Break Detection for a Class of Nonlinear Time Series Models

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

Department of Statistics, Colorado State University, Fort Collins, Colorado

June 12, 2005

Abstract

This paper considers the problem of detecting break points for a broad class of non-stationary time series models. In this formulation, the number and locations of the break points are assumed unknown. Each segment is assumed to be modeled from a class of parametric time series models for stationary processes. The minimum description length (MDL) principle is used as a criterion for estimating the number of break points, the location of break points, and the parametric model in each segment. The best segmentation found by minimizing the MDL criterion is obtained using a genetic algorithm. The implementation of this approach is illustrated using GARCH, stochastic volatility, and generalized state-space models as the parametric model for the segments. Empirical results show the good performance of the estimates of the number of breaks and their locations for these various models.

KEY WORDS: GARCH, genetic algorithm, minimum description length principle, model selection, multiple change points, non-stationary time series, state-space models, stochastic volatility model.

1 Introduction

In this paper the problem of modeling a class of non-stationary time series by segmenting the series into different stationary processes is considered. 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.

Testing for 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 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, Huang and Yao (1995) and Kitagawa, Takanami and Matsumoko (2001) studied the single change point problem in which the pieces are assumed to be autoregressive (AR) processes. Here a change occurs if one of the AR parameters, including the constant term, or white noise variance changes. Tong's threshold models (see Tong, 1990) also include AR models with changes in parameter values. However, these changes are triggered by lagged values of time series rather than at specified time points.

Multiple change points are considered in Kitagawa and Akaike (1978) and Davis, Lee and Rodriguez-Yam (2005) where the observed non-stationary "linear" series is decomposed into AR processes. A more general piecewise stationary process, for which the piecewise AR process is a particular case, is considered in Ombao, Raz, Von Sachs and Malow (2001). McCulloch and Tsay (1993); Djurić (1994), Lavielle (1998) and Punskaya et al. (2002) follow a Bayesian approach to the change point problem of time series. Csörgő and Horváth (1997) devote a chapter to the change point problem for dependent observations.

In this paper we consider the multiple change point problem for a class of non-stationary processes in which the pieces are modeled by a *specified* parametric class of stationary time series. More precisely, let τ_j, $j = 1, \ldots, m$, denote the breakpoints between the j-th and $(j+1)$-th segments respectively, and set $\tau_0 = 1$ and $\tau_{m+1} = n + 1$. It is assumed that the j-th piece of the time series $\{Y_t\}$ is modeled by a stationary time series $\{X_{t,j}\}$; i.e.,

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

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 parametric space Θ_j with $\boldsymbol{\theta}_j \neq \boldsymbol{\theta}_{j+1}$, $j = 1, \ldots, m$. The following examples illustrate this formulation.

Example 1 (*Segmented AR process*) Consider the case when $\{X_{t,j}\}$ is the AR(p_j) process

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

where the noise sequence $\{\varepsilon_t\}$ is iid N(0, 1). 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 setup has been considered by Kitagawa and Akaike (1978) and Davis, et al. (2005). 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 j-th piece of the process $\{Y_t\}$ is modeled as a generalized autoregressive conditionally heteroscedastic (GARCH) process introduced by Bollerslev (1986); i.e.,

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

where for each j, $\{X_{t,j}\}$ is the GARCH(p_j, q_j) model:

$$X_{tj} = \sigma_{tj}\varepsilon_t.$$

In the above $\{\varepsilon_t\}$ is iid N(0, 1) and σ_{tj} is a positive function of X_{tj} given by

$$\sigma_{tj}^2 = \alpha_{0,j} + \alpha_{j1}X_{t-1,j}^2 + \ldots + \alpha_{j,p_j}X_{t-p_j,j}^2 + \beta_{j1}\sigma_{t-1,j}^2 + \ldots + \beta_{j,q_j}\sigma_{t-q_j,j}^2, \quad \tau_{j-1} \le t < \tau_j, \tag{3}$$

subject to the constraints $\alpha_{0,j} > 0$, $\alpha_{i,j} \ge 0$, $i = 1, \ldots, m+1$, and $\alpha_{1,j}+\ldots+\alpha_{q_j,j}+\beta_{1,j}+\ldots+\beta_{q_j,j} < 1$. 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 vectors of α_j's and β_j's in (3), respectively.

Example 3 (*State space model*) The j-th piece of the time series $\{Y_t\}$ is modeled 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

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

where for each j, $\{X_{tj}\}$ is the AR(p_j) process in (2). Assuming the order p_j is unknown, the vector of parameters becomes $\boldsymbol{\theta}_j = (\boldsymbol{\delta}_j,\ \boldsymbol{\phi}_j,\ \sigma_j^2)$, where $\boldsymbol{\delta}_j$ is the vector of say q_j parameters associated

with the specification of $p(y_t|\alpha_t)$, $\tau_{j-1} \leq t < \tau_j$, and $\boldsymbol{\phi}_j$ is the vector of ϕ_j's associated with the AR model in (2).

Two state-space models considered in this paper are the stochastic volatility model (SVM) and the Poisson driven model (PDM); which belong to the *exponential family of distributions*. Durbin and Koopman (1997) and Kuk (1999) consider the following form for this family

$$p(y_t|\alpha_t) = e^{(\mathbf{z}_t^T\boldsymbol{\beta}+\alpha_t)y_t - b(\mathbf{z}_t^T\boldsymbol{\beta}+\alpha_t)+c(y_t)}, \tag{5}$$

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.

In this paper we focus on Examples 2 and 3; a more thorough treatment of Example 1 was given in Davis, et al. (2005). The SVM and GARCH are popular models for analyzing log returns of financial time series. The PDM is a frequently used model for time series of counts. For example, Zeger (1988); Harvey and Fernandes (1989) and Davis, Dunsmuir and Wang (1998) have used these models for modeling counts of individuals infected by a rare disease. Unlike Examples 1 and 2, the likelihood of the SVM and PDM models do not have a closed form expression, which makes the estimation of break points for these models computationally challenging.

The problem of finding a "best" combination of m, τ_j's 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 has been used in the literature for the change point problem. For example, Kitagawa and Akaike (1978) and Kitagawa, et al. (2001) used the AIC criterion; Yao (1988) used the Bayesian information criterion (BIC); and Lee (1995) and Liu, Wu and Zidek (1997) used modified versions of BIC. More recently Bai and Perron (1998, 2003) considered criteria based on squared-residuals, Lavielle (1998) and Gustaffson (2000) used maximum a posteriori (MAP) criterion; and Davis, et al. (2005) used the minimum description length (MDL) principle of Rissanen (1989).

In this paper we adopt the MDL principle. For even moderate values of n, optimization of this criterion is not easy task. To solve this optimization problem we develop a genetic algorithm (GA) to find nearly optimal values.

The rest of this paper 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 genetic algorithm 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 series.

2 Model Selection

Denote by $\mathcal{M}$ the class of piecewise processes defined in (1). In this section we find the code length of data associated with members $\mathcal{F} \in \mathcal{M}$. If $CL_{\mathcal{F}}(z)$ denote the code length of an object z associated with model $\mathcal{F}$, then by the two-part description length method of Rissanen (1989) (see also Lee, 2001) it is given by

$$CL_{\mathcal{F}}(\mathbf{y}) = CL_{\mathcal{F}}(\hat{\mathcal{F}}) + CL_{\mathcal{F}}(\hat{\mathbf{e}}|\hat{\mathcal{F}}),$$

where $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ is the observed data, $CL_{\mathcal{F}}(\hat{\mathcal{F}})$ denotes the code length of the fitted model $\hat{\mathcal{F}}$ and $CL_{\mathcal{F}}(\hat{\mathbf{e}}|\hat{\mathcal{F}})$ is the code length of the corresponding residuals (conditional on the fitted model $\hat{\mathcal{F}}$). The MDL criterion selects the model that yields the minimum length of a code used to encode the observed data $\mathbf{y}$.

Recall that $\boldsymbol{\theta}_j$ is the vector of all model parameters in the j-th piece. It is convenient to partition $\boldsymbol{\theta}_j$ as $\boldsymbol{\theta}_j = (\boldsymbol{\zeta}_j, \boldsymbol{\psi}_j)$, where $\boldsymbol{\zeta}_j$ and $\boldsymbol{\psi}_j$ are the integer-valued and real-valued parameters in $\boldsymbol{\theta}_j$, respectively. For example, for the piecewise AR models in Example 1, $\boldsymbol{\zeta}_j$ denotes the AR order of the j-th piece while $\boldsymbol{\psi}_j$ denotes the corresponding AR coefficients. We assume that once $\boldsymbol{\zeta}_j$ is specified, $\boldsymbol{\psi}_j$ can be estimated via maximum likelihood estimation. The resulting estimate will be denoted as $\hat{\boldsymbol{\psi}}_j$. Finally let c_j and d_j be the lengths of the vectors $\boldsymbol{\zeta}_j$ and $\boldsymbol{\psi}_j$, respectively.

Since $\hat{\mathcal{F}}$ is composed of m, τ_j's, $\boldsymbol{\zeta}_j$'s and $\hat{\boldsymbol{\psi}}_j$'s, we further decompose $CL_{\mathcal{F}}(\hat{\mathcal{F}})$ into

$$\begin{aligned} CL_{\mathcal{F}}(\hat{\mathcal{F}}) = {} & CL_{\mathcal{F}}(m) + CL_{\mathcal{F}}(\tau_1, \ldots, \tau_m) \\ & + CL_{\mathcal{F}}(\boldsymbol{\zeta}_1) + \ldots + CL_{\mathcal{F}}(\boldsymbol{\zeta}_{m+1}) + CL_{\mathcal{F}}(\hat{\boldsymbol{\psi}}_1) + \ldots + CL_{\mathcal{F}}(\hat{\boldsymbol{\psi}}_{m+1}). \end{aligned}$$

Let $n_j := \tau_j - \tau_{j-1}$ denote the number of observations in the j-th segment of $\hat{\mathcal{F}}$. Notice that complete knowledge of $(\tau_1, \ldots, \tau_m)$ implies complete knowledge of $(n_1, \ldots, n_{m+1})$, and vice versa,

hence $CL_{\mathcal{F}}(\tau_1, \ldots, \tau_m) = CL_{\mathcal{F}}(n_1, \ldots, n_{m+1})$. In general, to encode an integer I whose value is not upper bounded, approximately $\log_2 I$ bits are needed. Thus $CL_{\mathcal{F}}(m) = \log_2 m$ and $CL_{\mathcal{F}}(\boldsymbol{\zeta}_j) = \sum_{k=1}^{c_j} \log_2 \zeta_{kj}$, where ζ_{kj} is the k-th entry of $\boldsymbol{\zeta}_j$. If there is no integer parameter in $\boldsymbol{\theta}_j$ we define $CL_{\mathcal{F}}(\boldsymbol{\zeta}_j) := 0$. On the other hand, if the upper bound, say I_U, of I is known, approximately $\log_2 I_U$ bits are required. Since all n_j's are bounded by n, $CL_{\mathcal{F}}(n_j) = \log_2 n$ for all j. To calculate $CL_{\mathcal{F}}(\hat{\boldsymbol{\psi}}_j)$, 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,

$$CL_{\mathcal{F}}(\hat{\boldsymbol{\psi}}_j) = \frac{d_j}{2} \log_2 n_j.$$

Thus, we obtain

$$CL_{\mathcal{F}}(\hat{\mathcal{F}}) = \log_2 m + (m+1)\log_2 n + \sum_{j=1}^{m+1} \sum_{k=1}^{c_j} \log_2 \zeta_{kj} + \sum_{j=1}^{m+1} \frac{d_j}{2} \log_2 n_j.$$

Now, let $L_j(\boldsymbol{\psi}_j; \mathbf{y}_j)$ be the observed likelihood of the j-th piece. As demonstrated by Rissanen (1989), the code length for the residuals $\hat{\mathbf{e}}$ is given by the negative of the log likelihood of the fitted model $\hat{\mathcal{F}}$. This results in the following MDL expression for $CL_{\mathcal{F}}(\mathbf{y})$

$$\begin{aligned} \mathrm{MDL}(m, \tau_1, \ldots, \tau_m, \boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_{m+1}) &= \log m + (m+1)\log n + \sum_{j=1}^{m+1} \sum_{k=1}^{c_j} \log \zeta_{kj} \\ &\quad + \sum_{j=1}^{m+1} \frac{d_j}{2} \log n_j - \sum_{j=1}^{m+1} L(\hat{\boldsymbol{\psi}}_j; \mathbf{y}_j), \end{aligned} \tag{6}$$

where the last summand is obtained from the assumption that the pieces are independent. Notice that in (6) we have changed the logarithm to base e rather than base 2. 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 (6).

Example 4 (*State space model (SSM)*) 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 $\boldsymbol{\psi}_j = (\boldsymbol{\phi}_j, \boldsymbol{\delta}_j, \sigma_j^2)$. Thus, $c_j = 1$, $d_j = p_j + q_j + 2$, where q_j is defined in Example 3, and

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

Now, let $\mathbf{y}_j := (y_t, \ldots, y_{t+n_j-1})$ and $\boldsymbol{\alpha}_j := (\alpha_t, \ldots, \alpha_{t+n_j-1})$, $\tau_{j-1} \leq t < \tau_j$ be the vector of observations and states of the j-th piece of $\hat{\mathcal{F}}$. Also, let $\boldsymbol{\lambda}_j := (\boldsymbol{\phi}, \sigma_j^2)$. The likelihood corresponding to this piece based on the *complete data* $(\mathbf{y}_j, \boldsymbol{\alpha}_j)$ becomes

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

where $\mathbf{V}_j^{-1} := \mathrm{cov}\{\boldsymbol{\alpha}_j\}$, $\boldsymbol{\mu}_j = \gamma_j/(1 - \phi_{1,j} - \ldots - \phi_{p_j,j})\mathbf{1}$ is the vector of means of the state process, and $\mathbf{1}$ is a vector of ones. From (7) it follows that the likelihood of the observed data is given by the product of n_j-fold integrals

$$L(\boldsymbol{\psi}_1, \cdots, \boldsymbol{\psi}_{m+1}; \mathbf{y}) = \prod_{j=1}^{m+1} \int L(\boldsymbol{\psi}_j; \mathbf{y}_j, \boldsymbol{\alpha}_j) d\boldsymbol{\alpha}_j. \tag{7}$$

Except in simple cases, the integrals in (7) cannot be computed explicitly. In this paper we use the approximation $L_a(\boldsymbol{\psi}_j; \mathbf{y}_j)$ 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 neighborhood of the posterior mode of $p(\boldsymbol{\alpha}_j|\mathbf{y}_j; \boldsymbol{\psi}_j)$. To simplify notation, for the j-th piece we "drop" the subindex j that appears in $\mathbf{y}_j$, $\boldsymbol{\alpha}_j$, etc. Now, let $\ell(\boldsymbol{\theta}; \mathbf{y}|\boldsymbol{\alpha}) := \log p(\mathbf{y}|\boldsymbol{\alpha}; \boldsymbol{\theta})$ and $R(\boldsymbol{\alpha}, \boldsymbol{\alpha}^*)$ be the remainder of its second order Taylor series expansion. Also, let $p(\boldsymbol{\alpha}|\mathbf{y}; \boldsymbol{\psi})$ be the posterior distribution of $\boldsymbol{\alpha}$ and let $\boldsymbol{\alpha}^*$ the mode of this distribution. 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 of modes $\boldsymbol{\alpha}^*$ can be found by maximizing the complete likelihood. Davis and Rodriguez-Yam (2005) found the following approximation to the posterior distribution $p(\boldsymbol{\alpha}|\mathbf{y}; \boldsymbol{\psi})$

$$p_a(\boldsymbol{\alpha}|\mathbf{y}; \boldsymbol{\psi}) = \boldsymbol{\phi}(\boldsymbol{\alpha}; \boldsymbol{\alpha}^*, (\mathbf{K}^* + \mathbf{V})^{-1}), \tag{8}$$

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

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

The likelihood then admits the factorization

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

where $\mathrm{Er}_a(\boldsymbol{\psi}) := \int e^{R(\boldsymbol{\alpha}; \boldsymbol{\alpha}^*)} p_a(\boldsymbol{\alpha}|\mathbf{y}; \boldsymbol{\psi}) d\boldsymbol{\alpha}$ and

$$L_a(\boldsymbol{\psi}; \mathbf{y}) := \frac{|\mathbf{V}|^{1/2}}{|\mathbf{K}^* + \mathbf{V}|^{1/2}} e^{h^* - \frac{1}{2}(\boldsymbol{\alpha}^* - \boldsymbol{\mu})^T \mathbf{V} (\boldsymbol{\alpha}^* - \boldsymbol{\mu})}. \tag{9}$$

Here $h^* := \ell(\boldsymbol{\theta}; \mathbf{y}|\boldsymbol{\alpha})|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}^*}$. Ignoring the term $e^{R(\boldsymbol{\alpha};\boldsymbol{\alpha}^*)}$ in $\mathrm{Er}_a(\boldsymbol{\psi})$, an approximation to the likelihood is given by (9). For the SVM and PDM models, the estimates obtained by maximizing this approximation function 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).

Replacing $L(\boldsymbol{\psi}; \mathbf{y})$ with $L_a(\boldsymbol{\psi}; \mathbf{y})$, equation (6) then becomes

$$\begin{aligned} \mathrm{MDL}(m, \tau_1, \ldots, \tau_m, p_1, \ldots, p_{m+1}) &= \log m + (m+1)\log n + \sum_{j=1}^{m+1} \log p_j \\ &+ \sum_{j=1}^{m+1} \frac{p_j + q_j + 2}{2} \log n_j - \sum_{j=1}^{m+1} L_a(\hat{\boldsymbol{\psi}}_j; \mathbf{y}_j), \end{aligned} \tag{10}$$

where $\hat{\boldsymbol{\psi}}_j$ is the optimizer of (9). The best fitting model for $\mathbf{y}$ is then the minimizer of MDL(m, τ_1,..., τ_m, $p_1, \ldots, p_{m+1}$) in (10).

3 Optimization Using the Genetic Algorithm

3.1 Genetic Algorithm

To give an idea of how the genetic algorithm (GA) works for our segmentation problem, we describe how to optimize the MDL in (10) for the state space model from Example 4. The implementation details for other examples are similar.

Even for moderate values of n, the optimization of MDL(m, $\tau_1, \ldots,$ τ_m, $p_1, \ldots,$ p_{m+1}) with respect to m, τ_1, ..., τ_m, p_1, ..., p_{m+1} is not easy. A procedure that we will use to overcome this problem is the GAs, a class of evolutionary algorithms, first proposed by Holland (1975). Broadly speaking GAs are a randomized search technique that mimic natural selection to find the maximum or high values of an objective function. Among others, Chatterjee, et al. (1996), 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 (1975) and Eshelman (2000), we give a brief summary here. An initial population of M chromosomes are selected (usually at random) and to each individual a probability, which can be proportional to its observed fitness, is assigned. Then an offspring is created by mating individuals selected according

to the assigned probabilities. Two typical genetic operators (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 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 paper we implement the *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, Lienig and Cohoon 2000; Alba and Troya 2002). In this paper, we adopt the following migration policy: after every M_i generations, the worst M_N chromosomes from the j-th island are replaced by the best M_N chromosomes from the $(j-1)$-th island, $j = 1, \ldots, NI$. For $j = 1$ the best M_N chromosomes are migrated from the NI-th island. In all of our simulations, the number of islands (NI) was set to either 10 or 20, $M_i = 5$, $M_N = 2$ and a sub-population size of 10 or 20.

3.2 Implementation Details

This section provides details of our implementation of the GA tailored to the piecewise state space model.

Chromosome Representation: The chromosome representation is given by the vector $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)$ of length n with gene values

$$\delta_t = \begin{cases} -1, & \text{if there is no break at time } t, \\ p_j, & \text{the AR order of the } j\text{-th piece.} \end{cases}$$

Furthermore, the following "minimum span" constraint is imposed on $\boldsymbol{\delta}$: if the AR order of a certain piece in $\mathcal{F}$ is p, then this piece is made to have at least m_p observations. This predefined integer

m_p is chosen to guarantee that there are enough observations for obtaining quality estimates for the parameters of the segment modeled as a state space process with AR order equal to p. Also, in the implementation of the algorithm an upper bound P_0 on the order p_j's of the AR processes is imposed.

Initial Population Generation: Each individual of the initial population is selected randomly, accordingly to the following strategy: First, select a value for p_1 from $\{0, 1, \ldots, P_0\}$ with distribution π_p, $p = 0, 1, \ldots, P_0$ and set $\delta_1 = p_1$; i.e., the first AR piece is of order p_1. Then the next $m_{p_1} - 1$ genes δ_i's (i.e., δ_2 to $\delta_{m_{p_1}}$) are set to -1, so that the above minimum span constraint is imposed for this first piece. Now for the next gene $\delta_{m_{p_1}+1}$ in line, it will either be initialized as a break point, or it will be assigned -1 with probability $1 - \pi_B$. If $\delta_{m_{p_1}+1}$ is assigned the value p_2, where p_2 is randomly drawn from $\{0, 1, \ldots, P_0\}$ with distribution π_p, $p = 0, 1, \ldots, P_0$, then this implies that the second AR process is of order p_2, and the next $m_{p_2} - 1$ δ_i's will be assigned -1 so that the minimum span constraint is enforced. On the other hand, if $\delta_{m_{p_1}+1}$ is assigned -1, the initialization process will move to the next gene in line and a decision should be made if this gene should be a "break point" gene or a "-1" gene. This process continues in a similar fashion, and a random chromosome is generated when the process hits the last gene δ_n. We use $\pi_B = \min\{m_1, \ldots, m_{P_0}\}/n$.

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

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 a higher chance of being selected. From these two parents, the gene values δ_i's of the child chromosome will be inherited in the following manner. Firstly for $t = 1$, δ_t will take on the corresponding δ_t value from either the first or the 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., δ_{t+1}). If this value is not -1, then it is a non-negative integer p_j corresponding the AR order of the current piece. In this case the minimum span constraint will be imposed (i.e., the next $m_{p_j} - 1$

δ_t's will be set to -1), and the same gene–inheriting process will be applied to the next available δ_t.

For mutation one child is reproduced from one parent. Again, this process starts with $t = 1$, and every δ_t (subject to the minimum span constraint) can take on one of the following three possible values: (i) with probability π_P it will take the corresponding δ_t value from the parent, (ii) with probability π_N it will take the value -1, and (iii) with probability $1 - \pi_P - \pi_N$, it will take the a new randomly generated AR order p_j. In this paper we set $\pi_P = 0.3$ and $\pi_N = 0.3$.

Declaration of Convergence: Recall that we adopt the island model in which migration is allowed for every M_i generations. 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.

4 Simulations

4.1 Financial time series

In this section the performance of Auto-Seg is evaluated via simulation when the GARCH models introduced in Example 2 are used to study changes in the dynamics of returns of financial assets. The setup 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 $m = 1$, we have a two piece GARCH(1,1) model given by

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

where

$$\sigma^2_{t,k} = \omega_k + \alpha_k Y^2_{t-1,k} + \beta_k \sigma^2_{t-1,k}, \tag{11}$$

and $\{\varepsilon_t\} \sim$ iid $N(0, 1)$. Each two-piecewise process has a break at $\tau_1 = 501$ with a total sample size of n=1000. For each data generation process, only one of the β_k's or the ω_k's are modified from the GARCH model in the first segment while the other parameters remain unchanged (see column labeled as *Model parameters* in Table 1). For completeness, the case of no breaks (i.e., the second piece has the same parameters as the first piece) is included for each data generation process.

For a given two-piecewise process, let $\boldsymbol{\theta}_j$, denote the vector of parameters of the j-th 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

$$\text{MDL}(m, \tau_1, \ldots, \tau_m) = \log m + (m+1) \log n + \sum_{j=1}^{m+1} \log n_j - \sum_{j=1}^{m+1} L_q(\hat{\boldsymbol{\psi}}_j; \mathbf{y}_j), \tag{12}$$

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). Note the coefficient of $\log n_j$ in (12) should be 3/2 instead of 1. This is due to the strong correlation between w_j and the other parameters which suggests the number of free parameters should be $d_j = 2$ instead of 3.

Table 1 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 are based on the Lavielle and Moulines least-squares procedure (Lavielle and Moulines, 2000) applied to the squared values Y_t^2 and using the Bayesian Information (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 of these variances. The larger the change the higher the detection rate. For example, in Scenario C the increase in variance is 0.33, which is slightly larger than 0.25, the increase of variance of Scenario G. For Auto-Seg the detection rates are 0.122 and 0.192, respectively, while for AG, these values are 0.140 and 0.240, respectively.

For illustrative purposes, Figure 1 shows typical realizations of Scenarios C and D defined in Table 1. 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 modeled 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 than the "two-piece" volatilities. It

Table 1: *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. 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* $n = 1000$.

Piecewise GARCH(1,1) scenario		# 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) $\rightarrow$ (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) $\rightarrow$ (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) $\rightarrow$ (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) $\rightarrow$ (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) $\rightarrow$ (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) $\rightarrow$ (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) $\rightarrow$ (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) $\rightarrow$ (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	

is not the case for the realization of Scenario C. However, notice that the volatilities for t between 1 and 505 closely agree.

Next, we consider a different setup in which there is no break in the data generating process. The first row of Figure 3 contains a realization of Scenario A defined in Table 1 and Auto-Seg found no breaks. 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 row of Figure 3 with minimum MDL value of 1,410.0 at location 265. The horizontal dashed line in this row is the MDL with no breaks (1,404.8). In the third row the estimated volatilities based on a single piece are shown. In the last row we show the estimated volatilities based on two GARCH(1,1) models with break at location 265 (minimizer of two-piece MDL values shown in the second row). Notice that the one-piece estimated volatilities (third row) have smaller variance than that based on two piece GARCH fit (fourth row). In the latter, the pattern of the volatilities in the first piece is unexpected and does not agree with the realization in the first row. We also compared Auto-Seg to the sequential procedure proposed by Berkes, et al. (2004). To estimate changes in the GARCH model, Berkes, et al. (2004) construct a stopping time based on quasi-

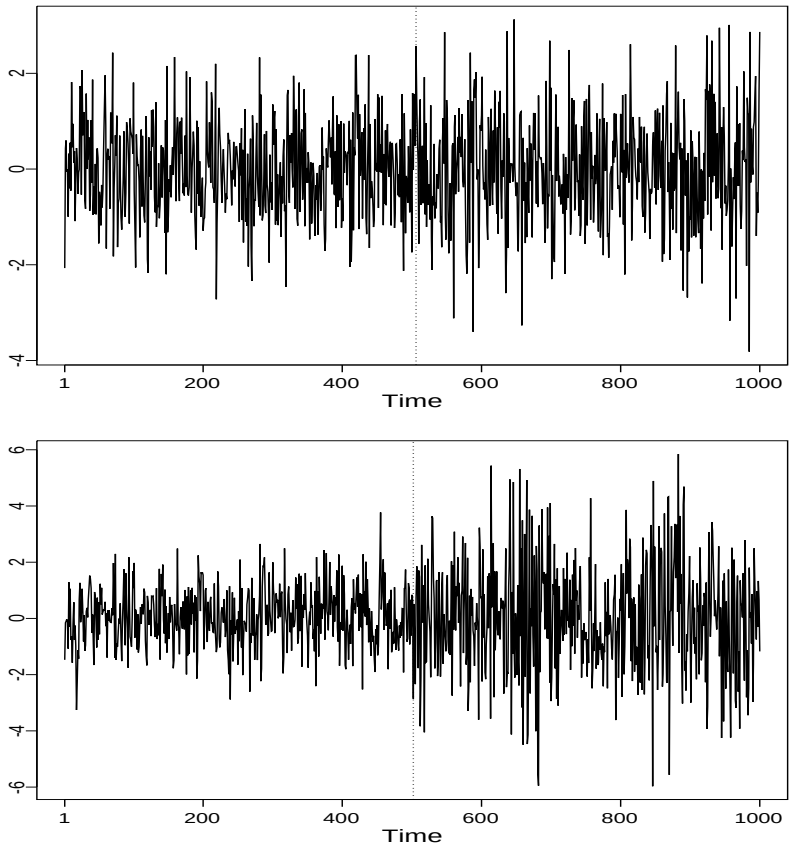

Figure 1: *Typical realizations from Scenarios C (top panel) and D (bottom panel) defined in Table 1. The vertical dotted lines are the break points found by Auto-Seg.*

maximum likelihood estimates. For their simulation study they use GARCH(1,1) models with the following sets of parameter values

Model 1: $\omega_1 = 0.05$, $\alpha_1 = 0.4$, $\beta_1 = 0.3$ ($\sigma^2_{Y_t}$=0.17),

Model 2: $\omega_1 = 0.05$, $\alpha_1 = 0.5$, $\beta_1 = 0.0$ ($\sigma^2_{Y_t}$=0.10),

Model 3: $\omega_1 = 1.00$, $\alpha_1 = 0.3$, $\beta_1 = 0.2$ ($\sigma^2_{Y_t}$=2.00),

where $\sigma^2_{Y_t}$ is the unconditional variance of Y_t. They also assume changes from Model 1 to Model 2 and from Model 1 to Model 3 at different time locations (see Table 2). 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 2 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. In the row labeled as BERKES, elementary statistics for the distribution of the first

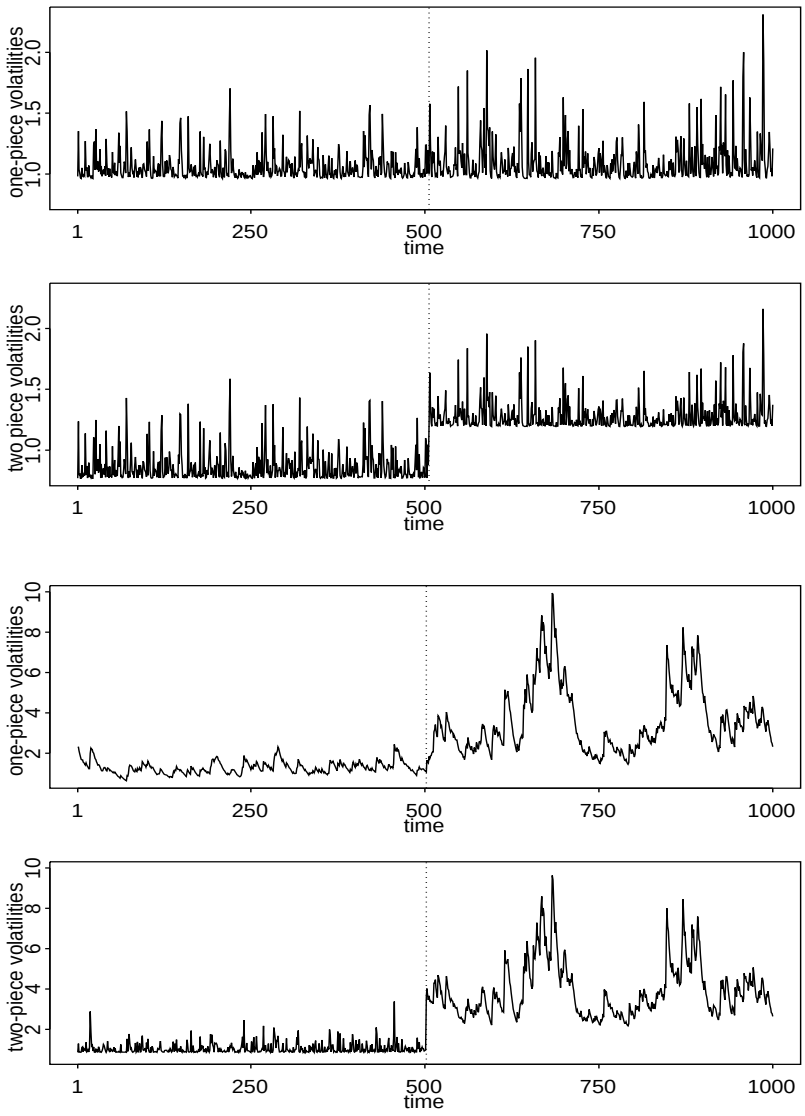

Figure 2: *Top two panels: Estimated volatilities of the realization of Scenario C shown in Figure 1 under the assumption of no break (first panel) and using the break (second panel) found by Auto-Seg. The last two panels are the analogous plots for the realization of Scenario D.*

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). Observe 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 3 of Berkes, et al., 2004). This is also in agreement with the results from Table 1. For this latter

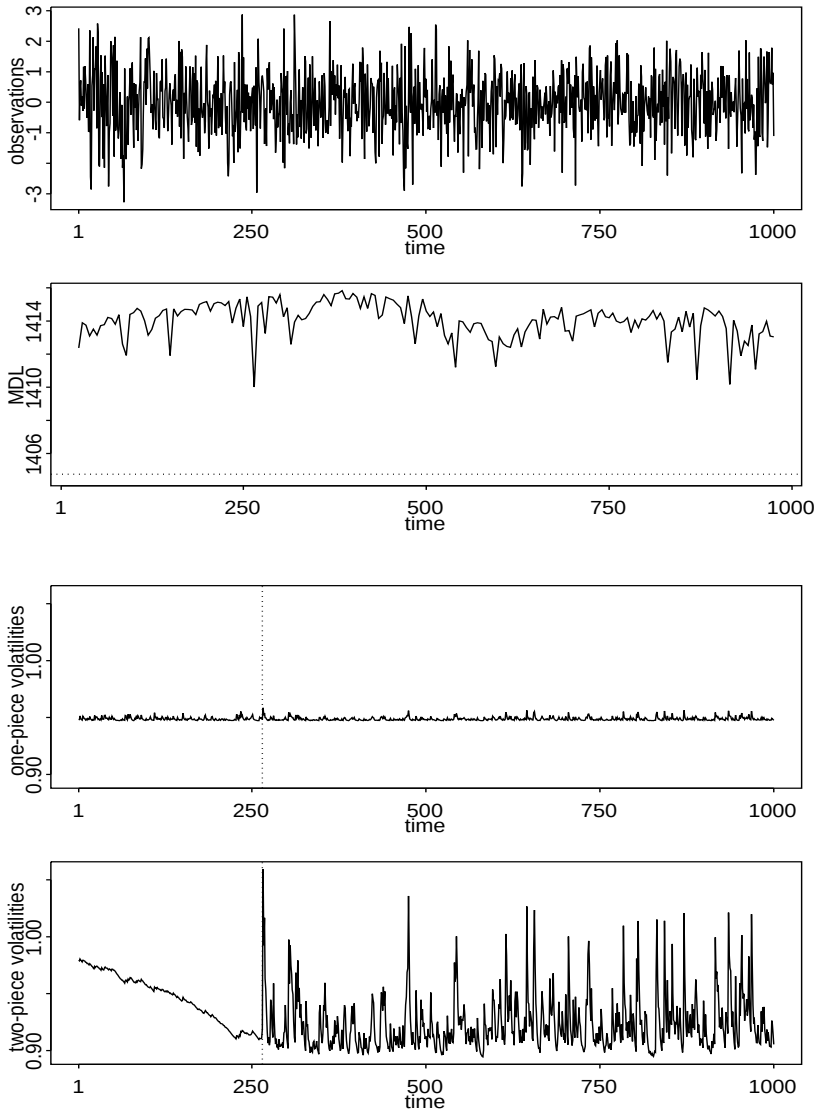

Figure 3: *First row: A typical realization of Scenario A defined in Table 6. Second row: two-piece MDL computed in a grid of points (solid line) and one-piece MDL (horizontal dashed line). Third panel: estimated volatilities based on a single piece. Fourth row: estimated volatilities based on two pieces by introducing an artificial break at location 265.*

configuration, the unconditional variance is 0.17 for the first piece and 0.10 for the second piece. Since the change of variance is small, a high detection rate is not expected.

For each scenario considered in Table 3, with the realizations considered in this table (i.e., realizations for which Auto-Seg found exactly one break), 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 3. Also included in this table is the average of the optimized MDL values. These values are shown in the column 8. Also, in the last column the average of the

Table 2: *Auto-Seg: elementary statistics for the distribution of the estimated location of break points (replications with only one break). BERKES: elementary statistics for the distribution of the first exceedance for the 10% critical level (from Table 4 of Berkes, et al. (2004)).*

	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	

MDL values obtained when only one piece is fitted to each realization is shown. In all cases, the two-pieces MDL average is considerable less than that of the one-piece MDL.

Table 3: *Parameters 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 → 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
std	0.04	0.26	0.26	0.21	0.05	0.12		
Model 1 → 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
std	0.02	0.15	0.15	0.23	0.06	0.13		
Model 1 → 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
std	0.01	0.13	0.11	0.27	0.08	0.16		
Model 1 → 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
std	0.03	0.17	0.17	0.01	0.01	0.04		

4.2 Stochastic Volatility Model

In the previous section 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 stochastic volatility model defined by the equation

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

where $\alpha_t = \gamma + \phi \alpha_{t-1} + \eta_t$, $\{\xi_t\} \sim$ iid, $N(0,1)$, and $\{\eta_t\} \sim$ iid $N(0, \sigma^2)$, $t = 1, \ldots, n$, and $|\phi| < 1$. This model can be written in the SSM framework given in Example 3 from Section 1. We are unaware of any work on the break point problem for the SV models. In this section, we consider

the performance of Auto-Seg on a two-piece model where each piece is the stochastic volatility model. The vector of parameters of this process is $\boldsymbol{\psi} = (\gamma, \phi, \sigma^2)$. Let us consider the models generated by the parameter values:

$$\text{Model 1 : } \gamma\text{=-0.8106703, } \phi = 0.90, \ \sigma^2 = 0.45560010,$$

$$\text{Model 2 : } \gamma\text{=-0.3738736, } \phi = 0.95, \ \sigma^2 = 0.06758185,$$

$$\text{Model 3 : } \gamma\text{=-0.3973738, } \phi = 0.95, \ \sigma^2 = 0.06758185.$$

The two piecewise processes considered in this section are listed in the last four lines of Table 4. The first piece of these processes is Model 1. Models 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 variance of the first piece is 0.0010, while the unconditional variances of the second piece of the processes B and D are 0.0008 (small decrease). The unconditional variance of the second pieces of the processes C 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 j-th piece. In the notation of Section 2, $\boldsymbol{\theta}_j = \boldsymbol{\psi}_j$. Then $c_j = 0$ and $d_j = 3$ and from (6) we obtain

$$\text{MDL}(m, \tau_1, \ldots, \tau_m) = \log m + (m+1) \log n + \sum_{j=1}^{m+1} \frac{3}{2} \log n_j - \sum_{j=1}^{m+1} L_a(\hat{\boldsymbol{\psi}}_j; \mathbf{y}_j, \boldsymbol{\alpha}_j), \tag{13}$$

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 columns 2 and 3 of Table 4.

Table 4: *Summary of Auto-Seg break points obtained from 500 replications. The length of the realizations is* $n = 2048$.

	# of break points		
Scenario	0	1	σ_Y^2
A: Model 1	100.0	0.0	0.0010
B: Model 1 $\rightarrow$ Model 2 at t= 513	18.2	81.8	0.0010, 0.0008
C: Model 1 $\rightarrow$ Model 2 at t=1025	0.4	99.6	0.0010, 0.0008
D: Model 1 $\rightarrow$ Model 3 at t= 513	17.2	82.8	0.0010, 0.0005
E: Model 1 $\rightarrow$ Model 3 at t=1025	1.2	98.8	0.0010, 0.0005

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 Auto-Seg found a break at location 550 (dashed vertical line) and for that of Scenario E it found a break at 1019 (dashed vertical line).

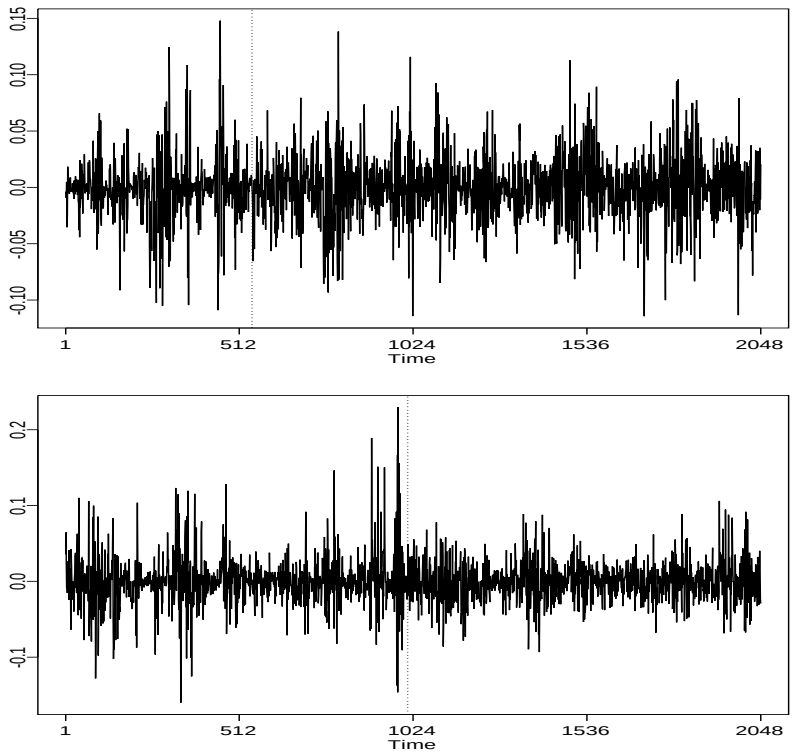

Figure 4: *Realizations from the piecewise stochastic volatility Scenarios B and E defined in Table 4. The vertical dotted lines are break points found by Auto-Seg.*

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 lines. The plots in the last two panels are the analogous modes for the realization of Scenario E for which Auto-Seg found a break 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.

Finally, elementary statistics for those replicates of Scenarios B through E from Table 4, for which Auto-Seg found exactly one break, are given in Table 5.

4.3 Poisson Parameter Driven process

In this section we consider the performance of Auto-Seg on a two-piecewise Poisson process. That is, for each piece, the observation equation $p(y_t|\alpha_t;\boldsymbol{\delta})$ has a Poisson distribution with rate $\lambda_t := e^{\beta+\alpha_t}$,

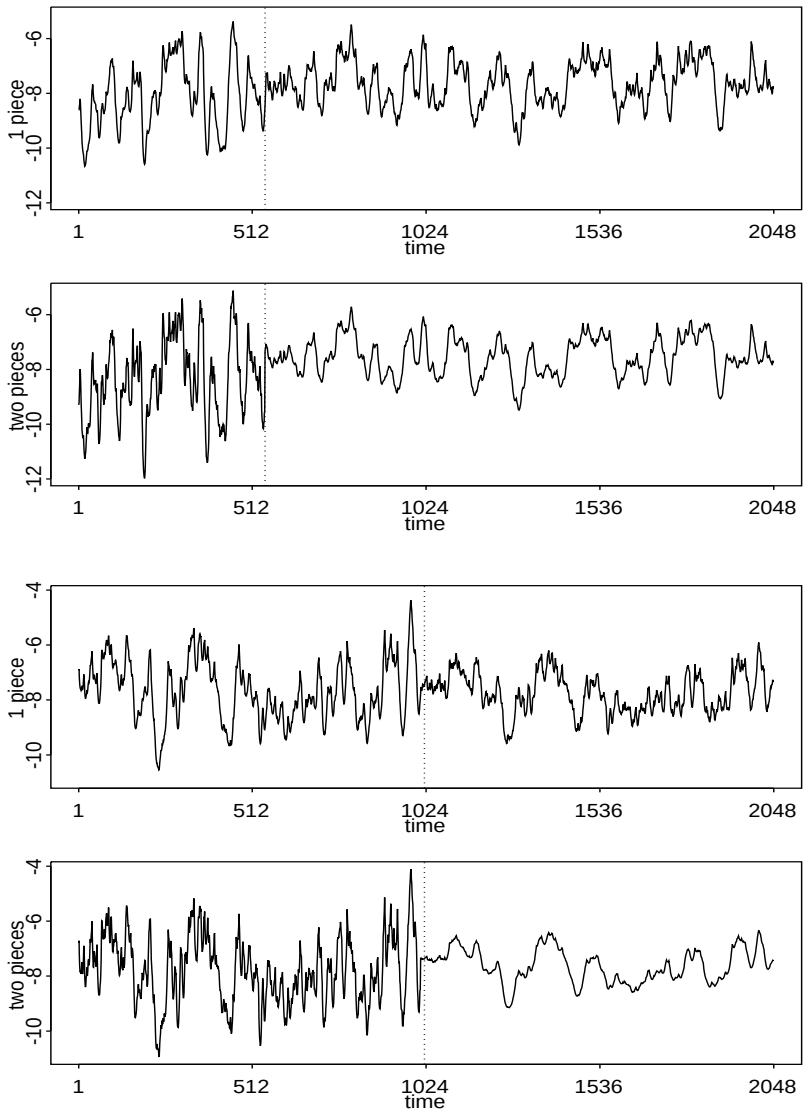

Figure 5: *Top two panels: Estimated posterior mode of the vector of states for the realization of Scenario C shown in Figure 4 under the assumption of no break (first panel) and using the break (second panel) found by Auto-Seg. The last two panels are the analogous plots for the realization from Scenario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

Table 5: *Elementary statistics for the distribution of the estimated location of break points of those replications with one break for the scenarios given in Table 4.*

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

and the state equation is $\alpha_t = \phi\alpha_{t-1} + \eta_t$, $\{\eta_t\} \sim$ iid $N(0,\ \sigma^2)$, $t = 1, \ldots, n$, and $|\phi| < 1$. The vector of parameters of this process is $\boldsymbol{\psi} = (\beta, \phi, \sigma^2)$. Let us consider the PDM models with the following set of parameter values:

Model 1: β = -1.5702, ϕ = 0.50, σ^2 = 1.9237,

Model 2: β = -1.3061, ϕ =-0.50, σ^2 = 1.5277,

Model 3: β =-1.3061, ϕ =0.90, σ^2 = 0.3870,

Model 4: β =-0.9373, ϕ =-0.50, σ^2 = 0.9745,

Model 5: β =-0.9373, ϕ =0.90, σ^2 = 0.2469.

The two piecewise PDM processes considered in this section are defined in the first column of Table 6. 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 variance of the first pieces is 7.5, while the unconditional variances of the second piece of the processes B, C, D and E are 4.5 (small decrease). The unconditional variance of the second pieces of the processes G, G, H and I are 2.25, which is a larger decrease. Notice that the MDL calculation of this piecewise process is identical to that for the SVM given in Section 4.2.

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 columns 2 and 3 of Table 6. As in

Table 6: *Summary of estimated Auto-Seg break points obtained from 500 replications. The length of the realizations is* $n = 1024$.

	# of break points		
Scenario	0	1	σ^2_Y
A: Model 1	100.0	0.0	7.5
B: Model 1 $\rightarrow$ Model 2 at t=257	34.0	66.0	7.5, 4.5
C: Model 1 $\rightarrow$ Model 2 at t=513	11.6	88.4	7.5, 4.5
D: Model 1 $\rightarrow$ Model 3 at t=257	31.0	69.0	7.5, 4.5
E: Model 1 $\rightarrow$ Model 3 at t=513	16.8	83.2	7.5, 4.5
F: Model 1 $\rightarrow$ Model 4 at t=257	13.4	86.6	7.5, 2.25
G: Model 1 $\rightarrow$ Model 4 at t=513	2.2	97.8	7.5, 2.25
H: Model 1 $\rightarrow$ Model 5 at t=257	16.0	84.0	7.5, 2.25
I: Model 1 $\rightarrow$ Model 5 at t=513	9.0	91.0	7.5, 2.25

the GARCH case, the performance of Auto-Seg improves when the change of variance between the pieces increases (e.g., the detection rate for Scenario E is better than that for Scenario C). Notice that the change of variances of the pieces of Scenario C is larger than that of the pieces of Scenario C. Another noteworthy comment from Table 6 is that the performance of Auto-Seg can vary when the change of variances are the same. For example, the change of 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., Scenario F and G have similar parameter values except the locations, which are 257 and 513 respectively. The fact that for Scenario A with no break Auto-Seg never finds a break point is remarkable. The detection rates for the scenarios with one break in this table vary from 66.0% to 97.8%. Taking in consideration that for all the scenarios the sample size (1024) is not large, the performance of Auto-Seg for these scenario 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 B Auto-Seg found a break at location

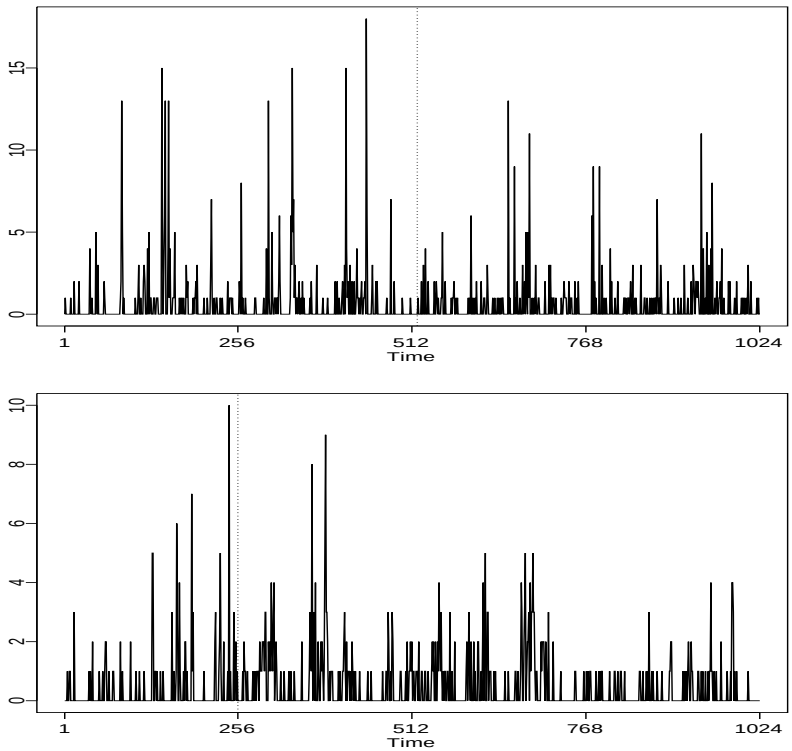

Figure 6: *Realizations from the piecewise Poisson processes B and H, respectively (defined in the first column of Table 6). Vertical dotted lines are break points found by Auto-Seg.*

520 (dashed vertical line) and for that of Scenario H it found a break at 256 (dashed vertical line). Like 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 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. The plots in the last two panels are the analogous modes for

the realization of Scenario H for which Auto-Seg found a break at location 256.

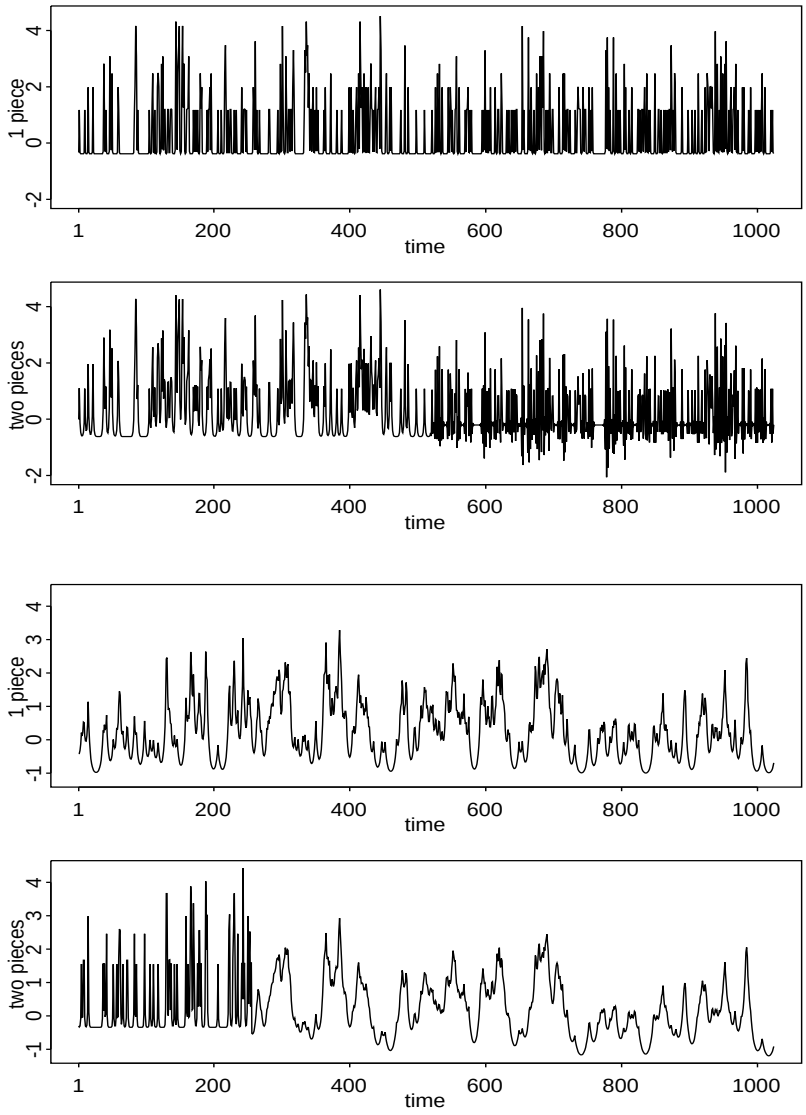

Figure 7: *Posterior mode of the realization of Scenario C shown in Figure 6 under the assumption of no break (first panel) and using the break (second panel) found by Auto-Seg. The last two panels are the analogous plots for the realization of Scenario H.*

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 (bottom two panels) the mode of the first piece is underestimated when no breaks are considered. Notice that the modes of the second piece look quite similar.

We include now the case where there is no break in the underlying scenario. In the first row of Figure 8 a realization of Scenario A defined in Table 6 is shown. Auto-Seg did not find any break for

this realization. Now, the MDL computed 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 row of Figure 8. Ignoring the last few MDL values on the right the minimum MDL value in this grid is 1,195.2 at 550. The horizontal dashed line in this row shows the MDL with no breaks (1,183.7). In the third row the estimated posterior mode of the vector of states based on a single piece is shown. In the last row 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 are shown in the second row of this figure. Notice

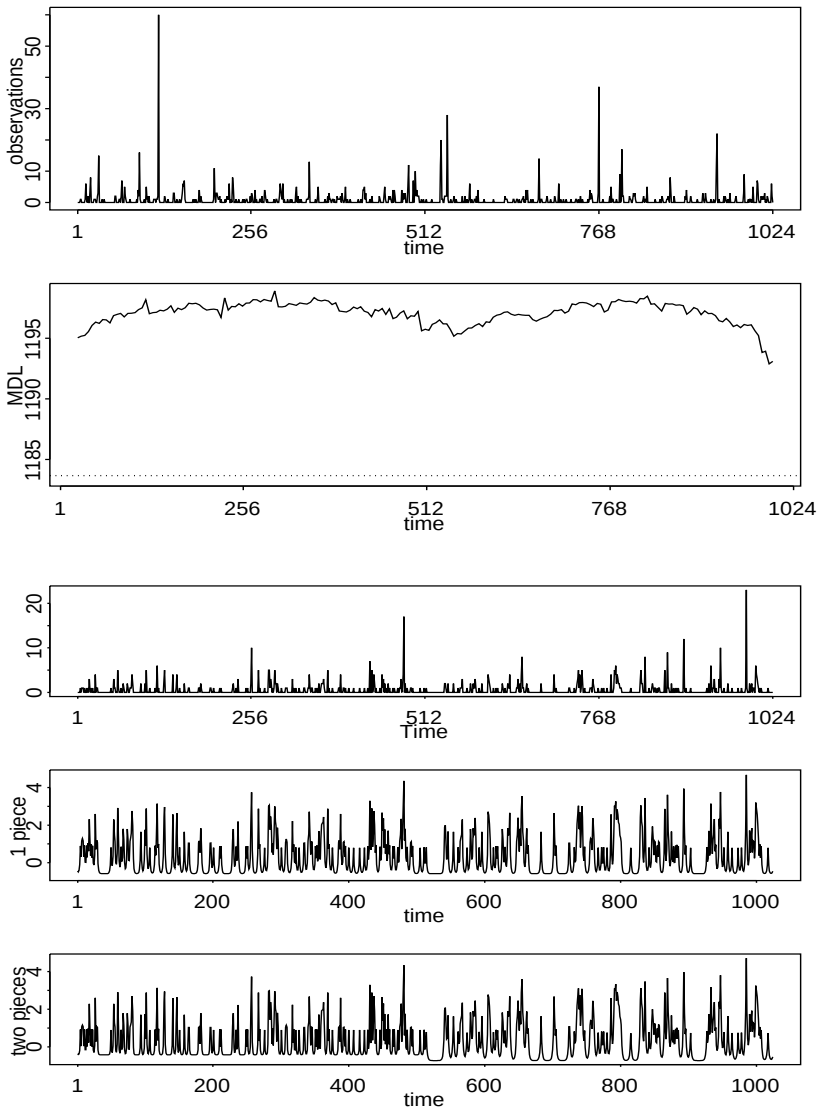

Figure 8: *First row: A typical realization of Scenario A defined in Table 6. Second row: two-piece MDL computed in a grid of points (solid line) and one-piece MDL (horizontal dashed line). Third panel: estimated posterior mode based on a single piece. Fourth row: estimated posterior mode based on two pieces with break at location 550.*

that the two sets of estimates agree.

Elementary statistics for the replicates of the two-piecewise models from Table 6 for which Auto-Seg found one break are given in Table 7. As seen in Table 7 the mean of the fitted breaks are

Table 7: *Elementary statistics for the distribution of the estimated location of break points of those replications with only one break for the models given in Table 6.*

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

generally close to the true value. The increase in change of variances tend 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.

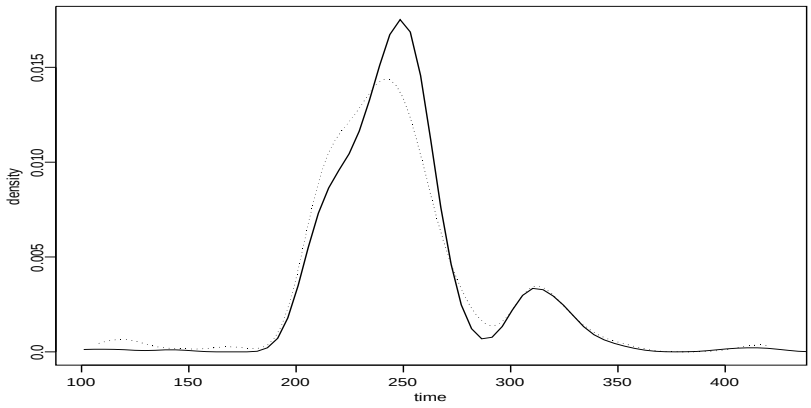

Figure 9: *Estimated densities of the locations of the breaks of Scenarios B (dotted line) and F (solid line).*

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

Now consider the two piece models from Table 6. For those realizations for which Auto-Seg found exactly one break, the parameters of each piece were estimated. The average and standard error of these estimates are shown in Table 8. Also in the last two columns in this table, the average of the minimized MDL and the average of the MDL values obtained when no breaks are considered are given.

Table 8: *Parameters estimates for the two-piecewise Poisson models from Table 6.*

		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
	std	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
	std	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
	std	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
	std	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
	std	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
	std	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
	std	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
	std	0.2336	0.11	0.4577	0.2483	0.04	0.0743		

In general, the estimates are sightly biased. This is true for the state-space Poisson model with no regime 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 and Davis and Rodriguez-Yam, 2005).

5 An Application

The Auto-Seg procedure was applied to analyze change points in the *Standard and Poors 500 index* (S&P 500) over the period Jan/4/1989 to Oct/19/2001 at daily frequency. This stock market series was also analyzed 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 into 1998. This section of S&P 500 consists of 3,230 observations. The log returns r_t of this series is shown in Figure 10. Auto-Seg was applied to the log returns series and 4 segments were found with break locations at 197, 726 and 2,229 which are shown as the vertical dotted lines in Figure 10. In Table 9 we show the breaks found by Andreou and Ghysels (2002)

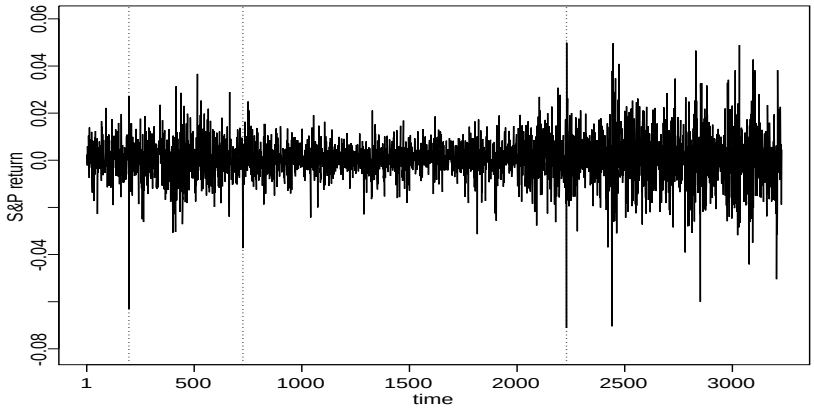

Figure 10: *Log returns of the S&P index over the period Jan/4/1989-Oct/19/2001. The dotted vertical lines are the breaks found by Auto-Seg.*

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 9, the last

Table 9: *Breaks of the S&P 500 index. 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.*

	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

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 fitted

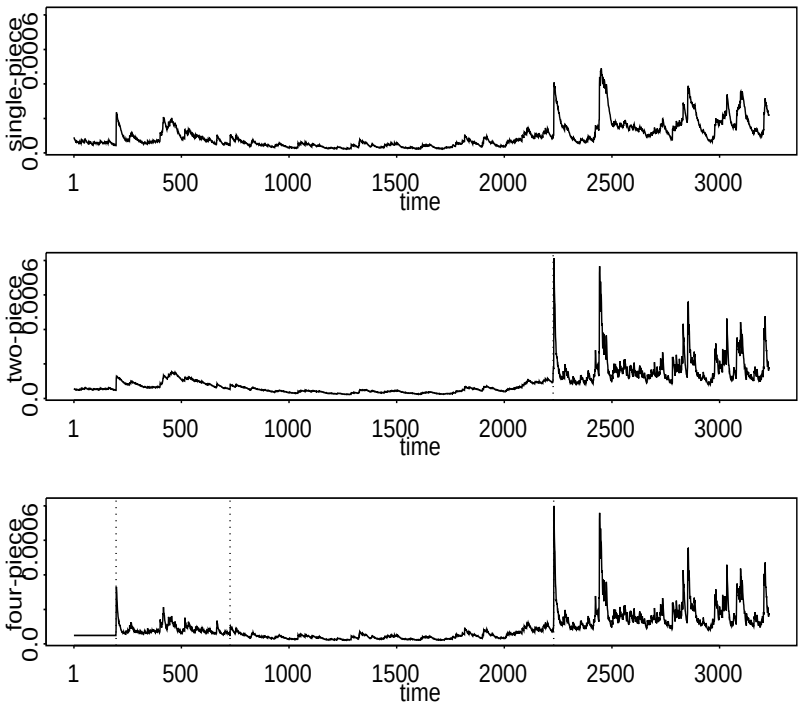

Figure 11: *Estimated volatilities of the log returns of the S&P 500 series. Top: estimated volatilities under no breaks. Middle panel: volatilities when a break in 27/10/97 is assumed. Bottom panel: Estimated volatilities based on the Auto-Seg breaks.*

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 greater than between the best Auto-Seg MDL and the single-break MDL model.

Acknowledgement

This research was supported in part by NSF grants DMS-0308109 (Davis) and DMS-0203901 (Lee) and by an IBM Faculty Research Award.

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, 1140-1167.

Bhattacharya, P. K. (1994), "Some Aspects of Change-point Analysis", in *Change-point problems* (edited by 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, 307-327.

Chatterjee, S., Laudato, M. and Lynch, L. (1996). "Genetic Algorithms and their Statistical Applications: an Introduction", *Computational Statistics & Data Analysis* 22, 633-651.

Csörgő, M., and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis* New-York, Wiley.

Davis, R. A., Dunsmuir, W. T. M. and Wang, Y. (1998). Modelling Time Series of Count Data. In *Asymptotics, Nonparametrics and Time Series* (edited by S. Ghosh), 63-112. Marcel Dekker, New York.

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., Lee, T.C.M., and Rodriguez-Yam, G. A. (2005). "Structural Breaks Estimation for Non-stationary Time Series Models". Unpublished manuscript.

Davis, R. A. and Rodriguez-Yam, G. A. (2005). "Estimation for state-space models based on a likelihood approximation", *Statistica Sinica* 15, 381-406.

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 (1994). "A MAP Solution to Off-line Segmentation of Signals", in *1992 IEEE International Conference on Accoustics, Speech and Signal Processing, March 23-26, 1992*, 505-508, New York, NY:IEEE.

Durbin, J. and Koopman, S. J. (1997). "Monte Carlo Maximum Likelihood Estimation for non-Gaussian State Space Models", *Biometrika* 84, 669-684.

Eshelman, L.J., (2000). "Genetic Algorithms", in *Evolutionary Computation. Vol 1, Basic Algorithms and Operatos* (edited by T. Bäck, D.B. Fogel, and T. Michalewicz), 64-80, Philadelphia: Bristol.

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, 559-570.

Gustafsson, F. (2000). *Adaptive Filtering and Change Detection*, New York: Wiley.

Harvey, A. C. and Fernandes, C. (1989). Time Series Models for Count or Qualitative Observations. *Journal of the American Statistical Association* 7, 407-417.

Holland, J. (1975). *Adaptation in Natural and Artificial Systems*, Ann Arbor, MI:University of Michigan 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, 351-363.

Kitagawa, G., Takanami, T. and Matsumoto, N. (2001). "Signal Extraction Problems in Seismology", *International Statistical Review* 69, 129-152.

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, 325-333.

Lavielle, M. (1998). "Optimal Segmentation of Random Processes", *IEEE Transactions on Signal Processing* 46, 1365-1373.

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, S-W. and Hansen, B.E. (1994). "Asymptotic Theory for the GARCH(1,1) quasi-maximum likelihood estimator", *Econometric Theory* 10, 29-52.

Lee, C-B. (1995). "Estimating the Number of Change Points in a Sequence of Independent Normal Random Variables", *Statistics & Probability Letters* 25, 241-248.

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, 823-842.

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*, (edited by T. Bäck, D.B. Fogel, and T. Michalewicz), 101-124, Philadelphia: Bristol.

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, 968-978

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.

Picard, D. (1985). "Testing and Estimating Change-Points in Time Series", *Advances in Applied Probability* 17, 841-867.

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, 747-758.

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.

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*, New York: Oxford.

Yao, Y-C. (1988), "Estimating the Number of Change-points via Schwarz' Criterion", *Statistics & Probability Letters* 6, 181-189.

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* (edited by M.H. Rivzi, J.S. Rustagi, and D. Sigmund), 245-269, Academic Press, New York.

Zeger, S. L. (1988). A regresion model for time series of counts. *Biometrika* 75, 621-629.