A New Algorithm for Multiple Change-points Estimation in Time Series

Ngai Hang Chan, Chun Yip Yau and Rong-Mao Zhang

Chinese University of Hong Kong and Zhejiang University

Abstract

We consider the structural break autoregressive process where a time series has an unknown number of break-points, and the time series follows a stationary AR model in between any two break-points. It is well-known that the estimation of the locations of the break-points involves huge computational challenges. By reformulating the problem in a regression variable selection context, we propose in this paper a group least absolute shrinkage and selection operator (LASSO) procedure to estimate the number and the locations of the break-points, where the computation can be efficiently performed. Simululation studies are conducted to assess the finite sample performance.

1 Introduction

Nonstationarity is a commonly found phenomenon in many practical situations. Although sophisticated nonstationary models have been developed in different fields, they are usually difficult to interpret. By partitioning the nonstationary data into several contiguous stationary segments, the notion of locally stationary models has become a popular device. As parsimonious models can be entertained in each stationary segment, locally stationary models offer a convenient and easy to interpret means to analyze nonstationary behavior. Among different types of locally stationary models, the so-called structural-break or change-point model has received particular attention. A well-known example is the change-in-mean model, which is a useful alternative to the long-memory models in financial time series, see e.g. Mikosch and Stărică (2000).

Because an autoregressive (AR) model is often used to describe simple stationary time series data, a particularly useful locally stationary model for describing the structural-break or change-point behavior is the so-called $(m+1)$-regime structural break autoregressive (SBAR) model given by

$$Y_t = \sum_{j=1}^{m+1} [\boldsymbol{\beta}_j^{\mathrm{T}} \boldsymbol{Y}_{t-1} + \sigma_j \varepsilon_t] I(t_{j-1} \le t < t_j), \tag{1.1}$$

where $\boldsymbol{Y}_{t-1} = (1, Y_{t-1}, \ldots, Y_{t-p})^{\mathrm{T}}$, $\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j1}, \ldots, \beta_{jp})^{\mathrm{T}} \in \mathbb{R}^{p+1}$, $j = 1, \ldots, m+1$, $1 = t_0 < t_1 < \ldots < t_{m+1} = n+1$. Herein, the time instincts $\{t_1, \ldots, t_m\}$ denote the change-points when the parameter $\boldsymbol{\beta}_j$ changes to $\boldsymbol{\beta}_{j+1}$ at time t_j. The number of change points m and the autoregressive order p are positive integers. The errors $\{\varepsilon_t\}$ are independent and identically distributed (i.i.d.) random variables with zero mean and unit variance, and ε_t is independent of the past information $\{y_{t-j} : j \ge 1\}$.

As the regime-changing autoregressive structure offers a simple and intuitive means for interpretation, the SBAR model has attracted considerable attention in diverse areas such as signal processing, biological sciences, econometrics, environmental sciences, finance, hydrology, physics and population dynamics. Information about these applications can be found in Basseville and Benveniste (1983), Scolve (1983), Adak and Sarkar (1996), Andreou and Ghysels (2008), Shao and Zhang (2010) and references therein.

At the same time, probabilistic properties and statistical inference for change-point models have also been extensively studied. For example, Andrews (1993) considered test of the break structure. Adak (1998) studied the spectral properties of the SBAR models. Omabao, Raz, Von Sachs and Malow (2001) considered locally stationary processes for change-points, Bai and Perron (1998, 2003) studied the estimation and test of multiple change point modeling for multiple linear regression. An excellent surveys on various applications of structure break models can be founded in Kim and Nelson (1999).

Although the interpretation is simple, the estimation of SBAR model constitutes a difficult task. The reason is that one needs to consider all possible combinations of the locations of m change-points, which requires extremely high computational burden for large m. In the literature, m is usually assumed to be known and small. To tackle the computational problem in estimation, Davis, Lee and Rodriguez-Yam (2006) apply the genetic algorithm

(GA) to estimate the location of the breaks based on minimum description length (MDL). In that paper, a consistency result for the breaks estimation is established only for a known m. The computational burden of the GA estimation procedure is formidable when $m \geq 10$, however. With the rapid growth of high frequency data analysis, long time series with possibly large number of change-points are often encountered. It is therefore imperative to develop efficient methods for multiple change-point estimations.

The main goal of this paper is to propose a computationally efficient procedure to estimate the change points and the auto-regressive parameters when m is large and unknown. Motivated by the well-known least absolute shrinkage and selection operator (LASSO) of Tibshirani (1996) and group LASSO of Yuan and Lin (2006) (for grouped variables cases), we reformulate the problem of estimating multiple-regime SBAR models in a model selection context and apply the efficient group LASSO algorithm to alleviate the computationally burden. The fast algorithm allows the estimation to be conducted in order $O(n)$ and the location of the breaks can be consistently estimated. One special case of this result is the multiple change-point model considered by Harchaoui and Lévy-Leduc (2010),

$$Y_t = \sum_{j=1}^{m+1} \mu_j I(t_{j-1} \leq t < t_j) + \varepsilon_t,$$

where $\mu_j, j = 1, \ldots, m+1$ are constants.

This paper is organized as follows. Section 2 presents the estimation procedure. Simulation studies and real data applications are given Section 3 and 4 respectively.

2 Estimation

2.1 Group LASSO Estimate

In this subsection, we introduce the group LASSO estimation procedure and the asymptotic theory for the estimate. Let $\boldsymbol{Y}_n^0 = (Y_1, Y_2, \cdots, Y_n)^{\mathrm{T}}$, $\boldsymbol{\eta}(n) = (\sigma_1\varepsilon_1, \sigma_2\varepsilon_2, \cdots, \sigma_n\varepsilon_n)^{\mathrm{T}}$, $\boldsymbol{\theta}(n) =$

$(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_n)^{\mathrm{T}}$ and $\boldsymbol{X}_n$ be an $n \times np$ matrix defined by

$$\boldsymbol{X}_n = \begin{pmatrix} \boldsymbol{Y}_0^{\mathrm{T}} & 0 & 0 & \ldots & 0 \\ \boldsymbol{Y}_1^{\mathrm{T}} & \boldsymbol{Y}_1^{\mathrm{T}} & 0 & \ldots & 0 \\ \boldsymbol{Y}_2^{\mathrm{T}} & \boldsymbol{Y}_2^{\mathrm{T}} & \boldsymbol{Y}_2^{\mathrm{T}} & \ldots & 0 \\ \vdots & & & & \\ \boldsymbol{Y}_{n-1}^{\mathrm{T}} & \boldsymbol{Y}_{n-1}^{\mathrm{T}} & \boldsymbol{Y}_{n-1}^{\mathrm{T}} & \ldots & \boldsymbol{Y}_{n-1}^{\mathrm{T}} \end{pmatrix},$$

where $\mathbf{Y}_k^{\mathrm{T}} = (Y_k, Y_{k-1}, \ldots, Y_{k-p+1})$. Let $\boldsymbol{\theta}_i = \boldsymbol{\beta}_1$ and $\boldsymbol{\theta}_{t_j} = \boldsymbol{\beta}_{j+1}$ for $j = 1, \ldots, m$. It can be seen that model (1.1) can be expressed as a high dimensional regression model

$$\boldsymbol{Y}_n^0 = \boldsymbol{X}_n \boldsymbol{\theta}(n) + \boldsymbol{\eta}(n)\,. \tag{2.1}$$

If $\{Y_n\}$ is generated from model (1.1) with $m+1$ regimes, then only $m+1$ of the $\boldsymbol{\theta}_j$s in (2.1) are non-zeros, which implies that

$$\sum_{j=1}^{n} ||\boldsymbol{\theta}_j|| \leq Cm$$

for some constant $C > 0$, where $||\cdot||$ denotes the l_2-norm. This inequality is similar to the high-dimensional variable selection in group linear regression. Thus, we propose to estimate $\boldsymbol{\theta}(n)$ by the following group LASSO equation:

$$\widehat{\boldsymbol{\theta}}(n) = \mathrm{argmin}_{\boldsymbol{\theta}(n)}\ \frac{1}{n}||\boldsymbol{Y}_n^0 - \boldsymbol{X}_n\boldsymbol{\theta}(n)||^2 + \lambda_n \sum_{i=1}^{n} ||\boldsymbol{\theta}_i||\,, \tag{2.2}$$

where λ_n is the regularization parameter. Rewrite (2.1) as

$$Y_t = \boldsymbol{\beta}_t^{\mathrm{T}} \boldsymbol{Y}_{t-1} + \sigma_t \varepsilon_t =: \boldsymbol{\beta}_t^{\mathrm{T}} \boldsymbol{Y}_{t-1} + \eta_t$$

and estimate $\boldsymbol{\beta}(n) = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_n)^{\mathrm{T}}$ by

$$\widehat{\boldsymbol{\beta}}_1 = \widehat{\boldsymbol{\theta}}_1 \quad \text{and} \quad \widehat{\boldsymbol{\beta}}_i = \sum_{j=1}^{i} \widehat{\boldsymbol{\theta}}_j. \tag{2.3}$$

Note that when $\widehat{\boldsymbol{\theta}}_j \neq 0$, $j \geq 2$, there is a change in the autoregressive parameter $\widehat{\boldsymbol{\beta}}_j$. Thus the structural breaks t_j, $j = 1, 2, \ldots, m$ can be estimated by identifying those

$\widehat{\boldsymbol{\theta}}_j, (j \geq 2)$ which are not zero. Denote the estimates of the location of change points by

$$\mathcal{A}_n = \{j \geq 2 : \widehat{\boldsymbol{\theta}}_j \neq 0\}. \tag{2.4}$$

Let $|\mathcal{A}_n|$ be the cardinality of the set $\mathcal{A}_n$ and denote the elements of $\mathcal{A}_n$ by $\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_{|\mathcal{A}_n|}$. Note that $\hat{m} := |\mathcal{A}_n|$ is the estimated number of change point and $\hat{t}_i$ is the i-th estimated change point. We impose the following assumptions for the time series:

H1: $\{\varepsilon_t\}$ is a white noise sequence with unit variance and $\mathrm{E}|\varepsilon_1|^{4+\delta} < \infty$ for some $\delta > 0$.

H2: $\min_{1 \leq i \leq m_0+1} ||\boldsymbol{\beta}_i^0 - \boldsymbol{\beta}_{i-1}^0|| > 0$, where m_0 is the true number of break points.

H3: $\min_{1 \leq i \leq m_0+1} |t_i^0 - t_{i-1}^0| > n\epsilon_c$ for some $\epsilon_c > 0$.

For the implementation of the group LASSO, an exact solution can be computed by the block coordinated descent algorithm. On the other hand, a computationally efficient approximation to the group LASSO solution can be achieved by the least angle regression (LARS) algorithm. Empirical results show that the LARS algorithm usually gives good approximation to the LASSO solution. The details of the two algorithms can be found in Yuan and Lin (2006). In our simulation studies and data applications in Section 3 and 4, the computationally efficient LARS algorithm is used. Simulation studies not given here show that the LARS algorithm gives very similar solution as the exact block coordinated descent algorithm.

Let $Y = (Y_{p+1}, Y_p, \ldots, Y_n)^{\mathrm{T}}$, $\mathbf{B}_j(r)$ be a p-dimensional vector with the i-th entry being $\sum_{t=j}^{n} Y_{t-i} r_t$, $i = 1, 2, \ldots, p$, where r is any n-dimensional vector. Define $\tilde{Y}_t = (\mathbf{0} \cdots \mathbf{0}\ \mathbf{Y}_t\ \mathbf{Y}_{t+1} \cdots \mathbf{Y}_{n-1})^{\mathrm{T}}$, where $\mathbf{Y}_k = (Y_k, Y_{k-1}, \ldots, Y_{k-p+1})^{\mathrm{T}}$. For any set $\mathcal{A}$ containing m integers $\{a_1, a_2, \ldots, a_m\}$, let $X_{\mathcal{A}} = (\tilde{Y}_{a_1}\ \tilde{Y}_{a_2} \cdots \tilde{Y}_{a_m})$. Note that $\tilde{Y}_t$ and $X_{\mathcal{A}}$ are matrices of dimension $(n-p) \times p$ and $(n-p) \times mp$ respectively. The Euclidean norm is defined by $\|z\| = \sqrt{\sum_{i=1}^{n} a_i^2}$ for $z = (z_1, \ldots, z_n)$. The implementation the group LARS algorithm on multiple change-points estimation is given below.

Algorithm

1. Initialization. Specify K, the maximum number of change-points, and Δ, the minimum distance between change-points. Set $\mu^{[0]}$, $k = 1$, $r^{[0]} = Y$, $\mathcal{A}_0 = \{\emptyset\}$ and $\mathcal{T}_1 = (p + \Delta, p + \Delta + 1, \ldots, n - \Delta)$.

2. Compute the "most correlated set"

$$\mathcal{A}_k = \arg\max_{j \in \mathcal{T}_k} \|\mathbf{B}_j(r^{[k-1]})\| .$$

3. Descent direction computation

$$\gamma_{\mathcal{A}_k} = (X'_{\mathcal{A}_k} X_{\mathcal{A}_k})^{-1} X'_{\mathcal{A}_k} r^{[k-1]} .$$

4. Descent step search: For $j \in \mathcal{T}_k \setminus \mathcal{A}_k$, define

$$a_j = \|\mathbf{B}_j(r^{[k-1]})\|^2 , \quad b_j = \mathbf{B}'_j(X_{\mathcal{A}_k}\gamma_{\mathcal{A}_k})\mathbf{B}_j(r^{[k-1]}) , \quad c_j = \|\mathbf{B}_j(X_{\mathcal{A}_k}\gamma_{\mathcal{A}_k})\|^2, \quad d_j = \max_{j \in \mathcal{T}_k \setminus \mathcal{A}_k} a_j .$$

Set $\alpha = \min_{j \in \mathcal{T}_k \setminus \mathcal{A}_k} \alpha_j \equiv \alpha_{j^*}$, where

$$\begin{aligned} \alpha_j^+ &= \frac{(b_j - d_j) + \sqrt{(b_j - d_j)^2 - (a_j - d_j)(c_j - d_j)}}{c_j - d_j} , \\ \alpha_j^- &= \frac{(b_j - d_j) - \sqrt{(b_j - d_j)^2 - (a_j - d_j)(c_j - d_j)}}{c_j - d_j} , \end{aligned}$$

and

$$\alpha_j = \begin{cases} \alpha_j^+ & \text{if } \alpha_j^+ \in [0,1] , \\ \alpha_j^- & \text{if } \alpha_j^- \in [0,1] . \end{cases}$$

5. If $\alpha \neq 1$ or $k < K$, update $\mathcal{A}_{k+1} = \mathcal{A}_k \cup \{j^*\}$, $\mathcal{T}_{k+1} = \mathcal{T}_k \setminus \{j^* - \Delta, j^* - \Delta + 1, \ldots, j^* + \Delta\}$, $\mu^{[k]} = \mu^{[k-1]} + \alpha X_{\mathcal{A}_k}\gamma_{\mathcal{A}_k}$ and $r^{[k]} = Y - \mu^{[k]}$. Set $k = k+1$ and repeat 3-6. Otherwise, return $\mathcal{A}_k$ as the estimated change-points.

By regarding each possible change-point as one explanatory variable to the time series (the response variable), the LARS algorithm begins by looking for the variable most correlated with the response and proceeds on this direction. The LARS takes the largest step in this direction until some other explanatory variable has as much correlation with the current residual. Then the algorithm keeps on searching for the most correlated set and go into the direction such that all the currently selected variables have the same correlation with the current residual. The procedure is repeated until the maximum number of variables are selected. For more details about the LARS and group LARS algorithm, see Yuan and Lin (2006) and Efron *et al.* (2004).

2.2 Two-step estimation procedure

As we do not know the number of change-points in advanced, in the estimation procedure we assumed that the number is upper bounded by K and searched for K most probable change-points. Thus, the number of estimated change-points is always overestimated. Also, simulation studies reveals that there tends to be more than one estimates clustering around a true change-point. This can be explained by the fact that when expressing the change-point estimation problem as a high-dimensional regression problem, the matrix of explanatory variables $\mathbf{X}_n$ has columns that are nearly identical. For example, the different between the first two columns is only the $\mathbf{Y}_0^{\mathrm{T}}$ term in the first entry. Thus, assigning a coefficient β to the first variable is similar to assigning $\beta/2$ to each of the first and the second variables. Note that this inconsistency is not contradict to the standard theory of high-dimensional regression since the standard assumption that ensures little dependency among explanatory variables is violated.

Two immediate issues arise: (i) how to estimate the true number of breaks m_0, and (ii) how to estimate the change points with a nearly optimal rate? These two issues are dealt with in this subsection.

Although the number of change-points is over-estimated, the estimated set $\mathcal{A}_n$ should identify all the true change-points within a small neighborhood. One way to achieve this mission is to choose the "best possible subset" of change-points in $\mathcal{A}_n$ according to some prescribed information criterion (IC). Given m and the change-points $\mathbf{t} = (t_1, \ldots, t_m)$, an information criterion $IC(m, \mathbf{t})$ typically consists of a sum of a goodness-of-fit measure and a penalty term that accounts for the model complexity. Specifically, let $\widehat{\widehat{\boldsymbol{\beta}}}_j = (\sum_{t=t_{j-1}}^{t_j-1} \boldsymbol{Y}_{t-1}\boldsymbol{Y}_{t-1}^{\mathrm{T}})^{-1} \sum_{t=t_{j-1}}^{t_j-1} \boldsymbol{Y}_{t-1}Y_t$ be the least squares estimator and $S_n(t_{j-1}, t_j) = \sum_{t=t_{j-1}}^{t_j-1}(Y_t - \widehat{\widehat{\boldsymbol{\beta}}}_j \boldsymbol{Y}_{t-1})^2$ be the residual sum of squares. Consider a general information criterion of the form

$$IC(m, \mathbf{t}) = S_n(t_1, t_2, \ldots, t_m) + m\omega_n\,, \tag{2.5}$$

where the least squares criterion $S_n(t_1, t_2, \ldots, t_m) = \sum_{j=1}^{m+1} S_n(t_{j-1}, t_j)$ is the goodness-of-fit measure and ω_n is the penalty term. We estimate the number and locations of the change

points by solving

$$(\hat{\hat{m}}, \hat{\hat{\mathbf{t}}}) = \arg \min_{\substack{m \in (0,1,\ldots,|\mathcal{A}_n|), \\ \mathbf{t}=(t_1,\ldots,t_m) \subset \mathcal{A}_n}} IC(m, \mathbf{t}) \,, \tag{2.6}$$

Some commonly used information criteria that take similar forms as (2.5) are the BIC of Yao (1988) and the MDL of Davis, Lee and Rodgridiz-Yam (2006). In these papers, the best subset of change-points is chosen over all possible locations, which could be computationally challenging when m_0 is large. In contrast, the minimizing domain in (2.6) is a much smaller set, namely over the set $\mathcal{A}_n$. In practice, all possible subsets of $\mathcal{A}$ have to be evaluated to yield the change-points estimates.

When $|\mathcal{A}_n|$ is large, it is possible to achieve further computational efficiency by using the following backward elimination algorithm (BEA). Intuitively, the BEA starts with the set of change-points $\mathcal{A}_n$, then removes the "most redundant" change-points that corresponds to the largest reduction of the IC. The preceding step is repeated successively until no further removal is possible. Specifically, the BEA goes as follows.

1. Set $K = |\mathcal{A}_n|$, $\mathbf{t}_K := (t_{K,1}, \ldots, t_{K,K}) = \mathcal{A}_n$ and $V_K^* = IC(K, \mathcal{A}_n)$.

2. For $i = 1, \ldots, K$, compute $V_{K,i} = IC(K-1, \mathbf{t}_K \backslash \{t_{K,i}\})$. Set $V_{K-1}^* = \min_i V_{K,i}$.

3. - If $V_{K-1}^* > V_K^*$, then the estimated locations of change-points are $\mathcal{A}_n^* = \mathbf{t}_K$.
   - If $V_{K-1}^* \leq V_K^*$ and $K = 1$, then $\mathcal{A}_n^* = \emptyset$. That is, there is no change-point in the time series.
   - If $V_{K-1}^* \leq V_K^*$ and $K > 1$, then set $j = \arg\min_i V_{K,i}$, $\mathbf{t}_{K-1} := \mathbf{t}_K \backslash \{t_{K-1,j}\}$ and $K = K - 1$. Go to step 2.

For example, suppose that from the first step, $\hat{m} = 3$ and $\mathcal{A}_n = (\hat{t}_1, \hat{t}_2, \hat{t}_3)$. The BEA works as follows. First, start with all the change points and compute $V_3^* = IC(3, \mathcal{A}_n) = 10$, say. Then consider removing one change point at a time, i.e., consider the three sets $(\hat{t}_1, \hat{t}_2)$, $(\hat{t}_1, \hat{t}_3)$ and $(\hat{t}_2, \hat{t}_3)$, and compute $V_{3,1} = IC(2, (\hat{t}_1, \hat{t}_2))$, $V_{3,2} = IC(2, (\hat{t}_1, \hat{t}_3))$ and $V_{3,3} = IC(2, (\hat{t}_2, \hat{t}_3))$, respectively. Suppose that $(V_{3,1}, V_{3,2}, V_{3,3}) = (11, 10.5, 12)$, then $V_2^* = \min_i V_{3,i} = 10.5 > 10 = V_3^*$, which means removing any one change point cannot reduce the IC. Thus all three change points are important and it can be concluded that $\hat{\hat{m}} = 3$,

$\hat{\hat{\mathbf{t}}} = (\hat{t}_1, \hat{t}_2, \hat{t}_3)$. On the other hand, if $(V_3^*, V_{3,1}, V_{3,2}, V_{3,3}) = (10, 9, 10.5, 8)$, then $V_2^* = 8 < 10 = V_3^*$, which indicates that removing $\hat{t}_1$ corresponds to the largest reduction of the IC. Thus we conclude that $(\hat{t}_2, \hat{t}_3)$ is a better subset, and proceed to check if further reduction is possible.

Using the BEA algorithm, we obtain the final estimate $\mathcal{A}_n^* =: (\hat{t}_i^*, \ldots, \hat{t}_{|\mathcal{A}_n^*|}^*)$, which is an accurate and computationally efficient estimates for the change-points.

In summary, when the true number of the change points m_0 is unknown, applying the one-step LASSO procedure, we cannot estimate m and the true set of change points $\mathcal{A}$ exactly. In fact we only obtain an estimate of $\mathcal{A}$, $\mathcal{A}_n$, say, which contains more points than the true number of change points. What we can say is that in the set $\mathcal{A}_n$, there exists a subset $\mathcal{A}_n^*$, which estimates $\mathcal{A}$ consistently. But we do not know exactly what $\mathcal{A}_n^*$ is. By going through a second-step selecting procedure, we are able to estimate the true number m_0.

Remark 2.1. *To simplify notations and to facilitate the presentation, we assumed that the autoregressive order p is known and is the same for all segments. This assumption can be relaxed so that the order of each segment is an unknown integer less than p^*. In this case each segment can be regarded as a p^*-th order autoregressive model with the last coefficients equaling zero. In practice, we first estimate the change-points by the two-step procedure using a sufficiently large order p^* for each segment, then identify the order of each segment by applying standard procedures such as BIC or C_p on each estimated segment.*

3 Simulation Results

In this section, we report the simulation results to assess the finite sample behavior of the procedure. Three sets of simulation studies are conducted. The first two sets are adapted from Davis, Lee and Rodriguez-Yam (2005) for comparisons. The third set of simulation investigates the performance of the procedure for long time series. In all simulation experiments, we used the penalty term of the Minimum Description Length principle (Davis, Lee and Rodgriduz-Yam (2006)) for the information criterion. Also, the autoregressive order p is fixed at 5 for each segment.

3.1 Piecewise Stationary Process with Dyadic Structure

The time series in this example is generated from the model

$$Y_t = \begin{cases} 0.9Y_{t-1} + \epsilon_t\,, & \text{if } 1 \le t \le 512\,, \\ 1.69Y_{t-1} - 0.81Y_{t-2} + \epsilon_t\,, & \text{if } 513 \le t \le 768\,, \\ 1.32Y_{t-1} - 0.81Y_{t-2} + \epsilon_t\,, & \text{if } 769 \le t \le 1024\,, \end{cases} \tag{3.7}$$

where $\epsilon_t \sim$ iid N(0,1). In the example, 200 realizations are simulated from model (3.7) and estimated by the two-step procedure. The estimation results for two-step procedure are summarized in Table 1, in which the performance of Auto-PARM reported in Davis, Lee and Rodriguez-Yam are compared. The percentage (%) of the estimated number of segments, the mean and standard error of the location estimates are reported. Note that the two-step procedure gives the correct number of segments in all the 200 realizations, whereas Auto-PARM gives the correct segmentation for 96% of the realizations.

Table 1: *Estimated breakpoints from Auto-PARM and two-step estimation procedure* (3.7).

Number of	Auto-PARM			Two-Step		
segments	(%)	Mean	SE	(%)	Mean	SE
3	96.0	0.500	0.007	100	0.500	0.012
		0.750	0.005		0.750	0.011
4	4.0	0.496	0.004	0		
		0.566	0.108			
		0.752	0.003			

3.2 Short Segments

This example compares the performance between the two-step procedure and the Auto-PARM of the following process.

$$Y_t = \begin{cases} 0.75Y_{t-1} + \epsilon_t\,, & \text{if } 1 \le t \le 50\,, \\ -0.5Y_{t-1} + \epsilon_t\,, & \text{if } 51 \le t \le 1024\,, \end{cases} \tag{3.8}$$

where $\epsilon_t \sim$ iid N(0,1). Similar to the preceding example, 200 realizations are simulated from model (3.8). It is reported in Davis, Lee and Rodriguez-Yam (2005) that the mean of the relative position estimates of the change-point is 0.042, with a standard error of 0.004. Using the two-step procedure, the mean of the relative position estimates of the change-point is 0.049, with a standard error of 0.004. The two-step procedure has a much smaller bias than the Auto-PARM in this example.

3.3 Long Time Series

In this section we demonstrate the performance of the two-step procedures in change-points estimation for several long time series with a large number of change-points. With the rapid growth of high frequency data analysis, time series with length over 10,000 are often encountered. As the time series becomes longer, there may be more change points. The two-step procedure developed in this paper, which inherits the computational efficiency from the LASSO, is well-suited for this situation. Consider three scenarios of long time series with 8 change points, with series lengths ranging from 10,000 to 50,000. Different patterns of the true change points locations are studied. In particular, the change points are evenly located (Scenario 1), located in the first half of the series (Scenario 2) and clustered near the beginning and the end of the sample. (Scenario 3). Since the main focus is about

estimating locations, for illustration, consider the following piecewise stationary segments.

$$Y_t = \begin{cases} 0.9Y_{t-1} + \epsilon_t, & \text{if } 1 \le t \le t_1, \\ 1.69Y_{t-1} - 0.81Y_{t-2} + \epsilon_t, & \text{if } t_1 \le t \le t_2, \\ 1.32Y_{t-1} - 0.81Y_{t-2} + \epsilon_t, & \text{if } t_2 \le t \le t_3, \\ 0.7Y_{t-1} - 0.2Y_{t-2} + \epsilon_t, & \text{if } t_3 \le t \le t_4, \\ 0.1Y_{t-1} - 0.3Y_{t-2} + \epsilon_t, & \text{if } t_4 \le t \le t_5, \\ 0.9Y_{t-1} + \epsilon_t, & \text{if } t_5 \le t \le t_6, \\ 1.32Y_{t-1} - 0.81Y_{t-2} + \epsilon_t, & \text{if } t_6 \le t \le t_7, \\ 0.25Y_{t-1} + \epsilon_t, & \text{if } t_7 \le t \le t_8, \\ -0.5Y_{t-1} + 0.1Y_{t-2} + \epsilon_t, & \text{if } t_8 \le t \le T, \end{cases} \tag{3.9}$$

where $\epsilon_t \sim$ iid N(0,1). In this example, 200 realizations are simulated from model (3.9) with different values of $\mathbf{t} = (t_1, t_2, \ldots, t_8)$ estimated by the two-step procedure. The estimation results are reported in Table 2. The percentage (%) of the estimated number of segments, the mean and standard error of the location estimates are reported. Despite the length of the time series, the computation of a two-step estimation procedure can be completed within 20 seconds. The estimation accuracy is also extremely high. On the other hand, the implementation of the AutoPARM of Davis, Lee and Rodriguez-Yam (2005) requires a large number of replications in computing the criterion function. In particular, their typical specification of 50 islands, 200 genes and 20 generations requires $50 \times 200 \times 20 = 200,000$ evaluations of the criterion function, which takes around 20 minutes for a time series of length $n = 10,000$. All computations are performed using the program **R** on a laptop with an Intel Core i5 480M processor.

Table 2: *Estimated breakpoints from two-step estimation procedure* (3.9).

	Scenario 1			Scenario 2			Scenario 3		
T	10000			20000			50000		
Computing Time	4s			7s			18s		
% of $\hat{m} = 8$	90			84			98		
	True	Mean	SE	True	Mean	SE	True	Mean	SE
t_1/T	0.1	0.1022	0.0091	0.1	0.1001	0.0010	0.01	0.0102	0.00264
t_2/T	0.2	0.2008	0.0012	0.2	0.1998	0.00042	0.1	0.1000	0.00025
t_3/T	0.3	0.3001	0.0010	0.25	0.2499	0.00048	0.15	0.1500	0.00058
t_4/T	0.4	0.3942	0.0088	0.3	0.2984	0.0032	0.2	0.1998	0.00035
t_5/T	0.5	0.4999	0.0012	0.35	0.3501	0.00090	0.8	0.8000	0.00014
t_6/T	0.6	0.5999	0.0010	0.4	0.4001	0.00081	0.85	0.8499	0.00024
t_7/T	0.75	0.7501	0.0011	0.45	0.4501	0.00057	0.9	0.9000	0.00014
t_8/T	0.8	0.7998	0.0016	0.5	0.4998	0.00070	0.99	0.9891	0.00493

4 Applications

4.1 Electroencephalogram Analysis

Electroencephalogram (EEG) displays the brain wave pattern measured by brain electrical potentials from two electrodes across the scalps of a subject. Figure 1 shows an electroencephalogram (EEGs) recorded from a female patient diagnosed with left temporal lobe epilepsy. The EEG was recorded with a sampling rate of 100Hz for a total of 5 minutes and 28 seconds, with the sample size n=32,768. This dataset has been modeled as a piecewise stationary time series in Ombao et al. (2001, 2005) and Davis, Lee and Rodriguez-Yam (2005). The estimated locations of the Lasso procedure to the EEG series is given in Table 3. The Lasso procedure results in 9 segments, corresponding to the vertical dash-lines in Figure 1. Note that the location estimates of the change points are in close agreement with those obtained by the Auto-PARM of Davis, Lee and Rodriguez-Yam (2005). In particular, the estimated starting time for the seizure is $t = 184.23$ seconds, which is very close to the neurologist's estimate of 185 seconds and the estimate of 185.8 seconds by Auto-PARM.

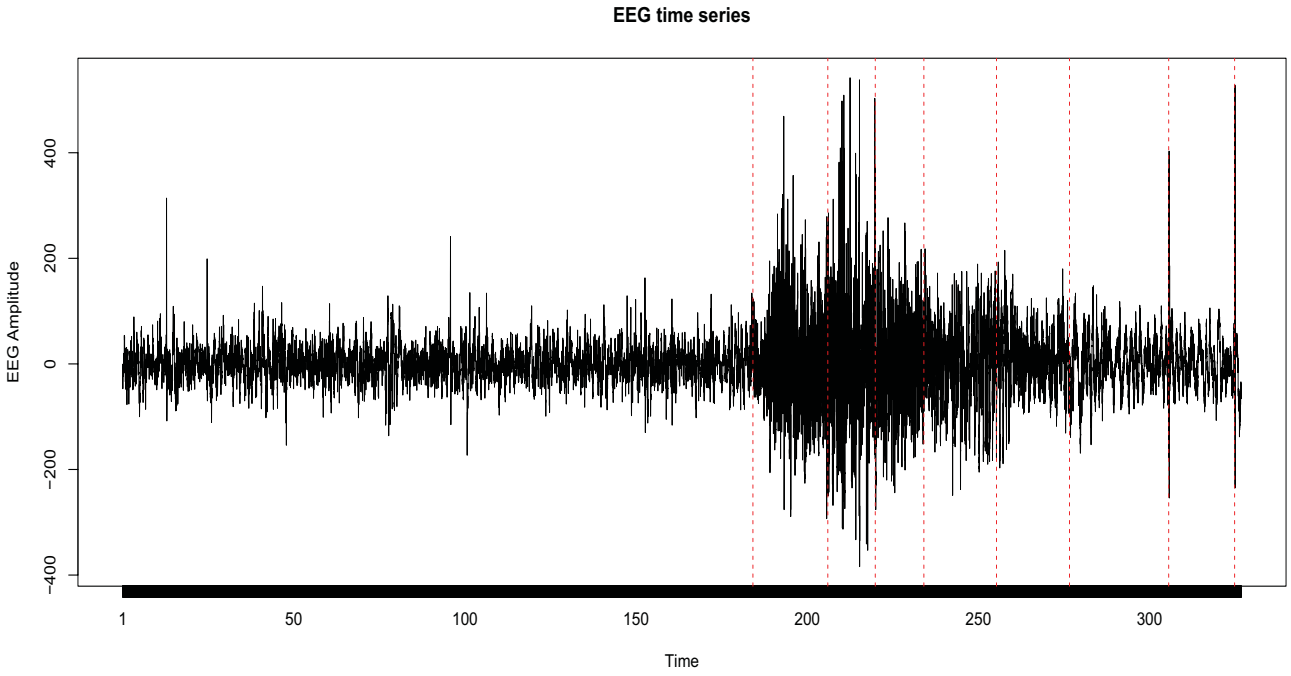

Figure 1: *EEG time series recorded from a female patient diagnosed with left temporal lobe epilepsy. The vertical dash lines correspond to the estimates of the change-point locations.*

4.2 Reventador Vocano Explosion

Lees *et al.* (2008) studied the association between seismic recordings and the explosive activities of the Reventador Volcano located in the Ecuadorian Andes, South America. Figure 2 displays the vertical component of the explosion tremor in 1,000 seconds recorded from a seismo-acoustic station on August 2, 2005. The sample size is $n = 100,000$. Lees *et al.* (2008) qualitatively categorize the seismic signals into seven phases of volcanic activities. The Lasso procedure results in 12 segments, corresponding to the vertical dash-lines in

Table 3: *Estimated breakpoints from the Lasso two-step estimation procedure and Auto-PARM.*

	Locations of change points (seconds)										
	1	2	3	4	5	6	7	8	9	10	11
Two-step	184.23	206.11	219.97	234.17	255.39	276.71	305.69	324.95	-	-	-
Auto-PARM	185.8	189.6	206.2	220.9	233.0	249.0	261.6	274.6	306.0	308.4	325.8

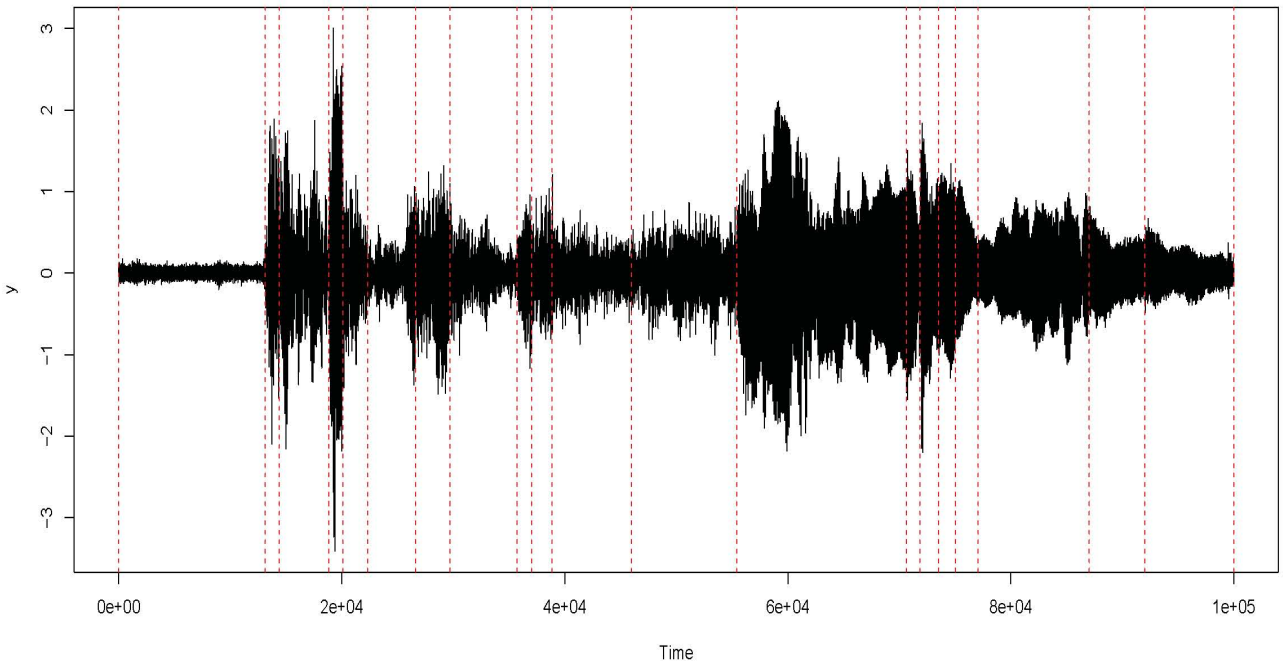

Figure 2: *Vertical component of the Reventador Volcano explosion tremor in 1000 seconds. The vertical dash lines correspond to the estimates of the change-point locations.*

Figure 1. The locations of phase-changes suggested by Lees *et al.* (2008) are fairly well identified by the Lasso procedure. Note that the Lasso procedure breaks this seismic signal into finer pieces, which suggests the possibility that more subtle volcanic activities exist.

4.3 Standard & Poor's 500 Index

We applied the Lasso procedure to analyze the returns of Standard & Poor's 500 index from Jan 2, 2004 to April 29, 2011. The log returns are shown in Figure 3a. It can be seen that the returns started becoming more volatile in 2007 and in the period from mid 2008 to mid 2009, the return fluctuates vigorously, which suggest structural changes in the volatility. Since the ARCH model is commonly used to model volatilities in log-returns and it is well-known that the square of a ARCH process can be regarded as an AR process (e.g., Chan (2010)), the structural changes in volatilities in the returns can be regarded as the structural changes in the autoregressive structure of the squared returns series. Thus

we can apply the Lasso change-point estimation procedure on the squared returns series. Figure 3b shows the squared log-returns of the S&P 500 index and the estimated change-points. The estimated change-points are located on July 10, 2007, September 15, 2008 and April 7, 2009. Referring to the history of the financial crisis, these estimated locations can be well interpreted. In particular, on July 11, 2007, Standard and Poors placed 612 securities backed by subprime residential mortgages on a credit watch, which preludes the panic of the market; on September 15, 2008, Lehman Brothers Holdings incorporated filed for bankruptcy protection and triggered the financial crisis. The last estimated change-point corresponds to the Quantitative Easing (QE) policy where the US Federal Reserve gradually purchased around $1 trillion debt, Mortgage-backed securities and Treasury notes in the early 2009, which stimulated the economy and reduced the volatility in the market.

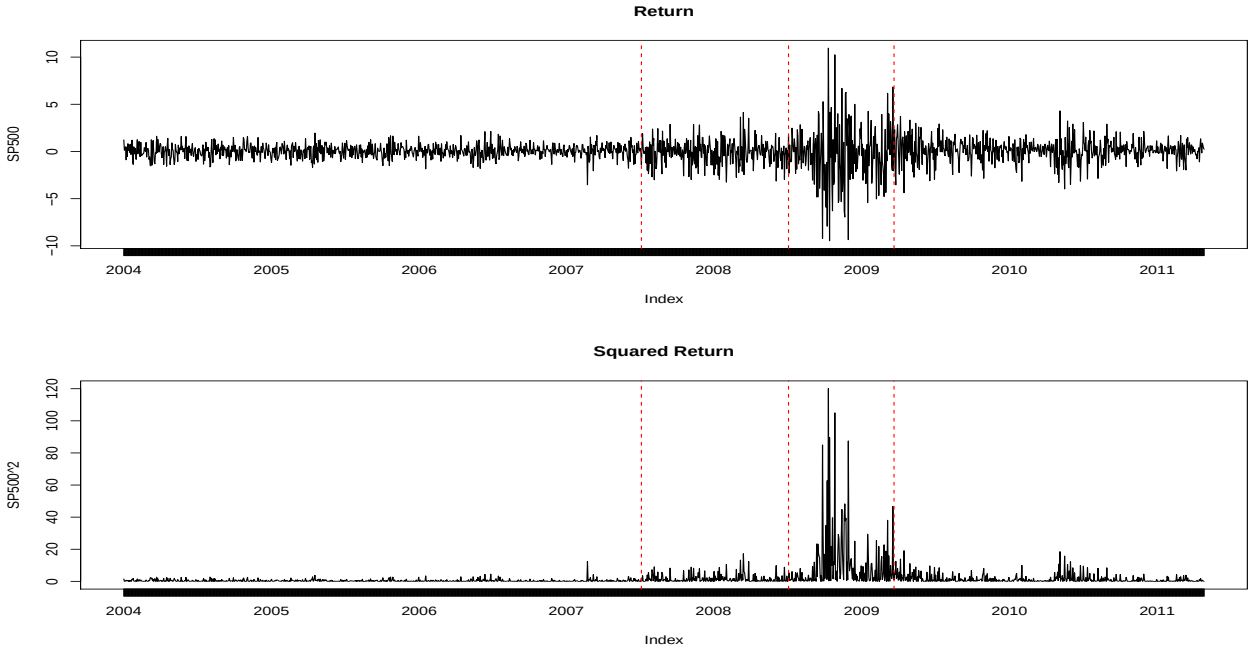

Figure 3: *a) Daily log-returns of the Standard and Poor's (S&P) 500 Index from January 2, 2004 to April 29, 2011. b) Squared valued of the S&P 500 Series. The vertical dash lines correspond to the estimates of the change-point locations.*

References

Adak, S. and Sarkar, A. (1996). A time-frequency search for stock market anomalies. In *Time-Frequency Representations: Algorithms and Applications*. Boston, MA: Birkhause.

Adak, S. (1998). Time-dependent spectral analysis of nonstationary time series. *Journal of the American Statistical Association* **93**, 1488–1501.

Andreou, E. and Ghysels, E. (2008). Structural breaks in financial time series. In *Handbook of Financial Time Series, (*Anderson, T. G., Davis, R. A., Kreiss, J. P. and Mikosch, T.*)* Berlin: Springer, pp. 839–866.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point, *Econometrica* **61**, 821C-856.

Bai, J. (1997). Estimation of a change point in multiple regression models. *The Review of Economics and Statistics* **4**, 551C-563.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structure changes. *Econometrica* **66**, 47–78.

Bai, J. and Perron, P. (2003). Computation and analysis of multiple structure change models. *Journal of Applied Econometrics* **18**, 1–22.

Basseville, M. and Benveniste, A. (1983). Sequential segmentation of nonstationary digital signals using spectral analysis. *Informatio Sciences* **29**, 57–73.

Boysen, L., Kempe, A., LiebsCher, V., Munk, A. and Wittich, O. (2009). Consistencies and rates of convergence of jump-penalized least squares estimators. *Annals of Statistics* **37**, 157–183.

Davis, R. A., Lee, T. C. M. and Rodriguez-Yam, G. A. (2006). Structure break estimation for nonstationary time series models. *Journal of the American Statistical Association* **101**, 223–239.

Dharmadhikari, S. W., Fabian, V. and Jogdeo, K. (1968). Bounds on the moments of martingales. *Ann. Math. Statist.* **39**, 1719–1723.

Freedman, D. A. (1975). On tail probabilities for martingales. *Annals of Probability* **3**, 100–118.

Harchaoui, Z. and Levy-Leduc, C. (2010). Multiple change-point estimation with a total variation penalty. *Journal of the American Statistical Association* **105**, 1480–1493.

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

Koul, H. L. and Ossiander, M. (1994). Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. *Annals of Statistics* **22**, 540–562.

Ombao, H. C., Raz, J. A., Von Sachs, R., and Malow, B. A. (2001). Automatic statistical analysis of bivariate nonstationary time series. *Journal of the American Statistical Association* **96**, 543–560.

Sclove, S. L. (1983). Time series segmentation: a model and a method. *Information Sciences* **29**, 7–25.

Shao, X. and Zhang, X. (2010). Testing for change points in time series. *Journal of the American Statistical Association* **105**, 1228–1240.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B* **58**, 267–288.

Yokoyama, R. (1980). Moment bounds for stationary mixing sequences. *Z. Wahrsch. Verw. Geb.* **52**, 45–57.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B* **68**, 49–67.