

BaSTA: consistent multiscale multiple change-point detection for piecewise-stationary ARCH processes

P. Fryzlewicz* S. Subba Rao †

May 6, 2009

Abstract

The emergence of the recent financial crisis, during which markets frequently underwent changes in their statistical structure over a short period of time, illustrates the importance of non-stationary modelling in financial time series. Motivated by this observation, we propose a fast, well-performing and theoretically tractable method for detecting multiple change-points in the structure of a piecewise-stationary ARCH model for financial returns. Our method, termed BaSTA (Binary Segmentation for Transformed ARCH), proceeds in two stages: process transformation and binary segmentation. The process transformation decorrelates the original process and lightens its tails, the binary segmentation consistently estimates the change-points. We propose and justify a particular choice of the transformation, and use simulation to fine-tune its parameter as

*Department of Statistics, Columbia House, London School of Economics, Houghton Street, London WC2A 2AE, UK.

†Department of Statistics, Blocker Building, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143, USA.

well as the threshold parameter for the binary segmentation stage. A comparative simulation study illustrates good performance in comparison with the state of the art. Although the method is easy to implement, ready-made R software is provided.

Key words: binary segmentation, non-stationary time series, CUSUM, mixing, process transformation, Unbalanced Haar wavelets.

1 Introduction

Log-returns on speculative prices, such as stock indices, currency exchange rates, share prices, and so on, often exhibit the following well known properties: the sample mean of the observed series is close to zero; the marginal distribution is roughly symmetric or slightly skewed, has a peak at zero, and is heavy-tailed; the sample autocorrelations are ‘small’ at almost all lags, although the sample autocorrelations of the absolute values and squares are significant for a large number of lags; and volatility is ‘clustered’, in that days of either large or small movements are likely to be followed by days with similar characteristics.

To capture the above properties, one needs to look beyond the stationary linear time series framework, and in order to preserve stationarity a large number of nonlinear models have been proposed. Among them, two branches are by far the most popular: the families of ARCH (Engle, 1982) and GARCH (Bollerslev, 1986; Taylor, 1986) models, as well as the family of ‘stochastic volatility’ models (Taylor, 1986). For a review of recent advances in ARCH, GARCH and stochastic volatility modelling, we refer the reader to Fan & Yao (2003) and Giraitis et al. (2005).

Although stationarity is an attractive assumption from the estimation point of view, some authors point out that the above properties can be better explained by resorting to non-stationary models. Dahlhaus & Subba Rao (2006) propose a time-varying

ARCH model, where the model parameters evolve over time in a continuous fashion. Mikosch & Stărică (2004) consider probabilistic properties of a piecewise-stationary GARCH model and show that it explains well the “long memory” effect in squared log-returns. Underlying these approaches is the observation that, given the changing pace of the world economy, it is unlikely that log-return series should stay stationary over long time intervals. For example, given the “explosion” of market volatility during the recent financial crisis, it is hardly plausible that the volatility dynamics before and during the crisis could be well described by the same stationary time series model. Indeed, Janeway (2009) goes further and argues that financial theorists’ belief that “statistical attributes of financial time series—such as variance, correlation and liquidity—are stable observables generated by stationary processes” might have been a contributing factor in the crisis.

In this paper, we focus on piecewise stationarity as the simplest form of departure from stationarity. Its appeal is in that it is easily interpretable as it provides segmentation of the data into time intervals where the parameters of the process remain constant. Also, the piecewise stationary approach can be of use in forecasting, where it is often of interest to obtain the “last” stationary segment of the data which can then be used to forecast the future behaviour. The model we consider is that of a piecewise stationary ARCH process. We note that Fryzlewicz *et al.* (2008) demonstrate that time-varying ARCH processes pick up well the empirical features of log-return series listed in the first paragraph. Since estimation for time-varying GARCH processes is a lot more challenging task (due to the fact that likelihood functions are typically “flat”), and time-varying ARCH processes describe typical log-returns sufficiently well, we do not consider time-varying GARCH processes in this paper.

In any piecewise-stationary model, one task of interest is to detect, *a posteriori*, the change-points, i.e. time instants when the parameters of the process changed. The problem of detecting a single change-point was studied, for example, by Chu (1995)

and Kulperger & Yu (2005) for the GARCH model and by Kokoszka & Leipus (2000) for the ARCH model. The problem of multiple change-point detection for ARCH-type processes is harder and has not been studied rigorously by many authors. The heuristic procedure of Andreou & Ghysels (2002) for the GARCH model was based on the work of Lavielle & Moulines (2000) for detecting multiple breaks in the mean of an otherwise stationary time series. We also mention the computationally intensive procedure of Davis *et al.* (2008) for nonlinear time series, based on the Minimum Description Length principle.

Our aim in this paper is to devise a statistically rigorous, well-performing and fast technique for multiple change-point detection in the piecewise stationary ARCH model, where neither the number nor the amplitudes of the changes are assumed to be known. Our method, termed BaSTA (Binary Segmentation for Transformed ARCH), proceeds in two stages: the *process transformation* and the *binary segmentation* stage, which we now briefly describe in turn.

Process transformation. Given a stretch of data from a piecewise-stationary ARCH process X_t , the initial step is to form a transformation of the data, $U_t = g(X_t^2, \dots, X_{t-\tau}^2)$, whose aim is twofold: to ensure that the marginal distribution of U_t is light-tailed (recall that X_t^2 is always heavy-tailed), and to ensure that the degree of autocorrelation in U_t is less than that in X_t^2 . Formally speaking, the aim of BaSTA will be to detect change-points in the mean value of U_t . We discuss a suitable choice of g . Our initial idea is to choose g in such a way that it corresponds to the sequence of empirical residuals of X_t under the null hypothesis of no change-points present. This then leads us to consider an entire family of transformations, $g_{\underline{C}}$, indexed by a vector constant \underline{C} , whose suitable default choice is discussed.

Binary segmentation. In the second stage of the BaSTA algorithm, we perform a binary segmentation procedure on the sequence U_t , with the aim of detecting the

change-points in $\mathbb{E}(U_t)$. Algorithmically, our binary segmentation procedure is performed similarly to Venkatraman’s (1993) method for detecting mean shifts in Gaussian white noise sequences, except that we use a more general form of the threshold, which we parameterise by a constant c . We demonstrate that for each fixed c , BaSTA leads to a consistent estimator of the number and location of change-points in $\mathbb{E}(U_t)$. We note that substantial modification in proof techniques are needed due to the fact that U_t is a highly structured time series rather than a Gaussian white noise sequence. We propose a default choice of the constant c , based on an extensive simulation study. The paper is organised as follows. Section 2 describes the model and the problem. Section 3 introduces our generic algorithm and shows its consistency. Section 4 discusses our particular proposed choice of function g and the threshold c . Section 5 describes the outcome of a comparative simulation study where we compare our method to state of the art. The proof of our consistency result appears in Appendix A.

R software implementing BaSTA can be obtained from

<http://stats.lse.ac.uk/fryzlewicz/basta/basta.html>

2 Model and problem set-up

The model we consider in this paper is as follows. Let $\{X_t^{(n)} : 1 \leq t \leq n < \infty\}$ be a triangular array of random variables. We will often skip the superscript (n) for economy of notation. Let the n th row of the array be distributed as a piecewise-stationary ARCH process with m change-points, where $m \leq n$. That is, let there exist $\nu_1^{(n)}, \dots, \nu_m^{(n)}$, where $0 = \nu_0^{(n)} = \nu_1^{(n)} < \dots < \nu_m^{(n)} < \nu_{m+1}^{(n)} = n$, such that

$X_t^{(n)} \sim \text{ARCH}(p)$ with parameters $\underline{a}^{(j)} = (a_0^{(j)}, \dots, a_p^{(j)})$ if $\nu_j^{(n)} < t \leq \nu_{j+1}^{(n)}, j = 0, \dots, m$.

Later, we will specify the degree to which we require the parameters $\underline{a}^{(j)}$ to differ for each segment j . As is often the case in change-point detection literature for time series, we assume that the stationary pieces of the process X_t are independent of each other.

We recall that a process Y_t follows the ARCH(p) model with parameters $\underline{a} = (a_0, \dots, a_p)$ if

$$\begin{aligned} Y_t &= \sigma_t Z_t \\ \sigma_t^2 &= a_0 + \sum_{i=1}^p a_i Y_{t-i}^2, \end{aligned} \tag{1}$$

where Z_t is an independent, identically distributed sequence with $\mathbb{E}Z_t = 0$ and $\mathbb{E}Z_t^2 = 1$.

We emphasise that neither the number m or the locations $\nu_j^{(n)}$ of the change-points are known, and the aim of this paper is to propose a consistent, well-performing and fast procedure for estimating these quantities. Naturally enough, we also do not assume that the parameter values $\underline{a}^{(j)}$ are known. In this paper, we do not study the issue of order selection for piecewise-stationary ARCH: whenever the order p is known, we use it; if it is not, we note that in our setting, both underfitting and overfitting the model is permitted in the sense that choosing p to be different from the true order does not affect the validity of either our theory or our algorithm, but may reduce the quality of the estimators. In what follows, we assume that we only observe the n th row of the triangular array: that is, a single time series X_t of length n .

3 Generic algorithm and consistency result

3.1 General approach and motivation

Our method for multiple change-point detection in the framework described in Section 2 is termed BaSTA: Binary Segmentation for Transformed ARCH. Its main ingredient is the *binary segmentation* procedure, suitably modified for use in the piecewise-stationary ARCH model. The binary segmentation procedure for detecting a change in the mean of normal random variables was first introduced by Sen and Srivastava (1975). Consistency of binary segmentation for a larger class of processes was shown by Vostrikova (1981); however, conditions for consistency were formulated in terms of submartingales (which is not a convenient class to use in our framework), under assumptions on change-points which do not hold in our case, and the procedure itself was not easily implementable due to the difficulty in computing the null distribution of the change-point detection statistic. Venkatraman (1993) proved consistency of the binary segmentation procedure in the Gaussian function+noise model, using a particularly simple form of the test statistic.

We now give a brief description of BaSTA; a more detailed description follows in the next section. BaSTA consists of two stages and proceeds as follows. In the first stage, which we term ‘data transformation’, a sequence U_1, \dots, U_n is produced, which is a suitable transformation of the piecewise-stationary ARCH process X_t from Section 2. The exact form of the transformation will be discussed later; for the moment, suffice it to say that U_t is constructed in such a way that its changing mean carries ‘good quality’ information about the time-varying ARCH parameters and the corresponding change-points.

The next stage of BaSTA is the actual binary segmentation. In the first step, the sequence U_t is tested against the null hypothesis of coming from a model whose

parameters do not change over time. The alternative hypothesis is that there is a ν such that U_1, \dots, U_ν and $U_{\nu+1}, \dots, U_n$ come from different time series models. If the null hypothesis is rejected, we call such a ν a change-point, estimate it, split the data into two segments at the estimated change-point, and continue the procedure at the two subsamples: hence the name “binary segmentation”. The procedure continues until no more change-points are detected in any of the segments. This yields an estimate of the number of change-points and estimates of their locations.

We note that Fryzlewicz’s (2007) Unbalanced Haar technique for function estimation in the Gaussian function+noise model is related to binary segmentation in that both proceed in a recursive fashion by iteratively fitting best step functions to the data (see the above work for a discussion of similarities and differences between the two approaches). Indeed, our choice of binary segmentation as a suitable methodology for change-point detection in the piecewise-stationary ARCH model is motivated by the good practical performance of the Unbalanced Haar estimation technique for the Gaussian function+noise model.

Since BaSTA proceeds in a recursive fashion by acting on subsamples determined by previously-detected change-points, it can be viewed as a “multiscale” procedure. The next section provides a more precise description of BaSTA and formulates a consistency result.

3.2 Algorithm and consistency result

As indicated above, BaSTA consists of two stages.

Stage I. In the first stage, a process $U_t = g(X_t, X_{t-1}, \dots, X_{t-\tau})$ is formed. A suitable choice of $g(\cdot)$ and τ will be discussed in Section 4. The process U_t is designed in such a way that its time-varying expectation carries information about the changing parameters of X_t and the corresponding change-points.

Stage II. Let $d = n$ and let $S_t = U_1 + \dots + U_t$. Form the statistic

$$Z_t = \frac{tS_d/d - S_t}{\{t(1 - t/d)\}^{1/2}}. \quad (2)$$

The above is a version of the well-known CUSUM test, and is described in more detail in Brodsky & Darkhovsky (1993), section 3.5. If U_t were a serially independent Gaussian sequence with one change-point in the means of otherwise identically distributed variables, $\operatorname{argmax}_{0 < t < d} |Z_t|$ would be the maximum likelihood statistic for detecting the likely location of the change-point, and would be optimal in the sense of Theorem 3.5.3 in Brodsky & Darkhovsky (1993). In our setting, it simply furnishes a least-squares-type estimator; note that since our U_t is a highly structured time series, exact maximum likelihood estimators of its change-points are not easy to obtain, and even if they were, their optimality (or otherwise) would not be easy to investigate.

Continuing with the interpretation, we note that, in the notation of Fryzlewicz (2007), Z_t is simply the inner product between the vector $\{U_t\}_{t=1}^d$ and the Unbalanced Haar vector $\psi_{1,t,d}$. We consider the statistic

$$S(U_1, \dots, U_d) = \max_{0 < t < d} |Z_t|,$$

which indicates the correlation between $\{U_t\}_{t=1}^d$ and the “best-fitting” Unbalanced Haar vector $\psi_{1,t,d}$. Thus $k := \operatorname{argmax}_{0 < t < d} |Z_t|$ is a likely location of a change-point. To test whether k corresponds to a significant change-point, we examine the magnitude of $S(U_1, \dots, U_d)$. If it is less than the critical level b_n , then we conclude that there are no change-points over the segment $t \in \{1, \dots, d\}$, and stop. Otherwise, we set the change-point estimate to be equal to k , and divide the sample into two subsamples: U_1, \dots, U_k and U_{k+1}, \dots, U_d . We repeat the procedure on the two subsamples, bearing in mind that the parameter d needs to be adjusted to reflect the length of the two subsamples. The procedure continues iteratively until no further change-points have

been found. We denote the number of the thus-obtained change-point estimates by \hat{m} and their locations, sorted in the increasing order, by $\hat{\mu}_1^{(n)}, \dots, \hat{\mu}_m^{(n)}$.

We note that the ‘‘threshold’’ b_n depends on the length n of the initial sample, and not on the changing d . Throughout the paper, we use $b_n = c n^{3/8}$, where the choice of the constant c will be discussed in Section 4.3. As in Venkatraman (1993), the rate of $n^{3/8}$ represents the ‘mid-point’ between the rate of $n^{1/4}$ (at which we are guaranteed to over-estimate change-points if there are a non-zero number of them) and the rate of $n^{1/2}$ (at which we are guaranteed to underestimate change-points if there are any). Note that our threshold is particularly simple to compute as it is a function of the sample size only.

The following notation prepares the ground for the main result of this paper: a consistency result for BaSTA. Let $\{\tilde{X}_t^j\}_{t=-\infty}^{\infty}$ denote a *stationary* ARCH(p) process with parameters as in the j th segment of the piecewise-stationary process X_t , i.e. $\underline{a}^{(j)} = (a_0^{(j)}, \dots, a_p^{(j)})$. We form the processes $\tilde{U}_t^j = g(\tilde{X}_t^j, \dots, \tilde{X}_{t-\tau}^j)$ and denote $\theta_j = \mathbb{E}(\tilde{U}_t^j)$.

Before we formulate a consistency result for BaSTA, we specify the following technical assumption.

Assumption 3.1 (i) $\inf_j \{\nu_{j+1}^{(n)} - \nu_j^{(n)}\} \geq \beta n$ where $\beta \in (0, 1)$,

(ii) $\inf_j |\theta_{j+1} - \theta_j| \geq \delta$ for some $\delta > 0$,

(iii) the function $g(\cdot)$ is bounded,

(iv) the constant c in the threshold b_n is arbitrary but fixed and positive,

(v) for some $\delta_1 > 0$, we have $\sup_j \sum_{i=1}^p a_i^{(j)} \leq 1 - \delta_1$,

(vi) for some $\delta_2 > 0$, we have $\inf_j a_0^{(j)} > \delta_2$ and $\sup_j a_0^{(j)} < \infty$,

(vii) Let f_Z denote the density of Z_t . For all $a > 0$ we have $\int |f_Z(u) - f_Z(u[1 + a])| du \leq K a$ for some K independent of a .

Assumption 3.1(i) specifies the minimum permitted distance between consecutive change-points, (ii) requires that the consecutive levels of the mean function $\mathbb{E}(\tilde{U}_t^j)$ (which obviously depend on the parameters $\underline{a}^{(j)}$) should be sufficiently well separated from their neighbours, and (iii) requires that the transformation function $g(\cdot)$ should be bounded. Unbounded functions g would also be possible if we were willing to impose extra constraints on the innovation sequence Z_t in (1), which we prefer to avoid. Section 4 discusses a suitable choice of g . Assumptions (v), (vi) and (vii) are required to guarantee that X_t is strongly mixing at a geometric rate, see Assumption 3.1 (and its discussion) as well as Theorem 3.1 in Fryzlewicz & Subba Rao (2008). Assumption (vii) is a mild one and is satisfied by many well-known distributions, as explained underneath Assumption 3.1 in Fryzlewicz & Subba Rao (2008).

The following theorem specifies a consistency result for BaSTA.

Theorem 3.1 *Suppose Assumption 3.1 holds. We have $P(\mathcal{A}_n) \rightarrow 1$, where*

$$\mathcal{A}_n = \{\hat{m} = m; |\hat{\nu}_j^{(n)} - \nu_j^{(n)}| \leq n^{3/4} \text{ for } 1 \leq j \leq m\}.$$

We note that the factor of $n^{3/4}$ appearing in the event \mathcal{A}_n is due to the fact that the change-points $\nu_j^{(n)}$ are measured in the “real” time $t \in \{1, \dots, n\}$, as opposed to the rescaled time $t/n \in [0, 1]$. Another way to interpret the above result is $|\hat{\nu}_j^{(n)}/n - \nu_j^{(n)}/n| \leq n^{-1/4}$. The proof of Theorem 3.1 appears in the Appendix.

4 Choice of the $g(\cdot)$ function and the threshold constant c

4.1 Family of transformations parameterised by \underline{C}

In this section, we discuss our recommended choice of the transformation function g . We start by recalling the desired properties of the transformed process $U_t = g(X_t, X_{t-1}, \dots, X_{t-\tau})$.

- (a) The time-varying expectation of U_t should carry information about the change-points, i.e. should change at the change-point locations.
- (b) A high degree of autocorrelation in U_t would not be desirable as it would have the potential to distort the inner products in (2), thus giving a false picture of the locations of the change-points. Thus, we aim at processes U_t with as little degree of autocorrelation as possible.
- (c) In addition, Assumption 3.1(iii) requires that the function g should be bounded.

Intuitively, the requirement (a) implies that the process U_t should be a function of *even* powers of X_t . This is because if Z_t has a symmetric distribution, then so does X_t , which means that for q odd, if $\mathbb{E}(X_t^q)$ exists, then it is equal to zero. Thus, the expectation of odd powers of X_t is “uninteresting” from the point of view of change-point detection.

The requirement (b) suggests that any “diagonal” transformation, where $g(X_t, X_{t-1}, \dots, X_{t-\tau})$ is only a function of X_t , should not be used. (Examples of such transformations include $U_t = g(X_t) = X_t^2$ or $U_t = g(X_t) = \log X_t^2$.) This is because by the definition of the ARCH process, X_t^2 has a high degree of autocorrelation, which is preserved in any diagonal transformation of the type $g(X_t^2)$.

Motivated by the above discussion, we construct the process U_t as follows. Under the null hypothesis of stationarity, the process

$$U_t^{(1)} = \frac{X_t^2}{a_0 + \sum_{i=1}^p a_i X_{t-i}^2} = Z_t^2$$

is stationary, and perfectly decorrelated as it is simply a sequence of squared residuals Z_t^2 . Obviously, in practice, the above transformation is impossible to effect as it involves the unknown parameter values a_i . Instead, we “approximate” it with a transformation

$$U_t^{(2)} = \frac{X_t^2}{C_0 + \sum_{i=1}^p C_i X_{t-i}^2},$$

which, under the null hypothesis, also results in a process which is stationary, and hopefully almost decorrelated due to its closeness to $U_t^{(1)}$. The parameter $\underline{C} = (C_0, \dots, C_p)$ will need to be chosen from the data and the latter part of this section will describe a method for selecting \underline{C} .

At this point however, we come back to the requirement (c) above, whereby the process U_t is required to be bounded. To ensure this, we add an extra term ϵX_t^2 in the denominator, which results in the following transformation:

$$U_t^{(3)} = \frac{X_t^2}{C_0 + \sum_{i=1}^p C_i X_{t-i}^2 + \epsilon X_t^2}. \quad (3)$$

In this paper, for simplicity, we do not dwell on the choice of ϵ : in fact, in the numerical experiments described later, we always assume that ϵ has the default value of 10^{-3} , with X_t being normalised in such a way that the sample variance of the data vector X_t equals one.

As discussed above, the hope is that the transformation $U_t^{(3)}$ will produce, with a suitable choice of \underline{C} , a sequence approximating the squared empirical residuals of the process under the null hypothesis of stationarity. Under the alternative hypothe-

sis, $U_t^{(3)}$ is still (by construction) a sequence of nonnegative random variables whose changing expectation from one stationary segment to another reflects the different regimes of stationarity. In practice, $U_t^{(3)}$ tends to have a distribution which is highly skewed to the right, which is illustrated later in the paper. This is unsurprising as $U_t^{(3)}$ is of the form $\tilde{\sigma}_t Z_t^2$, where Z_t^2 are the true squared residuals, and $\tilde{\sigma}_t$ is nonnegative. To alleviate the above rightward skew, and bring the model closer to additive, we consider our final transformation

$$U_t^{(4)} = \log(\epsilon + U_t^{(3)}), \quad (4)$$

where, for simplicity, the default value of ϵ is as in $U_t^{(3)}$ above. Note that we cannot simply use $\log(U_t^{(3)})$ as one of the requirements on the function $g(\cdot)$ is that it should be bounded (since $U_t^{(3)}$ is bounded and nonnegative and $\epsilon > 0$, it follows that $U_t^{(4)}$ is bounded).

In our simulations described later, both $U_t^{(3)}$ and $U_t^{(4)}$ are used.

4.2 Default choice of \underline{C}

We propose the following default choice of the vector constant \underline{C} in our transformations (3) and (4).

In the first stage, we (not necessarily correctly) act as if $\{X_t\}_{t=1}^n$ were a realisation of a stationary ARCH process with parameters a_0, \dots, a_p , and follow a normalized least-squares procedure (Fryzlewicz *et al.*, 2008) to estimate the values of a_0, \dots, a_p as $\hat{a}_0, \dots, \hat{a}_p$. If $\{X_t\}_{t=1}^n$ indeed happened to be stationary, i.e. contained no change-points, the computed values $\hat{a}_0, \dots, \hat{a}_p$ would then form meaningful estimates of the true parameters a_0, \dots, a_p .

Thus, in the null hypothesis of no change-points present, if we were to set $C_i := \hat{a}_i$ for

$i = 0, \dots, p$, the corresponding transformed sequences $U_t^{(3)}$ and $U_t^{(4)}$ would indeed be close to the (squared; squared and logged, respectively) empirical residuals from the model, as explained in the discussion of Section 4.1. The hope is that our change-point detection procedure would then correctly react to this construction by determining that no change-points were present in the model.

However, rather than directly setting $C_i := \hat{a}_i$ for $i = 0, \dots, p$, we add extra flexibility to our construction by introducing a positive factor $F \geq 1$, which we use to “dampen” the values of the constants C_1, \dots, C_p as follows:

$$\begin{aligned} C_0 &:= \hat{a}_0 \\ C_i &:= \frac{\hat{a}_i}{F}, \quad i = 1, \dots, p. \end{aligned}$$

The effect of the above dampening of the values of C_1, \dots, C_p is that as F increases, $U_t^{(3)}$ is, up to a multiplicative constant, closer and closer to X_t^2 itself. Indeed, in the limit as $F \rightarrow \infty$, we have

$$U_t^{(3)} \approx \frac{X_t^2}{C_0 + \epsilon X_t^2}$$

(bear in mind that the default value of ϵ is small). Empirical evidence suggests that larger values of F can lead to better exposure of change-points in the alternative hypothesis of change-points being present, at the expense of introducing a higher degree of autocorrelation and heavier tails in $U_t^{(3)}$. Naturally, this also applies to $U_t^{(4)}$, but to a far less extent. We now illustrate this through an example.

Figure 1 shows a sample path of length $n = 1000$ of a piecewise-stationary ARCH(1) process X_t where the parameter a_0 remains unchanged (and equal to 1), and the parameter a_1 changes, at time $t = 501$, from 0.2 to 0.7.

The left column of Figure 2 shows the corresponding sequences $U_t^{(3)}$, as F takes the values of 1, 2, 5 (from top to bottom). The right column shows the corresponding

sequences $U_t^{(4)}$. As F increases, $U_t^{(3)}$ becomes more and more heavy-tailed, but the change-point at $t = 501$ appears to be exposed more clearly. The impact of increasing F on $U_t^{(4)}$ is less pronounced, but the change-point also seems to be exposed more clearly as F increases.

Since, typically, higher values of F will lead to better exposure of change-points but also introduce higher autocorrelation, it is desirable to choose F so as to obtain a trade-off between these two trends. Section 4.3 will discuss the proposed default choice of F based on an extensive simulation study.

4.3 Default choice of F and c through simulation

A large-scale simulation study was performed in which we assessed the empirical performance of our procedure for a variety of piecewise-stationary ARCH(1) and ARCH(2) models and sample sizes, for the dampening constant F (see Section 4.2) ranging from 1 to 10, and the threshold constant c (see Section 3.2) ranging from 0.1 to 1. The number of change-points ranged from 0 to 2, and, if present, they were located $2/3$ and $1/3$ the way through the series. Sample sizes varied from $n = 750$ to $n = 3000$.

It was found that the algorithm based on the sequence $U_t^{(4)}$ performed better than that based on $U_t^{(3)}$: this was due to the “noise” in $U_t^{(4)}$ having a more homogeneous structure due to the use of the log transform. Performance was surprisingly robust across all models tested with respect to the choice of F . We found that the value of c ranging in the interval $[0.4, 0.6]$ was the best choice in terms of the probability of correctly detecting the true number of change-points. The obvious exception were “null hypothesis” models not containing change-points, for which, as expected, higher values of c resulted in better performance than lower ones.

Without going into details of the many models tested, we provide an overall illustra-

tion of the above discussion in the form of “maps” showing the empirical probability of correctly detecting the true number of change-points in the case of $U_t^{(4)}$, jointly across all models. This is shown in Figure 3. Motivated by this empirical result, we propose that the following default values should be used with our procedure based on $U_t^{(4)}$: $c = 0.5$, $F = 8$.

5 Performance evaluation

In this brief comparative simulation study, we use our algorithm to re-examine the three examples of difficult-to-segment GARCH processes, reported in Davis *et al.* (2008), which appears to be the state-of-the-art procedure for change-point detection in GARCH models. We recall that a process Y_t follows a GARCH(p, q) model if it is defined as in (1) except σ_t^2 is defined as $\sigma_t^2 = a_0 + \sum_{i=1}^p a_i Y_{t-i}^2 + \sum_{j=1}^q b_j \sigma_{t-j}^2$.

Amongst other models, Davis *et al.* (2008) consider three challenging (from the point of view of segmentation) GARCH(1,1) models with sample size $n = 1000$, and with one change-point occurring in the triple (a_0, a_1, b_1) at time $t = 501$ as follows:

(a) $(0.4, 0.1, 0.5) \rightarrow (0.4, 0.1, 0.6)$,

(b) $(0.1, 0.1, 0.8) \rightarrow (0.1, 0.1, 0.7)$,

(c) $(0.4, 0.1, 0.5) \rightarrow (0.5, 0.1, 0.5)$.

Method	(a)	(b)	(c)
Auto-Seg	0.19	0.63	0.12
AG	0.24	0.75	0.14
BaSTA	0.38	0.77	0.26

Table 1: Proportion of times the correct number of change-points ($= 1$) has been detected in models (a), (b), (c), for the three competing methods.

Table 1 shows the proportion of simulation runs for which the correct number of change-points ($= 1$) has been detected, for three competing methods: that of Andreou

& Ghysels (2002) (AG), that of Davis *et al.* (2008) (Auto-Seg), and ours (BaSTA). The results for AG and Auto-Seg have been taken from Davis *et al.* (2008). Our method used the following default values: $c = 0.5$, $F = 8$ (as recommended in Section 4.3), and was based on the sequence $U_t^{(4)}$ and used order $p = 1$. 100 simulation runs were performed.

As Table 1 demonstrates, our method was superior to its competitors in these 3 challenging case studies.

A Proof of Theorem 3.1

By Theorem 3.1 in Fryzlewicz & Subba Rao (2008), X_t is strongly mixing with a geometric rate ρ . By a well-known result (see e.g. Theorem 14.1 of Davidson (1994)), any process

$$U_t = g(X_t, X_{t-1}, \dots, X_{t-\tau})$$

is also strong-mixing with the same geometric rate ρ . We represent $U_t = g_t + \varepsilon_t$, where g_t is strictly piecewise-constant and coincides with $\mathbb{E}g(\cdot)$ except for $t = \nu_j^{(n)} + 1, \dots, \nu_j^{(n)} + \tau$ for each j (i.e. around the change-points), and $\mathbb{E}(\varepsilon_t) = 0$ except, again, for $t = \nu_j^{(n)} + 1, \dots, \nu_j^{(n)} + \tau$ for each j . Note that even for those indices around the change-points, we have $\varepsilon_t \leq 2\bar{g}$, where $\bar{g} = \sup g(\cdot)$ (bear in mind that g is bounded).

The proof of our consistency Theorem 3.1 proceeds like the proof of Theorem 2.1 in Venkatraman (1993) [V93], except that

- we use a more general threshold $cn^{3/8}$;
- in our setup, ε_t is a bounded, strongly-mixing process with a geometric rate, rather than an i.i.d Gaussian sequence as in V93.

Following the proof of Theorem 2.1 in V93, it is straightforward to see that it is

also valid for the more general threshold $cn^{3/8}$. The only result which needs to be re-proved is our equivalent of V93's Lemma 2.1, which we prove below.

We construct W_{k_1, k_2}^k in the same way as V93 (using, in our notation, ε_k rather than his U_k). We split W_{k_1, k_2}^k into two parts: that associated with positive weights ($W_{k_1, k}$) and that associated with negative ones ($-W_{k, k_2}$). In other words we have $W_{k_1, k_2}^k = W_{k_1, k} - W_{k, k_2}$. We have

$$P \left\{ \max_{0 \leq k_1 \leq k \leq k_2 \leq n} |W_{k_1, k_2}^k| > \lambda \right\} \leq P \left\{ \max_{0 \leq k_1 \leq k \leq k_2 \leq n} |W_{k_1, k}| + |W_{k, k_2}| > \lambda \right\} \leq \\ P \left\{ \max_{0 \leq k_1 \leq k \leq n} |W_{k_1, k}| > \lambda/2 \right\} + P \left\{ \max_{0 \leq k \leq k_2 \leq n} |W_{k, k_2}| > \lambda/2 \right\} \leq 2 P \left\{ \max_{0 \leq k_1 \leq k \leq n} |W_{k_1, k}| > \lambda/2 \right\}.$$

We now bound the above probability in either of two ways, depending on the difference $k - k_1$. If $k - k_1$ is “small”, then exponential inequality bounds based on the mixing properties of ε_t do not kick in. We consider two cases:

1. $k - k_1$ “small” (we will later make this more precise). We have $W_{k_1, k} = \sum_{l=k_1}^k w_l \varepsilon_l$ where the weights w_l are all equal and sum to less than one in l_2 . Thus they are at most $\eta^{-1/2}$ where $\eta = k - k_1 + 1$. We bound

$$|W_{k_1, k}| \leq 2\bar{g}\eta^{1/2},$$

which is less than $\lambda/2$ as long as

$$\eta \leq \frac{\lambda^2}{16\bar{g}^2},$$

where λ is logarithmic. This is what we mean by “small” η .

2. $k - k_1$ “large”. In this case, $\eta > \frac{\lambda^2}{16\bar{g}^2}$, where we have freedom in choosing λ as long as it is of the order $O(\log^b n)$. Here, the main tool is Theorem 1.3(i) in Bosq (1998). In order to obtain a mean-zero process we first need to centre

the ε_t 's. Denote $\varepsilon'_t = \varepsilon_t - \mathbb{E} \varepsilon_t$ and recall from the above discussion that only at most τm (where m is the number of change-points) terms $\mathbb{E} \varepsilon_t$ are non-zero. Those that are not, are bounded by $2\bar{g}$. We bound

$$P \left\{ \max_{0 \leq k_1 \leq k \leq n} |W_{k_1, k}| > \lambda/2 \right\} \leq \sum_{0 \leq k_1 \leq k \leq n} P \{|W_{k_1, k}| > \lambda/2\} \leq n^2 \max_{0 \leq k_1 \leq k \leq n} P \{|W_{k_1, k}| > \lambda/2\}.$$

Further,

$$P \{|W_{k_1, k}| > \lambda/2\} = P \left\{ \left| \sum_{l=k_1}^k \eta^{-1/2} (\varepsilon'_l + \mathbb{E} \varepsilon_l) \right| > \lambda/2 \right\} \leq P \left\{ \left| \sum_{l=k_1}^k \eta^{-1/2} \varepsilon'_l \right| > \lambda/2 - 2\eta^{-1/2} \tau m \bar{g} \right\}$$

Denote $\tilde{\lambda} = \lambda/2 - 2\eta^{-1/2} \tau m \bar{g}$. Using formula (1.25) of Bosq (1998), we have

$$P \left\{ \left| \sum_{l=k_1}^k \eta^{-1/2} \varepsilon'_l \right| > \tilde{\lambda} \right\} \leq 4 \exp \left(-\frac{\tilde{\lambda}^2}{C_1 \eta} q(\eta, n) \right) + 22 \left(1 + \frac{C_2 \eta^{1/2}}{\tilde{\lambda}} \right)^{1/2} q(\eta, n) \alpha \left(\left[\frac{\eta}{2q(\eta, n)} \right] \right), \quad (5)$$

where C_1, C_2 are positive constants, $q(\eta, n)$ is an *arbitrary* integer in $[1, \dots, \eta/2]$, $[a]$ is the integer part of a , and $\alpha(k)$ are the α -mixing coefficients of X_t which are of order ρ^k . Suitable choice of $q(\eta, n)$ is crucial. We set it to be

$$q(\eta, n) = \eta/h(n),$$

where $h(n)$ is of the same order as $\tilde{\lambda}$ (to keep things simple). Clearly $q(\eta, n) \leq \eta/2$ as $h(n) \rightarrow \infty$ and also $q(\eta, n) \geq 1$ as $\eta > O(\tilde{\lambda}^2)$. With this choice of $q(\eta, n)$, the bound in (5) becomes at most

$$4 \exp \left(-\frac{\tilde{\lambda}}{C_3} \right) + C_4 n^{5/4} \rho^{2\tilde{\lambda}},$$

which converges to zero exponentially fast for a suitable logarithmic choice of

$\tilde{\lambda}$. This completes the proof of our counterpart of Lemma 2.1 in V93. Note that, for simplicity, we keep the number of change-points m constant, but we could have allowed it to increase to infinity with n at a logarithmic rate. It is a simple exercise to generalise the above proof to this case.

REFERENCES

- Andreou, E. & Ghysels, E. (2002). Detecting multiple breaks in financial market volatility dynamics. *J. Applied Econometrics* **17**, 579–600.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *J. Economet.* **31**, 307–27.
- Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes*. New York: Springer-Verlag.
- Brodsky, B. & Darkhovsky, B. (1993) *Nonparametric Methods in Change-Point Problems*. Dordrecht: Kluwer Academic Publishers.
- Chu, C.-S. J. (1995) Detecting parameter shift in GARCH models. *Econometric Reviews* **14**, 241–266.
- Dahlhaus, R. & Subba Rao, S. (2006). Statistical inference for time-varying ARCH processes. *Ann. Statist.* **34**, 1075–1114.
- Davidson, J. (1994). *Stochastic Limit Theory*. Oxford: Oxford University Press.
- Davis, R., Lee, T. & Rodriguez-Yam, G. (2008). Break detection for a class of nonlinear time series models. *J. Time Ser. Anal.* **29**, 834–867.
- Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987–1008.

- Fan, J. & Yao, Q. (2003). *Nonlinear Time Series*. New York: Springer-Verlag.
- Fryzlewicz, P. (2007). Unbalanced Haar technique for nonparametric function estimation. *J. Amer. Stat. Assoc.* **102**, 1318–1327.
- Fryzlewicz, P., Sapatinas, T. and Subba Rao, S. (2008) Normalized least-squares estimation in time-varying ARCH models. *Ann. Stat.* **36**, 742–786.
- Fryzlewicz, P. and Subba Rao, S. (2008) On some mixing properties of ARCH and time-varying ARCH processes. *Preprint*, available from http://stats.lse.ac.uk/fryzlewicz/mixing/tvARCH_mixing.pdf.
- Giraitis, L., Leipus, R. & Surgailis, D. (2005). Recent advances in ARCH modelling. In *Long Memory in Economics*, Ed. A Kirman and G. Teyssiere, pp. 3–39, Berlin: Springer.
- Janeway, W. (2009). Six impossible things before breakfast: Lessons from the crisis. *Significance* **6**, 28–31.
- Kokoszka, P. & Leipus, R. (2000). Change-point estimation in ARCH models. *Bernoulli* **6**, 513–39.
- Kulperger, R. & Yu, H. (2005). High moment partial sum processes of residuals in GARCH models and their applications. *Ann. Stat.* **33**, 2395–2422.
- Lavielle, M. & Moulines, E. (2000). Least-squares estimation of an unknown number of shifts in a time series. *J. Time Series Anal.* **21**, 33–59.
- Mikosch, T. & Stărică, C. (2004). Non-stationarities in financial time series, the long-range dependence, and the IGARCH effects. *Rev. Econ. Statist.* **86**, 378–90.
- Sen, A. & Srivastava, M.S. (1975). Some one-sided tests for changes in level. *Technometrics* **17**, 61–65.

Taylor, S.J. (1986). *Modelling Financial Time Series*. Chichester: John Wiley & Sons.

Venkatraman, E.S. (1993) *Consistency results in multiple change-point problems*. PhD thesis, Stanford University.

Vostrikova, L.J. (1981) Detecting “disorder” in multidimensional random processes. *Soviet Math. Dokl.* **24**, 55–9.

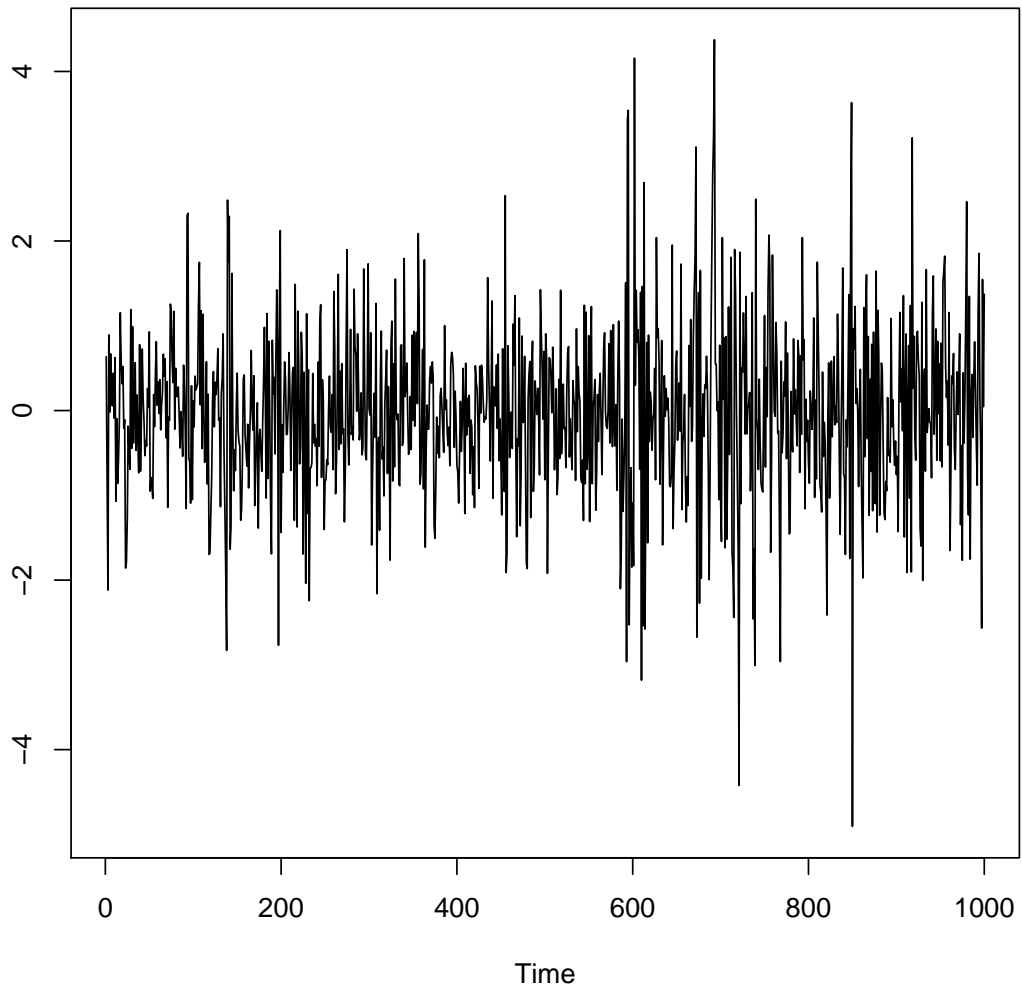


Figure 1: Realisation of a piecewise-stationary ARCH(1) process X_t of length $n = 1000$ where $a_0 \equiv 1$ and a_1 changes from 0.2 to 0.7 at $t = 501$.

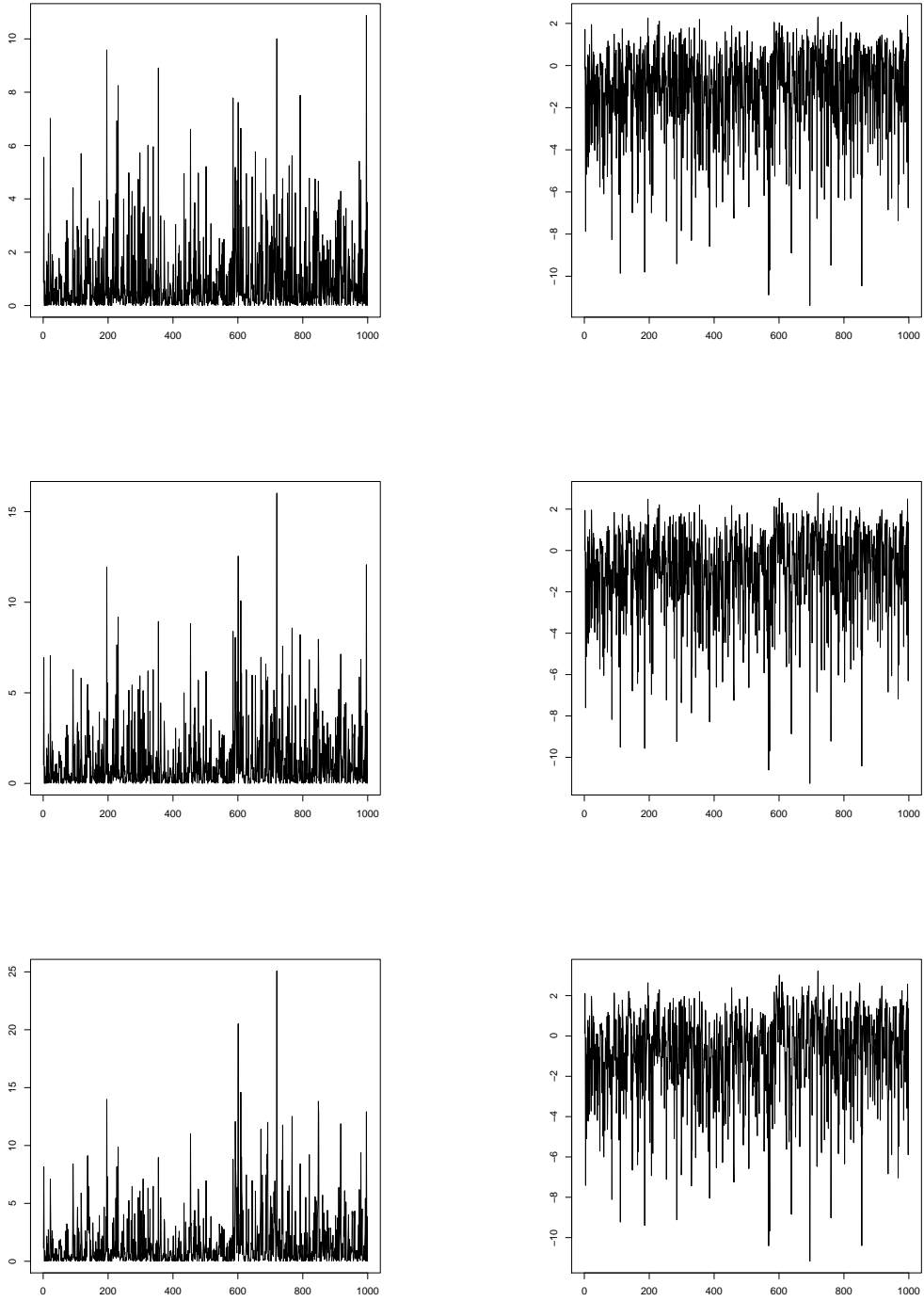


Figure 2: Sequences $U_t^{(3)}$ (left) and $U_t^{(4)}$ (right) corresponding to the process X_t from Figure 1. F takes the values of 1, 2, 5 (respectively from top to bottom).

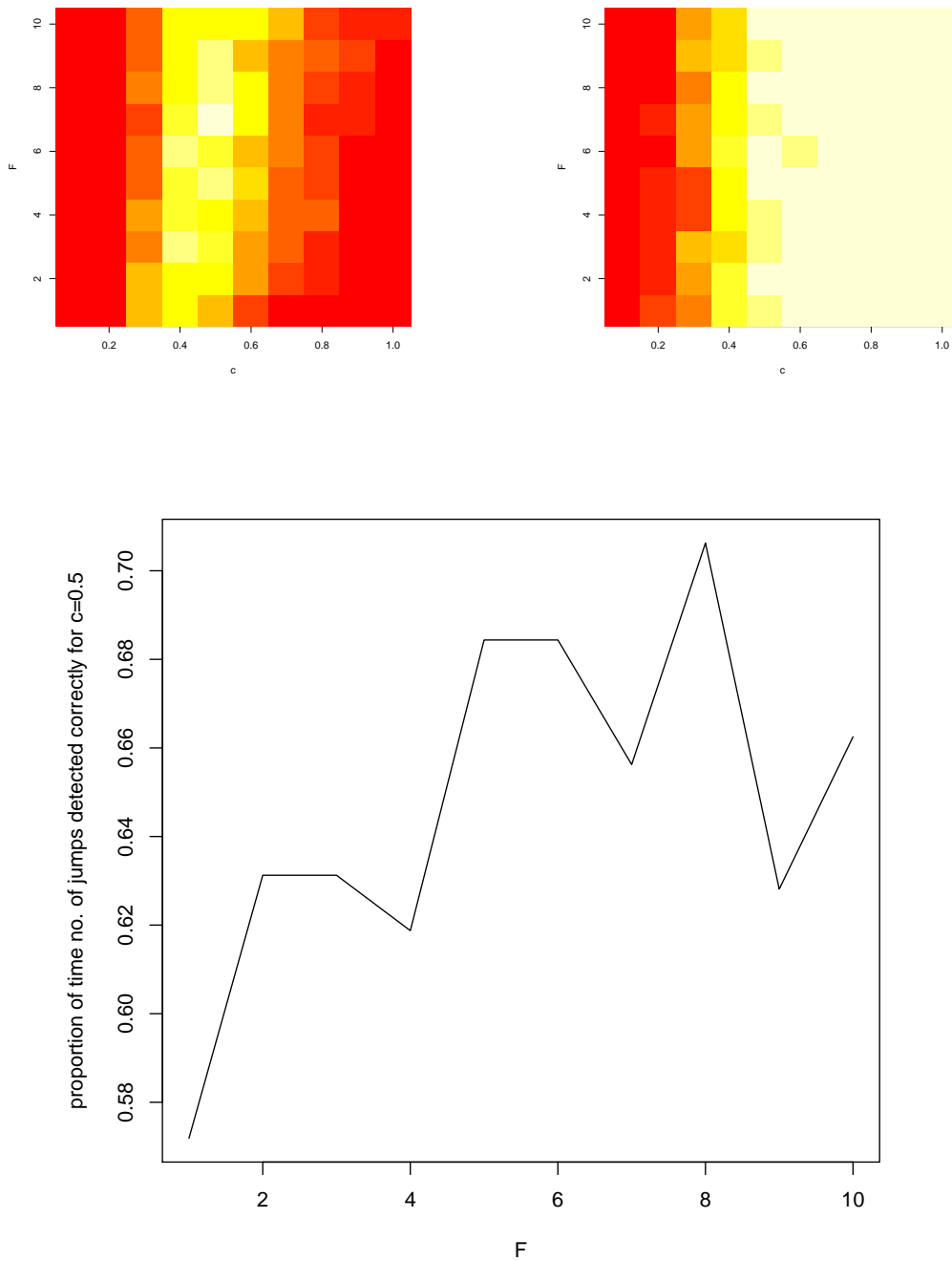


Figure 3: Top row: proportion of times number of change-points detected correctly across all models with (left) and without (right) change-points, as a function of F and c , when $U_t^{(4)}$ was used. Lighter colour means proportion higher. Bottom: proportion of times number of change-points detected correctly across all models, for $c = 0.5$ (default value), as a function of F .