{M}_1$ given the data Y_T

Result (Nile river)

- $\mathscr{M}_1$ has the largest log marginal likelihood among these models.
- There is one change point of d in the yearly minima of the Nile river.
- From the figure, the change point is around at $t = 120$, A.D.742.
- We can see the this estimation result is consistent with Beran and Terrin (1996).

Realized volatility

We use the realized volatility of Nikkei 225 made by five-minutes log-return.

Realized volatility is the sum of the square of intraday returns by Hansen and Lunde (2005) and Watanabe (2007).

Realized volatility (RV) by Hansen and Lunde (2005) is

$$RV_t = c\sum_{i=1}^{n} r_{i,t}^2, \tag{69}$$

$$c = \frac{\sum_{t=1}^{T}(R_t - \overline{R})^2}{\sum_{t=1}^{T} r_{i,t}^2}. \tag{70}$$

- $r_{i,t}$ is the ith intraday log-return at date t.
- R_t is a dairy log-return.
- $\overline{R}$ is the sample mean of dairy log-return.

The empirical result and conclusion will be given today.

References

Beran, J. (1994) *Statistics for Long-Memory Processes*, Chapman & Hall, London.

Beran, J. (1995) “Maximum Likelihood Estimation of the Differencing Parameter for Invertible Short and Long Memory Autoregressive Integrated Moving Average Models”, *Journal of the Royal Statistical Society. Ser. B* **57** 659–672.

Beran, J. and Terrin, N. (1996) “Testing for a change of the long-memory parameter”, *Biometrica* **83** 627–638.

Chan, N. H. and Palma, W. (1998) “State Space Modeling of Long-Memory Processes”, *Annals of Statistics* **26** 719–740.

References

Chib, S. (1996) “Calculating posterior distributions and modal estimates in Markov mixture models”, *Journal of Econometrics* **75** 79–97.

Chib, S. (1998) “Estimation and comparison of multiple change-point models”, *Journal of Econometrics* **86** 221–241.

Chib, S. and Greenberg, E. (1995) “Understanding the Metropolis-Hastings Algorithm”, *The American Statistician* **49** 327–335.

Doornik, J. A. (2011). *Ox: Object Oriented Matrix Programming Language*, Version 6.21, Timberlake Consultants Press, London.

Geweke, J. (1992) “Evaluating the accuracy of sampling-based approaches to calculating posterior moments”, In Bernado, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M. (eds.), *Bayesian Statistics*, volume 4. London: Clarendon Press.

References

Geweke, J. (1999) "Using simulation methods for bayesian economic models: inference, development, and communication", *Econometric Reviews* **18** 1–73.

Hansen, P. R. and Lunde, A. (2005) "A Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?", *Journal of Applied Econometrics* **20**, 873–889.

Nishino, H. (2010) "Tests on Long Memory Time Series (in Japanese)", *Journal of The Japan Statistical Society* **40**, 147–175.

Palma, W. (2007) *Long-Memory Time Series: Theory and Methods*, Wiley, New York.

References

Ray, B.K. and Tsay, R.S. (2002) "Bayesian methods for change-point detection in long-range dependent processes", *Journal of Time Series Analysis* **23** 687–705.

Robinson, P.M. (ed.) (2003) *Time Series with Long Memory*, Oxford, Oxford.

Robinson, P.M. (2006) "Conditional-Sum-of-Squares Estimation of Models for Stationary Time Series with Long Memory" In *Time Series and Related Topics: In Memory of Ching-Zong Wei.* (H.-C. Ho, C.-K. Ing and T.L. Lai, eds.). IMS Lecture Notes - Monograph Series, 52, 130–137.

Watanabe, T. (2001) "On Sampling the Degree-of-Freedom of Student's-t Disturbances", *Statistics & Probability Letters*, **52**, 177–181.

References

Watanabe, T. (2007) "Realized Volatility (in Japanese)", *The Economic Review*, **58**, 352–373.

Watanabe, T. (2010) "Bayesian analysis of structural changes in ARFIMA Models with an application to realized volatility", *International Workshop on Bayesian Econometrics and Statistics*.