Statistica Sinica **21** (2011), 1665-1686
doi:http://dx.doi.org/10.5705/ss.2009.237

ANALYSIS OF TIME SERIES WITH MULTIPLE SHIFTS OF LEVELS AND VOLATILITIES

Heping He

University of York

Abstract: A practical time series model is proposed with multiple shifts of levels and volatilities to overcome the intrinsic limitations of hidden Markov models used to capture change-point type behaviors of data. This model allows the set of level change points to be different from the set of volatility change points. Least square methods are then applied to the model to estimate level and volatility change points, those levels and volatilities. Asymptotic properties of the estimators, including their consistency, convergence rates and asymptotic distributions, are established under relatively weak conditions. Some simulations are carried out, showing that this model, its inference methods, and the asymptotic theory work quite well.

Key words and phrases: ARMA model, break fraction, change point, hidden Markov model, least square method.

1. Introduction

Hidden Markov models, first proposed by Baum and Petrie (1966), have been widely used to model structural changes in econometric contexts and others. See Goldfeld and Quandt (1973), Hamilton (1989), Engle and Hamilton (1990), and Rabiner (1989) for examples. One of the hidden Markov models proposed by Buckle, Haugh, and Thomson (2004) is

$$Y_t = \mu_{S_t} + \sigma_{S_t} X_t \qquad (t = 0, \pm 1, \pm 2, \ldots)$$

where the states S_t form an unobserved stationary Markov chain that takes the values $1, \ldots, N$ that index the states of the system; the level μ_{S_t} and the volatility σ_{S_t} switch among the N pairs $(\mu_1, \sigma_1), \ldots, (\mu_N, \sigma_N)$ according to S_t; the stochastic process X_t is assumed to be a zero mean stationary Gaussian process with unit variance independent of S_t, for example, the process $AR(1)$. By using maximum likelihood, Buckle, Haugh, and Thomson (2004) show that this model can capture change-point type behavior of means and volatilities in a wide range of contexts, for example, from GDP growth to asset prices to rainfall. However, there are some intrinsic limitations to this and other hidden Markov models. First, simulations in Buckle, Haugh, and Thomson (2004) and other papers show that the models may encourage the Markov chain to change states more often than is suggested by the data. Second, it is difficult to fit the models without making parametric assumptions about the error terms of the process X_t. However, contemporary inference for the process X_t, for example the autoregressive process that X_t may follow, is often nonparametric insofar as the distribution of the disturbances is concerned; the only parametric part is the structure defining the way in which the disturbances are built into the process X_t. Third, there seems to be no way to construct analogues of residuals when using the hidden Markov model approach. We need residuals to do nonparametric inference, for example, through simulating the process Y_t by resampling the residuals rather than simulating an assumed distribution of the disturbances.

This paper proposes a practical time series model with multiple changes of means and volatilities, through which these intrinsic limitations of the hidden Markov models can be fixed. Assume that observations $y_1, y_2, \ldots, y_T$ are generated by a model of the form

$$Y_t = \alpha_t(\theta_1) + \beta_t(\theta_2) X_t, \qquad t = 0, \pm 1, \pm 2, \ldots, \tag{1.1}$$

where

$$\begin{aligned}
&\alpha_t(\theta_1) = \alpha_i \ \text{ if } \ m_{i-1}^0 + 1 \le t \le m_i^0 \ \text{ for } \ i = 1, \ldots, k+1;\\
&\beta_t(\theta_2) = \beta_i \ \text{ if } \ n_{i-1}^0 + 1 \le t \le n_i^0 \quad \text{ for } \ i = 1, \ldots, l+1;\\
&\beta_t(\theta_2) \text{ is nonnegative}, \ m_0^0 = n_0^0 = 0, \ \text{ and } \ m_{k+1}^0 = n_{l+1}^0 = T;\\
&\theta_1 = (\alpha_1, \ldots, \alpha_k, \alpha_{k+1}, m_1^0, m_2^0, \ldots, m_k^0);\\
&\theta_2 = (\beta_1, \ldots, \beta_l, \beta_{l+1}, n_1^0, n_2^0, \ldots, n_l^0);\\
&X_t \text{ is stationary and } \ X_t = \sum_{i=1}^{p} a_i X_{t-i} + \epsilon_t + \sum_{i=1}^{q} b_i \epsilon_{t-i};\\
&\epsilon_t\text{'s } \text{ are i.i.d. with zero mean and variance } \ \sigma_0^2.
\end{aligned}$$

In order to guarantee identifiability of $\beta_t(\theta_2)$, we constrain it by

$$\sum_{t=1}^{T} \beta_t^2(\theta_2) = C_0, \tag{1.2}$$

where C_0 is a priori known constant. This paper shows that change-point type behaviors of means and volatilities of time series data can be accurately captured by this model without assumptions about distribution of disturbances ϵ_t, and that residuals can be estimated. In most practice, concerns are about means and variances of random phenomena. Thus, almost all technical analysis used in the finance industry and by investors is about the means and variances of financial asset prices. Our model simultaneously considers multiple shifts of levels and volatilities that can happen at two different sets of time points. The model is then quite general and useful; it can be used to simultaneously capture change-point type behaviors of levels and volatilities for a wide range of time series in economics, finance, climatology, sociology, and more.

The rest of this paper is organized as follows. Section 2 provides estimates of θ_1, θ_2, and the other parameters defining the model. Section 3 contains results on the consistency, convergence rates, and asymptotic distributions of estimates of parameters related to level shifts. Section 4 describes results on the consistency, convergence rates, and asymptotic distributions of estimates of parameters related to volatility shifts. It also provides brief proofs of some lemmas and theorems. Section 5 uses simulations to show that our model, its estimates, and asymptotic theories work quite well; it also contains some discussions and conclusions. Appendix A gives proofs of lemmas and theorems in Section 3.

2. Estimation Methodologies

Change-point problems may be estimated by maximum likelihood or least squares. Hinkley (1970), Bhattacharya (1987) and Yao (1987) use maximum likelihood in their change point problems having independent data. Picard (1985) uses maximum likelihood to estimate a shift in a Gaussian autoregressive process with a known order. Yao (1989) and Bai (1994) use least squares to estimate change points in independent and dependent data, respectively. Unlike maximum likelihood, least squares does not need to specify the error distribution function and is computationally much simpler. Least squares also allows a broader specification of correlation structure in the data than maximum likelihood method can typically permit. We use least squares to estimate parameters related to the multiple changes of levels and volatilities. The numbers k and l of change points of levels and volatilities are assumed to be known; however, they can be conveniently estimated by penalized least squares. Note that determination of number of change points is model selection problem, so various model selection criteria, such as those based on penalized likelihood, can be put to use. Interested readers may refer to Yao (1989), Yao (1988), and Schwarz (1978) for the use of penalized least squares and penalized likelihood in estimating the number of change points. We employ penalized least squares to estimate the numbers k and l of change points. Specifically, the estimates of k and l are obtained by minimizing $L + \alpha_T k$ and $M + \beta_T l$, respectively, where L and M are given at (3.1) and (4.1), α_T and β_T are positive and tend to zero as T tends to infinity. There is much literature on estimating parameters in ARMA processes, see Chapter 8 of Brockwell and Davis (1991), for example.

First consider estimation of the means $\alpha_1, \dots, \alpha_{k+1}$ and their shift points $m_1^0, \dots, m_k^0$. For any mean change point configuration $(m_1, \dots, m_k)$ satisfying $0 = m_0 < \cdots < m_k < m_{k+1} = T$, let

$$\tilde{\alpha}_j = \frac{1}{m_j - m_{j-1}} \sum_{t=m_{j-1}+1}^{m_j} Y_t \quad \text{for} \quad j = 1, \dots, k+1. \tag{2.1}$$

Then estimates $\hat{m}_1, \dots, \hat{m}_k$ of the mean change points $m_1^0, \dots, m_k^0$ are given by

$$(\hat{m}_1, \dots, \hat{m}_k) = \operatorname*{argmin}_{0=m_0<\cdots<m_{k+1}=T} \sum_{j=1}^{k+1} \sum_{i=m_{j-1}+1}^{m_j} (y_i - \tilde{\alpha}_j)^2. \tag{2.2}$$

Replacing $m_1, \dots, m_k$ with $\hat{m}_1, \dots, \hat{m}_k$ in (2.1) gives estimates $\hat{\alpha}_1, \dots, \hat{\alpha}_{k+1}$ of means $\alpha_1, \dots, \alpha_{k+1}$ by

$$\hat{\alpha}_j = \frac{1}{\hat{m}_j - \hat{m}_{j-1}} \sum_{t=\hat{m}_{j-1}+1}^{\hat{m}_j} Y_t \quad \text{for} \quad j = 1, \dots, k+1. \tag{2.3}$$

Next consider estimation of the vector parameter θ_2 of volatility shifts. Let $\sigma^2 = E(X_t^2)$. Since $\beta_t^2(\theta_2)\sigma^2$ is the mean of $(Y_t - \alpha_t(\theta_1))^2$, a similar method is adopted to estimate θ_2 using $(Y_t - \hat{\alpha}_t(\hat{\theta}_1))^2, t = 1, \dots, T$. For any volatility change point configuration $0 = n_0 < n_1 < \cdots < n_{l+1} = T$, let

$$\tilde{\beta}_j = \frac{1}{n_j - n_{j-1}} \sum_{t=n_{j-1}+1}^{n_j} (Y_t - \hat{\alpha}_t(\hat{\theta}_1))^2 \quad \text{for} \quad j = 1, \dots, l+1. \tag{2.4}$$

Then estimates $\hat{n}_1, \dots, \hat{n}_l$ of volatility change points $n_1^0, \dots, n_l^0$ are given by

$$(\hat{n}_1, \dots, \hat{n}_l) = \operatorname*{argmin}_{0=n_0<\cdots<n_{k+1}=T} \sum_{j=1}^{l+1} \sum_{i=n_{j-1}+1}^{n_j} [(y_i - \hat{\alpha}_i(\hat{\theta}_1))^2 - \tilde{\beta}_j]^2. \tag{2.5}$$

Replacing $n_1, \dots, n_l$ with $\hat{n}_1, \dots, \hat{n}_l$ in (2.4) gives quantities $\check{\beta}_1, \dots, \check{\beta}_{l+1}$ as

$$\check{\beta}_j = \frac{1}{\hat{n}_j - \hat{n}_{j-1}} \sum_{t=\hat{n}_{j-1}+1}^{\hat{n}_j} (Y_t - \hat{\alpha}_t(\hat{\theta}_1))^2 \quad \text{for} \quad j = 1, \dots, l+1. \tag{2.6}$$

Then we estimate σ^2 by

$$\hat{\sigma}^2 = \frac{1}{C_0} \sum_{j=1}^{l+1} (\hat{n}_j - \hat{n}_{j-1}) \check{\beta}_j,$$

where $\hat{n}_0 = 0$, $\hat{n}_{l+1} = T$. Thus estimate $\beta_t(\theta_2)$ by

$$\hat{\beta}_t(\hat{\theta}_2) = \Big(\frac{\check{\beta}_i}{\hat{\sigma}^2}\Big)^{1/2}, \quad \text{for} \quad \hat{n}_{i-1} + 1 \le t \le \hat{n}_i, i = 1, \ldots, l+1.$$

Finally we have the estimated data set $\hat{X}_t = (Y_t - \hat{\alpha}_t(\hat{\theta}_1))/\hat{\beta}_t(\hat{\theta}_2), t = 1, \ldots, T$ for the estimations of parameters in the ARMA model that X_t follows. This is not of interest here.

3. Asymptotic Theory for Mean Shifts

The problem of multiple structural changes and its theory have received some attention, mostly in the context of regression without any change of variances. Yao (1989) provides a comprehensive treatment of multiple changes of means, and uses penalized least squares to estimate their number. Liu, Wu, and Zidek (1997) consider multiple structural changes in the context of a general threshold model and propose an information criterion for the selection of the number of changes. Our analysis of multiple changes of means differs from existing literature because it is affected by the multiple changes of variances; furthermore, analysis of multiple variance changes is affected by fourth-order moments having multiple changes. Because of the multiple changes of variance, direct proofs of Lemmas 1 and 4 are difficult. We use induction, as in Móricz, Serfling, and Stout (1982), to obtain these important results. Because of the multiple changes of variances, the results in Sections 4 and 5 require some adjustments not found in existing literatures. This section builds asymptotic theory concerning mean shifts; the next section builds asymptotic theory about variance shifts.

Let $\lambda_i^0 = m_i^0/T$ and $\hat{\lambda}_i = \hat{m}_i/T$ for $i = 1, \ldots, k$, and $\lambda_i = m_i/T$ for $0 < m_1 < \cdots < m_k < T$. Note that the true break fractions λ_i^0, $i = 1, \ldots, k$, are supposed to be constants here. Define $\tilde{m}_{ij} = [m_{i-1}+1, m_{i-1}+2, \ldots, m_i] \cap [m_{j-1}^0 + 1, m_{j-1}^0 + 2, \ldots, m_j^0]$, and let m_{ij} be the number of observations in $\{Y_t | t \in \tilde{m}_{ij}\}$, $m_{i\cdot}$ the number of observations in $\{Y_t | t \in [m_{i-1}, m_i]\}$, and $m_{\cdot j}$ the number of observations in $\{Y_t | t \in [m_{j-1}^0, m_j^0]\}$. Define

$$L = \frac{1}{T}\sum_{j=1}^{k+1}\sum_{i=m_{j-1}+1}^{m_j}(y_i - \tilde{\alpha}_j)^2 - \frac{1}{T}\sum_{j=1}^{k+1}\sum_{i=m_{j-1}^0+1}^{m_j^0}(y_i - \alpha_j)^2; \tag{3.1}$$

$$L_1 = \frac{1}{T}\sum_{j=1}^{k+1}\sum_{i=1}^{k+1} m_{ji}(\alpha_i - \tilde{\alpha}_j)^2; \tag{3.2}$$

$$L_2 = \frac{1}{T}\sum_{j=1}^{k+1}\sum_{i=1}^{k+1}[(-2\tilde{\alpha}_j \sum_{t\in\tilde{m}_{ji}} \beta_t(\theta_2)X_t) - (-2\alpha_i \sum_{t\in\tilde{m}_{ji}} \beta_t(\theta_2)X_t)]. \tag{3.3}$$

It is straightforward to show that $L = L_1 + L_2$. The following result is a type of Hájek-Rényi inequality for this model, and Lemma 2 provides a lower bound for L_1.

Lemma 1. *For any $0 \le T_1 \le t_1 < t_2 \le T_2 \le T$, let $d(t_2, t_1)$ be a positive function which is non-increasing in t_2 and non-decreasing in t_1. For any constant $0 \le C_0 \le T_2 - T_1 - 1$, and any $\delta > 0$, there exists a constant $C < \infty$ such that*

$$P_r\Big(\max_{T_1+C_0 \le t_1+C_0 < t_2 \le T_2} d(t_2,t_1)\Big|\sum_{i=t_1+1}^{t_2}\beta_i(\theta_2)X_i\Big| > \delta\Big) \le C\frac{C_0+1}{\delta^2}\sum_{i=T_1+1}^{T_2} d(i,T_1)^2. \tag{3.4}$$

Define

$$\underline{\alpha} = \min_{1\le i\le k}|\alpha_{i+1}-\alpha_i|, \quad \overline{\alpha} = \max_{1\le i\le k}|\alpha_{i+1}-\alpha_i|, \quad \lambda = (\lambda_1,\ldots,\lambda_k);$$
$$\hat\lambda = (\hat\lambda_1,\hat\lambda_2,\ldots,\hat\lambda_k), \quad \lambda^0 = (\lambda_1^0,\ldots,\lambda_k^0), \quad \triangle_\lambda^0 = \min_{1\le i\le k}|\lambda_{i+1}^0-\lambda_i^0|;$$
$$\|\lambda-\lambda^0\|_\infty = \max_{1\le i\le k}|\lambda_i-\lambda_i^0|, \quad \|\hat\lambda-\lambda^0\|_\infty = \max_{1\le i\le k}|\hat\lambda_i-\lambda_i^0|;$$
$$\|\tilde\alpha-\alpha\|_\infty^2 = \max_{1\le i\le k+1}|\tilde\alpha_i-\alpha_i|^2.$$

Lemma 2. *There exists a constant $C > 0$ such that, for all $T \ge 1$ and all $0 < m_1 < m_2 < \cdots < m_k < T$, we have $L_1 \ge C\|\lambda-\lambda^0\|_\infty$.*

A first result establishes the consistency of the estimates of mean break fractions, a second shows that their convergence rates are of order T. Then, with the convergence rates, we derive the asymptotic distributions of the estimates $(\hat\alpha_1,\ldots,\hat\alpha_{k+1})$ of means.

Theorem 1. *The estimated break fractions of means converge to their true values in probability, $\hat\lambda_i \xrightarrow{p} \lambda_i^0$ for $i = 1,\ldots,k$.*

Theorem 2. *Suppose that X_t is a stationary ARMA model, then we have for every $\eta > 0$, there exists a positive number $\delta < \infty$ such that, for all large T, $P_r(|T(\hat\lambda_j-\lambda_j^0)| > \delta) < \eta$ for $j = 1,\ldots,k$.*

Theorem 3. *Suppose that X_i is stationary ARMA process, then $\sqrt{T}(\hat\alpha_j-\alpha_j) \xrightarrow{d} N_j(0,\upsilon_j^2)$, independently in the limit, and where υ_j^2 is defined at* (A.10), *$j = 1,\ldots,k+1$.*

When the mean shift is constant independent of T, the results of Hinkley (1970) and Hinkley and Hinkley (1970) for the i.i.d. case indicate that the limiting distribution of the mean break fraction depends on the underlying distribution of the innovations and also on the mean shift, in an intricate way; the limit distributions are of little practical use. In order to obtain useful asymptotic distributions of mean break fractions, consider the following.

Assumption 1. *Let* $\delta_j = \alpha_{j+1} - \alpha_j$ *for* $j = 1, \ldots, k$. *Then* $\delta_j \to 0$ *and* $T^{1/2}\delta_j \to \infty$ *as* $T \to \infty$, *for all* $j = 1, \ldots, k$.

Theorem 4. *Under Assumption* 1, (i) $\hat{\lambda}_j \xrightarrow{p} \lambda_j^0$ *for* $j = 1, \ldots, k$. (ii) *for every* $\eta > 0$, *there exists a* $C < \infty$ *such that for all large* T, $P_r(|T\delta_j^2(\hat{\lambda}_j - \lambda_j^0)| > C) < \eta$ *for* $j = 1, \ldots, k$.

The conclusion (i) can be proved in a way similar to Theorem 1, and (ii) can be proved in a way similar to Theorem 2. We omit these proofs to avoid repetition.

Let

$$L_3 = \sum_{j=1}^{k+1} \sum_{i=m_{j-1}+1}^{m_j} \Big(\alpha_i(\theta_1) - \frac{1}{m_j - m_{j-1}} \sum_{t=m_{j-1}+1}^{m_j} \alpha_t(\theta_1)\Big)^2,$$

$$L_4 = \sum_{j=1}^{k+1} [\frac{1}{m_j^0 - m_{j-1}^0}\Big(\sum_{i=m_{j-1}^0+1}^{m_j^0} \beta_i(\theta_2)X_i\Big)^2 - \frac{1}{m_j - m_{j-1}}(\sum_{i=m_{j-1}+1}^{m_j} \beta_i(\theta_2)X_i)^2],$$

$$L_5 = 2\sum_{j=1}^{k+1}\Big[\Big(\sum_{i=m_{j-1}^0}^{m_j^0} \beta_i(\theta_2)X_i\Big)\alpha_j - \Big(\sum_{i=m_{j-1}+1}^{m_j} \beta_i(\theta_2)X_i\Big)\frac{1}{m_j - m_{j-1}} \sum_{t=m_{j-1}+1}^{m_j} \alpha_t(\theta_1)\Big].$$

It is then straightforward to show that

$$TL + \sum_{j=1}^{k+1}[\frac{1}{m_j^0 - m_{j-1}^0}(\sum_{i=m_{j-1}^0+1}^{m_j^0} \beta_i(\theta_2)X_i)^2] = L_3 + L_4 + L_5.$$

Let

$$\Omega_{i,1} = \lim_{T\to\infty} E[\frac{1}{m_i^0 - m_{i-1}^0}(\sum_{t=m_{i-1}^0+1}^{m_i^0} [y_t - \alpha_t(\theta_1)])^2],$$

$$\Omega_{i,2} = \lim_{T\to\infty} E[\frac{1}{m_{i+1}^0 - m_i^0}(\sum_{t=m_i^0+1}^{m_{i+1}^0} [y_t - \alpha_t(\theta_1)])^2].$$

For any $i = 1, \ldots, k$, let $W_{ij}(s), j = 1, 2$ be two independent standard Wiener processes defined on $[0, \infty)$, starting at the origin when $s = 0$. Let

$$Z^{(i)}(s) = \begin{cases} \Omega_{i,1}^{1/2} W_{i1}(-s) - \frac{|s|}{2}, & \text{if } s \leq 0 \\ \Omega_{i,2}^{1/2} W_{i2}(s) - \frac{|s|}{2}, & \text{if } s > 0. \end{cases}$$

Lemma 3. *For $j = 1, \ldots, k$,*

$$\delta_j \sum_{i=m_j^0+1}^{m_j^0+s_j\delta_j^{-2}} \beta_i(\theta_2) X_i \Rightarrow \Omega_{j,2}^{1/2} W_{i2}(s_j), \tag{3.5}$$

$$\delta_j \sum_{i=m_j^0-s_j\delta_j^{-2}}^{m_j^0} \beta_i(\theta_2) X_i \Rightarrow \Omega_{j,1}^{1/2} W_{i1}(s_j), \tag{3.6}$$

where $0 \leq s_j \leq D$, a constant, and "$\Rightarrow$" denotes weak convergence in the space of continuous function on $[0, D]$ equipped with the uniform metric.

Theorem 5. *Under Assumption 1, $\delta_i^2(\hat{m}_i - m_i^0) \xrightarrow{d} \operatorname{argmax}_s Z^{(i)}(s)$ for $i = 1, \ldots, k$.*

4. Asymptotic Theory for Volatility Shifts

Let $\tau_i^0 = n_i^0/T$ and $\hat{\tau}_i = \hat{n}_i/T$ for $i = 1, \ldots, l$, and $\lambda_i = m_i/T$ for $0 < m_1 < \cdots < m_k < T$. Note that the true break fractions τ_i^0, $i = 1, \ldots, l$, are supposed to be constants here. Define $\tilde{n}_{ij} = [n_{i-1}+1, n_{i-1}+2, \ldots, n_i] \cap [n_{j-1}^0+1, n_{j-1}^0+2, \ldots, n_j^0]$, and let n_{ij} be the number of observations in $\{Y_t | t \in \tilde{n}_{ij}\}$, $n_{i\cdot}$ the number of observations in $\{Y_t | t \in [n_{i-1}, n_i]\}$, and $n_{\cdot j}$ the number of observations in $\{Y_t | t \in [n_{j-1}^0, n_j^0]\}$. Define

$$\tilde{\beta}_j^0 = \frac{1}{n_j - n_{j-1}} \sum_{i=n_{j-1}+1}^{n_j} \beta_i^2(\theta_2) X_i^2 \ \text{ for } \ j = 1, \ldots, l+1;$$

$$\begin{aligned} M = {} & \frac{1}{T} \sum_{j=1}^{l+1} \sum_{i=n_{j-1}+1}^{n_j} [(y_i - \hat{\alpha}_i(\hat{\theta}_1))^2 - \tilde{\beta}_j]^2 \\ & - \frac{1}{T} \sum_{j=1}^{l+1} \sum_{i=n_{j-1}^0+1}^{n_j^0} [(y_i - \hat{\alpha}_i(\hat{\theta}_1))^2 - \beta_j^2 \sigma^2]^2; \end{aligned} \tag{4.1}$$

$$M_1 = \frac{1}{T} \sum_{j=1}^{l+1} \sum_{i=1}^{l+1} n_{ji} (\beta_i^2 \sigma^2 - \tilde{\beta}_j^0)^2;$$

$$M_2 = \frac{1}{T}\sum_{j=1}^{l+1}\sum_{i=1}^{l+1}\Big[\Big(-2\tilde{\beta}_j^0 \sum_{t\in\tilde{n}_{ji}} \beta_t^2(\theta_2)(X_t^2-\sigma^2)\Big)-\Big(-2\beta_i^2\sigma^2 \sum_{t\in\tilde{n}_{ji}} \beta_t^2(\theta_2)(X_t^2-\sigma^2)\Big)\Big].$$

With our assumptions, we have

$$M=\frac{1}{T}\sum_{j=1}^{l+1}\sum_{i=n_{j-1}+1}^{n_j}[(\beta_i(\theta_2)X_i)^2-\tilde{\beta}_j^0]^2-\frac{1}{T}\sum_{j=1}^{l+1}\sum_{i=n_{j-1}^0+1}^{n_j^0}[(\beta_i(\theta_2)X_i)^2-\beta_j^2\sigma^2]^2+o_p(1),$$

so $M = M_1 + M_2 + o_p(1)$. The following lemma gives a type of the Hájek-Rényi inequality for this model, while Lemma 5 provides a lower bound on M_1.

Lemma 4. *Assume that $E(\epsilon_t^3)$ and $E(\epsilon_t^4)$ exist for any t. For any $0 \le T_1 \le t_1 < t_2 \le T_2 \le T$, let $d(t_2, t_1)$ be a positive function that is non-increasing in t_2, and non-decreasing in t_1. For any constant C_0 with $0 \le C_0 \le T_2 - T_1 - 1$, and any $\delta > 0$, there exists a constant $C < \infty$ such that*

$$\begin{aligned} P_r\Big(&\max_{T_1+C_0\le t_1+C_0<t_2\le T_2} d(t_2,t_1)\Big|\sum_{i=t_1+1}^{t_2}\beta_i^2(\theta_2)(X_i^2-\sigma^2)\Big| > \delta\Big)\\ &\le C\frac{C_0+1}{\delta^2}\sum_{i=T_1+1}^{T_2} d(i,T_1)^2. \end{aligned}$$

Proof. Since $E(\epsilon_i^3)$ and $E(\epsilon_i^4)$ exist, there exists a constant C_1 such that

$$E[(\sum_{i=t_1+1}^{t_2}\beta_i^2(\theta_2)(X_i^2-\sigma^2))^2] \le C_1(t_2-t_1)$$

for any $0 \le t_1 < t_2 \le T$. The rest of this proof is similar to that of Lemma 1.

Let

$$\begin{aligned} &\underline{\beta} = \sigma^2 \min_{1\le i\le l}|\beta_{i+1}^2-\beta_i^2|, \quad \overline{\beta} = \sigma^2 \max_{1\le i\le l}|\beta_{i+1}^2-\beta_i^2|, \quad \tau=(\tau_1,\ldots,\tau_l);\\ &\hat{\tau}=(\hat{\tau}_1,\hat{\tau}_2,\ldots,\hat{\tau}_l), \quad \tau^0=(\tau_1^0,\ldots,\tau_l^0), \quad \triangle_\tau^0=\min_{1\le i\le l}|\tau_{i+1}^0-\tau_i^0|;\\ &\|\tau-\tau^0\|_\infty = \max_{1\le i\le l}|\tau_i-\tau_i^0|, \quad \|\hat{\tau}-\tau^0\|_\infty=\max_{1\le i\le l}|\hat{\tau}_i-\tau_i^0|;\\ &\|\tilde{\beta}^0-\beta\|_\infty^2 = \max_{1\le i\le l+1}|\tilde{\beta}_i^0-\beta_i^2\sigma^2|^2. \end{aligned}$$

Lemma 5. *There exists a positive constant C such that, for all $T \ge 1$ and all $0 < n_1 < n_2 < \cdots < n_l < T$, we have $M_1 \ge C\|\tau-\tau^0\|_\infty$.*

This proof is similar to that of Lemma 2, and is omitted here.

A first result establishes the consistency of the estimates of volatility break fractions, a second shows that their convergence rates are of order T. Then, with the convergence rates, we derive the asymptotic distributions of the estimates $(\check{\beta}_1, \ldots, \check{\beta}_{l+1})$ of volatility shifts.

Theorem 6. *The estimated break fractions of volatilities converge to their true values in probability,* $\hat{\tau}_i \xrightarrow{p} \tau_i^0$ *for* $i = 1, \ldots, l$.

The proof here is similar to that of Theorem 1, and is omitted here.

Theorem 7. *Suppose* X_t *is a stationary ARMA process. For every* $\eta > 0$, *there exists a positive number* $\delta < \infty$ *such that for all large* T, $P_r(|T(\hat{\tau}_j - \tau_j^0)| > \delta) < \eta$ *for* $j = 1, \ldots, l$.

This proof is similar to that of Theorem 2, we omit it here.

Theorem 8. *Assume* X_i *is a stationary ARMA process. Then* $\sqrt{T}(\check{\beta}_j - \beta_j^2\sigma^2) \xrightarrow{d} N_j(0, \nu_j^2)$, *independently in the limit, and where* ν_j^2 *is defined at* (4.2).

Proof. For $j = 1, \ldots, l+1$, formula manipulations give that

$$\sqrt{T}(\check{\beta}_j - \beta_j^2\sigma^2) = \sqrt{T}\{\frac{1}{\hat{n}_j - \hat{n}_{j-1}} \sum_{t=\hat{n}_{j-1}+1}^{\hat{n}_j} \beta_t^2(\theta_2)X_t^2 - \frac{1}{n_j^0 - n_{j-1}^0} \sum_{t=n_{j-1}^0}^{n_j^0} \beta_t^2(\theta_2)X_t^2$$
$$+\frac{1}{n_j^0 - n_{j-1}^0} \sum_{t=n_{j-1}^0+1}^{n_j^0} \beta_t^2(\theta_2)(X_t^2 - \sigma^2)\}$$
$$= \sqrt{T}\frac{1}{n_j^0 - n_{j-1}^0} \sum_{t=n_{j-1}^0+1}^{n_j^0} \beta_j^2(X_t^2 - \sigma^2) + o_p(1).$$

Let

$$\nu_j^2 = \lim_{T\to\infty} (\tau_j^0 - \tau_{j-1}^0)^{-1} E\Big[(n_j^0 - n_{j-1}^0)^{-1}\Big(\sum_{i=n_{j-1}^0+1}^{n_j^0} \beta_j^2(X_i^2 - \sigma^2)\Big)^2\Big]. \qquad (4.2)$$

Then methods similar to those in the proof of Lemma 6 can finish the proof.

Assumption 2. *If* $\eta_j = (\beta_{j+1}^2 - \beta_j^2)\sigma^2$ *for* $j = 1, \ldots, l$, *then* $\eta_j \to 0$ *and* $T^{1/2}\eta_j \to \infty$ *as* $T \to \infty$, *for all* $j = 1, \ldots, l$.

Theorem 9. *Under Assumption* 2, (i) $\hat{\tau}_j \xrightarrow{p} \tau_j^0$ *for* $j = 1, \ldots, l$, (ii) *for every* $\eta > 0$, *there exists a* $C < \infty$ *such that for all large* T, $P_r(|T\eta_j^2(\hat{\tau}_j - \tau_j^0)| > C) < \eta$ *for* $j = 1, \ldots, l$.

Proof. (i) can be proved similarly to Theorem 1, (ii) can be proved similarly to Theorem 2. We omit details.

Let

$$M_3 = \sum_{j=1}^{l+1} \sum_{i=n_{j-1}+1}^{n_j} (\beta_i^2(\theta_2)\sigma^2 - \frac{1}{n_j - n_{j-1}} \sum_{t=n_{j-1}+1}^{n_j} \beta_t^2(\theta_2)\sigma^2)^2,$$

$$M_4 = \sum_{j=1}^{l+1} \Big[\frac{1}{n_j^0 - n_{j-1}^0} (\sum_{i=n_{j-1}^0+1}^{n_j^0} \beta_i^2(\theta_2)(X_i^2 - \sigma^2))^2 - \frac{1}{n_j - n_{j-1}} \Big(\sum_{i=n_{j-1}+1}^{n_j} \beta_i^2(\theta_2)(X_i^2 - \sigma^2)\Big)^2\Big],$$

$$M_5 = 2\sum_{j=1}^{l+1} \Big[(\sum_{i=n_{j-1}^0}^{n_j^0} \beta_i^2(\theta_2)(X_i^2 - \sigma^2))\beta_j^2\sigma^2 - \Big(\sum_{i=n_{j-1}+1}^{n_j} \beta_i^2(\theta_2)(X_i^2 - \sigma^2)\Big) \frac{1}{n_j - n_{j-1}} \sum_{t=n_{j-1}+1}^{n_j} \beta_t^2(\theta_2)\sigma^2\Big].$$

Under Assumption 2, it is straightforward to show that

$$TM + \sum_{j=1}^{l+1} \Big[\frac{1}{n_j^0 - n_{j-1}^0} \Big(\sum_{i=n_{j-1}^0}^{n_j^0} \beta_i^2(\theta_2)(X_i^2 - \sigma^2)\Big)^2\Big] = M_3 + M_4 + M_5 + o_p(1).$$

Let

$$\Pi_{i,1} = \lim_{T\to\infty} E\Big[\frac{1}{n_i^0 - n_{i-1}^0} \Big(\sum_{t=n_{i-1}^0+1}^{n_i^0} \beta_t^2(\theta_2)(X_t^2 - \sigma^2)\Big)^2\Big],$$

$$\Pi_{i,2} = \lim_{T\to\infty} E\Big[\frac{1}{n_{i+1}^0 - n_i^0} \Big(\sum_{t=n_i^0+1}^{n_{i+1}^0} \beta_t^2(\theta_2)(X_t^2 - \sigma^2)\Big)^2\Big].$$

For any $i = 1, \ldots, l$, let $W_{ij}(s), j = 1, 2$ be independent standard Wiener processes defined on $[0, \infty)$, starting at the origin when $s = 0$, and let

$$Q^{(i)}(s) = \begin{cases} \Pi_{i,1}^{1/2} W_{i1}(-s) - \frac{|s|}{2}, & \text{if } s \le 0, \\ \Pi_{i,2}^{1/2} W_{i2}(s) - \frac{|s|}{2}, & \text{if } s > 0. \end{cases}$$

Lemma 6. *Suppose $E(\epsilon_t^3)$ and $E(\epsilon_t^4)$ exist for any t, and that there is some constant K_2 such that for any interger j, $E\{(\epsilon_j^2 - \sigma_0^2)^2\} = K_2\sigma_0^4$. For $0 \le s_j \le D$,*

$j = 1, \ldots, l$,

$$\eta_j \sum_{i=n_j^0+1}^{n_j^0+s_j\eta_j^{-2}} \beta_{j+1}^2 (X_i^2 - \sigma^2) \Rightarrow \Pi_{j,2}^{1/2} W_{j2}(s_j), \tag{4.3}$$

$$\eta_j \sum_{i=n_j^0-s_j\eta_j^{-2}}^{n_j^0} \beta_j^2 (X_i^2 - \sigma^2) \Rightarrow \Pi_{j,1}^{1/2} W_{j1}(s_j). \tag{4.4}$$

Proof. We just give a brief proof because it is similar to that of Lemma 3. First we have

$$\eta_j \sum_{i=n_j^0+1}^{n_j^0+s_j\eta_j^{-2}} \beta_{j+1}^2 (X_i^2 - \sigma^2) = \sqrt{D}\beta_{j+1}^2 \sum_{i=n_j^0+1}^{n_j^0+(s_j/D)(\sqrt{D}/\eta_j)^2} \frac{\eta_j}{\sqrt{D}} (X_i^2 - \sigma^2),$$

so that $s_j/D \in [0, 1]$. Using Theorem 3.1.3 in Brockwell and Davis (1991), Corollary 2 in Truong-Van (1995) gives (4.3). Equation (4.4) can be proved in a similar fashion.

Next we prove the independence between W_{i1} and W_{i2}. Write $\psi_j^* = \sum_{i\geq j+1} \psi_i$ and $X_i^* = \sum_{j=0}^{+\infty} \psi_j^* \epsilon_{i-j}$ to get $X_i = \psi(1)\epsilon_i - X_i^* + X_{i-1}^*$, where $\psi(1) = \sum_{j\geq 0} \psi_j \neq 0$. Then

$$\eta_j \sum_{i=n_j^0+1}^{n_j^0+s_j\eta_j^{-2}} \beta_{j+1}^2 (X_i^2 - \sigma^2) = \eta_j \sum_{i=n_j^0+1}^{n_j^0+s_j\eta_j^{-2}} \beta_{j+1}^2 \psi(1)^2 (\epsilon_i^2 - \sigma_0^2) + o_p(1),$$

so W_{j2} is determined by $\epsilon_i, i > n_j^0$. Similarly W_{j1} is determined by $\epsilon_i, i \leq n_j^0$, so they are independent.

Theorem 10. *Under Assumption* 2, $\eta_i^2(\hat{n}_i - n_i^0) \xrightarrow{d} \operatorname{argmax}_s Q^{(i)}(s)$ *for* $i = 1, \ldots, l$.

Proof. A proof is only suggested here since it is similar to that of Theorem 5. Because of (ii) of Theorem 9, we can assume that $(n_1, \ldots, n_k)$ fall into the configuration

$$\{(n_1, \ldots, n_k) | n_j = n_j^0 + s_j\eta_j^{-2}, |s_j| \leq D_j, j = 1, \ldots, l; \}.$$

By Lemma 6 and the method of proof of Theorem 5, we can show that

$$M_3 \to_p \sum_{j=1}^{l} |s_j|,\ M_4 \to_p 0,\ M_5 \Rightarrow \sum_{j\in\{i|s_i\geq 0\}} 2\Pi_{j,2}^{1/2} W_{j,2}(s_j) + \sum_{j\in\{i|s_i<0\}} 2\Pi_{j,1}^{1/2} W_{j,1}(-s_j).$$

Therefore it follows from the continuity of the minimization functional that

$$\eta_j^2(\hat{n}_j - n_j^0) \xrightarrow{d} \underset{s}{\operatorname{argmax}}\, Q^{(j)}(s) \quad \text{for} \quad j = 1, \ldots, l.$$

5. Simulations and Conclusions

The data set $y_1, \ldots, y_T$ was generated according to the model $Y_t = \alpha_t(\theta_1) + \beta_t(\theta_2)X_t$, where

$$\alpha_t(\theta_1) = \begin{cases} 4, \text{ if } 1 \le t \le \lambda_1^0 T, \\ 8, \text{ if } \lambda_1^0 T + 1 \le t \le \lambda_2^0 T, \\ 2, \text{ if } \lambda_2^0 T + 1 \le t \le \lambda_3^0 T, \\ 6, \text{ if } \lambda_3^0 T + 1 \le t \le T, \end{cases} \qquad \beta_t(\theta_2) = \begin{cases} 3, \text{ if } 1 \le t \le \tau_1^0 T, \\ 2, \text{ if } \tau_1^0 T + 1 \le t \le \tau_2^0 T, \\ 5, \text{ if } \tau_2^0 T + 1 \le t \le T, \end{cases}$$

X_t was the autoregressive process of order 2 (AR(2)) given by $X_t = 0.4X_{t-1} - 0.04X_{t-2} + \epsilon_t$, the $\epsilon_t' s$ were i.i.d. $N(0, \sigma_0)$, and $C_0 = (5\tau_1^0 - 21\tau_2^0 + 25)T$. Simulations were carried out for two sets of specific values of parameters. Simulation I took $T = 160, \lambda_1^0 = 0.25$, $\lambda_2^0 = 0.5$, $\lambda_3^0 = 0.75$, $\tau_1^0 = 3/8$, $\tau_2^0 = 3/4$, $\sigma_0 = 1$, and $C_0 = 1{,}780$. After 1,000 replications, we took corresponding averages as the estimates of means and volatilities and their break fractions; these are listed in **Table B.1** of Appendix B. Simulation II took $T = 1{,}600$ and kept the other parameters unchanged. After 100 replications, we took averages as before; these are listed in **Table B.2** of Appendix B.

Our estimates are easy to implement because they only involve minimizations of simple functions over a finite number of change point configurations. Simulations suggest that the estimates and asymptotic theory work well for reasonable sample sizes.

In practice, one might first try to use the penalized least square method described in Section 2 to estimate the number of change points of means and volatilities, then apply our model to those data sets to capture their change-point behavior. One can then proceed to estimate the residuals of the process X_t and make statistical inferences using resampling techniques.

Acknowledgement

The author finished this paper in the Department of Mathematics and Statistics, University of Melbourne, as a postdoc of Professor Peter Hall, with financial support from MASCOS.

Appendix A: Proofs

Proof of Lemma 1. We use induction similarly to Móricz, Serfling, and Stout (1982). First consider that $T_2 = T_1 + C_0 + 1$. Since X_i follows a stationary ARMA process, it follows that

$$E\Big(\Big|\sum_{i=T_1+1}^{T_1+C_0+1} \beta_i(\theta_2)X_i\Big|\Big)^2 \le C(C_0+1).$$

Thus we obtain that

$$P_r\Big(d(T_1+C_0+1,T_1)\Big|\sum_{i=T_1+1}^{T_1+C_0+1} \beta_i(\theta_2)X_i\Big| > \delta\Big) \le C\frac{C_0+1}{\delta^2}\sum_{i=T_1+1}^{T_2} d^2(i,T_1).$$

Assume that the result holds for all integers T_2 satisfying $T_1+C_0+1 \le T_2 < N$. We will show that the lemma then holds for $T_2 = N$. We have that

$$\begin{aligned}
&P_r\Big(\max_{T_1+C_0\le t_1+C_0<t_2\le N} d(t_2,t_1)|\sum_{i=t_1+1}^{t_2}\beta_i(\theta_2)X_i| > \delta\Big)\\
&\le P_r\Big(\max_{T_1+C_0\le t_1+C_0<t_2\le N-1} d(t_2,t_1)|\sum_{i=t_1+1}^{t_2}\beta_i(\theta_2)X_i| > \delta\Big)\\
&\quad +P_r\Big(\max_{T_1+C_0\le t_1+C_0<N-1} d(N-1,t_1)|\sum_{i=t_1+1}^{N-1}\beta_i(\theta_2)X_i| > \delta\Big)\\
&\quad +P_r\Big(d(N-1,N-C_0-1)|\beta_N(\theta_2)X_N| > \delta\Big)\\
&\quad +P_r\Big(d(N-1,N-C_0-1)|\sum_{i=N-C_0}^{N}\beta_i(\theta_2)X_i| > \delta\Big)\\
&\le 2C_1\frac{C_0+1}{\delta^2}\sum_{i=T_1+1}^{N-1} d^2(i,T_1) + \frac{C_2}{\delta^2}d^2(N-1,N-C_0-1)\\
&\quad +C_3\frac{C_0+1}{\delta^2}d^2(N-1,N-C_0-1)\\
&\le C\frac{C_0+1}{\delta^2}\sum_{i=T_1+1}^{N} d^2(i,T_1),
\end{aligned}$$

which completes this proof by induction.

Proof of Lemma 2. We divide the problem into two cases: $\|\lambda-\lambda^0\|_\infty \le \triangle_\lambda^0/4$, and $\|\lambda-\lambda^0\|_\infty > \triangle_\lambda^0/4$. First assume that $\|\lambda-\lambda^0\|_\infty \le \triangle_\lambda^0/4$. A change point m_j can be left or right of the true change point m_j^0.

For any j such that $\lambda_{j-1} \le \lambda_j^0 \le \lambda_j$, by (3.2), we have

$$\begin{aligned} L_1 &\ge \frac{m_{j,j+1}}{T}(\alpha_{j+1} - \tilde{\alpha}_j)^2 + \frac{m_{jj}}{T}(\alpha_j - \tilde{\alpha}_j)^2 \\ &\ge \frac{1}{2}(\alpha_{j+1} - \alpha_j)^2 \frac{m_{j,j+1}}{T} = \frac{1}{2}(\alpha_{j+1} - \alpha_j)^2(\lambda_j - \lambda_j^0). \end{aligned}$$

For any j such that $\lambda_j \le \lambda_j^0 \le \lambda_{j+1}$, by (3.2), we have

$$\begin{aligned} L_1 &\ge \frac{m_{j+1,j}}{T}(\alpha_j - \tilde{\alpha}_{j+1})^2 + \frac{m_{j+1,j+1}}{T}(\alpha_{j+1} - \tilde{\alpha}_{j+1})^2 \\ &\ge \frac{1}{2}(\alpha_{j+1} - \alpha_j)^2 \frac{m_{j+1,j+1}}{T} = \frac{1}{2}(\alpha_{j+1} - \alpha_j)^2(\lambda_j^0 - \lambda_j). \end{aligned}$$

Thus, if $\|\lambda - \lambda^0\|_\infty \le \triangle_\lambda^0/4$, then $L_1 \ge C_1(\alpha)\|\lambda - \lambda^0\|_\infty$.

If now $\|\lambda - \lambda^0\|_\infty > \triangle_\lambda^0/4$, there clearly exists a pair (i, j) such that $m_{ij} \ge T\triangle_\lambda^0/4$ and $m_{i,j+1} \ge T\triangle_\lambda^0/4$. Then we have

$$\begin{aligned} L_1 &\ge \frac{m_{i,j+1}}{T}(\alpha_{j+1} - \tilde{\alpha}_i)^2 + \frac{m_{i,j}}{T}(\alpha_j - \tilde{\alpha}_i)^2 \\ &\ge \frac{m_{ij}m_{i,j+1}}{T(m_{i,j+1} + m_{ij})}(\alpha_{j+1} - \alpha_j)^2 \ge \frac{\triangle_\lambda^0}{8}(\alpha_{j+1} - \alpha_j)^2 = C_2(\alpha)\triangle_\lambda^0. \end{aligned}$$

Therefore,

$$L_1 \ge \min(C_2(\alpha)\triangle_\lambda^0, C_1(\alpha)\|\lambda - \lambda^0\|_\infty) \ge \min(C_1(\alpha), C_2(\alpha))\triangle_\lambda^0\|\lambda - \lambda^0\|_\infty.$$

Proof of Theorem 1. For any $\delta > 0$, we have

$$\begin{aligned} P_r(\|\hat{\lambda} - \lambda^0\|_\infty \ge \delta) &\le P_r(\min_{\|\lambda-\lambda^0\|_\infty \ge \delta} L \le 0) \\ &\le P_r(\max_{\|\lambda-\lambda^0\|_\infty \ge \delta} |L_2| \ge \min_{\|\lambda-\lambda^0\|_\infty \ge \delta} L_1). \end{aligned}$$

It then follows from Lemma 2, (3.2), and (3.3) that

$$\begin{aligned} &P_r\Big(\|\hat{\lambda} - \lambda^0\|_\infty \ge \delta\Big) \\ &\quad\le P_r\Big(\max_{0\le t_1<t_2\le T}\Big[\frac{1}{(t_2 - t_1)}\Big|\Big(\sum_{t=t_1+1}^{t_2}\beta_t(\theta_2)X_t\Big)\Big(\sum_{t=t_1+1}^{t_2}Y_t)|\Big] \ge \frac{C\delta}{4k}T\Big) \\ &\quad\quad+P_r\Big(\max_{1\le t_1<t_2\le T}\Big|\sum_{t=t_1+1}^{t_2}\alpha_t(\theta_1)\beta_t(\theta_2)X_t| \ge \frac{C\delta}{4k}T\Big) \\ &\quad\le P_r\Big(\max_{0\le t_1<t_2\le T}[\frac{1}{(t_2 - t_1)}(\sum_{t=t_1+1}^{t_2}\beta_t(\theta_2)X_t)^2] \ge \frac{C\delta}{8k}T\Big) \end{aligned}$$

$$+P_r\Big(\max_{1\le t_1<t_2\le T}\Big|\sum_{t=t_1+1}^{t_2}\beta_t(\theta_2)X_t|\ge\frac{C_1\delta}{8k}T\Big)$$
$$+P_r\Big(\max_{1\le t_1<t_2\le T}\Big|\sum_{t=t_1+1}^{t_2}\alpha_t(\theta_1)\beta_t(\theta_2)X_t|\ge\frac{C\delta}{4k}T\Big).$$

This concludes the proof by Lemma 1, noting that $\sum_{i=1}^{T}1/T=O(\log T)$.

Proof of Theorem 2 Because of the consistency of $\hat\lambda_i$ for all i, we only need to consider the change point configurations with nonzero $m_{j,j-1},m_{jj},m_{j,j+1}$ for all j. Then we have

$$L_1=\frac{1}{T}\sum_{j=1}^{k+1}\sum_{i=j-1}^{j+1}m_{ji}(\alpha_i-\tilde\alpha_j)^2,$$
$$L_2=\frac{1}{T}\sum_{j=1}^{k+1}\sum_{i=j-1}^{j+1}[(-2\tilde\alpha_j\sum_{t\in\tilde m_{ji}}\beta_t(\theta_2)X_t)-(-2\alpha_i\sum_{t\in\tilde m_{ji}}\beta_t(\theta_2)X_t)].$$

For any $\delta>0$, we obtain that

$$P_r(T\|\hat\lambda-\lambda^0\|_\infty>\delta)\le P_r(\min_{\|\lambda-\lambda^0\|_\infty>\delta T^{-1}}L<0)$$
$$\le\sum_{j=1}^{k+1}P_r(\min_{\|\lambda-\lambda^0\|_\infty>\delta T^{-1}}[\frac{2}{T}(\alpha_j-\tilde\alpha_j)\sum_{t\in\tilde m_{jj}}\beta_t(\theta_2)X_t+\frac{1}{3(k+1)}L_1]<0)\tag{A.1}$$
$$+\sum_{j=2}^{k+1}P_r(\min_{\|\lambda-\lambda^0\|_\infty>\delta T^{-1}}[\frac{2}{T}(\alpha_{j-1}-\tilde\alpha_j)\sum_{t\in\tilde m_{j,j-1}}\beta_t(\theta_2)X_t+\frac{1}{3(k+1)}L_1]<0)\tag{A.2}$$
$$+\sum_{j=1}^{k}P_r(\min_{\|\lambda-\lambda^0\|_\infty>\delta T^{-1}}[\frac{2}{T}(\alpha_{j+1}-\tilde\alpha_j)\sum_{t\in\tilde m_{j,j+1}}\beta_t(\theta_2)X_t+\frac{1}{3(k+1)}L_1]<0).\tag{A.3}$$

First consider the probability in (A.2) for general $j\in[2,\dots,k+1]$. Here $\lambda_{j-1}<\lambda_{j-1}^0$, otherwise the probability of the term corresponding to j in the (A.2) is zero by definition. We have that

$$L_1\ge\frac{m_{j-1}^0-m_{j-1}}{T}(\alpha_{j-1}-\tilde\alpha_j)^2.$$

Then it follows from Lemma 2 that

$$L_1\ge\max\Big(C\|\lambda-\lambda^0\|_\infty,\frac{m_{j-1}^0-m_{j-1}}{T}(\alpha_{j-1}-\tilde\alpha_j)^2\Big).$$

If $0 < \lambda^0_{j-1} - \lambda_{j-1} \leq \delta T^{-1}$, that is, $0 < m^0_{j-1} - m_{j-1} \leq \delta$, then

$$P_r\Big(\min_{\|\lambda-\lambda^0\|_\infty > \delta T^{-1}} \Big[\frac{2}{T}(\alpha_{j-1} - \tilde{\alpha}_j) \sum_{t \in \tilde{m}_{j,j-1}} \beta_t(\theta_2) X_t + \frac{1}{3(k+1)} L_1\Big] < 0\Big) \tag{A.4}$$

$$\leq P_r\Big(\frac{2}{T}|\alpha_{j-1} - \alpha^*_j| \max_{1 \leq m^0_{j-1} - m_{j-1} \leq \delta} \Big|\sum_{t=m_{j-1}+1}^{m^0_{j-1}} \beta_t(\theta_2) X_t\Big| >$$

$$\frac{1}{3(k+1)} \max\Big(C\delta T^{-1}, \frac{1}{T}(\alpha_{j-1} - \alpha^*_j)^2\Big)$$

$$\leq P_r\Big(\max_{1 \leq m^0_{j-1} - m_{j-1} \leq \delta} \Big|\sum_{t=m_{j-1}+1}^{m^0_{j-1}} \beta_t(\theta_2) X_t\Big| > \frac{C\delta}{6(k+1)}\Big), \tag{A.5}$$

where α^*_j is the corresponding value of $\tilde{\alpha}_j$ obtained through minimization at (A.4), and we have used the fact that $\max(b/c, ca) \geq \sqrt{bc}$ for $a, b, c > 0$.

If $\lambda^0_{j-1} - \lambda_{j-1} > \delta T^{-1}$, that is, $m^0_{j-1} - m_{j-1} > \delta$, then

$$P_r\Big(\min_{\|\lambda-\lambda^0\|_\infty > \delta T^{-1}} \Big[\frac{2}{T}(\alpha_{j-1} - \tilde{\alpha}_j) \sum_{t \in \tilde{m}_{j,j-1}} \beta_t(\theta_2) X_t + \frac{1}{3(k+1)} L_1\Big] < 0\Big) \tag{A.6}$$

$$\leq P_r\Big(\frac{2}{T}|\alpha_{j-1} - \alpha^*_j| \Big|\sum_{t=m^*_{j-1}+1}^{m^0_{j-1}} \beta_t(\theta_2) X_t\Big| >$$

$$\frac{1}{3(k+1)} \max\Big(C\delta T^{-1}, \frac{m^0_{j-1} - m^*_{j-1}}{T}(\alpha_{j-1} - \alpha^*_j)^2\Big)\Big)$$

$$\leq P_r\Big(\max_{m^0_{j-1} - m_{j-1} > \delta} \Big|\frac{1}{m^0_{j-1} - m_{j-1}} \sum_{t=m_{j-1}+1}^{m^0_{j-1}} \beta_t(\theta_2) X_t\Big| > \frac{C\delta}{6(k+1)}\Big), \tag{A.7}$$

where α^*_j and m^*_{j-1} are the corresponding values of $\tilde{\alpha}_j$ and m_{j-1} obtained through minimization at (A.6), and we have used the fact that $\max(b/c, ca) \geq \sqrt{bc}$ for $a, b, c > 0$.

Similar methods can be applied to handle the probability terms in (A.3) for $j = 1, \ldots, k$. When considering the probability in (A.1), we can restrict the configurations to $(m_1, \ldots, m_k)$ satisfying $m_{jj} > (m^0_j - m^0_{j-1})/2$ because of the consistency of $\hat{\lambda}$. Then similar methods can be used to handle the probability terms. Then the proof follows from Lemma 1.

Proof of Theorem 3. For $j = 1, \ldots, k+1$, formula manipulations give that

$$\sqrt{T}(\hat{\alpha}_j - \alpha_j) = \sqrt{T}\{\frac{1}{\hat{m}_j - \hat{m}_{j-1}} \sum_{i=\hat{m}_{j-1}+1}^{\hat{m}_j} y_j - \frac{1}{m_j^0 - m_{j-1}^0} \sum_{i=m_{j-1}^0}^{m_j^0} y_i$$
$$+ \frac{1}{m_j^0 - m_{j-1}^0} \sum_{i=m_{j-1}^0+1}^{m_j^0} \beta_i(\theta_2) X_i\}.$$

It is easy to prove by Theorem 2 that

$$\sqrt{T}\{\frac{1}{\hat{m}_j - \hat{m}_{j-1}} \sum_{i=\hat{m}_{j-1}+1}^{\hat{m}_j} y_j - \frac{1}{m_j^0 - m_{j-1}^0} \sum_{i=m_{j-1}^0}^{m_j^0} y_i\} \xrightarrow{p} 0,$$

so we just have to consider the limiting distribution of the rest of $\sqrt{T}(\hat{\alpha}_j - \alpha_j)$.

Theorem 3.1.3 in Brockwell and Davis (1991) has that $X_t = \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}$, where $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ and $\sum_{j=-\infty}^{\infty} \psi_j \neq 0$, $\epsilon_t' s$ are I.I.D. with mean zero. Let

$$\upsilon_j^2 = \frac{1}{\lambda_j^0 - \lambda_{j-1}^0} \lim_{T \to \infty} E[\frac{1}{m_j^0 - m_{j-1}^0} (\sum_{i=m_{j-1}^0+1}^{m_j^0} \beta_i(\theta_2) X_i)^2]. \tag{A.8}$$

Similar methods to those in Lemma 3 complete this proof.

Proof of Lemma 3. Theorem 2 in Truong-Van (1995) is used to prove this result. Equation (3.6) can be proved in a way similar to that of (3.5), so we just prove equation (3.5) here. First we have that

$$\delta_j \sum_{i=m_j^0+1}^{m_j^0+s_j\delta_j^{-2}} \beta_i(\theta_2) X_i = \sum_{i=m_j^0+1}^{m_j^0+s_j/D(\sqrt{D}/\delta_j)^2} \frac{\delta_j}{\sqrt{D}} \sqrt{D} \beta_i(\theta_2) X_i,$$

so that $s_j/D \in [0, 1]$.

If $\mathcal{F}_i$ be the σ-algebra generated by $\{\epsilon_j, j \leq i\}$, then $(\delta_j \beta_i(\theta_2) X_i, \mathcal{F}_t; 1 \leq i \leq T, t \in Z)$ is a quasi-stationary sequence of linear processes according to the definition 2 in Truong-Van (1995). It is easy to check that (H1), (H2), and (H4.1) are all satisfied for this case, and γ^2 defined at (2.8) of Truong-Van (1995) is equal to $\Omega_{j,2}$. Therefore Theorem 2 in Truong-Van (1995) gives (3.5).

Next we prove the independence between W_{i1} and W_{i2}. Write $\psi_j^* = \sum_{i \geq j+1} \psi_i$ and $X_i^* = \sum_{j=0}^{+\infty} \psi_j^* \epsilon_{i-j}$, to get $X_i = \psi(1)\epsilon_i - X_i^* + X_{i-1}^*$, where $\psi(1) = \sum_{j \geq 0} \psi_j \neq$

0. Thus

$$\delta_j \sum_{i=m_j^0+1}^{m_j^0+s_j\delta_j^{-2}} \beta_i(\theta_2)X_i = \delta_j \sum_{i=m_j^0+1}^{m_j^0+s_j\delta_j^{-2}} \beta_i(\theta_2)\psi(1)\epsilon_i + \delta_j \sum_{i=m_j^0+1}^{m_j^0+s_j\delta_j^{-2}} \beta_i(\theta_2)(X_{i-1}^* - X_i^*),$$

and it is straightforward to show that

$$\delta_j \sum_{i=m_j^0+1}^{m_j^0+s_j\delta_j^{-2}} \beta_i(\theta_2)(X_{i-1}^* - X_i^*) \xrightarrow{p} 0.$$

So W_{i2} is determined by $\epsilon_i, i > m_j^0$. Similarly W_{i1} is determined by $\epsilon_i, i \le m_j^0$, hence independent of W_{i2}.

Proof of Theorem 5. Because of (ii) of Theorem 4, we can assume that $(m_1, \ldots, m_k)$ falls into the configuration

$$\{(m_1, \ldots, m_k) | m_j^0 - D\delta_j^{-2} \le m_j \le m_j^0 + D\delta_j^{-2}, j = 1, \ldots, k; D > 0\}.$$

For $j = 1, \ldots, k$, m_j may fall to the left or the right of m_j^0. By symmetry, we just need to consider the configuration

$$\{(m_1, \ldots, m_k) | m_j = m_j^0 + s_j\delta_j^{-2}, j = 1, \ldots, k; 0 \le s_j \le D\}.$$

Since $s_j \ge 0$ and $T\delta_j^2 \to \infty$, we have $m_{j+1}^0 > m_j \ge m_j^0$ for $j = 1, \ldots, k$.

Under the above configuration, we have

$$\begin{aligned} L_3 &= \sum_{j=1}^{k}\{ \sum_{i=m_{j-1}+1}^{m_j^0} [\alpha_j - \frac{1}{m_{j\cdot}}(m_{jj}\alpha_j + m_{j,j+1}\alpha_{j+1})]^2 \\ &\quad + \sum_{i=m_j^0+1}^{m_j} [\alpha_{j+1} - \frac{1}{m_{j\cdot}}(m_{jj}\alpha_j + m_{j,j+1}\alpha_{j+1})]^2\} \\ &= \sum_{j=1}^{k} \frac{m_{jj}m_{j,j+1}}{m_{j\cdot}}(\alpha_{j+1} - \alpha_j)^2 = \sum_{j=1}^{k} \frac{m_j^0 - m_{j-1}}{m_j - m_{j-1}} s_j. \end{aligned}$$

Hence L_3 converges to $\sum_{j=1}^{k} s_j$ since $(m_j^0 - m_{j-1})/(m_j - m_{j-1})$ converges to 1.

Under above configuration, we have that

$$L_4 = \sum_{j=1}^{k} \Big[\frac{m_{j,j+1}}{m_{j\cdot}m_{jj}} S_{jj}^2 + (\frac{m_{jj}}{m_{j\cdot}m_{j,j+1}} - \frac{m_{j+1,j+1}}{m_{\cdot j+1}m_{j,j+1}}) S_{j,j+1}^2 - \frac{2}{m_{j\cdot}} S_{jj}S_{j,j+1}$$

$$-\frac{m_{j,j+1}}{m_{\cdot j+1}m_{j+1,j+1}}S^2_{j+1,j+1}+\frac{2}{m_{\cdot j+1}}S_{j+1,j+1}S_{j,j+1}\Big],$$

where $S_{ij}=\sum_{t\in\tilde{m}_{ij}}\beta_t(\theta_2)X_t$. Then

$$|L_4|\le\sum_{j=1}^{k}\Big[\frac{m_{j,j+1}}{m_{j\cdot}m_{jj}}S^2_{jj}+|\frac{m_{jj}}{m_{j\cdot}}-\frac{m_{j+1,j+1}}{n_{\cdot j+1}}|\frac{S^2_{j,j+1}}{m_{j,j+1}}$$
$$+2|S_{j,j+1}|(\frac{|S_{jj}|}{m_{jj}}+\frac{|S_{j+1,j+1}|}{m_{j+1,j+1}})+\frac{m_{j,j+1}}{m_{\cdot j+1}m_{j+1,j+1}}S^2_{j+1,j+1}\Big]. \quad \text{(A.9)}$$

It follows from $m_{j,j+1}/m_{j\cdot}=o_p(1)$ and $S^2_{jj}/m_{jj}=O_p(1)$ that $(m_{j,j+1}/(m_{j\cdot}m_{jj}))\times S^2_{jj}=o_p(1)$. Similarly we can have that the other terms in the above equation is also $o_p(1)$, so that $L_4=o_p(1)$.

Algebraic manipulations give

$$L_5=\sum_{j=1}^{k}2(\alpha_{j+1}-\alpha_j)[\frac{m_{jj}}{m_{j\cdot}}\sum_{i=m_j^0+1}^{m_j}\beta_i(\theta_2)X_i-\frac{m_{j,j+1}}{m_{j\cdot}}\sum_{i=m_{j-1}+1}^{m_j^0}\beta_i(\theta_2)X_i]. \quad \text{(A.10)}$$

For the first term on the right, we have that

$$|2(\alpha_{j+1}-\alpha_j)\frac{m_{j,j+1}}{m_{j\cdot}}\sum_{i=m_{j-1}+1}^{m_j^0}\beta_i(\theta_2)X_i|$$
$$\le\frac{2s_j}{\delta_jT^{1/2}(\lambda_j^0-\lambda_{j-1})^{1/2}}[\frac{1}{(m_j^0-m_{j-1})^{1/2}}|\sum_{i=m_{j-1}+1}^{m_j^0}\beta_i(\theta_2)X_i|]=o_p(1). \quad \text{(A.11)}$$

It is obvious that $m_{jj}/m_{j\cdot}\to 1$. Therefore (3.5) of Lemma 3 gives that

$$L_5\Rightarrow\sum_{j=1}^{k}2\Omega_{j,2}^{1/2}W_{j,2}(s_j).$$

When we consider other configurations with some $s_j's<0$, we similarly have that

$$L_3\xrightarrow{p}\sum_{j=1}^{k}|s_j|,\ L_4\xrightarrow{p}0,\ L_5\Rightarrow\sum_{j\in\{i|s_i\ge0\}}2\Omega_{j,2}^{1/2}W_{j,2}(s_j)+\sum_{j\in\{i|s_i<0\}}2\Omega_{j,1}^{1/2}W_{j,1}(-s_j).$$

Therefore it follows from the continuity of minimization functional that

$$\delta_j^2(\hat{m}_j-m_j^0)\xrightarrow{d}\operatorname*{argmax}_s\ Z^{(j)}(s)\quad\text{for}\quad j=1,\dots,k.$$

Appendix B: Simulation Results

Table B.1. Change-point simulation I.

	Mean							Volatility				
Comparison	mean				Break fraction			Volatility			Break fraction	
True value	4	8	2	6	0.25	0.50	0.75	3	2	5	0.375	0.750
Estimate	3.92	8.10	1.95	6.20	0.25	0.50	0.75	3.02	2.02	5.08	0.366	0.766

Table B.2. Change-point simulation II.

	Mean							Volatility				
Comparison	mean				Break fraction			Volatility			Break fraction	
True value	4	8	2	6	0.25	0.50	0.75	3	2	5	0.375	0.750
Estimate	4.00	8.05	1.98	6.02	0.25	0.50	0.75	3.01	2.00	5.00	0.375	0.752

References

Bai, J. (1994). Least square estimation of a shift in linear processes. *J. Time Ser. Anal.* **15**, 453-472.

Bhattacharya, P. (1987). Maximum likelihood estimation of a change-point in the distribution of independent random variables: general multiparameter case. *J. Multivariate Anal.* **32**, 183-208.

Baum, L. E. and Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Statist.* **37**, 1554-1563.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods.* Springer-Verlag, New York.

Buckle, R. A., Haugh, D. and Thomson, P. J. (2004). Markov switching models for GDP growth in a small open economy: the New Zealand experience. *J. Business Cycle Measurement and Analysis* **1**, 227-257.

Engle, C. and Hamilton, J. D. (1990). Long swings in the dollar: are they in the data and do markets know it? *Amer. Economic Rev.* **80**, 689-713.

Goldfeld, S. M. and Quandt, R. E. (1973). A Markov model for switching regressions. *J. Econometrics* **1**, 3-16.

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

Hinkley, D. (1970). Inference about the change point in a sequence of random variables. *Biometrika* **57**, 1-17.

Hinkley, D. and Hinkley, E. (1970). Inference about the change point in a sequence of binomial variables. *Biometrika* **57**, 477-488.

Liu, J., Wu, S. and Zidek, J. V. (1997). On segmented multivariate regressions, *Statist. Sinica* **7**, 497-525.

Móricz, F., Serfling, R. and Stout, W. (1982). Moment and probability bounds with quasi-superadditive structure for the maximum partial sum. *Ann. Probab.* **10**, 1032-1040.

Picard, D. (1985). Testing and estimating change-points in time series. *J. Appl. Probab.* **14**, 411-415.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* **77**, 257-285.

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6**, 461-464.

Truong-Van, B. (1995). Invariance principles for semi-stationary sequence of linear processes and applications to ARMA process. *Stochastic Process. Appl.* **58**, 155-72.

Yao, Y. C. (1987). Approximating the distribution of the ML estimate of the change-point in a sequence of independent r.v.'s. *Ann. Statist.* **3**, 1321-1328.

Yao, Y. C. (1988). Estimating the number of change-points via Schwartz criterion. *Statist. Probab. Lett.* **6**, 181-189.

Yao, Y. C. (1989). Least-squares estimation of a step function. *Sankhyā A* **51**, 370-381.

Department of Mathematics, University of York, Heslington, York YO10 5DD, United Kingdom.

E-mail: hhe@math.ku.edu

(Received September 2009; accepted June 2010)