doi:10.1111/j.1467-9892.2008.00585.x

BREAK DETECTION FOR A CLASS OF NONLINEAR TIME SERIES MODELS

By Richard A. Davis, Thomas C. M. Lee and Gabriel A. Rodriguez-Yam

Columbia University, The Chinese University of Hong Kong and Universidad Autónoma Chapingo

First Version received January 2006

Abstract. This article considers the problem of detecting break points for a nonstationary time series. Specifically, the time series is assumed to follow a parametric nonlinear time-series model in which the parameters may change values at fixed times. In this formulation, the number and locations of the break points are assumed unknown. The minimum description length (MDL) is used as a criterion for estimating the number of break points, the locations of break points and the parametric model in each segment. The best segmentation found by minimizing MDL is obtained using a genetic algorithm. The implementation of this approach is illustrated using generalized autoregressive conditionally heteroscedastic (GARCH) models, stochastic volatility models and generalized state-space models as the parametric model for the segments. Empirical results show good performance of the estimates of the number of breaks and their locations for these various models.

Keywords. Generalized autoregressive conditionally heteroscedastic process; genetic algorithm; minimum description length principle; model selection; multiple change point; non-stationary time series; state-space models; stochastic volatility model.

1. INTRODUCTION

The problem of modelling a class of nonstationary time series by segmenting the series into different linear or nonlinear stationary pieces is considered. Specifically, the time series is assumed to follow an underlying parametric nonlinear model in which the parameters (and possibly order of the model) change values at fixed time points. Here, the number of break points and their locations are assumed to be unknown. An automatic procedure, termed Auto-Seg for automatic segmentation, is developed for obtaining an *optimal segmentation*.

A single change point in the distribution for independent observations has been broadly studied in the literature. The multiple change point case, a much more difficult problem, has also been considered. A review and an extensive list of references for the multiple change point problem can be found in Shaban (1980), Zacks (1983), Krishnaiah and Miao (1988), Bhattacharya (1994) and Csörgő and Horváth (1997).

In time series, various versions of the change point problem has also been studied. Picard (1985), Davis *et al.* (1995) and Kitagawa *et al.* (2001) studied the single change point problem in which the pieces are assumed to be autoregressive

0143-9782/08/05 834–867 JOURNAL OF TIME SERIES ANALYSIS Vol. 29, No. 5

(AR) processes. Here, a change occurs if one of the parameters, including the constant term, or white noise variance changes.

Multiple change points for dependent data are considered in Kitagawa and Akaike (1978), Fearnhead (2005) and Davis *et al.* (2006) where the observed nonstationary time series is decomposed into segments of AR processes. A more general piecewise stationary process, for which the piecewise AR process is a particular case, is considered in Ombao *et al.* (2001). McCulloch and Tsay (1993), Djurić (1994) and Lavielle (1998). Punskaya *et al.* (2002) consider a Bayesian approach to the change point problem in time series, while Csörgő and Horváth (1997) devote a chapter to the change point problem for dependent observations.

In this article, we consider the multiple change point problem for a class of nonlinear processes in which the segments are modelled by a *specified* parametric class of nonlinear time-series models. More precisely, let m be the unknown number of break points of the observed time series of length n. In addition let $\tau_j, j = 1, \ldots, m$ be the break points between the jth and $(j + 1)$th segments, and set $\tau_0 = 1$ and $\tau_{m+1} = n + 1$. It is assumed that the jth piece of the time series $\{Y_t\}$ is modelled by a stationary time series $\{X_{t,\,j}\}$, i.e.

$$Y_t = X_{t+1-\tau_{j-1},\,j}, \quad \tau_{j-1} \leq t < \tau_j, \tag{1}$$

where the pieces $\{X_{t,\,j}\}, j = 1, \ldots, m + 1$ are independent, $\{X_{t,\,j} : t = 0, \pm 1, \pm 2, \ldots\}$ has stationary distribution $p_{\boldsymbol{\theta}_j}(\cdot)$, and $\boldsymbol{\theta}_j$ is a member of a parameter space Θ_j with $\boldsymbol{\theta}_j \neq \boldsymbol{\theta}_{j+1}$, $j = 1, \ldots, m$. The dimension of $\boldsymbol{\theta}_j$ and its parameter space Θ_j may not only vary with j, but, as in Examples 1 and 2, it can be unknown.

While the independence of the segments in this model framework may seem unduly restrictive, it is used primarily for convenience and numerical stability in the formulation of the segmentation procedure Auto-Seg. After considering a few examples, we will show in Remark 1 how our formulation can be viewed as an approximation to a modelling scheme in which dependence in the observations is allowed to leak across the segmented boundaries.

Example 1 (Segmented AR process). Consider the case when, for all $j = 1, \ldots, m + 1$, $\{X_{t,\,j}\}$ is the AR(p_j) process

$$X_{t,j} = \phi_{j0} + \phi_{j1} X_{t-1,j} + \cdots + \phi_{j,p_j} X_{t-p_j,j} + \sigma_j \varepsilon_{t,j}, \quad t = \ldots, -1, 0, 1, \ldots, \tag{2}$$

where $\{\varepsilon_{t,\,j} : t = 0, \pm 1, \pm 2, \ldots\}$ is a sequence of independently and identically distributed (i.i.d.) $N(0,1)$ noise. If the autoregressive order p_j is assumed unknown, then the parameter $\boldsymbol{\theta}_j$ becomes $(p_j, \boldsymbol{\phi}_j, \sigma_j^2)$, where $\boldsymbol{\phi}_j = (\phi_{j0}, \ldots, \phi_{j,\,p_j})$ is the vector of AR parameters. This set-up has been considered, for example, by Kitagawa and Akaike (1978), Fearnhead (2005) and Davis *et al.* (2006). If p_j is known, then $\boldsymbol{\theta}_j = (\boldsymbol{\phi}_j, \sigma_j^2)$.

Example 2 (GARCH(p,q) process). In this example, the jth piece of the process $\{Y_t\}_{t=1}^n$ is modelled as a generalized autoregressive conditionally heteroscedastic (GARCH) process introduced by Bollerslev (1986); i.e.

$$Y_t = X_{t+1-\tau_{j-1},\,j}, \quad \tau_{j-1} \le t < \tau_j,$$

where for each $j = 1, \ldots, m+1$, $\{X_{t,\,j}\}$ is a GARCH(p_j, q_j) model. That is,

$$X_{t,\,j} = \sigma_{t,\,j}\varepsilon_{t,\,j}, \quad t = \ldots, -1, 0, 1, \ldots,$$

where $\{\varepsilon_{t,\,j}\}$ is i.i.d. $N(0,1)$ and $\sigma_{t,\,j}$ is a positive function of $X_{t,\,j}$ given by

$$\sigma^2_{t,\,j} = \alpha_{0,\,j} + \alpha_{j1}X^2_{t-1,\,j} + \cdots + \alpha_{j,\,p_j}X^2_{t-p_j,\,j} + \beta_{j1}\sigma^2_{t-1,\,j} + \cdots + \beta_{j,\,q_j}\sigma^2_{t-q_j,\,j}, \quad t = \ldots, -1, 0, 1, \ldots, \tag{3}$$

subject to the constraints $\alpha_{0,\,j} > 0$, $\alpha_{i,\,j} \ge 0$, $\beta_{i,\,j} \ge 0$, $i = 1, \ldots, m+1$ and $\alpha_{1,\,j} + \cdots + \alpha_{q_j,\,j} + \beta_{1,\,j} + \cdots + \beta_{q_j,\,j} < 1$. Assuming that the orders p_j and q_j are unknown, then $\boldsymbol{\theta}_j = (p_j, q_j, \alpha_{0,\,j}, \boldsymbol{\alpha}_j, \boldsymbol{\beta}_j)$, where $\boldsymbol{\alpha}_j$ and $\boldsymbol{\beta}_j$ are the vector of α_js and β_js in eqn (3) respectively.

EXAMPLE 3 (STATE-SPACE MODEL). The jth piece of the time series $\{Y_t\}$ is modelled by a state-space model (SSM). If $\{\alpha_t\}$ is the *state process*, then the conditional distribution

$$p(y_t|\alpha_t, \alpha_{t-1}, \ldots, \alpha_1, y_{t-1}, \ldots, y_1) = p(y_t|\alpha_t), \quad \tau_{j-1} \le t < \tau_j, \tag{4}$$

is assumed to belong to a known parametric family of distributions and the state process $\{\alpha_t\}$ is given by

$$\alpha_t = X_{t+1-\tau_{j-1},\,j}, \quad \tau_{j-1} \le t < \tau_j,$$

where for each j, $\{X_{t,\,j}\}$ is the AR(p_j) process in eqn (2). Assuming the order p_j is unknown, the vector of parameters becomes $\boldsymbol{\theta}_j = (p_j, \boldsymbol{\delta}_j, \boldsymbol{\phi}_j, \sigma_j^2)$, where $\boldsymbol{\delta}_j$ is the vector of say q_j parameters associated with the specification of $p(y_t|\alpha_t)$, $\tau_{j-1} \le t < \tau_j$, and $\boldsymbol{\phi}_j$ is the vector of ϕ_js associated with the AR model in eqn (2).

Two SSMs considered in this article that are widely used in the literature are the stochastic volatility model (SVM) and the Poisson-driven model (PDM). These processes have observation equations (4) that belong to the *exponential family of distributions*. Durbin and Koopman (1997) and Kuk (1999) consider the following form for this family

$$p(y_t|\alpha_t) = \exp\left[(\mathbf{z}_t^{\mathrm{T}}\boldsymbol{\beta} + \alpha_t)y_t - b(\mathbf{z}_t^{\mathrm{T}}\boldsymbol{\beta} + \alpha_t) + c(y_t)\right],$$

where $\mathbf{z}_t$ is a vector of covariates observed at time t, $\boldsymbol{\beta}$ is a vector of parameters and $b(\cdot)$ and $c(\cdot)$ are known real functions.

REMARK 1. While at first glance, the assumption of independence in eqn (1) may seem restrictive, it can be viewed as an approximating model in which dependence is allowed across segments. For example, a natural definition of a segmented autoregression could be specified by the recursions,

$$Y_t = \phi_{j0} + \phi_{j1}Y_{t-1} + \cdots + \phi_{j,p_j}Y_{t-p_j} + \sigma_j\varepsilon_t, \quad \tau_{j-1} \le t < \tau_j, \tag{5}$$

where $\{\varepsilon_t\}$ is i.i.d. $N(0,1)$. Notice that the first p_j values of the jth piece are written in terms of the last p_j values of the preceding piece. The log likelihood of the data $(y_1, \ldots, y_n)$ based on this model, assuming that the τ_j and p_j are known, is

$$\sum_{j=1}^{m+1} L_{j-1}(\boldsymbol{\psi}_j, \sigma_j^2; \mathbf{y}_j), \tag{6}$$

where $\mathbf{y}_j = (y_{\tau_{j-1}}, \ldots, y_{\tau_j - 1})$, $\boldsymbol{\psi}_j = (\phi_{0,j}, \ldots, \phi_{p_j,j})$ and L_{j-1} is the conditional likelihood of $\mathbf{y}_j$ given $(y_{\tau_{j-1}-p_j}, \ldots, y_{\tau_{j-1}-1})$. Here, we assume that $(y_{1-p_1}, \ldots, y_0)$ are pre-observed values and that the length $n_j = \tau_j - \tau_{j-1}$ of the jth segment is greater than p_j. The conditional likelihood can then be computed as

$$\begin{aligned} L_{j-1}(\boldsymbol{\psi}_j, \sigma_j^2; \mathbf{y}_j) = &-\frac{n_j}{2}\ln(2\pi) - \frac{n_j}{2}\ln\sigma_j^2 \\ &- \frac{1}{2\sigma_j^2}\sum_{k=\tau_{j-1}}^{\tau_j - 1}(y_k - \phi_{j,0} - \phi_{j,1}y_{k-1} - \cdots - \phi_{j,p_j}y_{k-p_j})^2, \end{aligned}$$

which, upon replacing σ_j^2 with its maximum likelihood estimator $\hat{\sigma}_j^2$, gives the log-profile likelihood

$$L_{j-1}(\boldsymbol{\psi}_j, \hat{\sigma}_j^2; \mathbf{y}_j) = -\frac{n_j}{2}\ln(2\pi) - \frac{n_j}{2}\ln\hat{\sigma}_j^2 - \frac{n_j}{2}.$$

Now, under the model formulation in eqn (1), the log likelihood is given by

$$\sum_{j=1}^{m+1} L(\boldsymbol{\psi}_j, \sigma_j^2; \mathbf{y}_j), \tag{7}$$

where $L(\boldsymbol{\psi}_j, \sigma_j^2; \mathbf{y}_j)$ is the log-likelihood function based on the model specified by the jth piece. For AR processes, the log-profile likelihood $L(\boldsymbol{\psi}_j, \tilde{\sigma}_j^2; \mathbf{y}_j)$ can be well approximated (see Brockwell and Davis, 1991) by

$$-\frac{n_j}{2}\ln(2\pi) - \frac{n_j}{2}\ln\tilde{\sigma}_j^2 - \frac{n_j}{2},$$

where $\tilde{\sigma}_j^2$ is the maximum likelihood estimator of σ_j^2 based *only* on the observations in the jth segment $\mathbf{y}_j$. Since there is often little difference between the estimates $\hat{\sigma}_j^2$ and $\tilde{\sigma}_j^2$, the likelihoods given by eqn (6) under models (5) and (7) under the independent segment model are nearly the same. A similar argument can be made for the models in Examples 2 and 3 where the conditional likelihood L_{j-1} is conditional on an appropriately chosen sigma field. Furthermore, the likelihood given by eqn (7) tends to be more numerically stable at the maximum likelihood estimates, as a function of the configuration of change points and the orders of the models. For these reasons, we have adopted the model framework as specified in eqn (1) which can be viewed as an approximation to a more general formulation.

In this article, we focus on Examples 2 and 3; a more thorough treatment of Example 1 can be found in Davis *et al.* (2006). The SVM and GARCH models are

popular models to analyse log returns of financial time series. The PDM is commonly used for modelling time series of counts. For example, Zeger (1988), Harvey and Fernandes (1989) and Davis *et al.* (1998) have used PDMs for modelling counts of individuals infected by a rare disease. Unlike Example 1, the likelihood of the GARCH, SVM and PDMs do not have a closed form, which makes the estimation of break points for these models, especially for the SVMs and PDMs, computationally challenging.

The problem of finding a 'best' combination of m, τ_js and possibly the orders of the segmented models can be treated as a model selection problem of non-nested models. The best combination of these values are then found by optimizing a desired objective function. Various selection criteria have been used in the literature for the change point problem. For example, Kitagawa and Akaike (1978) and Kitagawa *et al.* (2001) used the Akaike information criterion (AIC); Yao (1988) used the Bayesian information criterion (BIC); Lee (1995) and Liu *et al.* (1997) used modified versions of BIC; Davis *et al.* (2006) used minimum description length (MDL) principle of Rissanen (1989) and Lavielle (1998) and Gustaffson (2000) used maximum *a posteriori* (MAP) criterion. Bai and Perron (1998, 2003) and Ombao *et al.* (2001) designed a criterion tailored to their procedures.

In this article, we adopt the MDL principle as in Davis *et al.* (2006) to solve our model selection problem. For even moderate values of n, optimization of this criterion is not an easy task. To solve this optimization problem, a genetic algorithm (GA) is used to find optimal or near optimal values of the MDL criterion.

The rest of this article is organized as follows. In Section 2, we derive a general expression for the MDL and apply it to the piecewise state-space model. In Section 3, we give an overview of the GA and discuss its implementation to the segmentation problem. In Section 4, we study the performance of Auto-Seg via simulation and in Section 5 the Auto-Seg procedure is applied to the S&P 500 index and results are compared with other change point analyses. Finally, concluding remarks are given in Section 6.

2. MODEL SELECTION

As suggested before, the problem of finding a 'best' segmentation of the data can be posed as a model selection problem and we shall use the two-part MDL method of Rissanen (1989) (see also Hansen and Yu, 2001; Lee, 2001) to solve it. In general, the idea of MDL for solving a model selection problem is to select the model $\mathcal{M} \in \mathcal{F}$ that achieves the best compression of the data, where $\mathcal{F}$ is a family of candidate models. This can be performed using the following steps. First, given any model $\mathcal{M}$ in $\mathcal{F}$, the data $\mathbf{y}$ is split into two parts: $\hat{\mathcal{M}}$ plus the corresponding residuals $\hat{\mathbf{e}}_{\hat{\mathcal{M}}}$, where $\hat{\mathcal{M}}$ denotes a fitted version of $\mathcal{M}$. Then the total code length (i.e. the amount of hardware memory, typically in terms of number of bits) for

storing $\mathbf{y}$ is calculated as the sum of the individual code lengths of $\hat{\mathcal{M}}$ and $\hat{\mathbf{e}}_{\hat{\mathcal{M}}}$. Notice that this total code length for $\mathbf{y}$ is a function of $\hat{\mathcal{M}}$. Lastly, the model $\hat{\mathcal{M}}$ that gives the smallest such total code length of $\mathbf{y}$ is chosen (and defined) as the best fitting model for the problem. As a good compression method and a good statistical model share the same characteristic of capturing the regularities in the data, one should expect that this idea of code length minimization will lead to good choices of statistical models.

To apply this MDL idea to the current problem, we need to derive various expressions for code length comparison. Let $\mathrm{CL}(z)$ denote the code length of an object z and $\mathcal{F}$ be the class of piecewise processes defined in eqn (1). The goal is to find the code length $\mathrm{CL}(\mathbf{y})$ of the data $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ associated with any member $\mathcal{M} \in \mathcal{F}$. Following the convention of Rissanen (1989), $\mathrm{CL}(\mathbf{y})$ is decomposed into

$$\mathrm{CL}(\mathbf{y}) = \mathrm{CL}(\hat{\mathcal{M}}) + \mathrm{CL}(\hat{\mathbf{e}}|\hat{\mathcal{M}}), \tag{8}$$

where $\mathrm{CL}(\hat{\mathcal{M}})$ denotes the code length of the fitted model $\hat{\mathcal{M}}$ and $\mathrm{CL}(\hat{\mathbf{e}}|\hat{\mathcal{M}})$ is the code length of the corresponding residuals (conditional on the fitted model $\hat{\mathcal{M}}$). The MDL defines the best segmentation for our problem as the one that minimizes $\mathrm{CL}(\mathbf{y})$. Therefore, we need to derive expressions for $\mathrm{CL}(\hat{\mathcal{M}})$ and $\mathrm{CL}(\hat{\mathbf{e}}|\hat{\mathcal{M}})$ and we begin with the former.

Before proceeding, we introduce some notation. First, recall that all the model parameters in the jth segments are collected in $\boldsymbol{\theta}_j$. For a reason to be made clear below, we partition $\boldsymbol{\theta}_j$ as $\boldsymbol{\theta}_j = (\boldsymbol{\zeta}_j, \boldsymbol{\psi}_j)$, where $\boldsymbol{\zeta}_j$ collects all the integer-valued parameters, such as those unknown model orders (e.g. AR order) as in Examples 1–3, while $\boldsymbol{\psi}_j$ contains all the real-valued parameters. We shall use c_j and d_j to denote the dimensionality of $\boldsymbol{\zeta}_j$ and $\boldsymbol{\psi}_j$ respectively. We shall also assume that given m, the τ_js and the $\boldsymbol{\zeta}_j$s, the unique maximum likelihood estimate $\hat{\boldsymbol{\psi}}_j$ of the real-valued parameters $\boldsymbol{\psi}_j$ can be obtained. Lastly, let $n_j = \tau_j - \tau_{j-1}$ be the number of observations in the jth segment of $\mathcal{M}$; thus $(\tau_1, \ldots, \tau_m)$ and $(n_1, \ldots, n_{m+1})$ provide equivalent information. With this set-up, any fitted model $\hat{\mathcal{M}}$ can be completely characterized by m, the n_js, the $\boldsymbol{\zeta}_j$s and the $\hat{\boldsymbol{\psi}}_j$s. Therefore, $\mathrm{CL}(\hat{\mathcal{M}})$ can be decomposed as

$$\begin{aligned}\mathrm{CL}(\hat{\mathcal{M}}) = {} & \mathrm{CL}(m) + \mathrm{CL}(n_1) + \cdots + \mathrm{CL}(n_{m+1}) \\ & + \mathrm{CL}(\boldsymbol{\zeta}_1) + \cdots + \mathrm{CL}(\boldsymbol{\zeta}_{m+1}) + \mathrm{CL}(\hat{\boldsymbol{\psi}}_1) + \cdots + \mathrm{CL}(\hat{\boldsymbol{\psi}}_{m+1}),\end{aligned} \tag{9}$$

which, as shown in the Appendix, admits the following expression:

$$\mathrm{CL}(\hat{\mathcal{M}}) = \log_2 m + (m+1)\log_2 n + \sum_{j=1}^{m+1}\sum_{k=1}^{c_j} \log_2 \zeta_{kj} + \sum_{j=1}^{m+1} \frac{d_j}{2} \log_2 n_j, \tag{10}$$

where ζ_{kj} is the kth entry of $\boldsymbol{\zeta}_j$.

The next step is to derive an expression for the second term $\mathrm{CL}(\hat{\mathbf{e}}|\hat{\mathcal{M}})$ in eqn (8). This second term is the code length of the residuals of the fitted model $\hat{\mathcal{M}}$, and, as demonstrated by Rissanen (1989), it is given by the negative of the log (base 2)

likelihood of $\hat{\mathcal{M}}$. If we denote $L(\boldsymbol{\psi}_j; \mathbf{y}_j)$ the observed likelihood of the jth piece and switch the log base from 2 to e, then we have the following as our final total code length expression for CL($\mathbf{y}$):

$$\begin{aligned}&\mathrm{MDL}(m, \tau_1, \ldots, \tau_m, \boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_{m+1}) \\ &\quad = \mathrm{CL}(\mathbf{y}) \\ &\quad = \log m + (m+1)\log n + \sum_{j=1}^{m+1}\sum_{k=1}^{c_j}\log\zeta_{kj} + \sum_{j=1}^{m+1}\frac{d_j}{2}\log n_j - \sum_{j=1}^{m+1} L(\hat{\boldsymbol{\psi}}_j; \mathbf{y}_j), \quad (11)\end{aligned}$$

where the last summand is obtained from the assumption that the pieces are independent. The best fitting model for $\mathbf{y}$ is then the minimizer of $\mathrm{MDL}(m,\tau_1, \ldots, \tau_m,\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_{m+1})$ in eqn (11). Since the values of $\hat{\boldsymbol{\psi}}_j$s are uniquely determined once m, τ_js and $\boldsymbol{\zeta}_j$s are specified, they are suppressed in the notation of $\mathrm{MDL}(m,\tau_1, \ldots, \tau_m,\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_{m+1})$.

Example 4 (State-space model). Recall from Example 3 that $\boldsymbol{\theta}_j = (p_j, \boldsymbol{\delta}_j, \boldsymbol{\phi}_j, \sigma_j^2)$. Let us assume that p_j is the only integer parameter in $\boldsymbol{\theta}_j$. Then $\boldsymbol{\zeta}_j = p_j$ and $\boldsymbol{\psi}_j = (\boldsymbol{\delta}_j, \boldsymbol{\phi}_j, \sigma_j^2)$. Thus, $c_j = 1$, $d_j = p_j + q_j + 2$, where q_j is defined in Example 3. Then

$$\sum_{j=1}^{m+1}\sum_{k=1}^{c_j}\log\zeta_{kj} = \sum_{j=1}^{m+1}\log p_j, \quad \text{and} \quad \sum_{j=1}^{m+1}\frac{d_j}{2}\log n_j = \sum_{j=1}^{m+1}\frac{p_j + q_j + 2}{2}\log n_j. \quad (12)$$

Now, let $\mathbf{y}_j := (y_{\tau_{j-1}}, \ldots, y_{\tau_j - 1})$ and $\boldsymbol{\alpha}_j := (\alpha_{\tau_{j-1}}, \ldots, \alpha_{\tau_j - 1})$ be the vector of observations and states of the jth piece of $\mathcal{M}$. In addition, let $\boldsymbol{\lambda}_j := (\boldsymbol{\phi}, \sigma_j^2)$. For a fixed (known) value of p_j, the likelihood of this piece based on the *complete data* $(\mathbf{y}_j, \boldsymbol{\alpha}_j)$ becomes

$$\begin{aligned}L(\boldsymbol{\psi}_j; \mathbf{y}_j, \boldsymbol{\alpha}_j) &= p(\mathbf{y}_j|\boldsymbol{\alpha}_j, \boldsymbol{\delta}_j)p(\boldsymbol{\alpha}_j|\boldsymbol{\lambda}_j) \\ &= \left(\prod_{t=1}^{n_j} p(y_{t,j}|\alpha_{t,j}, \boldsymbol{\delta}_j)\right)|\mathbf{V}_j|^{1/2}\exp\left[-(\boldsymbol{\alpha}_j - \boldsymbol{\mu}_j)^{\mathrm{T}}\mathbf{V}_j(\boldsymbol{\alpha}_j - \boldsymbol{\mu}_j)/2\right]/(2\pi)^{n_j/2}, \quad (13)\end{aligned}$$

where $\mathbf{V}_j^{-1} := \mathrm{cov}\{\boldsymbol{\alpha}_j\}$, $\boldsymbol{\mu}_j = \phi_{0,j}/(1 - \phi_{1,j} - \cdots - \phi_{p_j,j})\mathbf{1}$ is the vector of means of the state process, and $\mathbf{1}$ is a vector of ones. From eqn (13), it follows that the likelihood of the observed data is given by the product of n_j-fold integrals

$$L(\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_{m+1}; \mathbf{y}) = \prod_{j=1}^{m+1}\int L(\boldsymbol{\psi}_j; \mathbf{y}_j, \boldsymbol{\alpha}_j)\mathrm{d}\boldsymbol{\alpha}_j. \quad (14)$$

Except in simple cases, the integrals in eqn (14) cannot be computed explicitly. In this article, we use the approximation $L_a(\boldsymbol{\psi}_j; \mathbf{y}_j)$ [see eqn (15)] to the likelihood given in Davis and Rodriguez-Yam (2005). Briefly, this approximation is based on a second-order Taylor series expansion of $\log p(\mathbf{y}_j|\boldsymbol{\alpha}_j; \boldsymbol{\delta}_j)$ in a neighbourhood

of the posterior mode of $p(\boldsymbol{\alpha}_j|\mathbf{y}_j; \boldsymbol{\psi}_j)$. To simplify notation, for the jth piece we 'drop' the sub-index j that appears in $\mathbf{y}_j$, $\boldsymbol{\alpha}_j$, etc. Now, let $\boldsymbol{\alpha}^*$ be the mode of the posterior distribution $p(\boldsymbol{\alpha}|\mathbf{y}; \boldsymbol{\psi})$. In addition, let $l(\boldsymbol{\delta}; \mathbf{y}|\boldsymbol{\alpha}) := \log p(\mathbf{y}|\boldsymbol{\alpha}; \boldsymbol{\delta})$ and $R(\boldsymbol{\alpha}, \boldsymbol{\alpha}^*)$ be the remainder of its second-order Taylor series expansion around $\boldsymbol{\alpha}^*$. Since $p(\boldsymbol{\alpha}|\mathbf{y}; \boldsymbol{\psi}) \propto p(\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\delta})p(\boldsymbol{\alpha}|\boldsymbol{\lambda}) = L(\boldsymbol{\psi}; \mathbf{y}, \boldsymbol{\alpha})$, the vector $\boldsymbol{\alpha}^*$ can be found by maximizing the complete likelihood $L(\boldsymbol{\psi}; \mathbf{y}, \boldsymbol{\alpha})$. For the jth segment, the posterior distribution $p(\boldsymbol{\alpha}|\mathbf{y};\boldsymbol{\psi})$ is approximated, as described in Davis and Rodriguez-Yam (2005), by

$$p_a(\boldsymbol{\alpha}|\mathbf{y};\boldsymbol{\psi}) = \boldsymbol{\phi}(\boldsymbol{\alpha}; \boldsymbol{\alpha}^*, (\mathbf{K}^* + \mathbf{V})^{-1}),$$

where $\boldsymbol{\phi}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the multivariate normal density with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ and

$$\mathbf{K}^* := -\frac{\partial^2}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^{\mathrm{T}}} l(\boldsymbol{\delta}; \mathbf{y}|\boldsymbol{\alpha})|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}^*}.$$

This is the same density that is used in the importance sampling procedure of Durbin and Koopman (1997). In the importance sampling context, the ability to draw random samples easily is a key advantage of this density. Moreover, closed formed expressions can be used to give a reasonably close approximation to the exact likelihood. On the jth segment, the likelihood has the factorization

$$L(\boldsymbol{\psi}; \mathbf{y}) = L_a(\boldsymbol{\psi}; \mathbf{y})\mathrm{Er}_a(\boldsymbol{\psi}),$$

where

$$\mathrm{Er}_a(\boldsymbol{\psi}) := \int_{\Re^n} \mathrm{e}^{R(\boldsymbol{\alpha};\boldsymbol{\alpha}^*)} p_a(\boldsymbol{\alpha}|\mathbf{y};\boldsymbol{\psi})\mathrm{d}\boldsymbol{\alpha},$$

and $R(\boldsymbol{\alpha}; \boldsymbol{\alpha}^*)$ is the second-order Taylor series expansion of $l(\boldsymbol{\delta}; \mathbf{y}|\boldsymbol{\alpha})$ around $\boldsymbol{\alpha}^*$, and

$$L_a(\boldsymbol{\psi}; \mathbf{y}) := \frac{|\mathbf{V}|^{1/2}}{|\mathbf{K}^* + \mathbf{V}|^{1/2}} \exp\left[h^* - \frac{1}{2}(\boldsymbol{\alpha}^* - \boldsymbol{\mu})^{\mathrm{T}}\mathbf{V}(\boldsymbol{\alpha}^* - \boldsymbol{\mu})\right]. \tag{15}$$

Here, $h^* := l(\boldsymbol{\delta}; \mathbf{y}|\boldsymbol{\alpha})|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}^*}$. Ignoring the term $\mathrm{e}^{R(\boldsymbol{\alpha};\boldsymbol{\alpha}^*)}$ in the integrand of $\mathrm{Er}_a(\boldsymbol{\psi})$, an approximation to the likelihood $L_a(\boldsymbol{\psi}; \mathbf{y})$ given by eqn (15) is obtained. For SVMs and PDMs, the estimates obtained by maximizing $L_a(\boldsymbol{\psi}; \mathbf{y})$ were found to be close to the Monte Carlo maximum likelihood estimates given for example by Durbin and Koopman (1997) and Sandmann and Koopman (1998). A plausible explanation for this approximation to work is the following. Under the assumption that the second-order Taylor series expansion of $l(\boldsymbol{\delta}; \mathbf{y}|\boldsymbol{\alpha})$ around $\boldsymbol{\alpha}^*$ is a reasonable approximation to $l(\boldsymbol{\delta}; \mathbf{y}|\boldsymbol{\alpha})$ for values of $\boldsymbol{\alpha}$ not too far from $\boldsymbol{\alpha}^*$, the value $\mathrm{e}^{R(\boldsymbol{\alpha},\boldsymbol{\alpha}^*)}$ must be reasonably close to 1. Since the probability that $\boldsymbol{\alpha}$ is far away from $\boldsymbol{\alpha}^*$ must be negligible, the term $\mathrm{Er}_a(\boldsymbol{\psi})$ should not be much different from 1 and hence $L_a(\boldsymbol{\psi}; \mathbf{y})$ is expected to be a good approximation to $L(\boldsymbol{\psi}; \mathbf{y})$.

Now by using eqn (12), replacing $L(\hat{\boldsymbol{\psi}}_j; \mathbf{y}_j)$ with its approximation $L_a(\hat{\boldsymbol{\psi}}_j; \mathbf{y}_j)$, and writing $\boldsymbol{\zeta}_j = p_j$, eqn (11) becomes

$$\begin{aligned}\mathrm{MDL}(m, \tau_1, \ldots, \tau_m, p_1, \ldots, p_{m+1}) = \log m + (m+1)\log n + \sum_{j=1}^{m+1} \log p_j \\ + \sum_{j=1}^{m+1} \frac{p_j + q_j + 2}{2} \log n_j - \sum_{j=1}^{m+1} L_a(\hat{\boldsymbol{\psi}}_j; \mathbf{y}_j),\end{aligned} \tag{16}$$

where $\hat{\boldsymbol{\psi}}_j$ is the optimizer of eqn (15). The best fitting model for $\mathbf{y}$ is then the minimizer of $\mathrm{MDL}(m, \tau_1, \ldots, \tau_m, p_1, \ldots, p_{m+1})$ in eqn (16).

3. OPTIMIZATION USING THE GENETIC ALGORITHM

3.1. *The genetic algorithm*

As we will see later in this section, the optimization of eqn (11) is 'model dependent'. To give an idea of how to proceed, we describe how to optimize the MDL in eqn (16) for the state-space model from Example 4. Optimization for other examples can be achieved in a similar manner. Even for moderate values of n, the minimization of $\mathrm{MDL}(m, \tau_1, \ldots, \tau_m, p_1, \ldots, p_{m+1})$ with respect to m, $\tau_1, \ldots, \tau_m, p_1, \ldots, p_{m+1}$ is not trivial. A procedure that we will use to overcome this problem is the GA, a class of evolutionary algorithms, first proposed by Holland (1992). GAs are randomized search techniques that mimic natural selection to find the maximum (or minimum) of an objective function. Among others, Chatterjee *et al.* (1996) and Gaetan (2000) and Lee (2002) have applied GAs to statistical problems with good results.

The basic component of the GA are structures typically named *chromosomes*, which are usually represented as vectors. While the basics of the canonical GA can be found in Holland (1992) and Eshelman (2000), we give a brief summary. An initial population of M chromosomes are selected (usually at random) and to each individual a probability, which can be proportional to its objective function value, is assigned; that is, chromosomes having better objective function values would have higher probabilities. Then an offspring is created by mating individuals selected according to the assigned probabilities. Two typical genetic operators for mating are crossover and mutation. The new offspring and the parents are merged to create a new population (generation) of size M. The process is iterated to create new generations. The iterations are stopped once a convergence criterion is met. De Jong (1975) suggests to return the best individual found in successive generations. This is referred to as an *elitist* step which guarantees monotonicity of the algorithm.

There are many variations of the above canonical GA. For example, parallel implementations can be applied to speed up the convergence rate as well as to

reduce the chance of converging to sub-optimal solutions (Forrest, 1991; Alba and Troya, 1999). In this article, we implement the following *Island Model*. Instead of running only one search in one giant population, the island model simultaneously runs NI (number of islands) canonical GAs in NI different sub-populations. Periodically, a number of individuals are allowed to migrate amongst the islands according to some migration policy. The migration can be implemented in numerous ways (Martin *et al.*, 2000; Alba and Troya, 2002). In this article, we adopt the following migration policy: after every M_i generations, the worst M_N chromosomes from the jth island are replaced by the best M_N chromosomes from the $(j-1)$th island, $j = 1, \ldots, \text{NI}$. For $j = 1$, the best M_N chromosomes are migrated from the NI-th island. In all our simulations, we set NI to either 10 or 20, $M_i = 5$, $M_N = 2$ and the sub-population size to either 10 or 20.

Declaration of convergence. At the end of each migration, the overall best chromosome is noted. If this best chromosome does not change for 10 consecutive migrations, or the total number of migrations exceeds 20, this best chromosome is taken as the solution to this optimization problem.

3.2. *Implementation details*

This section provides an overview of our implementation of the canonical GA tailored in the special case when $c_j = 1$ (the number of integer-valued parameters is equal to 1) for all j. More complete details of the implementation can be gleaned from the AR case described in Davis *et al.* (2006). Recall that d_j is the number of real-valued parameters in the jth segment. For some of the nonlinear models considered here, such as the GARCH(1,1) or the stochastic volatility process, d_j corresponds to the order of the model which is usually fixed at 2 or 1 for each j.

Chromosome representation. A chromosome should carry all the information that completely characterizes a fitted model $\hat{\mathcal{M}}$. In our implementation, a chromosome $\mathbf{g} = (g_1, \ldots, g_n)$ is a vector of length n, the number of observations of the time series, for which its 'genes' g_t take on the values of -1, if there is no break at time t or the value of d_j, the dimension of the real-valued parameter in the jth segment. That is,

$$g_t = \begin{cases} d_{j+1}, & \text{if } t = \tau_j, j = 0, 1, \ldots, m, \\ -1, & \text{otherwise.} \end{cases}$$

EXAMPLE 5. For a segmented autoregression model of Example 1 with $n = 10$ observations, the chromosome $\mathbf{g} = (2,-1,-1,-1,1,-1,-1,0,-1,-1)$ corresponds to a model having $m = 2$ breaks, the first at $\tau_1 = 5$ and the second at $\tau_2 = 8$. The AR order of the first piece is $p_1 = 2$, of the second piece is $p_2 = 1$, and the last piece is $p_3 = 0$.

In the implementation of the algorithm, a discrete random variable D with values $0,1,\ldots,D_0$, is used to select the *order* of the model in a segment, where D_0

is the largest order model allowed. The probabilities $\pi_j := P[D = j]$, $j = 0, 1, \ldots, D_0$ are predetermined and, in the absence of further information, are often set to be $1/(D_0 + 1)$. To ensure that there are enough observations for obtaining quality estimates for the parameters in each segment, a 'minimum span' constraint is imposed on $\mathbf{g}$. That is, a change point will not be allowed until there are a sufficient number, say m_j, of observations to estimate a model of order d_j.

EXAMPLE 6. For the SVM in Section 4.2, the distribution considered for D is 0 with probability $\pi_0 = 0.5$ and 1 with probability $\pi_1 = 0.5$. Thus, $D_0 = 1$ in this case. Moreover, the minimum span values are $m_0 = 20$ and $m_1 = 100$.

Initial population generation. Each chromosome in the initial population is generated according to the following strategy. First, g_1 is generated from the distribution of D. To fulfil the minimum span constraint, the next $m_{g_1} - 1$ genes $(g_2, g_3, \ldots, g_{m_{g_1}})$ are all set to -1 indicating that no structural break is allowed for at least m_{g_1} observations. For the next freely chosen gene, $g_{m_{p_1}+1}$, it will be either initialized as a break point with a preassigned probability π_{break} or assigned -1. If a break point is selected, then $g_{m_{g_1}+1}$ is chosen by another independent draw from the distribution of D and the process is repeated. We use $\pi_{\text{break}} = \min\{m_1, \ldots, m_{D_0}\}/n$.

Crossover and mutation. Once a set of initial random chromosomes is generated, new chromosomes are generated by either a *crossover* or *mutation* operation. We set the probability for conducting a crossover operation as $1 - \min_k(m_k)/n$.

Crossover. For the crossover operation, two parent chromosomes are chosen from the current population of chromosomes. These two parents are chosen with probabilities inversely proportional to their ranks sorted by their MDL values. In other words, chromosomes that have smaller MDL values will have higher chances of getting selected. From these two parents, the gene values of the child's chromosome will be inherited in the following manner. For the first gene, g_1 will take on the corresponding value from either the first or second parent with equal probabilities. If this value is -1, then the same gene-inheriting process will be repeated for the next gene in line (i.e. g_2). If this value is not -1, then it is a non-negative integer d_j corresponding to the order of the model in the current segment. In this case, the minimum span constraint will be imposed (i.e. the next $m_{p_j} - 1$ g_ts will be set to -1) and the same gene-inheriting process will be applied to the next available g_t.

Mutation. For mutation one child is reproduced from one parent. Again, this process starts with $t = 1$, and every g_t (subject to the minimum span constraint) can take on one of the following three possible values: (i) with probability π_{parent} it will take the corresponding g_t value from the parent; (ii) with probability π_{nobreak} it will take the value -1; and (iii) with probability $1 - \pi_{\text{parent}} - \pi_{\text{nobreak}}$, it will take a new randomly generated model of order d_j. In this article, we set $\pi_{\text{parent}} = 0.3$ and $\pi_{\text{nobreak}} = 0.3$.

4. SIMULATIONS

4.1. *Financial time series*

The performance of Auto-Seg is evaluated via simulation when GARCH models introduced in Example 2 are used to study changes in the dynamics of returns of financial assets. The set-up of this simulation is similar to that of Andreou and Ghysels (2002), who consider piecewise processes. For these models, the pieces are considered to be GARCH(1,1) models. When $m = 1$, the first piece is one of the following two GARCH(1,1) processes

$$Y_{t,k} = \sigma_{t,k}\varepsilon_t, \quad k = 1, 2$$

where

$$\sigma^2_{t,k} = \omega_k + \alpha_k Y^2_{t-1,k} + \beta_k \sigma^2_{t-1,k},$$

and $\varepsilon_t \sim$ i.i.d. $N(0,1)$. Each two-piecewise process has a break at $\tau_1 = 501$ with a total sample size of $n = 1000$. The parameters of the second piece are obtained as follows: for each data generation process (DGP), only one of the β_ks or the ω_ks is modified, while the other parameters remain unchanged (see column labelled as *Piecewise GARCH (1,1) Scenario* in Table I). For completeness, the case of no breaks (i.e. the second piece has the same parameters as the first piece) is included for each DGP.

For a given two-piecewise process, let $\boldsymbol{\theta}_j$ denote the vector of parameters of the jth piece, $j = 1, 2$. In the notation of Section 2, $\boldsymbol{\theta}_j = \boldsymbol{\psi}_j$, and since the orders are fixed at $p_j = q_j = 1$, the MDL is given by

$$\mathrm{MDL}(m, \tau_1, \ldots, \tau_m) = \log m + (m+1)\log n + c\sum_{j=1}^{m+1} \log n_j - \sum_{j=1}^{m+1} L_q(\hat{\boldsymbol{\psi}}_j; \mathbf{y}_j) \quad (17)$$

where $L_q(\boldsymbol{\psi}_j; \mathbf{y}_j)$ is the quasi-likelihood function. The estimation of the parameters $\boldsymbol{\psi}_j$ are obtained using the quasi-maximum likelihood method (Lee and Hansen, 1994). Notice that if one follows the derivation in Section 2, the coefficient c of $\log n_j$ in eqn (17) should be 3/2. However, as there is a strong correlation between the estimates of w_j with the other parameter estimates, this value of $c = 3/2$ is over-penalizing the model complexity. To overcome this problem, we recommend using $c = 1$, which is equivalent to suggesting that the number of free parameters is $d_j = 2$ instead of 3.

Table I lists the relative frequencies of the number of breaks estimated by Auto-Seg obtained from 500 replicates. The AG values were taken from Table III of Andreou and Ghysels (2002) and are also based on 500 replicates. Their estimates were obtained by applying the least-squares procedure of Lavielle and Moulines (2000) to the squared values Y_t^2 with the BIC as a penalty function criterion. In the last column in this table, the unconditional variances of $Y_{t,\,j}$, $j = 1, 2$, are shown. As a general rule, the 'detection rate' is influenced by the size of the change in these variances. The larger the change the higher the detection rate. For example,

TABLE I

SUMMARY OF AUTO-SEG ESTIMATED BREAK POINTS BASED ON 500 REPLICATIONS WHEN THERE IS A BREAK AT 501 OF THE SAMPLE IN THE GARCH PROCESS

Piecewise GARCH(1,1) scenario		No. of break points 0	1	≥2	Unconditional Variance
No break points					
A: (0.4, 0.1, 0.5)	Auto-Seg	0.958	0.042	0.000	1.00
	AG	0.960	0.030	0.010	
B: (0.1, 0.1, 0.8)	Auto-Seg	0.956	0.045	0.00	1.00
	AG	0.880	0.070	0.050	
Break in the dynamics of volatility					
C: (0.4, 0.1, 0.5) → (0.4, 0.1, 0.6)	Auto-Seg	0.804	0.192	0.004	1.00, 1.33
	AG	0.720	0.240	0.040	
D: (0.4, 0.1, 0.5) → (0.4, 0.1, 0.8)	Auto-Seg	0.000	0.964	0.036	1.00, 4.00
	AG	0.000	0.950	0.050	
E: (0.1, 0.1, 0.8) → (0.1, 0.1, 0.7)	Auto-Seg	0.370	0.626	0.004	1.00, 0.50
	AG	0.210	0.750	0.030	
F: (0.1, 0.1, 0.8) → (0.1, 0.1, 0.4)	Auto-Seg	0.004	0.978	0.018	1.00, 0.20
	AG	0.000	0.720	0.280	
Break in the constant of volatility					
G: (0.4, 0.1, 0.5) → (0.5, 0.1, 0.5)	Auto-Seg	0.878	0.122	0.000	1.00, 1.25
	AG	0.850	0.140	0.010	
H: (0.4, 0.1, 0.5) → (0.8, 0.1, 0.5)	Auto-Seg	0.072	0.912	0.016	1.00, 2.00
	AG	0.000	0.940	0.060	
I: (0.1, 0.1, 0.8) → (0.3, 0.1, 0.8)	Auto-Seg	0.068	0.910	0.022	1.00, 3.00
	AG	0.000	0.940	0.060	
J: (0.1, 0.1, 0.8) → (0.5, 0.1, 0.8)	Auto-Seg	0.008	0.952	0.040	1.00, 5.00
	AG	0.000	0.860	0.140	

In the last column, the unconditional variances of both pieces (when applies) are shown. The AG values were taken from table III, Andreou and Ghysels (2002). The length of the realizations is $n = 1000$.

in scenario C, the increase in variance is 0.33, which is slightly larger than 0.25, the increase in variance of scenario G. For Auto-Seg, the detection rates are 0.192 and 0.122, respectively, while, for AG, these values are 0.240 and 0.140.

For illustrative purposes, Figure 1 shows typical realizations of scenarios C and D defined in Table I. Realizations of scenario C/D are shown in the top/bottom panels of this figure. In Figure 1, the dotted vertical lines at 506 and 502 are the breaks found by Auto-Seg for these two realizations. In Figure 2, two 'versions' of volatilities ($\hat{\sigma}_t^2$s) are shown for these realizations. In the top panel, the estimated volatilities were obtained when the realization of scenario C is modelled as a single segment. The volatilities shown in the second panel were obtained using a two-piece GARCH(1,1) process with a break at 506 found by Auto-Seg. In both panels, the Auto-Seg break is shown as the vertical dotted line. The plots in the last two panels are the analogous volatilities for the realization of scenario D (the break is at 502). From Figure 2 we notice that for the realization of scenario D the 'one-piece' volatilities are not much different from the 'two-piece' volatilities. It is not the case for the realization of scenario C. However, notice that the volatilities for t between 1 and 505 are in close agreement.

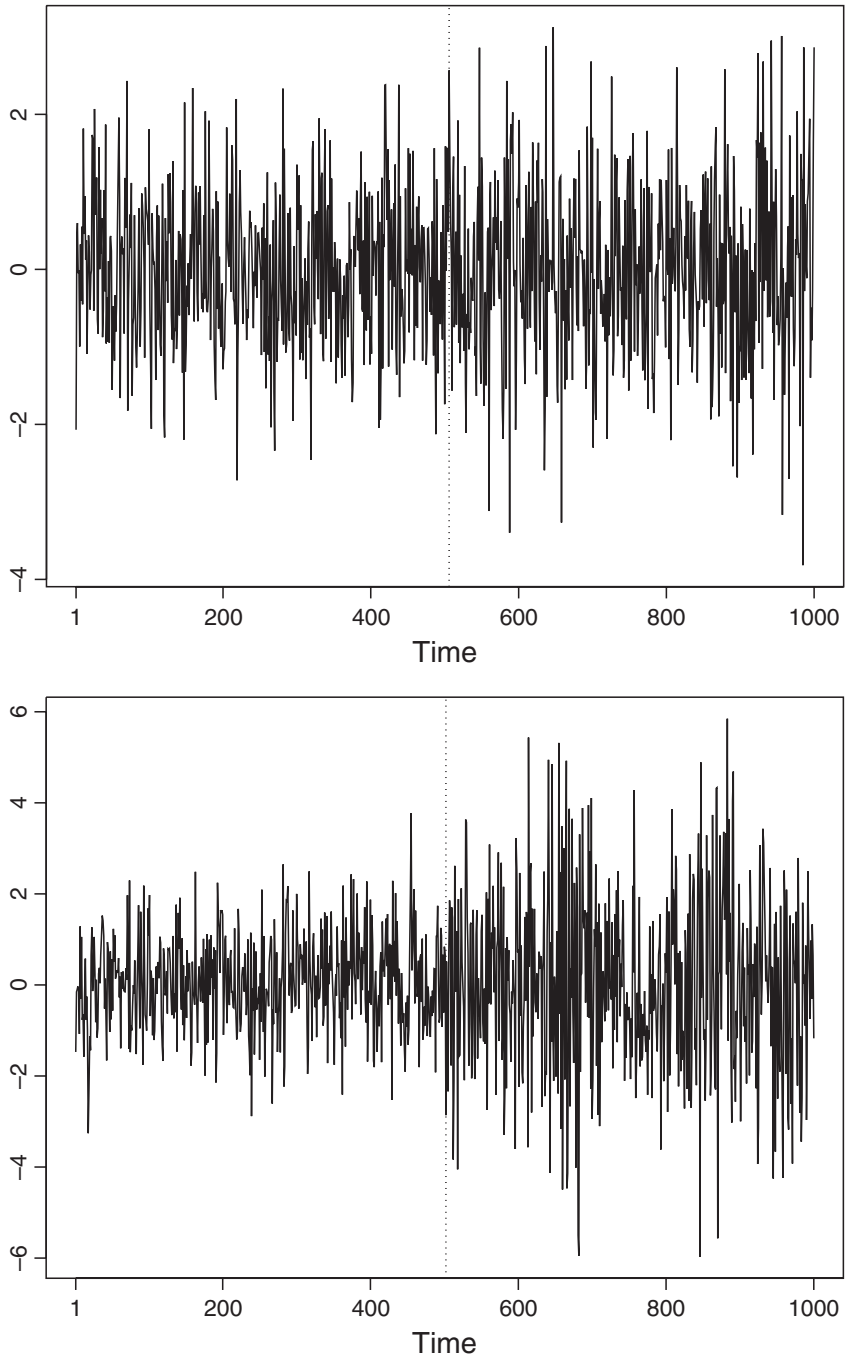

FIGURE 1. Typical realizations of scenarios C (top panel) and D (bottom panel) defined in Table I. The vertical dotted lines are the break points found by Auto-Seg.

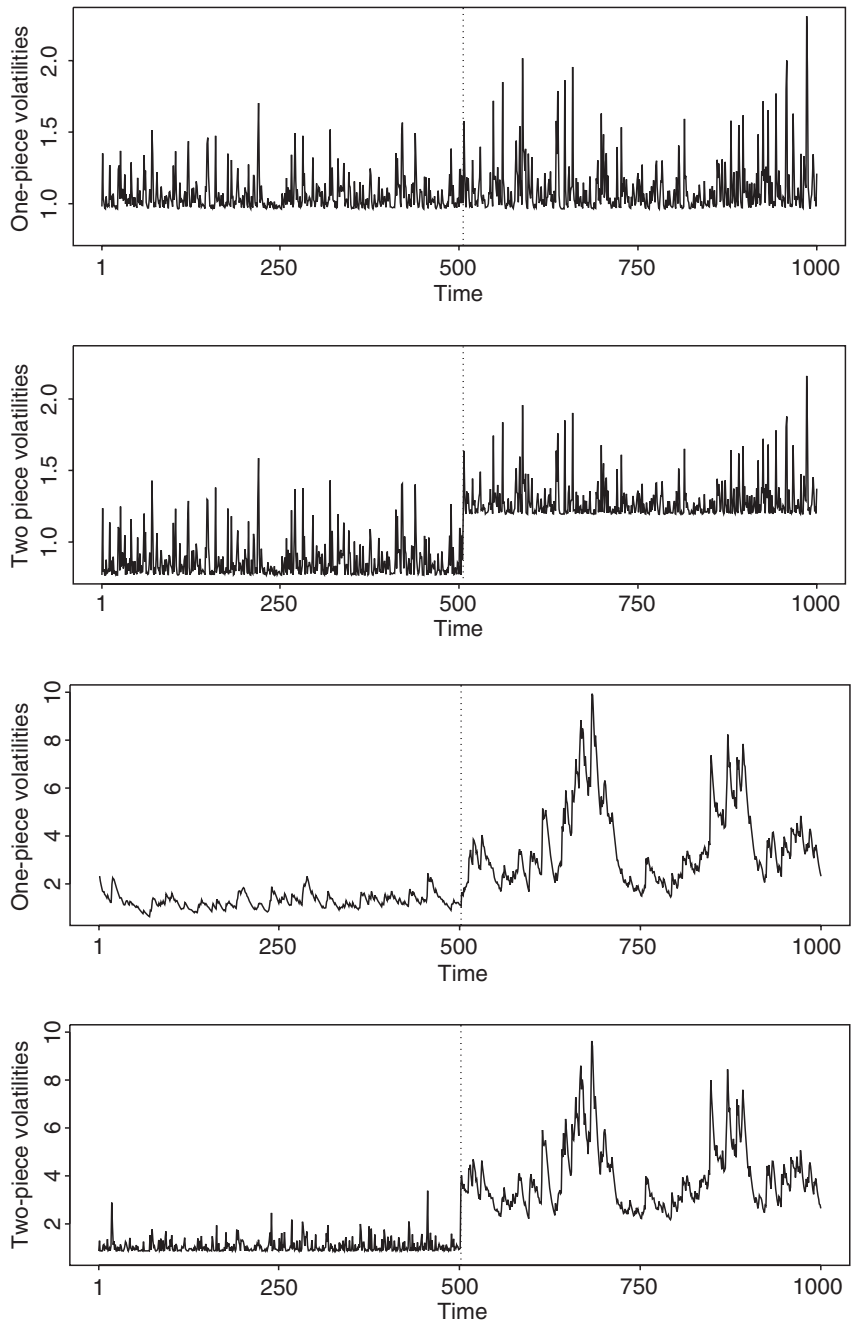

FIGURE 2. Top two panels: estimated volatilities of the realization of scenario C shown in Figure 1 under the assumption of no break (first panel) and using the break (second panel) found by Auto-Seg. The last two panels are the analogous plots for the realization of scenario D.

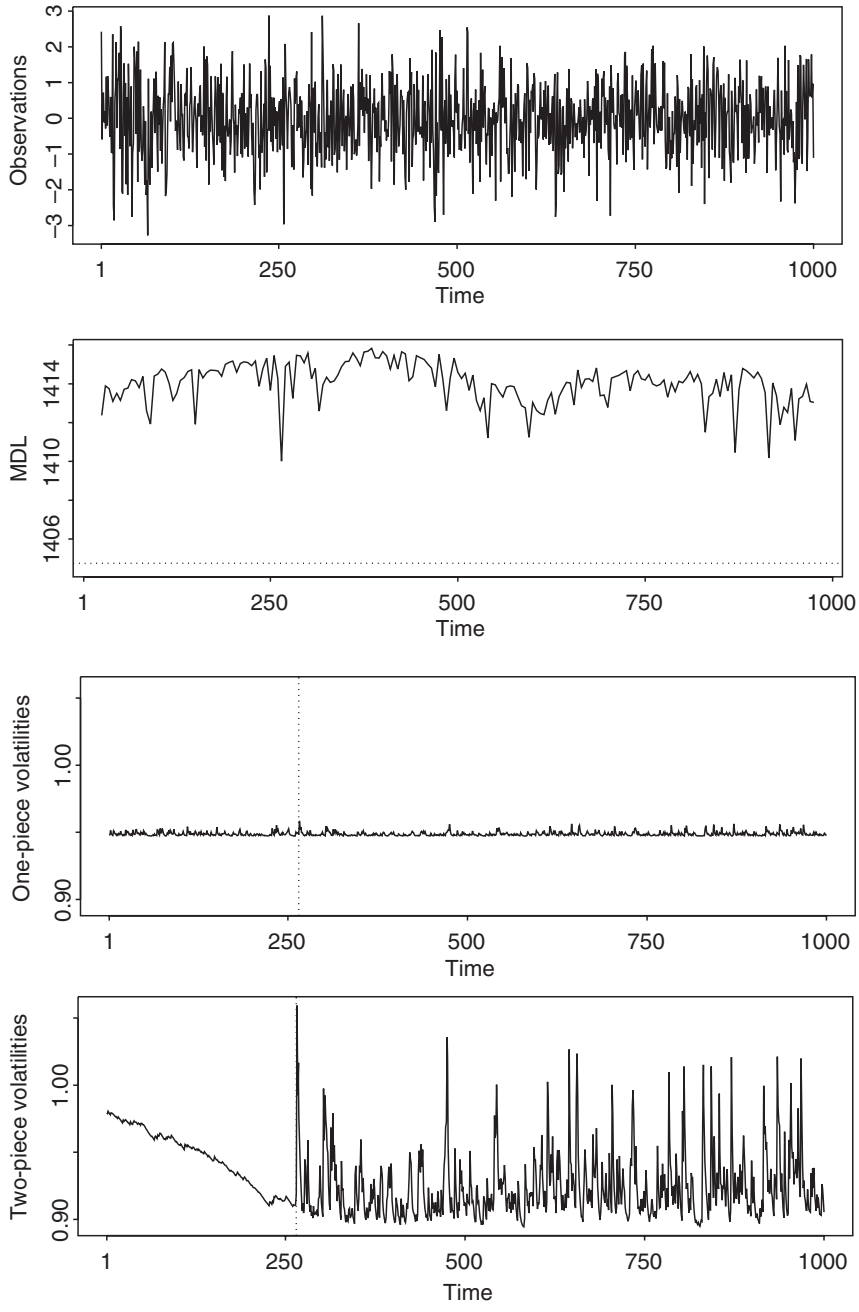

FIGURE 3. First panel: a typical realization of scenario A defined in Table VI. Second panel: two-piece MDL computed in a grid of points (solid line) and one-piece MDL (horizontal dashed line). Third panel: estimated volatilities based on a single piece. Fourth panel: estimated volatilities based on two pieces by introducing an artificial break at location 265.

Next, we consider a different set-up in which there is no break in the DGP. The top panel of Figure 3 contains a realization of scenario A defined in Table I and using Auto-Seg, no breaks were found. For this realization, the MDL was computed for a two-piece model with breaks at true locations t, $t = 25, 30, \ldots, 975$. These MDL values are shown as the solid line in the second panel of Figure 3 with an MDL value of 1410.02 at location 265. The horizontal dashed line in this plot is the MDL with no breaks (1404.75). In the third panel, the estimated volatilities based on a single piece are shown. In the last panel, we show the estimated volatilities based on two GARCH(1,1) models with a break at location 265 (minimizer of two-piece MDL values shown in the second panel). Notice that the one-piece estimated volatilities (third panel) have smaller variance than that based on a two-piece GARCH fit (fourth panel). In the latter, the pattern of the volatilities in the first piece is unexpected and does not agree with the realization in the first panel. We also compared Auto-Seg with the sequential procedure proposed by Berkes *et al.* (2004). To estimate changes in the GARCH model, Berkes *et al.* (2004) constructed a stopping time based on quasi-maximum likelihood estimates. For their simulation study, they used GARCH(1,1) models with the following sets of parameter values

$$\begin{aligned}
&\text{model 1: } \omega_1 = 0.05, \alpha_1 = 0.4, \beta_1 = 0.3 (\sigma^2_{Y_t} = 0.17),\\
&\text{model 2: } \omega_1 = 0.05, \alpha_1 = 0.5, \beta_1 = 0.0 (\sigma^2_{Y_t} = 0.10),\\
&\text{model 3: } \omega_1 = 1.00, \alpha_1 = 0.3, \beta_1 = 0.2 (\sigma^2_{Y_t} = 2.00),
\end{aligned}$$

where $\sigma^2_{Y_t}$ is the unconditional variance of Y_t, and assumed changes from model 1 to model 2 and from model 1 to model 3 at different time locations (see Table II). Notice that unlike Andreou and Ghysels (2002), this simulation study of Berkes *et al.* (2004) allows for changes to more than one parameter. For example, when model 1 changes to model 3 at $t = 250$, all three parameters are altered. In Table II,

TABLE II

AUTO-SEG: SUMMARY STATISTICS FOR THE DISTRIBUTION OF THE ESTIMATED LOCATION OF BREAK POINTS (REPLICATIONS WITH ONLY ONE BREAK)

	Mean	SE	Min	Q1	Med	Q3	Max	Freq
Model 1 → model 3 at $t = 50$								
Auto-Seg	52.62	11.70	37	50	50	52	233	0.98
BERKES	71.40	12.40	50	63	71	79	135	
Model 1 → model 3 at $t = 250$								
Auto-Seg	251.18	4.50	228	250	250	252	271	0.99
BERKES	272.30	18.10	89	262	271	282	338	
Model 1 → model 3 at $t = 500$								
Auto-Seg	501.22	4.76	481	500	500	502	551	0.98
BERKES	516.40	54.70	121	511	523	538	618	
Model 1 → model 2 at $t = 250$								
Auto-Seg	237.28	85.68	38	204.5	237.5	263.0	918	0.52
BERKES	612.90	66.50	89	498.0	589.0	710.0	1000	

BERKES summary statistics for the distribution of the first exceedance for the 10% critical level (from Table 4 of Berkes *et al.*, 2004).

TABLE III

PARAMETER ESTIMATES FOR THE SCENARIOS B, C, D AND E BASED ON THE REPLICATES WITH TWO FITTED PIECES

	Piece 1			Piece 2			MDL	
	ω	α	β	ω	α	β	Two-piece	One-piece
Model 1 $\rightarrow$ model 3 at $t = 50$								
True	0.05	0.40	0.30	1.00	0.30	0.20		
Mean	0.07	0.37	0.23	1.02	0.30	0.19	1677.40	1702.50
SD	0.04	0.26	0.26	0.21	0.05	0.12		
Model 1 $\rightarrow$ model 3 at $t = 250$								
True	0.05	0.40	0.30	1.00	0.30	0.20		
Mean	0.05	0.39	0.28	1.02	0.30	0.19	1418.53	1574.03
SD	0.02	0.15	0.15	0.23	0.06	0.13		
Model 1 $\rightarrow$ model 3 at $t = 500$								
True	0.05	0.40	0.30	1.00	0.30	0.20		
Mean	0.05	0.39	0.29	1.01	0.29	0.20	1094.64	1143.83
SD	0.01	0.13	0.11	0.27	0.08	0.16		
Model 1 $\rightarrow$ model 2 at $t = 250$								
True	0.05	0.40	0.30	0.05	0.50	0.00		
Mean	0.06	0.37	0.31	0.05	0.49	0.02	250.90	255.24
SD	0.03	0.17	0.17	0.01	0.01	0.04		

we show some basic statistics for both the breaks from Auto-Seg and the sequential method. For Auto-Seg, the statistics are from the estimated break points based on 500 replicates of size $n = 1000$. In the row labelled as BERKES, summary statistics for the distribution of the first exceedance of the 10% critical level, taken from Table 4 of Berkes *et al.* (2004) are shown. For Auto-Seg estimates, the proportion of replicates that contain one break point is shown in the last column (Freq). Notice that for the first three configurations, the proportion of replicates with one break is large, while, for the last configuration, this proportion is small. This is in agreement with Berkes *et al.* (2004) results, where the proportion of trajectories that crossed the 10% critical level at $t \leq 400$ is only 0.071, while for $t \leq 500$ this proportion is 0.252 (values taken from Table III of Berkes *et al.* (2004)). This is also in agreement with the results from Table I. For this latter configuration, the unconditional variance is 0.17 for the first piece and 0.10 for the second piece. Since the change in variance is small, a high detection rate is not expected.

For each scenario considered in Table II, with the realizations considered in this table (i.e. realizations for which exactly one break was found using Auto-Seg), the parameters of each piece defined by the Auto-Seg break point were computed. For each scenario, the average and standard error of these estimates are shown in Table III. Also included in this table is the average of the optimized MDL values. These values are shown in column 8. In addition, in the last column, the average of the MDL values obtained when only one piece is fitted to each realization is shown. In all cases, the two-piece MDL average is considerably less than that of the one-piece MDL.

4.2. *Stochastic volatility model*

In Section 4.1, the performance of Auto-Seg on a piecewise GARCH(1,1) model was studied. Another competing model that is often used for financial time series is the SVM defined by the equation

$$y_t = \sigma_t \xi_t = e^{\alpha_t/2} \xi_t,$$

where $\alpha_t = \gamma + \phi \alpha_{t-1} + \eta_t$, $\{\xi_t\} \sim$ i.i.d. $N(0,1)$ and $\{\eta_t\} \sim$ i.i.d. $N(0, \sigma^2)$, $t = 1, \ldots, n$ and $|\phi| < 1$. This model can be written in the SSM framework given in Example 3 in Section 1. We are unaware of any work on break points in the SVM. In this section, we consider the performance of Auto-Seg on a two-piece model where each piece is the SVM. The vector of parameters of this process is $\boldsymbol{\psi} = (\gamma, \phi, \sigma^2)$. Let us consider the models generated by the parameter values:

model 1: $\gamma = -0.8106703$, $\phi = 0.90$, $\sigma^2 = 0.45560010$
model 2: $\gamma = -0.3738736$, $\phi = 0.95$, $\sigma^2 = 0.06758185$
model 3: $\gamma = -0.3973738$, $\phi = 0.95$, $\sigma^2 = 0.06758185$

The two-piecewise processes considered in this section are listed in the last four lines of Table IV. The first piece of these processes is model 1. Scenarios B and D have one true break at 513 and scenarios C and E have true breaks at 1025. The number of observations for each scenario is 2048. In the last column of this table, the true unconditional variances of each piece are displayed. The unconditional variances of the first piece are 0.0010, while the unconditional variances of the second piece of processes B and C are 0.0008 (small decrease). The unconditional variances of the second pieces of the processes D and E are 0.0005, which is half the variance of the first piece.

For each of these piecewise processes, let $\boldsymbol{\theta}_j$ be the vector of parameters of the jth piece. In the notation of Section 2, $\boldsymbol{\theta}_j = \boldsymbol{\psi}_j$. Then $c_j = 0$ and $d_j = 3$ and from eqn (11) we obtain

$$\mathrm{MDL}(m, \tau_1, \ldots, \tau_m) = \log m + (m+1) \log n + \sum_{j=1}^{m+1} \frac{3}{2} \log n_j - \sum_{j=1}^{m+1} L_a(\hat{\boldsymbol{\psi}}_j; \mathbf{y}_j, \boldsymbol{\alpha}_j),$$

TABLE IV

SUMMARY OF AUTO-SEG BREAK POINTS OBTAINED FROM 500 REPLICATIONS

Scenario	No. of break points			σ_Y^2
	0	1	≥ 2	
A: model 1	100.0	0.0	0.0	0.0010
B: model 1 $\rightarrow$ model 2 at $t = 513$	18.2	81.8	0.0	0.0010, 0.0008
C: model 1 $\rightarrow$ model 2 at $t = 1025$	0.4	99.6	0.0	0.0010, 0.0008
D: model 1 $\rightarrow$ model 3 at $t = 513$	17.2	82.8	0.0	0.0010, 0.0005
E: model 1 $\rightarrow$ model 3 at $t = 1025$	1.2	98.8	0.0	0.0010, 0.0005

The length of each realization is $n = 2048$.

where $L_a(\hat{\boldsymbol{\psi}}_j; \mathbf{y}_j, \boldsymbol{\alpha}_j)$ is defined in Example 4. For each scenario, Auto-Seg procedure was applied to 500 realizations. The relative frequencies of the number of breaks estimated by Auto-Seg are displayed in Table IV.

As an illustration, in Figure 4, we show typical realizations of scenarios B (top panel) and E (bottom panel). In Figure 4, for the realization of scenario B, using Auto-Seg a break was found at location 550 (dashed vertical line) and for that of scenario E a break at 1019 (dashed vertical line). In Figure 5, two estimates of the posterior mode $\boldsymbol{\alpha}^*$ of the vector of states described in Example 4 are shown for these realizations. In the top panel, the estimated modes were obtained when a single (unsegmented) model fitted to a realization of scenario B. The estimated modes shown in the second panel were obtained using the two-piece SVM found by Auto-Seg (i.e. there is a break at 550). In both panels, the Auto-Seg break is shown as the vertical dotted line. The plots in the last two panels are the analogous modes for the realization of scenario E for which using Auto-Seg a break was found at 1019. Although in Figure 5 there are differences between both estimates of the posterior mode (i.e. without and with the Auto-Seg break), the agreement of the 'shapes' between these estimates is remarkable.

Now, summary statistics for those replicates of scenarios B–E from Table IV, for which using Auto-Seg exactly one break was found, are given in Table V.

4.3. *Poisson parameter-driven process*

We consider the performance of Auto-Seg on a two-piecewise Poisson process, i.e. for each piece, the observation equation $p(y_t|\alpha_t; \boldsymbol{\delta})$ has a Poisson distribution with rate $\lambda_t := e^{\beta+\alpha_t}$, and the state equation is $\alpha_t = \phi\alpha_{t-1} + \eta_t$, $\{\eta_t\} \sim$ i.i.d. $N(0, \sigma^2)$, $t = 1, \ldots, n$ and $|\phi| < 1$. The vector of parameters of this process is $\boldsymbol{\psi} = (\beta, \phi, \sigma^2)$. In particular, we consider the PDMs with the following set of parameter values:

model 1: $\beta = -1.5702$, $\phi = 0.50$, $\sigma^2 = 1.9237$
model 2: $\beta = -1.3061$, $\phi = -0.50$, $\sigma^2 = 1.5277$
model 3: $\beta = -1.3061$, $\phi = 0.90$, $\sigma^2 = 0.3870$
model 4: $\beta = -0.9373$, $\phi = -0.50$, $\sigma^2 = 0.9745$
model 5: $\beta = -0.9373$, $\phi = 0.90$, $\sigma^2 = 0.2469$

The two-piecewise PDM processes considered in this section are defined in the first column of Table VI. The first piece of these processes is model 1 with a true break at either 257 or 513. The total number of observations for all models is 1024. In the last column of this table, the true unconditional variances of each piece are displayed. The unconditional variances of the first pieces are 7.5, while the unconditional variances of the second pieces of processes B–E are 4.5 (small decrease). The unconditional variances of the second pieces of the processes F–I are 2.25, which consist of a larger decrease. Notice that the MDL calculation of this piecewise process is identical to that for the SVM given in Section 4.2.

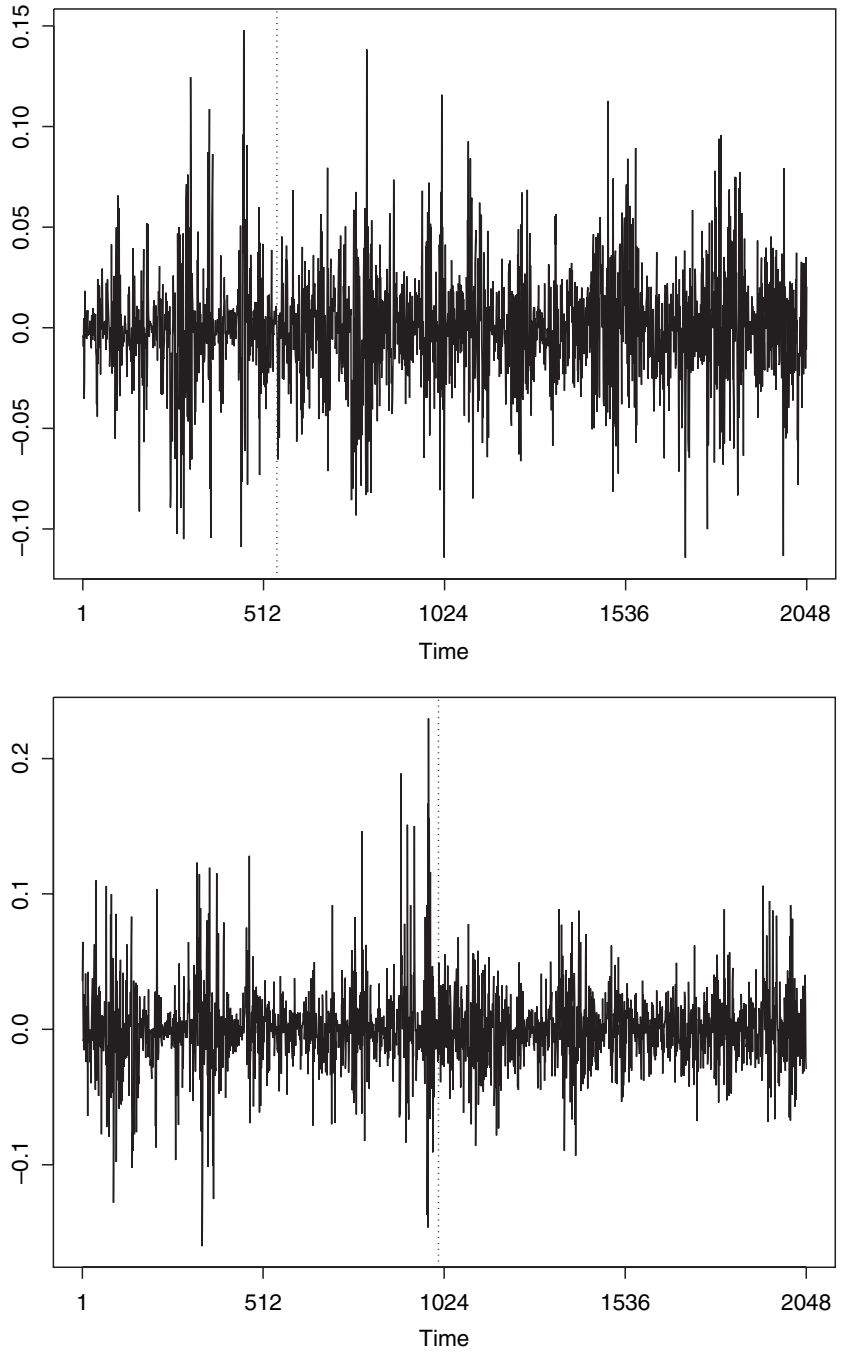

FIGURE 4. Realizations from the piecewise stochastic volatility scenarios B and E defined in Table IV. The vertical dotted lines are break points found by Auto-Seg.

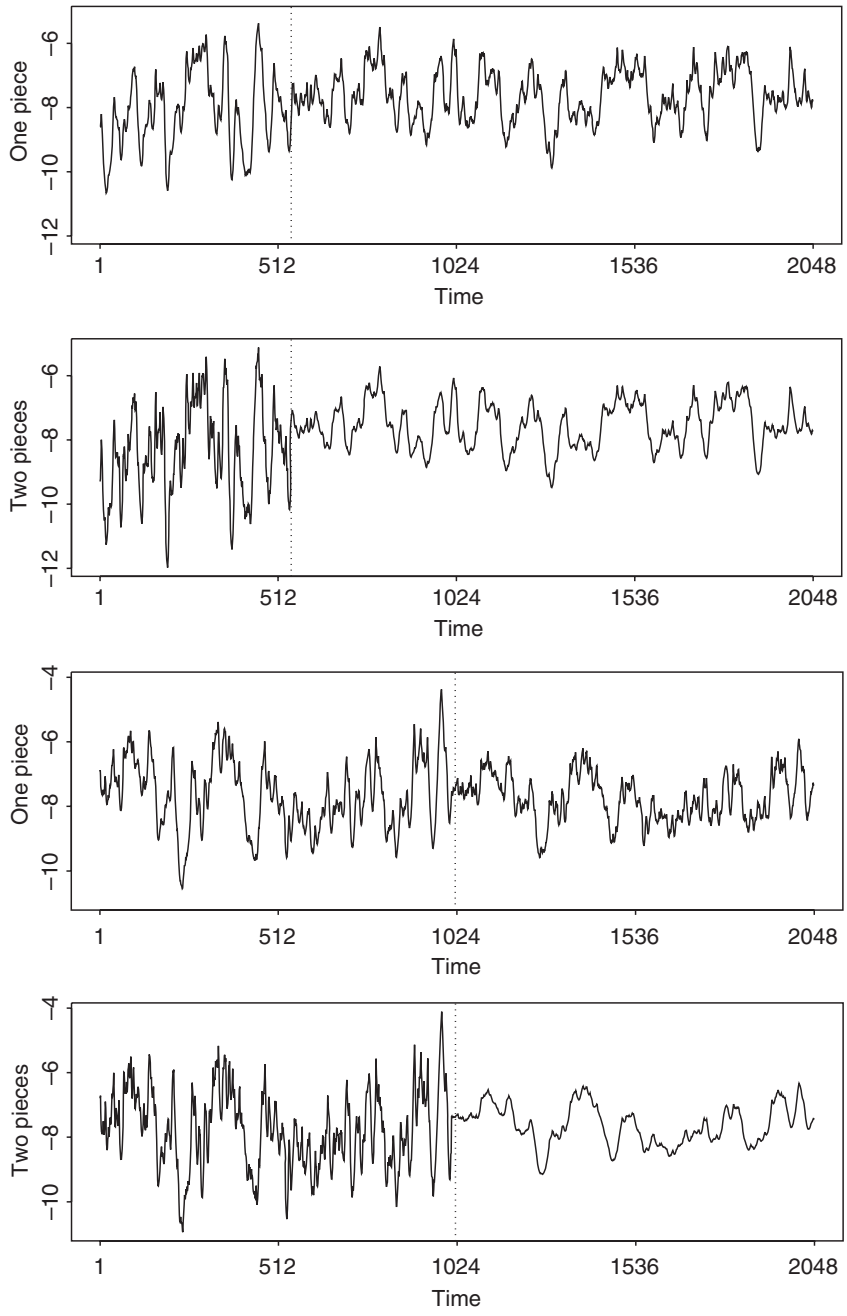

FIGURE 5. Top two panels: estimated posterior mode of the vector of states for the realization of scenario C shown in Figure 4 under the assumption of no break (first panel) and using the break (second panel) found by Auto-Seg. The last two panels are the analogous plots for the realization of scenario E.

TABLE V

SUMMARY STATISTICS FOR THE DISTRIBUTION OF THE ESTIMATED LOCATION OF BREAK POINTS OF THOSE REPLICATIONS WITH ONE BREAK FOR THE SCENARIOS GIVEN IN TABLE IV

Scenario	Mean	SE	Min	Q1	Med	Q3	Max	Freq
Unconditional variance decreases from 0.0010 to 0.0008								
B	506.83	90.44	207	481	509	535	1239	409
C	1020.84	80.68	657	993	1023	1047	1525	498
Unconditional variance decreases from 0.0010 to 0.0005								
D	502.59	72.04	203	479	507	527	831	414
E	1018.37	79.44	685	985	1023	1047	1469	494

TABLE VI

SUMMARY OF ESTIMATED AUTO-SEG BREAK POINTS OBTAINED FROM 500 REPLICATIONS

	No. of break points			
Scenario	0	1	≥ 2	σ_Y^2
A: model 1	100.0	0.0	0.0	7.5
B: model 1 $\rightarrow$ model 2 at $t = 257$	34.0	66.0	0.0	7.5, 4.5
C: model 1 $\rightarrow$ model 2 at $t = 513$	11.6	88.4	0.0	7.5, 4.5
D: model 1 $\rightarrow$ model 3 at $t = 257$	31.0	69.0	0.0	7.5, 4.5
E: model 1 $\rightarrow$ model 3 at $t = 513$	16.8	83.2	0.0	7.5, 4.5
F: model 1 $\rightarrow$ model 4 at $t = 257$	13.4	86.6	0.0	7.5, 2.25
G: model 1 $\rightarrow$ model 4 at $t = 513$	2.2	97.8	0.0	7.5, 2.25
H: model 1 $\rightarrow$ model 5 at $t = 257$	16.0	84.0	0.0	7.5, 2.25
I: model 1 $\rightarrow$ model 5 at $t = 513$	9.0	91.0	0.0	7.5, 2.25

The length of the realizations is $n = 1024$.

For each scenario, Auto-Seg was applied to 500 realizations. The relative frequencies of the number of breaks estimated by Auto-Seg are displayed in Table VI. As in the GARCH case, the performance of Auto-Seg improves when the change in variance between the pieces increases. A noteworthy comment from Table VI is that the performance of Auto-Seg can vary for two scenarios when the change in variances are the same. For example, the change in variances of scenarios B and D are the same; however, the performance of Auto-Seg is better for scenario D. In addition, the detection rate depends on the location of the break; e.g. scenarios F and G have similar parameter values except for the locations, which are at 257 and 513 respectively. The fact that for scenario A with no break, an incorrect break was never found using Auto-Seg is remarkable. The detection rates for the scenarios with one break in this table vary from 66.0% to 97.8%. Taking into consideration that for all the scenarios the sample size (1024) is not large, the performance of Auto-Seg for these scenarios is good.

As an illustration, in Figure 6, we show typical realizations of scenarios C (top panel) and H (bottom panel). In this figure, for the realization of scenario C using Auto-Seg, a break was found at location 520 (dashed vertical line) and for that of scenario H a break at 256 (dashed vertical line). As in the case for the SVM, we computed two estimates of the posterior mode of the vector of states. In the top panel of Figure 7, the estimated modes were obtained when the realization of scenario C is not segmented. The estimated modes shown in the second panel of this

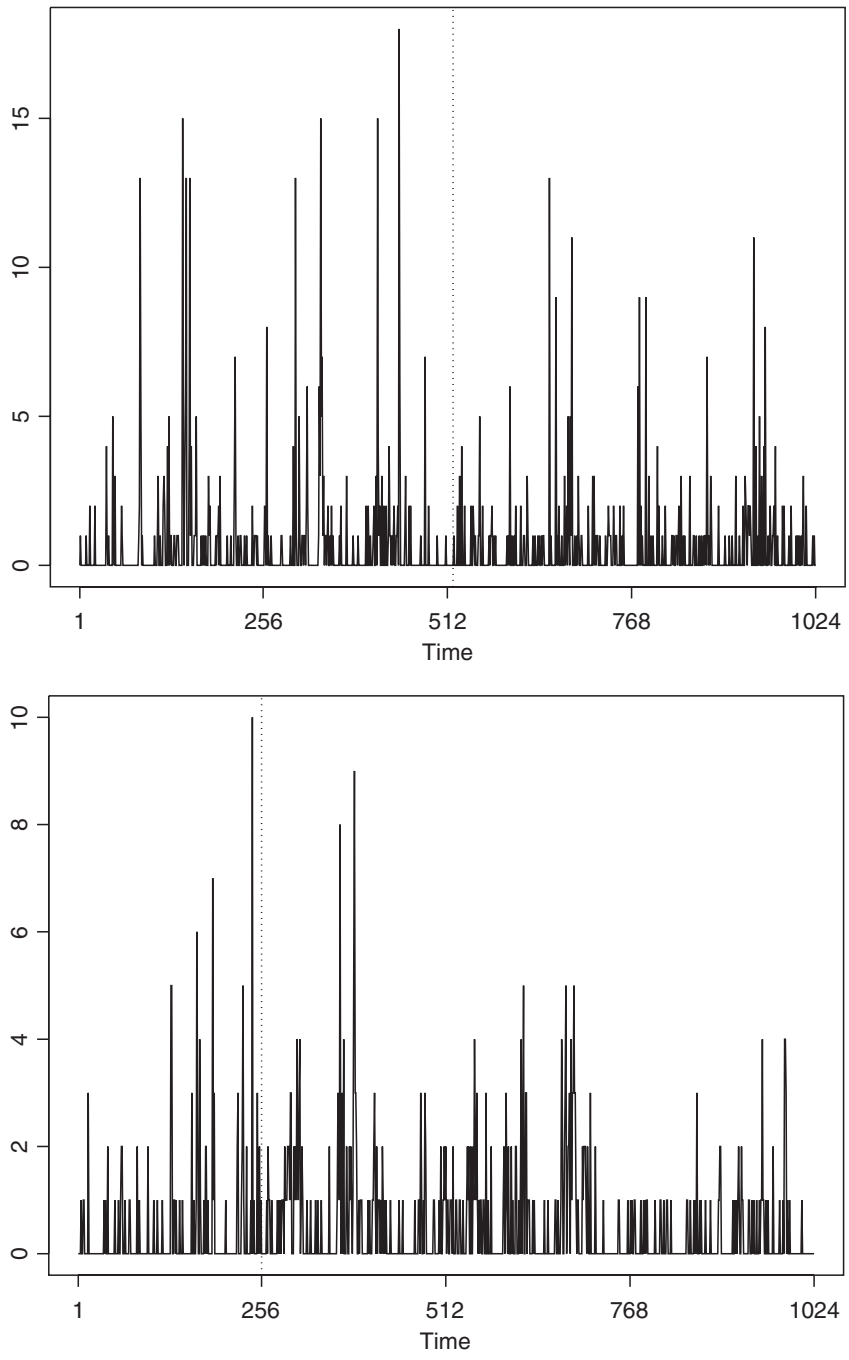

FIGURE 6. Realizations from the piecewise Poisson processes C and H respectively (defined in the first column of Table VI). Vertical dotted lines are break points found by Auto-Seg.

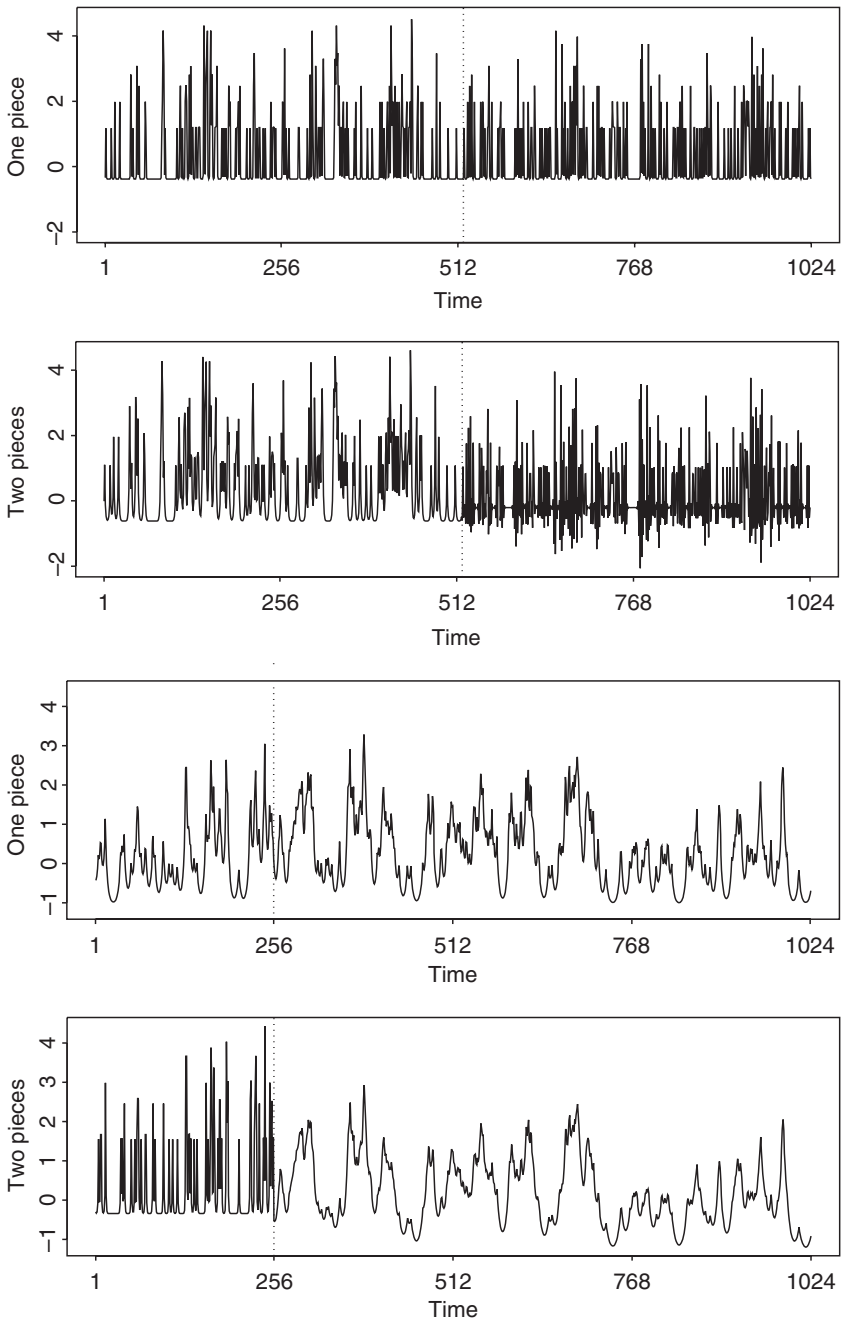

FIGURE 7. Posterior mode of the realization of scenario C shown in Figure 6 under the assumption of no break (first panel) and using the break (second panel) found by Auto-Seg. The last two panels are the analogous plots for the realization of scenario H.

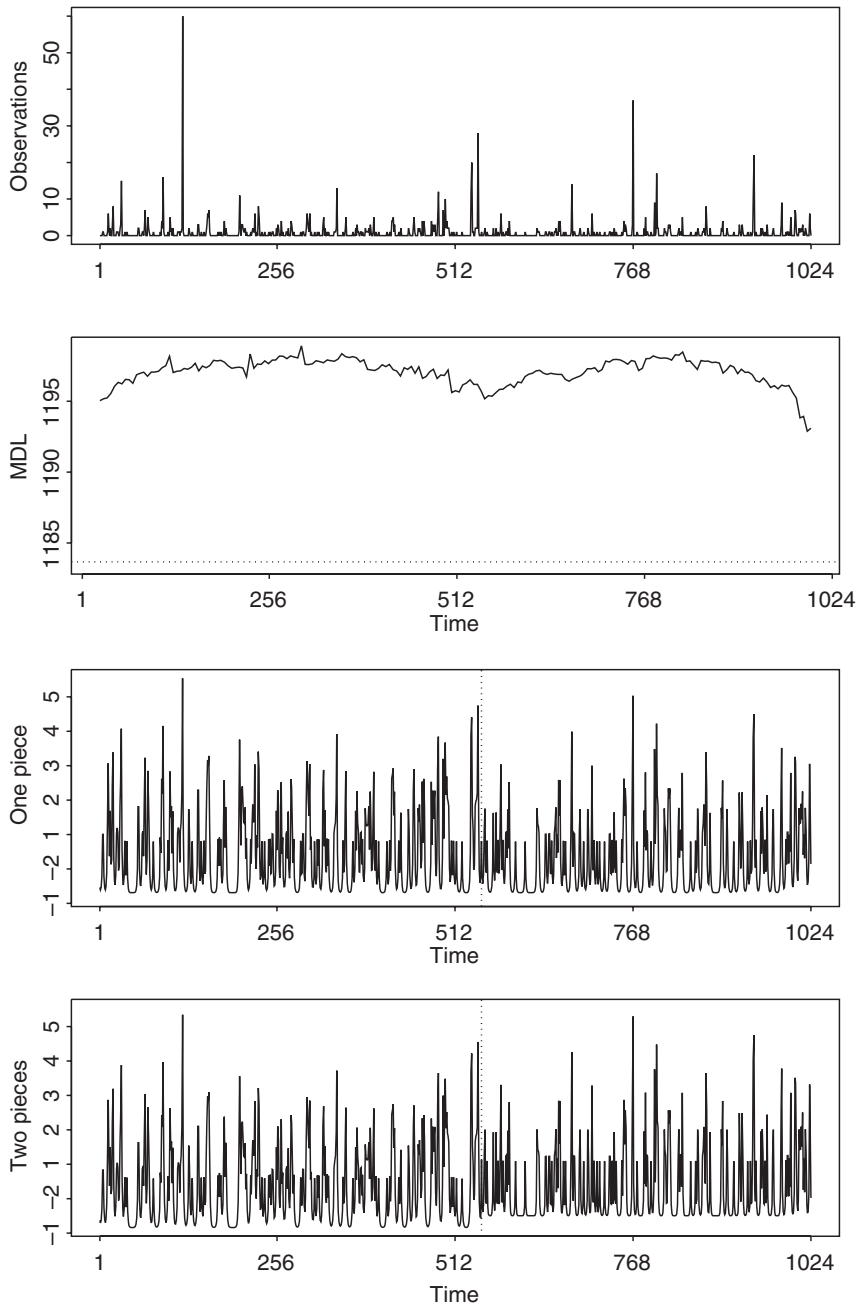

FIGURE 8. First panel: a typical realization of scenario A defined in Table VI. Second panel: two-piece MDL computed in a grid of points (solid line) and one-piece MDL (horizontal dashed line). Third panel: estimated posterior mode based on a single piece. Fourth panel: estimated posterior mode based on two pieces with break at location 550.

figure were obtained using the two-piecewise PDM found by Auto-Seg (i.e. with a break at 520). In both panels, the Auto-Seg break is shown as the vertical dotted line. The plots in the last two panels are the analogous modes for the realization of scenario H for which using Auto-Seg a break was found at location 256.

From Figure 7, we notice that for the realization of scenario C the estimated modes of the vector of states does not differ too much. That is not the case for the realization of scenario H. In this case (last two panels), the mode of the first piece is underestimated when no breaks are considered. Notice that the modes of the second piece look much similar.

We include now the case when there is no break in the underlying scenario. In the first panel of Figure 8 a realization of scenario A defined in Table VI is shown. Using Auto-Seg no break was found for this realization. Now, the MDL values at the break with location at time t, $t = 25, 30, \ldots, 995$, was computed. These MDL values are shown as the solid line in the second panel of Figure 8. Ignoring the last few MDL values on the right, the MDL value in this grid is 1195.172 at 550. The horizontal dashed line in this panel shows the MDL with no breaks (1183.677). In the third panel, the estimated posterior mode of the vector of states based on a single piece is shown. In the last panel, we show the estimate of the posterior mode based on two PDM scenarios having a break at the minimizer of the two-piece MDL values shown in the second panel of this figure. Notice that the two sets of estimates agree.

Summary statistics for the replicates of the two-piecewise models from Table VI for which using Auto-Seg one break was found are given in Table VII. In general, in Table VII the mean of the fitted breaks are close to the true value. The increase in change in variances tends to decrease the standard error of the locations of the breaks; e.g. the standard errors of the breaks of scenarios C and G are 49.1 and 34.9 respectively.

For illustration purposes, we obtain the densities of the estimated breaks of scenarios B and F. The variances change from 7.5 to 4.5 for the first scenario and from 7.5 to 2.25 for the second scenario. In Figure 9, the estimated densities are shown as a dotted line for the density of the breaks of scenario B and as a solid line for the density of the breaks of scenario F.

TABLE VII

SUMMARY STATISTICS FOR THE DISTRIBUTION OF THE ESTIMATED LOCATION OF BREAK POINTS OF THOSE REPLICATIONS WITH ONLY ONE BREAK FOR THE MODELS GIVEN IN TABLE VI

Scenario	Mean	SE	Min	Q1	Med	Q3	Max	Freq
Unconditional variance decreases from 7.5 to 4.5								
B	245.6	40.9	107	222	240	260.0	419	66.0
C	505.9	49.1	213	487	515	526.0	772	88.4
D	265.0	66.9	117	226	246	305.0	881	69.0
E	520.8	67.7	312	491	520	540.3	905	83.2
Unconditional variance decreases from 7.5 to 2.25								
F	250.1	40.2	100	229	246	258.0	571	86.6
G	509.6	34.9	318	501	516	528.0	641	97.8
H	265.8	67.0	103	224	249	308.0	747	84.0
I	522.1	60.6	136	509	522	541.0	857	91.0

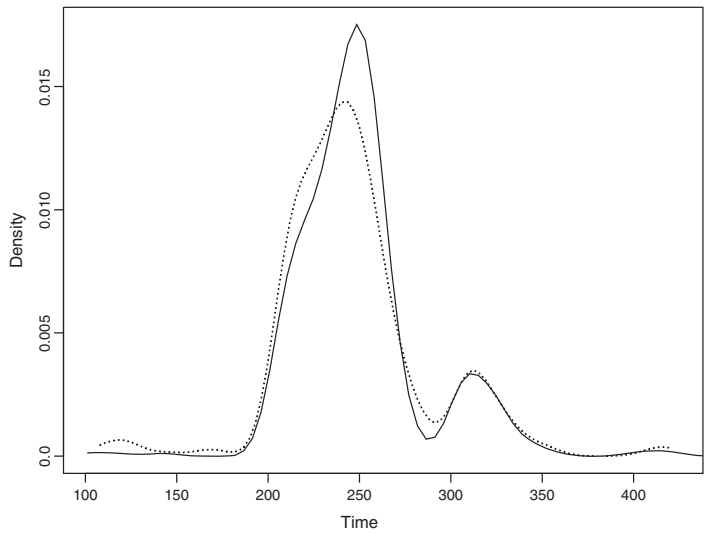

FIGURE 9. Estimated densities of the locations of the breaks of scenarios B (dotted line) and F (solid line).

TABLE VIII

PARAMETER ESTIMATES FOR THE TWO-PIECEWISE POISSON MODELS FROM TABLE VI

		Piece 1			Piece 2			MDL	
Scenario		β	ϕ	σ^2	β	ϕ	σ^2	Two-piece	One-piece
Unconditional variance decreases from 7.5 to 4.5									
B	True	−1.5702	0.50	1.9237	−1.3061	−0.50	1.5277		
	Mean	−1.6535	0.45	2.1635	−1.3918	−0.46	1.8113	1122.37	1130.12
	SD	0.3547	0.12	0.6467	0.1154	0.06	0.2854		
C	True	−1.5702	0.50	1.9237	−1.3061	−0.50	1.5277		
	Mean	−1.6442	0.41	2.1400	−1.3878	−0.46	1.8067	1107.32	1117.21
	SD	0.2327	0.10	0.4576	0.1472	0.07	0.3395		
D	True	−1.5702	0.50	1.9237	−1.3061	0.90	0.3870		
	Mean	−1.7020	0.32	2.3957	−1.2580	0.90	0.3692	1046.14	1053.28
	SD	0.3572	0.15	0.6200	0.2578	0.03	0.0833		
E	True	−1.5702	0.50	1.9237	−1.3061	0.90	0.3870		
	Mean	−1.6641	0.37	2.2258	−1.2717	0.90	0.3667	1052.57	1061.99
	SD	0.2480	0.11	0.4715	0.3246	0.03	0.0979		
Unconditional variance decreases from 7.5 to 2.25									
F	True	−1.5702	0.50	1.9237	−0.9373	−0.50	0.9745		
	Mean	−1.7097	0.40	2.2389	−0.9866	−0.47	1.1224	1166.49	1176.60
	SD	0.3437	0.15	0.6569	0.0875	0.06	0.1626		
G	True	−1.5702	0.50	1.9237	−0.9373	−0.50	0.9745		
	Mean	−1.6528	0.39	2.1683	−0.9875	−0.47	1.1115	1137.65	1151.41
	SD	0.2291	0.11	0.4589	0.1105	0.08	0.2135		
H	True	−1.5702	0.50	1.9237	−0.9373	0.90	0.2469		
	Mean	−1.6967	0.35	2.3187	−0.9309	0.89	0.2453	1092.90	1102.29
	SD	0.3521	0.16	0.6283	0.2037	0.03	0.0571		
I	True	−1.5702	0.50	1.9237	−0.9373	0.90	0.2469		
	Mean	−1.6633	0.38	2.2065	−0.9171	0.89	0.2419	1089.66	1101.82
	SD	0.2336	0.11	0.4577	0.2483	0.04	0.0743		

Notice that both densities are multimodal in spite of the fact that scenario F has a large change in variance between the pieces. We believe that the multimodality in these densities is due to the small sample size of the realizations of the process.

Now, let us consider the two-piecewise models from Table VI. For those realizations for which using Auto-Seg exactly one break was found, the parameters of each piece were estimated. The average and standard error of these estimates are shown in Table VIII. In addition, last two columns in this table the average of the minimized MDL values and the average of the MDL values obtained when there are no breaks assumed.

In general, the estimates are slightly biased. This is true for the state-space Poisson model with no structural change even when the Monte Carlo approximation of the likelihood is used to estimate the parameters of this model (see, for example Sandman and Koopman, 1998; Davis and Rodriguez-Yam, 2005).

5. AN APPLICATION

The Auto-Seg procedure was applied to analyse change points in the *Standard and Poors 500 index* (S&P 500) over the period 4 January 1989 to 19 October

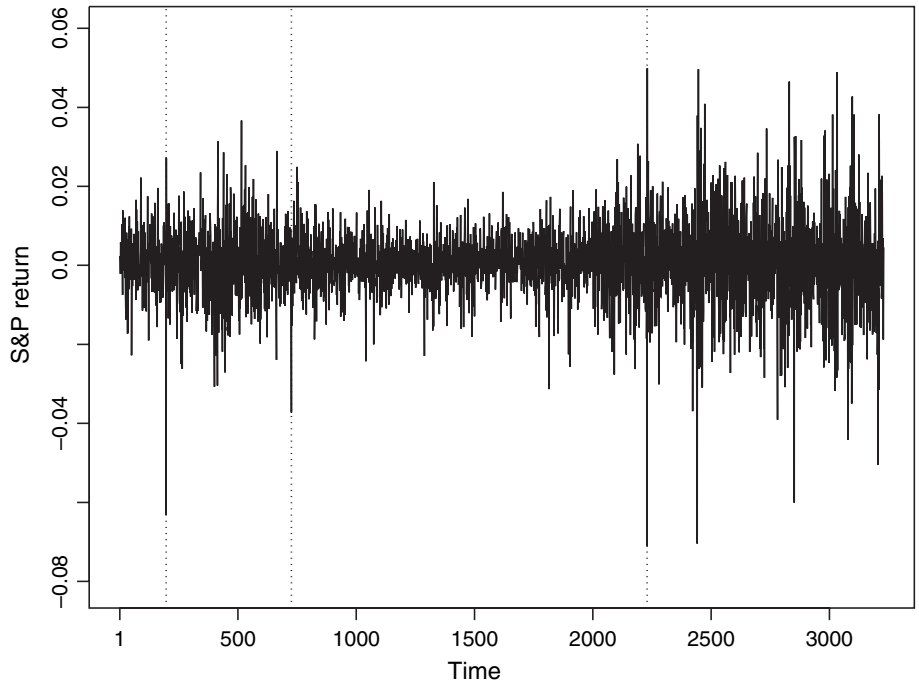

FIGURE 10. Log returns of the S&P index over the period 4 January 1989 to 19 October 2001. The dotted vertical lines are the breaks found by Auto-Seg.

2001 at daily frequency. This stock market series was also analysed by Andreou and Ghysels (2002) during this same period. They were interested in studying the impact, if any, on the Asian and Russian financial crises which started in July 1997 and continued through 1998. This section of S&P 500 consists of 3230 observations. The time series of log returns r_t for this data is shown in Figure 10. Applying Auto-Seg to the log returns series using a segmented GARCH(1,1) model, four segments were found with break locations at 197, 726

TABLE IX

BREAKS OF THE S&P 500 INDEX

	Process	Selection criterion	Number and location of breaks	
Auto-Seg	r_t	MDL	3	13/10/89, 15/11/91, 27/10/97
AG	$\|r_t\|$	BIC	3	27/12/91, 5/1/96, 28/7/98
		LWZ	2	20/8/91, 3/2/97
	$(r_t)^2$	BIC	1	14/10/97
		LWZ	1	14/10/97

The AG values are taken from table VII of Andreou and Ghysels (2002). Auto-Seg: best piece-wise GARCH(1,1) process found by Auto-Seg.

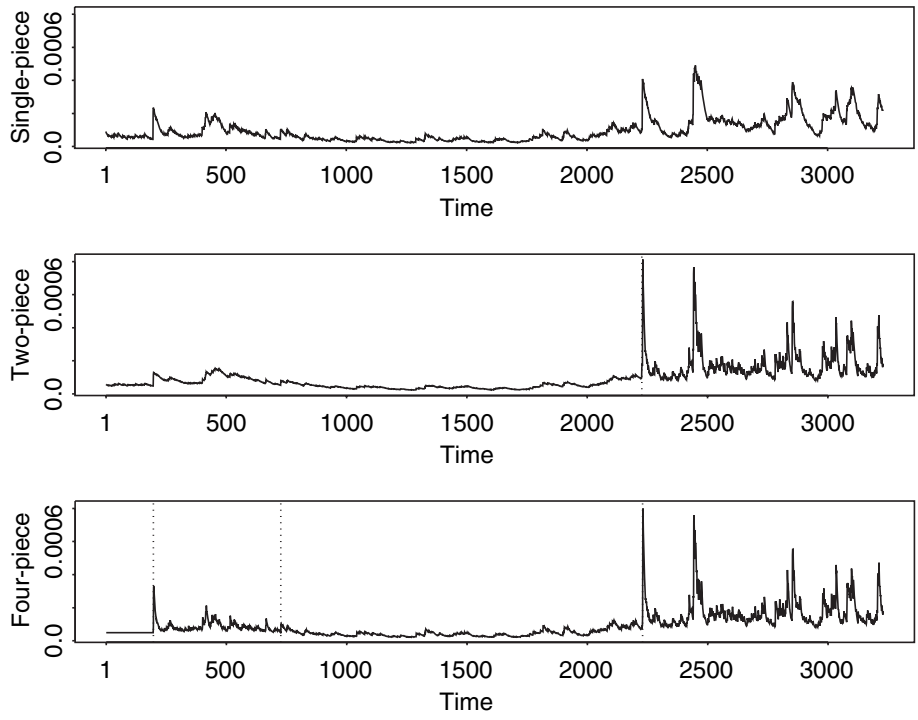

FIGURE 11. Estimated volatilities of the log returns of the S&P 500 series. Top: estimated volatilities under no breaks. Middle panel: volatilities when a break in 27/10/97 is assumed. Bottom panel: estimated volatilities based on the Auto-Seg breaks.

and 2229, which are shown as the vertical dotted lines in Figure 10. In Table IX, we show the breaks found by Andreou and Ghysels (2002) when the Lavielle and Moulines procedure is applied to the absolute and squared returns using the BIC and LWZ. The latter is a modified BIC proposed in Liu *et al.* (1997). In Table IX, the last break found by Auto-Seg is in close agreement with the single break found by Andreou and Ghysels (2002) when squared returns are used in the Lavielle and Moulines procedure. In Figure 11, three sets of volatilities are shown. In the top panel the volatilities were obtained by fitting a single GARCH(1,1) model to the log returns of the S&P 500 series. In the middle panel, the volatilities were obtained fitting a model based on a break at 27/10/97 that is close to the single break found by Andreou and Ghysels (2002). In the bottom panel, the volatilities were obtained using the Auto-Seg breaks. Notice in Figure 11 that the single-break volatilities (middle panel) resemble the estimated volatilities based on Auto-Seg (bottom panel). As a reference, the MDL values of the fitted models in this figure are $-10{,}688$, $-10{,}752$ and $-10{,}705$ respectively. As expected, the difference between the best Auto-Seg MDL and the single-piece MDL is much higher than between the best Auto-Seg MDL and the single-break MDL model.

6. CONCLUSION

In this article, we considered the problem of partitioning a nonstationary time series into segments of stationary series. The original series was assumed to follow a parametric nonlinear time-series model in which the parameters may change values at fixed times. Particular attention was given to the GARCH models, SVMs and the generalized SSMs. The MDL principle was adopted to simultaneously estimate the number of segments, the locations of break points and the parametric model in each segment. A new GA was developed to solve the hard optimization problem involved. Numerical experiments were conducted to demonstrate the good performances of the proposed approach.

APPENDIX

This appendix derives eqn (10) from eqn (9). First, we define that $\mathrm{CL}(\boldsymbol{\zeta}_j) = 0$ if there is no integer parameter in $\boldsymbol{\theta}_j$. Moreover, recall that c_j and d_j are the lengths of $\boldsymbol{\zeta}_j$ and $\boldsymbol{\psi}_j$ respectively. Now, in general, approximately $\log_2 I$, bits are needed to encode an integer I whose value is not bounded above. Thus,

$$\mathrm{CL}(m) = \log_2 m \quad \text{and} \quad \mathrm{CL}(\boldsymbol{\zeta}_j) = \sum_{k=1}^{c_j} \log_2 \zeta_{kj},$$

where ζ_{kj} is the kth entry of $\boldsymbol{\zeta}_j$. On the other hand, if the upper bound, say I_{U}, of I is known, approximately $\log_2 I_{\mathrm{U}}$ bits are required. Since all n_js are bounded by n,

$$\mathrm{CL}(n_j) = \log_2 n$$

for all j. To calculate $\mathrm{CL}(\hat{\boldsymbol{\psi}}_j)$, we use the following result of Rissanen: a maximum likelihood estimate of a real parameter computed from N observations can be effectively encoded with $\frac{1}{2}\log_2 N$ bits. Since each of the d_j parameters of $\hat{\boldsymbol{\psi}}_j$ is computed from n_j observations, for all j we have

$$\mathrm{CL}(\hat{\boldsymbol{\psi}}_j) = \frac{d_j}{2}\log_2 n_j.$$

Substituting these expressions back into eqn (9) we obtain eqn (10).

ACKNOWLEDGEMENTS

The authors are grateful to the reviewer for a set of most useful comments, many of which are incorporated in the current version of the article. The work of Davis is partially supported by NSF grants DMS 0308109 and 0743459, and EPA STAR grant CR-829095. The work of Lee is partially supported by NSF grants 0203901 and 0707037, and HKRGC Competitive Earmarked Research Grant 401507.

NOTE

Corresponding author: Department of Statistics, The Chinese University of Hong Kong, Shatin, N. T., Hong Kong. Tel.: +852 2609 7941; Fax: +852 2603 5188; E-mail: tlee@sta.cuhk.edu.hk

REFERENCES

Alba, E. and Troya, J. M. (1999) A survey of parallel distributed genetic algorithm. *Complexity* 4, 31–52.

Alba, E. and Troya, J. M. (2002) Improving flexibility and efficiency by adding parallelism to genetic algorithms. *Statistics and Computing* 12, 91–114.

Andreou, E. and Ghysels, E. (2002) Detecting multiple breaks in financial market volatility dynamics. *Journal of Applied Econometrics* 17, 579–600.

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

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

Berkes, I., Gombay, E., Horváth, L. and Kokoszka, P. (2004) Sequential change-point detection in GARCH(p,q) Models. *Econometric Theory* 20, 1140–67.

Bhattacharya, P. K. (1994) Some aspects of change-point analysis. In *Change-point Problems* (eds E. Carlstein, H.-G. Müller and D. Siegmund). Institute of Mathematical Statistics. Lecture Notes–Monograph Series 23, 28–56.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–27.

BROCKWELL, P. J. and DAVIS, R. A. (1991) *Time Series: Theory and Methods*. New York: Springer-Verlag.
CHATTERJEE, S., LAUDATO, M. and LYNCH, L. (1996) Genetic algorithms and their statistical applications: an introduction. *Computational Statistics & Data Analysis* 22, 633–51.
CSÖRGŐ, M. and HORVÁTH, L. (1997) *Limit Theorems in Change-Point Analysis*. Wiley: New York.
DAVIS, R. A. and RODRIGUEZ-YAM, G. A. (2005) Estimation for state-space models based on a likelihood approximation. *Statistica Sinica* 15, 381–406.
DAVIS, R. A., HUANG, D. and YAO, Y.-C. (1995) Testing for a change in the parameter values and order of an autoregressive model. *The Annals of Statistics* 23, 282–304.
DAVIS, R. A., DUNSMUIR, W. T. M. and WANG, Y. (1998) Modelling time series of count data. In: *Asymptotics, Nonparametrics and Time Series* (ed. S. GHOSH). New York: Marcel Dekker, pp. 63–112.
DAVIS, R. A., LEE, T. C. M. and RODRIGUEZ-YAM, G. A. (2006) Structural breaks estimation for non-stationary time series models. *Journal of the American Statistical Association* 101, 223–39.
DE JONG, K. (1975) *An Analysis of the Behaviour of a Class of Genetic Adaptive Systems*. Doctoral Thesis, Department of Computer and Communication Sciences, University of Michigan.
DJURIĆ, P. M. (1994) A MAP solution to off-line segmentation of signals. In: *1994 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 4. New York: IEEE, pp. 505–8.
DURBIN, J. and KOOPMAN, S. J. (1997) Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika* 84, 669–84.
ESHELMAN, L. J. (2000) Genetic algorithms. In: *Evolutionary Computation*, Vol. 1, *Basic Algorithms and Operators* (eds T. BÄCK, D. B. FOGEL and T. MICHALEWICZ). Philadelphia, PA: Bristol, pp. 64–80.
FEARNHEAD, P. (2005) Exact Bayesian curve fitting and signal segmentation. *IEEE Transactions on Signal Processing* 53, 2160–6.
FORREST, S. (1991) *Emergent Computation*. Cambridge, MA: MIT Press.
GAETAN, C. (2000) Subset ARMA model identification using genetic algorithms. *Journal of Time Series Analysis* 21(5), 559–70.
GUSTAFSSON, F. (2000) *Adaptive Filtering and Change Detection*. New York: Wiley.
HANSEN, M. H. and YU, B. (2001) Model selection and the principle of minimum description length. *Journal of the American Statistical Association* 96, 746–74.
HARVEY, A. C. and FERNANDES, C. (1989) Time series models for count or qualitative observations. *Journal of the American Statistical Association* 7, 407–17.
HOLLAND, J. (1992) *Adaptation in Natural and Artificial Systems*. Cambridge, MA: MIT Press.
KITAGAWA, G. and AKAIKE, H. (1978) A procedure for the modeling of non-stationary time series. *Annals of the Institute of Statistical Mathematics Part B* 30, 351–63.
KITAGAWA, G., TAKANAMI, T. and MATSUMOTO, N. (2001) Signal extraction problems in seismology. *International Statistical Review* 69, 129–52.
KRISHNAIAH, P. R. and MIAO, B. Q. (1988) Review about estimation of change points. In: *Handbook of Statistics 7* (eds P. R. KRISHNAIAH and C. R. RAO). New York: North Holland, pp. 375–402.
KUK, A. Y. (1999) The use of approximating models in monte carlo maximum likelihood estimation. *Statistics & Probability Letters* 45, 325–33.
LAVIELLE, M. (1998) Optimal segmentation of random processes. *IEEE Transactions on Signal Processing* 46, 1365–73.
LAVIELLE, M. and MOULINES, E. (2000) Least-squares estimation of an unknown number of shifts in a time series. *Journal of Time Series Analysis* 21, 33–59.
LEE, C.-B. (1995) Estimating the number of change points in a sequence of independent normal random variables. *Statistics & Probability Letters* 25, 241–8.
LEE, T. C. M. (2001) An introduction to coding theory and the two-part minimum description length principle. *International Statistical Review* 69, 169–183.
LEE, T. C. M. (2002) Automatic smoothing for discontinuous regression functions. *Statistica Sinica* 12, 823–42.
LEE, S.-W. and HANSEN, B. E. (1994) Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory* 10, 29–52.
LIU, J., WU, S. and ZIDEK, J. (1997) On segmented multivariate regression. *Statistica Sinica* 7, 497–525.
MARTIN, W. N., LIENIG, J. and COHOON, J. P. (2000) Island (migration) models: evolutionary algorithm based on punctuated equilibria. In *Evolutionary Computation, Vol. 2, Advanced Algorithms and Operators* (eds T. BÄCK, D. B. FOGEL and T. MICHALEWICZ). Philadelphia, PA: Bristol, pp. 101–24.

McCulloch, R. E. and Tsay, R. S. (1993) Bayesian inference and prediction for mean and variance shifts in autoregressive time series. *Journal of the American Statistical Association* 88, 968–78.
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, 543–60.
Picard, D. (1985) Testing and estimating change-points in time series. *Advances in Applied Probability* 17, 841–67.
Punskaya, E., Andrieu, C., Doucet, A. and Fitzgerald, W. J. (2002) Bayesian curve fitting using MCMC with applications to signal segmentation. *IEEE Transactions on Signal Processing* 50, 747–58.
Rissanen, J. (1989) *Stochastic Complexity in Statistical Inquiry*. Singapore: World Scientific.
Sandmann, G. and Koopman, S. J. (1998) Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *Journal of Econometrics* 87, 271–301.
Shaban, S. A. (1980) Change point problem and two-phase regression: an annotated bibliography. *International Statistics Review* 48, 83–93.
Yao, Y.-C. (1988), Estimating the number of change-points via schwarz criterion. *Statistics & Probability Letters* 6, 181–9.
Zacks, S. (1983) Survey of classical and Bayesian approaches to the change-point problem: fixed sample and sequential procedures of testing and estimation. In: *Recent Advances in Statistics: Papers in Honor of Herman Chernoff on His Sixtieth Birthday* (eds M. H. Rivzi, J. S. Rustagi and D. Sigmund). New York: Academic Press, pp. 245–69.
Zeger, S. L. (1988) A regression model for time series of counts. *Biometrika* 75, 621–9.