The Econometrics Journal

Econometrics Journal (2009), volume **12**, pp. 248–271.
doi: 10.1111/j.1368-423X.2009.00292.x

Adaptive pointwise estimation in time-inhomogeneous conditional heteroscedasticity models

P. ČÍŽEK[†], W. HÄRDLE[‡] AND V. SPOKOINY[§]

[†]*Department of Econometrics & OR, Tilburg University, P.O. Box 90153, 5000LE Tilburg, The Netherlands*
E-mail: P.Cizek@uvt.nl

[‡]*Humboldt-Universität zu Berlin and CASE, Spandauerstrasse 1, 10178 Berlin, Germany*
E-mail: haerdle@wiwi.hu-berlin.de

[§]*Weierstrass-Institute, Humboldt-Universität zu Berlin and CASE, Mohrenstrasse 39, 10117 Berlin, Germany*
E-mail: spokoiny@wias-berlin.de

First version received: April 2008; final version accepted: April 2009

Summary This paper offers a new method for estimation and forecasting of the volatility of financial time series when the stationarity assumption is violated. Our general, local parametric approach particularly applies to general varying-coefficient parametric models, such as GARCH, whose coefficients may arbitrarily vary with time. Global parametric, smooth transition and change-point models are special cases. The method is based on an adaptive pointwise selection of the largest interval of homogeneity with a given right-end point by a local change-point analysis. We construct locally adaptive estimates that can perform this task and investigate them both from the theoretical point of view and by Monte Carlo simulations. In the particular case of GARCH estimation, the proposed method is applied to stock-index series and is shown to outperform the standard parametric GARCH model.

Keywords: *Adaptive pointwise estimation*, *Autoregressive models*, *Conditional heteroscedasticity models*, *Local time-homogeneity*.

1. INTRODUCTION

A growing amount of econometrical and statistical research is devoted to modelling financial time series and their volatility, which measures dispersion at a point in time (i.e. conditional variance). Although many economies and financial markets have been recently experiencing many shorter and longer periods of instability or uncertainty such as the Asian crisis (1997), the Russian crisis (1998), the start of the European currency (1999), the 'dot-Com' technology-bubble crash (2000–02) or the terrorist attacks (September, 2001), the war in Iraq (2003) and the current global recession (2008), mostly used econometric models are based on the assumption of time homogeneity. This includes linear and non-linear autoregressive (AR) and moving-average models and conditional heteroscedasticity (CH) models such as ARCH (Engel, 1982)

and GARCH (Bollerslev, 1986), stochastic volatility models (Taylor, 1986), as well as their combinations such as AR-GARCH.

On the other hand, the market and institutional changes have long been assumed to cause structural breaks in financial time series, which was confirmed, e.g. in data on stock prices (Andreou and Ghysels, 2002, and Beltratti and Morana, 2004) and exchange rates (Herwatz and Reimers, 2001). Moreover, ignoring these breaks can adversely affect the modelling, estimation and forecasting of volatility as suggested e.g. by Diebold and Inoue (2001), Mikosch and Starica (2004), Pesaran and Timmermann (2004) and Hillebrand (2005). Such findings led to the development of the change-point analysis in the context of CH models; see e.g. Chen and Gupta (1997), Kokoszka and Leipus (2000) and Andreou and Ghysels (2006).

An alternative approach lies in relaxing the assumption of time homogeneity and allowing some or all model parameters to vary over time (Chen and Tsay, 1993, Cai et al., 2000, and Fan and Zhang, 2008). Without structural assumptions about the transition of model parameters over time, time-varying coefficient models have to be estimated non-parametrically, e.g. under the identification condition that their parameters are smooth functions of time (Cai et al., 2000). In this paper, we follow a different strategy based on the assumption that a time series can be locally, i.e. over short periods of time, approximated by a parametric model. As suggested by Spokoiny (1998), such a local approximation can form a starting point in the search for the longest period of stability (homogeneity), i.e. for the longest time interval in which the series is described well by the parametric model. In the context of the local constant approximation, this strategy was employed for volatility modelling by Härdle et al. (2003), Mercurio and Spokoiny (2004) and Spokoiny (2009a). Our aim is to generalize this approach so that it can identify intervals of homogeneity for any parametric CH model regardless of its complexity.

In contrast to the local constant approximation of the volatility of a process (Mercurio and Spokoiny, 2004), the main benefit of the proposed generalization consists in the possibility to apply the methodology to a much wider class of models and to forecast over a longer time horizon. The reason is that approximating the mean or volatility process by a constant is in many cases too restrictive or even inappropriate and it is fulfilled only for short time intervals, which precludes its use for longer-term forecasting. On the contrary, parametric models like GARCH mimic the majority of stylized facts about financial time series and can reasonably fit the data over rather long periods of time in many practical situations. Allowing for time dependence of model parameters offers then much more flexibility in modelling real-life time series, which can be both with or without structural breaks since global parametric models are included as a special case.

Moreover, the proposed adaptive local parametric modelling unifies the change-point and varying-coefficient models. First, since finding the longest time-homogeneous interval for a parametric model at any point in time corresponds to detecting the most recent change-point in a time series, this approach resembles the change-point modelling as in Bai and Perron (1998) or Mikosch and Starica (1999, 2004), for instance, but it does not require prior information such as the number of changes. Additionally, the traditional structural-change tests require that the number of observations before each break point is large (and can grow to infinity) as these tests rely on asymptotic results. On the contrary, the proposed pointwise adaptive estimation does not rely on asymptotic results and does not thus place any requirements on the number of observations before, between or after any break point. Second, since the adaptively selected time-homogeneous interval used for estimation necessarily differs at each time point, the model coefficients can arbitrarily vary over time. In comparison to varying-coefficient models assuming

smooth development of parameters over time (Cai et al., 2000), our approach however allows for structural breaks in the form of sudden jumps in parameter values.

Although seemingly straightforward, extending Mercurio and Spokoiny's (2004) procedure to the local parametric modelling is a non-trivial problem, which requires new tools and techniques. We concentrate here on the change-point estimation of financial time series, which are often modelled by data-demanding models such as GARCH. While the benefits of a flexible change-point analysis for time series spanning several years are well known, its feasibility (which stands in the focus of this work) is much more difficult to achieve. The reason is thus that, at each time point, the procedure starts from a small interval, where a local parametric approximation holds, and then iteratively extends this interval and tests it for time-homogeneity until a structural break is found or data exhausted. Hence, a model has to be initially estimated on very short time intervals (e.g. 10 observations). Using standard testing methods, such a procedure might be feasible for simple parametric models, but it is hardly possible for more complex parametric models such as GARCH that generally require rather large samples for reasonably good estimates.

Therefore, we use an alternative and more robust approach to local change-point analysis that relies on a finite-sample theory of testing a growing sequence of historical time intervals on homogeneity against a change-point alternative. The proposed adaptive pointwise estimation procedure applies to a wide class of time-series models, including AR and CH models. Concentrating on the latter, we describe in details the adaptive procedure, derive its basic properties, and focusing on the feasibility of adaptive estimation for CH models, study the performance in comparison to the parametric (G)ARCH by means of simulations and real-data applications. The main conclusion is two-fold: on one hand, the adaptive pointwise estimation is feasible and beneficial also in the case of data-demanding models such as GARCH; on the other hand, the adaptive estimates based on various parametric models such as constant, ARCH or GARCH models are much closer to each other (while being better than the usual parametric estimates), which eliminates to some extent the need for using too complex models in adaptive estimation.

The rest of the paper is organized as follows. In Section 2, the parametric estimation of CH models and its finite-sample properties are introduced. In Section 3, we define the adaptive pointwise estimation procedure and discuss the choice of its parameters. Theoretical properties of the method are discussed in Section 4. In the specific case of the ARCH(1) and GARCH(1,1) models, a simulation study illustrates the performance of the new methodology with respect to the standard parametric and change-point models in Section 5. Applications to real stock-index series data are presented in Section 6. The proofs are provided in the Appendix.

2. PARAMETRIC CONDITIONAL HETEROSCEDASTICITY MODELS

Consider a time series Y_t in discrete time, $t \in N$. The CH assumption means that $Y_t = \sigma_t \varepsilon_t$, where $\{\varepsilon_t\}_{t \in N}$ is a white noise process and $\{\sigma_t\}_{t \in N}$ is a predictable volatility (conditional variance) process. Modelling of the volatility process σ_t typically relies on some parametric CH specification such as the ARCH (Engle, 1982) and GARCH (Bollerslev, 1986) models:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{2.1}$$

where $p \in N, q \in N$ and $\boldsymbol{\theta} = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)^\top$ is the parameter vector. An attractive feature of this model is that, even with very few coefficients, one can model most stylized facts of financial time series like volatility clustering or excessive kurtosis, for instance. A number of (G)ARCH extensions were proposed to make the model even more flexible; e.g. EGARCH (Nelson, 1991), QGARCH (Sentana, 1995) and TGARCH (Glosten et al., 1993) that account for asymmetries in a volatility process.

All such CH models can be put into a common class of generalized linear volatility models:

$$Y_t = \sigma_t \varepsilon_t = \sqrt{g(X_t)}\varepsilon_t, \tag{2.2}$$

$$X_t = \omega + \sum_{i=1}^{p} \alpha_i h(Y_{t-i}) + \sum_{j=1}^{q} \beta_j X_{t-j}\,, \tag{2.3}$$

where g and h are known functions and X_t is a (partially) unobserved process (structural variable) that models the volatility coefficient σ_t^2 via transformation $g : \sigma_t^2 = g(X_t)$. For example, the GARCH model (2.1) is described by $g(u) = u$ and $h(r) = r^2$.

Models (2.2)–(2.3) are time homogeneous in the sense that the process Y_t follows the same structural equation at each time point. In other words, the parameter $\boldsymbol{\theta}$ and hence the structural dependence in Y_t is constant over time. Even though models like (2.2)–(2.3) can often fit data well over a longer period of time, the assumption of homogeneity is too restrictive in practical applications: to guarantee a sufficient amount of data for sufficiently precise estimation, these models are often applied over time spans of many years. On the contrary, the strategy pursued here requires only local time homogeneity, which means that at each time point t there is a (possibly rather short) interval $[t - m, t]$, where the process Y_t is well described by models (2.2)–(2.3). This strategy aims then both at finding an interval of homogeneity (preferably as long as possible) and at the estimation of the corresponding parameter values $\boldsymbol{\theta}$, which then enable predicting Y_t and X_t.

Next, we discuss the parameter estimation for models (2.2)–(2.3) using observations Y_t from some time interval $I = [t_0, t_1]$. The conditional distribution of each observation Y_t given the past $\mathscr{F}_{t-1}$ is determined by the structural variable X_t, whose dynamics are described by the parameter vector $\boldsymbol{\theta}$: $X_t = X_t(\boldsymbol{\theta})$ for $t \in I$ due to (2.3). We denote the underlying value of $\boldsymbol{\theta}$ by $\boldsymbol{\theta}_0$.

For estimating $\boldsymbol{\theta}_0$, we apply the quasi-maximum likelihood (quasi-MLE) approach using the estimating equations generated under the assumption of Gaussian errors ε_t. This guarantees efficiency under the normality of innovations and consistency under rather general moment conditions (Hansen and Lee, 1994, and Francq and Zakoian, 2007). The log-likelihood for models (2.2)–(2.3) on an interval I can be represented in the form

$$L_I(\boldsymbol{\theta}) = \sum_{t \in I} \ell\{Y_t, g[X_t(\boldsymbol{\theta})]\}$$

with log-likelihood function $\ell(y, \upsilon) = -0.5\{\log(\upsilon) + y^2/\upsilon\}$. We define the quasi-MLE estimate $\widetilde{\boldsymbol{\theta}}_I$ of the parameter $\boldsymbol{\theta}$ by maximizing the log-likelihood $L_I(\boldsymbol{\theta})$,

$$\widetilde{\boldsymbol{\theta}}_I = \operatorname*{argmax}_{\boldsymbol{\theta} \in \Theta} L_I(\boldsymbol{\theta}) = \operatorname*{argmax}_{\boldsymbol{\theta} \in \Theta} \sum_{t \in I} \ell\{Y_t, g[X_t(\boldsymbol{\theta})]\}, \tag{2.4}$$

and denote by $L_I(\widetilde{\boldsymbol{\theta}}_I)$ the corresponding maximum.

To characterize the quality of estimating the parameter vector $\boldsymbol{\theta}_0 = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)^\top$ by $\widetilde{\boldsymbol{\theta}}_I$, we now present an exact (non-asymptotic) exponential risk bound. This bound concerns the value of maximum $L_I(\widetilde{\boldsymbol{\theta}}_I) = \max_{\boldsymbol{\theta}\in\Theta} L_I(\boldsymbol{\theta})$ rather than the point of maximum $\widetilde{\boldsymbol{\theta}}_I$. More precisely, we consider the difference $L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta}_0) = L_I(\widetilde{\boldsymbol{\theta}}_I) - L_I(\boldsymbol{\theta}_0)$. By definition, this value is non-negative and represents the deviation of the maximum of the log-likelihood process from its value at the 'true' point $\boldsymbol{\theta}_0$. Later, we comment on how the accuracy of estimation of the parameter $\boldsymbol{\theta}_0$ by $\widetilde{\boldsymbol{\theta}}_I$ relates to the value $L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta}_0)$. We will also see that the bound for $L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta}_0)$ yields the confidence set for the parameter $\boldsymbol{\theta}_0$, which will be used for the proposed change-point test. Now, the non-asymptotic risk bound is specified in the following theorem, which formulates corollaries 4.2 and 4.3 of Spokoiny (2009b) for the case of the quasi-MLE estimation of a CH model (2.2)–(2.3) at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. The result can be viewed as an extension of the Wilks phenomenon that the distribution of $L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta}_0)$ for a linear Gaussian model is $\chi^2_p/2$, where p is the number of estimated parameters in the model.

THEOREM 2.1. *Assume that the process Y_t follows models (2.2)–(2.3) with the parameter $\boldsymbol{\theta}_0 \in \Theta$, where the set Θ is compact. The function $g(\cdot)$ is assumed to be continuously differentiable with the uniformly bounded first derivative and $g(x) \geq \delta > 0$ for all x. Further, let the process $X_t(\boldsymbol{\theta})$ be sub-ergodic in the sense that for any smooth function $f(\cdot)$ there exists f^* such that for any time interval I*

$$\boldsymbol{E}_{\boldsymbol{\theta}_0}\left|\sum_I \left\{f(X_t(\boldsymbol{\theta})) - \boldsymbol{E}_{\boldsymbol{\theta}_0} f(X_t(\boldsymbol{\theta}))\right\}\right|^2 \leq f^*|I|, \quad \boldsymbol{\theta} \in \Theta.$$

Finally, let $\boldsymbol{E}\exp\{\varkappa(\varepsilon_t^2 - 1)|\mathscr{F}_{t-1}\} \leq c(\varkappa)$ for some $\varkappa > 0, c(\varkappa) > 0$, and all $t \in N$. Then there are $\lambda > 0$ and $\mathfrak{e}(\lambda, \boldsymbol{\theta}_0) > 0$ such that for any interval I and $\mathfrak{z} > 0$

$$\boldsymbol{P}_{\boldsymbol{\theta}_0}\big(L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta}_0) > \mathfrak{z}\big) \leq \exp\{\mathfrak{e}(\lambda, \boldsymbol{\theta}_0) - \lambda\mathfrak{z}\}. \tag{2.5}$$

Moreover, for any $r > 0$, there is a constant $\mathfrak{R}_r(\boldsymbol{\theta}_0)$ such that

$$\boldsymbol{E}_{\boldsymbol{\theta}_0}\big|L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta}_0)\big|^r \leq \mathfrak{R}_r(\boldsymbol{\theta}_0). \tag{2.6}$$

REMARK 2.1. The condition $g(x) \geq \delta > 0$ guarantees that the variance process cannot reach zero. In the case of GARCH, it is sufficient to assume $\omega > 0$, for instance.

One attractive feature of Theorem 2.1, formulated in the following corollary, is that it enables constructing the non-asymptotic confidence sets and testing the parametric hypothesis on the basis of the fitted log-likelihood $L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta})$. This feature is especially important for our procedure presented in Section 3.

COROLLARY 2.1. *Under the assumptions of Theorem 2.1, let the value $\mathfrak{z}_\alpha$ fulfil $\mathfrak{e}(\lambda, \boldsymbol{\theta}_0) - \lambda\mathfrak{z}_\alpha < \log\alpha$ for some $\alpha < 1$. Then the random set $\mathscr{E}_I(\mathfrak{z}_\alpha) = \{\boldsymbol{\theta} : L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta}) \leq \mathfrak{z}_\alpha\}$ is an α-confidence set for $\boldsymbol{\theta}_0$ in the sense that $\boldsymbol{P}_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0 \notin \mathscr{E}_I(\mathfrak{z}_\alpha)) \leq \alpha$.*

Theorem 2.1 also gives a non-asymptotic and fixed upper bound for the risk of estimation $L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta}_0)$ that applies to an arbitrary sample size $|I|$. To understand the relation of this result to the classical rate result, we can apply the standard arguments based on the quadratic expansion

of the log-likelihood $L(\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta})$. Let $\nabla^2 L(\boldsymbol{\theta})$ denote the Hessian matrix of the second derivatives of $L(\boldsymbol{\theta})$ with respect to the parameter $\boldsymbol{\theta}$. Then

$$L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta}_0) = 0.5(\widetilde{\boldsymbol{\theta}}_I - \boldsymbol{\theta}_0)^\top \nabla^2 L_I(\boldsymbol{\theta}'_I)(\widetilde{\boldsymbol{\theta}}_I - \boldsymbol{\theta}_0), \tag{2.7}$$

where $\boldsymbol{\theta}'_I$ is a convex combination of $\boldsymbol{\theta}_0$ and $\widetilde{\boldsymbol{\theta}}_I$. Under usual regularity assumptions and for sufficiently large $|I|$, the normalized matrix $|I|^{-1}\nabla^2 L_I(\boldsymbol{\theta})$ is close to some matrix $V(\boldsymbol{\theta})$, which depends only on the stationary distribution of Y_t and is continuous in $\boldsymbol{\theta}$. Then (2.5) approximately means that $\|\sqrt{V(\boldsymbol{\theta}_0)}(\widetilde{\boldsymbol{\theta}}_I - \boldsymbol{\theta}_0)\|^2 \le \mathfrak{z}/|I|$ with probability close to 1 for large $\mathfrak{z}$. Hence, the large deviation result of Theorem 2.1 yields the root-$|I|$ consistency of the MLE estimate $\widetilde{\boldsymbol{\theta}}_I$. See Spokoiny (2009b) for further details.

3. POINTWISE ADAPTIVE NON-PARAMETRIC ESTIMATION

An obvious feature of models (2.2)–(2.3) is that the parametric structure of the process is assumed constant over the whole sample and cannot thus incorporate changes and structural breaks at unknown times in the models. A natural generalization leads to models whose coefficients may change over time (Fan and Zhang, 2008). One can then assume that the structural process X_t satisfies the relation (2.3) at any time, but the vector of coefficients $\boldsymbol{\theta}$ may vary with the time t, $\boldsymbol{\theta} = \boldsymbol{\theta}(t)$. The estimation of the coefficients as general functions of time is possible only under some additional assumptions on these functions. Typical assumptions are (i) varying coefficients are smooth functions of time (Cai et al., 2000) and (ii) varying coefficients are piecewise constant functions (Bai and Perron, 1998, and Mikosch and Starica, 1999, 2004).

Our local parametric approach differs from the commonly used identification assumptions (i) and (ii). We assume that the observed data Y_t are described by a (partially) unobserved process X_t due to (2.2), and at each point T, there exists a historical interval $I(T) = [t_0, T]$ in which the process X_t 'nearly' follows the parametric specification (2.3) (see Section 4 for details on what 'nearly' means). This local structural assumption enables us to apply well-developed parametric estimation for data $\{Y_t\}_{t \in I(T)}$ to estimate the underlying parameter $\boldsymbol{\theta} = \boldsymbol{\theta}(T)$ by $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(T)$. (The estimate $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(T)$ can then be used for estimating the value $\hat{X}_T$ of the process X_t at T from equation (2.3) and for further modelling such as forecasting Y_{T+1}.) Moreover, this assumption includes the above-mentioned 'smooth transition' and 'switching regime' assumptions (i) and (ii) as special cases: parameters $\hat{\boldsymbol{\theta}}(T)$ vary over time as the interval $I(T)$ changes with T and, at the same time, discontinuities and jumps in $\hat{\boldsymbol{\theta}}(T)$ as a function of time are possible.

To estimate $\hat{\boldsymbol{\theta}}(T)$, we have to find the historical interval of homogeneity $I(T)$, i.e. the longest interval I with the right-end point T, where data do not contradict a specified parametric model with fixed parameter values. Starting at each time T with a very short interval $I = [t_0, T]$, we search by successive extending and testing of interval I on homogeneity against a change-point alternative: if the hypothesis of homogeneity is not rejected for a given I, a larger interval is taken and tested again. Contrary to Bai and Perron (1998) and Mikosch and Starica (1999), who detect all change points in a given time series, our approach is local: it focuses on the local change-point analysis near point T of estimation and tries to find only one change closest to the reference point.

In the rest of this section, we first discuss the test statistics employed to test the time-homogeneity of an interval I against a change-point alternative in Section 3.1. Later, we rigorously describe the pointwise adaptive estimation procedure in Section 3.2. Its

implementation and the choice of parameters entering the adaptive procedure are described in Sections 3.2–3.4. Theoretical properties of the method are studied in Section 4.

3.1. Test of homogeneity against a change-point alternative

The pointwise adaptive estimation procedure crucially relies on the test of local time-homogeneity of an interval $I = [t_0, T]$. The null hypothesis for I means that the observations $\{Y_t\}_{t \in I}$ follow the parametric models (2.2)–(2.3) with a fixed parameter $\boldsymbol{\theta}_0$, leading to the quasi-MLE estimate $\widetilde{\boldsymbol{\theta}}_I$ from (2.4) and the corresponding fitted log-likelihood $L_I(\widetilde{\boldsymbol{\theta}}_I)$.

The change-point alternative for a given change-point location $\tau \in I$ can be described as follows: process Y_t follows the parametric models (2.2)–(2.3) with a parameter $\boldsymbol{\theta}_J$ for $t \in J = [t_0, \tau]$ and with a different parameter $\boldsymbol{\theta}_{J^c}$ for $t \in J^c = [\tau + 1, T]; \boldsymbol{\theta}_J \neq \boldsymbol{\theta}_{J^c}$. The fitted log-likelihood under this alternative reads as $L_J(\widetilde{\boldsymbol{\theta}}_J) + L_{J^c}(\widetilde{\boldsymbol{\theta}}_{J^c})$. The test of homogeneity can be performed using the likelihood ratio (LR) test statistic $T_{I,\tau}$:

$$T_{I,\tau} = \max_{\boldsymbol{\theta}_J, \boldsymbol{\theta}_{J^c} \in \Theta} \{L_J(\boldsymbol{\theta}_J) + L_{J^c}(\boldsymbol{\theta}_{J^c})\} - \max_{\boldsymbol{\theta} \in \Theta} L_I(\boldsymbol{\theta}) = \left\{L_J(\widetilde{\boldsymbol{\theta}}_J) + L_{J^c}(\widetilde{\boldsymbol{\theta}}_{J^c}) - L_I(\widetilde{\boldsymbol{\theta}}_I)\right\}.$$

Since the change-point location τ is generally not known, we consider the supremum of the LR statistics $T_{I,\tau}$ over some subset $\tau \in \mathscr{T}(I)$; cf. Andrews (1993):

$$T_{I,\mathscr{T}(I)} = \sup_{\tau \in \mathscr{T}(I)} T_{I,\tau}. \tag{3.1}$$

A typical example of a set $\mathscr{T}(I)$ is $\mathscr{T}(I) = \{\tau : t_0 + m' \leq \tau \leq T - m''\}$ for some fixed $m', m'' > 0$.

3.2. Adaptive search for the longest interval of homogeneity

This section presents the proposed adaptive pointwise estimation procedure. At each point T, we aim at estimating the unknown parameters $\boldsymbol{\theta}(T)$ from historical data $Y_t, t \leq T$; this procedure repeats for every current time point T as new data arrive. At the first step, the procedure selects on the base of historical data an interval $\hat{I}(T)$ of homogeneity in which the data do not contradict the parametric models (2.2)–(2.3). Afterwards, the quasi-MLE estimation is applied using the selected historical interval $\hat{I}(T)$ to obtain estimate $\hat{\boldsymbol{\theta}}(T) = \widetilde{\boldsymbol{\theta}}_{\hat{I}(T)}$. From now on, we consider an arbitrary, but fixed time point T.

Suppose that a growing set $I_0 \subset I_1 \subset \cdots \subset I_K$ of historical interval-candidates $I_k = [T - m_k + 1, T]$ with the right-end point T is fixed. The smallest interval I_0 is accepted automatically as homogeneous. Then the procedure successively checks every larger interval I_k on homogeneity using the test statistic $T_{I_k,\mathscr{T}(I_k)}$ from (3.1). The selected interval $\hat{I}$ corresponds to the largest accepted interval $I_{\hat{k}}$ with index $\hat{k}$ such that

$$T_{I_k,\mathscr{T}(I_k)} \leq \mathfrak{z}_k, \quad k \leq \hat{k}, \tag{3.2}$$

and $T_{I_{\hat{k}+1},\mathscr{T}(I_{\hat{k}+1})} > \mathfrak{z}_{\hat{k}+1}$, where the critical values $\mathfrak{z}_k$ are discussed later in this section and specified in Section 3.3. This procedure then leads to the adaptive estimate $\hat{\boldsymbol{\theta}} = \widetilde{\boldsymbol{\theta}}_{\hat{I}}$ corresponding to the selected interval $\hat{I} = I_{\hat{k}}$.

The complete description of the procedure includes two steps. (A) Fixing the set-up and the parameters of the procedure. (B) Data-driven search for the longest interval of homogeneity.

(A) Set-up and parameters:

1 Select specific parametric models (2.2)–(2.3) [e.g. constant volatility, ARCH(1), GARCH(1,1)].
2 Select the set $\mathscr{I} = (I_0, \ldots, I_K)$ of interval-candidates, and for each $I_k \in \mathscr{I}$, the set $\mathscr{T}(I_k)$ of possible change points $\tau \in I_k$ used in the LR test (3.1).
3 Select the critical values $\mathfrak{z}_1, \ldots, \mathfrak{z}_K$ in (3.2) as described in Section 3.3.

(B) Adaptive search and estimation: Set $k = 1$, $\hat{I} = I_0$ and $\hat{\boldsymbol{\theta}} = \widetilde{\boldsymbol{\theta}}_{I_0}$.

1 Test the hypothesis $H_{0,k}$ of no change point within the interval I_k using test statistics (3.1) and the critical values $\mathfrak{z}_k$ obtained in (A3). If a change point is detected ($H_{0,k}$ is rejected), go to (B3). Otherwise proceed with (B2).
2 Set $\hat{\boldsymbol{\theta}} = \widetilde{\boldsymbol{\theta}}_{I_k}$ and $\hat{\boldsymbol{\theta}}_{I_k} = \widetilde{\boldsymbol{\theta}}_{I_k}$. Further, set $k := k + 1$. If $k \le K$, repeat (B1); otherwise go to (B3).
3 Define $\hat{I} = I_{k-1} =$ 'the last accepted interval' and $\hat{\boldsymbol{\theta}} = \widetilde{\boldsymbol{\theta}}_{\hat{I}}$. Additionally, set $\hat{\boldsymbol{\theta}}_{I_k} = \cdots = \hat{\boldsymbol{\theta}}_{I_K} = \hat{\boldsymbol{\theta}}$ if $k \le K$.

In step (A), one has to select three main ingredients of the procedure. First, the parametric model used locally to approximate the process Y_t has to be specified in (A1), e.g. the constant volatility or GARCH(1,1) in our context. Next, in step (A2), the set of intervals $\mathscr{I} = \{I_k\}_{k=0}^K$ is fixed, each interval with the right-end point T, length $m_k = |I_k|$, and the set $\mathscr{T}(I_k)$ of tested change points. Our default proposal is to use a geometric grid $m_k = [m_0 a^k]$, $a > 1$, and to set $I_k = [T - m_k + 1, T]$ and $\mathscr{T}(I_k) = [T - m_{k-1} + 1, T - m_{k-2}]$. Although our experiments show that the procedure is rather insensitive to the choice of m_0 and a (e.g. we use $m_0 = 10$ and $a = 1.25$ in simulations), the length m_0 of interval I_0 should take into account the parametric model selected in (A1). The reason is that I_0 is always assumed to be time-homogeneous and m_0 thus has to reflect flexibility of the parametric model; e.g. while $m_0 = 20$ might be reasonable for the GARCH(1,1) model, $m_0 = 5$ could be a reasonable choice for the locally constant approximation of a volatility process. Finally, in step (A3), one has to select the K critical values $\mathfrak{z}_k$ in (3.2) for the LR test statistics $T_{I_k, \mathscr{T}(I_k)}$ from (3.1). The critical values $\mathfrak{z}_k$ will generally depend on the parametric model describing the null hypothesis of time-homogeneity, the set $\mathscr{I}$ of intervals I_k and corresponding sets of considered change points $\mathscr{T}(I_k)$, $k \le K$, and additionally, on two constants r and ρ that are counterparts of the usual significance level. All these determinants of the critical values can be selected in step (A) and the critical values are thus obtained before the actual estimation takes place in step (B). Due to its importance, the method of constructing critical values $\{\mathfrak{z}_k\}_{k=1}^K$ is discussed separately in Section 3.3.

The main step (B) performs the search for the longest time-homogeneous interval. Initially, I_0 is assumed to be homogeneous. If I_{k-1} is negatively tested on the presence of a change point, one continues with I_k by employing test (3.1) in step (B1), which checks for a potential change point in I_k. If no change point is found, then I_k is accepted as time-homogeneous in step (B2); otherwise the procedure terminates in step (B3). We sequentially repeat these tests until we find a change point or exhaust all intervals. The latest (longest) interval accepted as time-homogeneous is used for estimation in step (B3). Note that the estimate $\hat{\boldsymbol{\theta}}_{I_k}$ defined in (B2) and (B3) corresponds to the latest accepted interval $\hat{I}_k$ after the first k steps, or equivalently, the interval selected out of $I_1, \ldots, I_k$.

Moreover, the whole search and estimation step (B) can be repeated at different time points T without reiterating the initial step (A) as the critical values $\mathfrak{z}_k$ depend only on the approximating parametric model and interval lengths $m_k = |I_k|$, not on the time point T (see Section 3.3).

3.3. Choice of critical values $\mathfrak{z}_k$

The presented method of choosing the interval of homogeneity $\hat{I}$ can be viewed as multiple testing procedure. The critical values for this procedure are selected using the general approach of testing theory: to provide a prescribed performance of the procedure under the null hypothesis, i.e. in the pure parametric situation. This means that the procedure is trained on the data generated from the pure parametric time-homogeneous model from step (A1). The correct choice in this situation is the largest considered interval I_K and a choice $I_{\hat{k}}$ with $\hat{k} < K$ can be interpreted as a 'false alarm'. We select the minimal critical values ensuring a small probability of such a false alarm. Our condition slightly differs though from the classical level condition because we focus on parameter estimation rather than on hypothesis testing.

In the pure parametric case, the 'ideal' estimate corresponds to the largest considered interval I_K. Due to Theorem 2.1, the quality of estimation of the parameter $\boldsymbol{\theta}_0$ by $\widetilde{\boldsymbol{\theta}}_{I_K}$ can be measured by the log-likelihood 'loss' $L_{I_K}(\widetilde{\boldsymbol{\theta}}_{I_K}, \boldsymbol{\theta}_0)$, which is stochastically bounded with exponential and polynomial moments: $\boldsymbol{E}_{\boldsymbol{\theta}_0}|L_{I_K}(\widetilde{\boldsymbol{\theta}}_{I_K}, \boldsymbol{\theta}_0)|^r \le \mathfrak{R}_r(\boldsymbol{\theta}_0)$. If the adaptive procedure stops earlier at some intermediate step $k < K$, we select instead of $\widetilde{\boldsymbol{\theta}}_{I_K}$ another estimate $\hat{\boldsymbol{\theta}} = \widetilde{\boldsymbol{\theta}}_{I_k}$ with a larger variability. The loss associated with such a false alarm can be measured by the value $L_{I_K}(\widetilde{\boldsymbol{\theta}}_{I_K}, \hat{\boldsymbol{\theta}}) = L_{I_K}(\widetilde{\boldsymbol{\theta}}_{I_K}) - L_{I_K}(\hat{\boldsymbol{\theta}})$. The corresponding condition bounding the loss due to the adaptive estimation reads as

$$\boldsymbol{E}_{\boldsymbol{\theta}_0}\big|L_{I_K}(\widetilde{\boldsymbol{\theta}}_{I_K}, \hat{\boldsymbol{\theta}})\big|^r \le \rho\mathfrak{R}_r(\boldsymbol{\theta}_0). \tag{3.3}$$

This is in fact an implicit condition on the critical values $\{\mathfrak{z}_k\}_{k=1}^K$, which ensures that the loss associated with the false alarm is at most the ρ-fraction of the log-likelihood loss of the 'ideal' or 'oracle' estimate $\widetilde{\boldsymbol{\theta}}_{I_K}$ for the parametric situation. The constant r corresponds to the power of the loss in (3.3), while ρ is similar in meaning to the test level. In the limit case when r tends to zero, this condition (3.3) becomes the usual level condition: $\boldsymbol{P}_{\boldsymbol{\theta}_0}(I_K \text{ is rejected}) = \boldsymbol{P}_{\boldsymbol{\theta}_0}(\widetilde{\boldsymbol{\theta}}_{I_K} \ne \hat{\boldsymbol{\theta}}) \le \rho$. The choice of the metaparameters r and ρ is discussed in Section 3.4.

A condition similar to (3.3) is imposed at each step of the adaptive procedure. The estimate $\hat{\boldsymbol{\theta}}_{I_k}$ coming after the k steps of the procedure should satisfy

$$\boldsymbol{E}_{\boldsymbol{\theta}_0}\big|L_{I_k}(\widetilde{\boldsymbol{\theta}}_{I_k}, \hat{\boldsymbol{\theta}}_{I_k})\big|^r \le \rho_k\mathfrak{R}_r(\boldsymbol{\theta}_0), \quad k = 1, \dots, K, \tag{3.4}$$

where $\rho_k = \rho\, k/K \le \rho$. The following theorem presents some sufficient conditions on the critical values $\{\mathfrak{z}_k\}_{k=1}^K$ ensuring (3.4); recall that $m_k = |I_k|$ denotes the length of I_k.

THEOREM 3.1. *Suppose that $r > 0, \rho > 0$. Under the assumptions of Theorem 2.1, there are constants a_0, a_1, a_2 such that the condition (3.4) is fulfilled with the choice*

$$\mathfrak{z}_k = a_0 r \log(\rho^{-1}) + a_1 r \log(m_K/m_{k-1}) + a_2 \log(m_k), \quad k = 1, \dots, K.$$

Since K and $\{m_k\}_{k=1}^K$ are fixed, the $\mathfrak{z}_k$'s in Theorem 3.1 have a form $\mathfrak{z}_k = C + D\log(m_k)$ for $k = 1, \dots, K$ with some constant C and D. However, a practically relevant choice of these constants has to be done by Monte Carlo simulations. Note first that every particular choice of the coefficients C and D determines the whole set of the critical values $\{\mathfrak{z}_k\}_{k=1}^K$ and thus the local change-point procedure. For the critical values given by fixed (C, D), one can run the procedure and observe its performance on the simulated data using the data-generating process (2.2)–(2.3); in particular, one can check whether the condition (3.4) is fulfilled. For any (sufficiently large) fixed value of C, one can thus find the minimal value $D(C) < 0$ of D that ensures (3.4).

Every corresponding set of critical values in the form $\mathfrak{z}_k = C + D(C)\log(m_k)$ is admissible. The condition $D(C) < 0$ ensures that the critical values decreases with k. This reflects the fact that a false alarm at an early stage of the algorithm is more crucial because it leads to the choice of a highly variable estimate. The critical values $\mathfrak{z}_k$ for small k should thus be rather conservative to provide the stability of the algorithm in the parametric situation. To determine C, the value $\mathfrak{z}_1$ can be fixed by considering the false alarm at the first step of the procedure, which leads to estimation using the smallest interval I_0 instead of the 'ideal' largest interval I_K. The related condition (used in Section 5.1) reads as

$$\boldsymbol{E}_{\boldsymbol{\theta}_0}\big|L_{I_K}(\widetilde{\boldsymbol{\theta}}_{I_K}, \widetilde{\boldsymbol{\theta}}_{I_0})\big|^r \mathbf{1}(T_{I_1,\mathscr{T}(I_1)} > \mathfrak{z}_1) \le \rho\mathfrak{R}_r(\boldsymbol{\theta}_0)/K. \tag{3.5}$$

Alternatively, one could select a pair (C, D) that minimizes the resulting prediction error; see Section 3.4.

3.4. Selecting parameters r and ρ

The choice of critical values using inequality (3.4) additionally depends on two 'metaparameters' r and ρ. A simple strategy is to use conservative values for these parameters and the corresponding set of critical values (e.g. our default is $r = 1$ and $\rho = 1$). On the other hand, the two parameters are global in the sense that they are independent of T. Hence, one can also determine them in a data-driven way by minimizing some global forecasting error (Cheng et al., 2003). Different values of r and ρ may lead to different sets of critical values and hence to different estimates $\hat{\boldsymbol{\theta}}^{(r,\rho)}(T)$ and to different forecasts $\hat{Y}^{(r,\rho)}_{T+h|T}$ of the future values Y_{T+h}, where h is the forecasting horizon. Now, a data-driven choice of r and ρ can be done by minimizing the following objective function:

$$(\hat{r}, \hat{\rho}) = \operatorname*{arg\,min}_{r\ge 0, \rho>0} PE_{\Lambda,\mathscr{H}}(r, \rho) = \operatorname*{arg\,min}_{r,\rho} \sum_T \sum_{h\in\mathscr{H}} \Lambda\big(Y_{T+h}, \hat{Y}^{(r,\rho)}_{T+h|T}\big), \tag{3.6}$$

where Λ is a loss function and $\mathscr{H}$ is the forecasting horizon set. For example, one can take $\Lambda_r(\upsilon, \upsilon') = |\upsilon - \upsilon'|^r$ for $r \in [1/2, 2]$. For daily data, the forecasting horizon could be one day, $\mathscr{H} = \{1\}$, or two weeks, $\mathscr{H} = \{1, \ldots, 10\}$.

4. THEORETIC PROPERTIES

In this section, we collect basic results describing the quality of the proposed adaptive procedure. First, the definition of the procedure ensures the performance prescribed by (3.4) in the parametric situation. We however claimed that the adaptive pointwise estimation applies even if the process Y_t is only locally approximated by a parametric model. Therefore, we now define a locally 'nearly parametric' process, for which we derive an analogy of Theorem 2.1 (Section 4.1). Later, we prove certain 'oracle' properties of the proposed method (Section 4.2).

4.1. Small modelling bias condition

This section discusses the concept of a 'nearly parametric' case. To define it rigorously, we have to quantify the quality of approximating the true latent process X_t, which drives the observed data Y_t due to (2.2), by the parametric process $X_t(\boldsymbol{\theta})$ described by (2.3) for some $\boldsymbol{\theta} \in \Theta$. Below

we assume that the innovations ε_t in the model (2.2) are independent and identically distributed and denote the distribution of $\sqrt{\upsilon}\varepsilon_t$ by P_υ so that the conditional distribution of Y_t given $\mathscr{F}_{t-1}$ is $P_{g(X_t)}$. To measure the distance of a data-generating process from a parametric model, we introduce for every interval $I_k \in \mathscr{I}$ and every parameter $\boldsymbol{\theta} \in \Theta$ the random quantity

$$\Delta_{I_k}(\boldsymbol{\theta}) = \sum_{t \in I_k} \mathscr{K}\{g(X_t), g[X_t(\boldsymbol{\theta})]\},$$

where $\mathscr{K}(\upsilon, \upsilon')$ denotes the Kullback–Leibler distance between P_υ and $P_{\upsilon'}$. For CH models with Gaussian innovations ε_t, $\mathscr{K}(\upsilon, \upsilon') = -0.5\{\log(\upsilon/\upsilon') + 1 - \upsilon/\upsilon'\}$. In the parametric case with $X_t = X_t(\boldsymbol{\theta}_0)$, we clearly have $\Delta_{I_k}(\boldsymbol{\theta}_0) = 0$. To characterize the 'nearly parametric case', we introduce a {small modelling bias} (SMB) condition, which simply means that, for some $\boldsymbol{\theta} \in \Theta$, $\Delta_{I_k}(\boldsymbol{\theta})$ is bounded by a small constant with a high probability. Informally, this means that the 'true' model can be well approximated on the interval I_k by the parametric one with the parameter $\boldsymbol{\theta}$. The best parametric fit (2.3) to the underlying model (2.2) on I_k can be defined by minimizing the value $\boldsymbol{E}\Delta_{I_k}(\boldsymbol{\theta})$ over $\boldsymbol{\theta} \in \Theta$ and $\widetilde{\boldsymbol{\theta}}_{I_k}$ can be viewed as its estimate.

The following theorem claims that the results on the accuracy of estimation given in Theorem 2.1 can be extended from the parametric case to the general non-parametric situation under the SMB condition. Let $\varrho(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})$ be any loss function for an estimate $\hat{\boldsymbol{\theta}}$.

THEOREM 4.1. *Let for some $\boldsymbol{\theta} \in \Theta$ and some $\Delta \geq 0$*

$$\boldsymbol{E}\Delta_{I_k}(\boldsymbol{\theta}) \leq \Delta. \tag{4.1}$$

Then it holds for an estimate $\hat{\boldsymbol{\theta}}$ constructed from the observations $\{Y_t\}_{t \in I_k}$ that

$$\boldsymbol{E}\log\bigl(1 + \varrho(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})/\boldsymbol{E}_{\boldsymbol{\theta}}\varrho(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})\bigr) \leq 1 + \Delta.$$

This general result applied to the quasi-MLE estimation with the loss function $L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta})$ yields the following corollary.

COROLLARY 4.1. *Let the SMB condition (4.1) hold for some interval I_k and $\boldsymbol{\theta} \in \Theta$. Then*

$$\boldsymbol{E}\log\Bigl(1 + \bigl|L_{I_k}(\widetilde{\boldsymbol{\theta}}_{I_k}, \boldsymbol{\theta})\bigr|^r/\mathfrak{R}_r(\boldsymbol{\theta})\Bigr) \leq 1 + \Delta,$$

where $\mathfrak{R}_r(\boldsymbol{\theta})$ is the parametric risk bound from (2.6).

This result shows that the estimation loss $|L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta})|^r$ normalized by the parametric risk $\mathfrak{R}_r(\boldsymbol{\theta})$ is stochastically bounded by a constant proportional to e^Δ. If Δ is not large, this result extends the parametric risk bound (Theorem 2.1) to the non-parametric situation under the SMB condition. Another implication of Corollary 4.1 is that the confidence set built for the parametric model (Corollary 2.1) continues to hold, with a slightly smaller coverage probability, under SMB.

4.2. The 'oracle' choice and the 'oracle' result

Corollary 4.1 suggests that the 'optimal' or 'oracle' choice of the interval I_k from the set $I_1, \ldots, I_K$ can be defined as the largest interval for which the SMB condition (4.1) still holds (for a given small $\Delta > 0$). For such an interval, one can neglect deviations of the underlying

process from a parametric model with a fixed parameter $\boldsymbol{\theta}$. Therefore, we say that the choice k^* is the 'oracle' choice if there exists $\boldsymbol{\theta} \in \Theta$ such that

$$\boldsymbol{E}\,\Delta_{I_{k^*}}(\boldsymbol{\theta}) \le \Delta \tag{4.2}$$

for a fixed $\Delta > 0$ and that (4.2) does not hold for $k > k^*$. Unfortunately, the underlying process X_t and, hence, the value Δ_{I_k} is unknown and the oracle choice cannot be implemented. The proposed adaptive procedure tries to mimic this oracle on the basis of available data using the sequential test of homogeneity. The final oracle result claims that the adaptive estimate provides the same (in order) accuracy as the oracle one.

By construction, the pointwise adaptive procedure described in Section 3 provides the prescribed performance if the underlying process follows the parametric model (2.2). Now, condition (3.4) combined with Theorem 4.1 implies similar performance in the first k^* steps of the adaptive estimation procedure.

THEOREM 4.2. *Let $\boldsymbol{\theta} \in \Theta$ and $\Delta > 0$ be such that $\boldsymbol{E}\,\Delta_{I_{k^*}}(\boldsymbol{\theta}) \le \Delta$ for some $k^* \le K$. Also let $\max_{k \le k^*} \boldsymbol{E}_{\boldsymbol{\theta}} |L_{I_k}(\widetilde{\boldsymbol{\theta}}_{I_k}, \boldsymbol{\theta})|^r \le \mathfrak{R}_r(\boldsymbol{\theta})$. Then*

$$\boldsymbol{E} \log\left(1 + \frac{\left|L_{I_{k^*}}\left(\widetilde{\boldsymbol{\theta}}_{I_{k^*}}, \boldsymbol{\theta}\right)\right|^r}{\mathfrak{R}_r(\boldsymbol{\theta})}\right) \le 1 + \Delta \quad \textit{and} \quad \boldsymbol{E} \log\left(1 + \frac{\left|L_{I_{k^*}}\left(\widetilde{\boldsymbol{\theta}}_{I_{k^*}}, \hat{\boldsymbol{\theta}}_{I_{k^*}}\right)\right|^r}{\mathfrak{R}_r(\boldsymbol{\theta})}\right) \le \rho + \Delta.$$

Similarly to the parametric case, under the SMB condition $\boldsymbol{E}\,\Delta_{I_{k^*}}(\boldsymbol{\theta}) \le \Delta$, any choice $\hat{k} < k^*$ can be viewed as a false alarm. Theorem 4.2 documents that the loss induced by such a false alarm at the first k^* steps and measured by $L_{I_{k^*}}(\widetilde{\boldsymbol{\theta}}_{I_{k^*}}, \hat{\boldsymbol{\theta}}_{I_{k^*}})$ is of the same magnitude as the loss $L_{I_{k^*}}(\widetilde{\boldsymbol{\theta}}_{I_{k^*}}, \boldsymbol{\theta})$ of estimating the parameter $\boldsymbol{\theta}$ from the SMB (4.2) by $\widetilde{\boldsymbol{\theta}}_{I_{k^*}}$. Thus, under (4.2) the adaptive estimation during steps $k \le k^*$ does not induce larger errors into estimation than the quasi-MLE estimation itself.

For further steps of the algorithm with $k > k^*$, where (4.2) does not hold, the value $\Delta' = \boldsymbol{E}\,\Delta_{I_k}(\boldsymbol{\theta})$ can be large and the bound for the risk becomes meaningless due to the factor $e^{\Delta'}$. To establish the result about the quality of the final estimate, we thus have to show that the quality of estimation cannot be destroyed at the steps $k > k^*$. The next 'oracle' result states the final quality of our adaptive estimate $\hat{\boldsymbol{\theta}}$.

THEOREM 4.3. *Let $\boldsymbol{E}\,\Delta_{I_{k^*}}(\boldsymbol{\theta}) \le \Delta$ for some $k^* \le K$. Then $L_{I_{k^*}}(\widetilde{\boldsymbol{\theta}}_{I_{k^*}}, \hat{\boldsymbol{\theta}})\mathbf{1}(\hat{k} \ge k^*) \le \mathfrak{z}_{k^*}$ yielding*

$$\boldsymbol{E} \log\left(1 + \frac{\left|L_{I_{k^*}}\left(\widetilde{\boldsymbol{\theta}}_{I_{k^*}}, \hat{\boldsymbol{\theta}}\right)\right|^r}{\mathfrak{R}_r(\boldsymbol{\theta})}\right) \le \rho + \Delta + \log\left(1 + \frac{\mathfrak{z}_{k^*}^r}{\mathfrak{R}_r(\boldsymbol{\theta})}\right).$$

Due to this result, the value $L_{I_{k^*}}(\widetilde{\boldsymbol{\theta}}_{I_{k^*}}, \hat{\boldsymbol{\theta}})$ is stochastically bounded. This can be interpreted as the oracle property of $\hat{\boldsymbol{\theta}}$ because it means that the adaptive estimate $\hat{\boldsymbol{\theta}}$ belongs with a high probability to the confidence set of the oracle estimate $\widetilde{\boldsymbol{\theta}}_{I_{k^*}}$.

5. SIMULATION STUDY

In the last two sections, we present simulation study (Section 5) and real data applications (Section 6) documenting the performance of the proposed adaptive estimation procedure. To verify the practical applicability of the method in a complex setting, we concentrate on the volatility estimation using parametric and adaptive pointwise estimation of constant volatility, ARCH(1) and GARCH(1,1) models (for the sake of brevity, referred to as the local constant,

local ARCH and local GARCH). The reason is that the estimation of GARCH models requires generally hundreds of observations for a reasonable quality of estimation, which puts the adaptive procedure working with samples as small as 10 or 20 observations to a hard test. Additionally, the critical values obtained as described in Section 3.3 depend on the underlying parameter values in the case of (G)ARCH.

Here we first study the finite-sample critical values for the test of homogeneity by means of Monte Carlo simulations and discuss practical implementation details (Section 5.1). Later, we demonstrate the performance of the proposed adaptive pointwise estimation procedure in simulated samples (Section 5.2). Note that, throughout this section, we identify the GARCH(1,1) models by triplets (ω, α, β): e.g. (1, 0.1, 0.3)-model. Constant volatility and ARCH(1) are then indicated by $\alpha = \beta = 0$ and $\beta = 0$, respectively. The GARCH estimation is done using the GARCH 3.0 package (Laurent and Peters, 2006) and Ox 3.30 (Doornik, 2002). Finally, since the focus is on modelling the volatility σ_t^2 in (2.2), the performance measurement and comparison of all models at time t is done by the absolute prediction error (PE) of the volatility process over a prediction horizon $\mathscr{H}$: $\mathrm{APE}(t) = \sum_{h\in\mathscr{H}} |\sigma_{t+h}^2 - \hat{\sigma}_{t+h|t}^2|/|\mathscr{H}|$, where $\hat{\sigma}_{t+h|t}^2$ represents the volatility prediction by a particular model.

5.1. *Finite-sample critical values for the test of homogeneity*

A practical application of the pointwise adaptive procedure requires critical values for the test of local homogeneity of a time series. Since they are obtained under the null hypothesis that a chosen parametric model (locally) describes the data, see Section 3, we need to obtain the critical values for the constant volatility, ARCH(1) and GARCH(1,1) models. Furthermore, for given r and ρ, the average risk (3.4) between the adaptive and oracle estimates can be bounded for critical values that linearly depend on the logarithm of interval length $|I_k| : \mathfrak{z}(|I_k|) = \mathfrak{z}_k = C + D\log(|I_k|)$ (see Theorem 3.1). As described in Section 3.3, we choose here the smallest C satisfying (3.5) and the corresponding minimum admissible value $D = D(C) < 0$ that guarantees the conditions (3.4).

We simulated the critical values for ARCH(1) and GARCH(1,1) models with different values of underlying parameters; see Table 1 for the critical values corresponding to $r = 1$ and $\rho = 1$. Their simulation was performed sequentially on intervals with lengths ranging from $|I_0| = m_0 = 10$ to $|I_K| = 570$ observations using a geometric grid with multiplier $a = 1.25$; see Section 3.2. (The results are, however, not sensitive to the choice of a.)

Unfortunately, the critical values depend on the parameters of the underlying (G)ARCH model (in contrast to the constant-volatility model). They generally seem to increase with the values of the ARCH and GARCH parameters keeping the other one fixed; see Table 1. To deal with this dependence on the underlying model parameters, we propose to choose the largest (most conservative) critical values corresponding to any estimated parameter in the analysed data. For example, if the largest estimated parameters of GARCH(1,1) are $\hat{\alpha} = 0.3$ and $\hat{\beta} = 0.8$, one should use $\mathfrak{z}(10) = 26.4$ and $\mathfrak{z}(570) = 14.5$, which are the largest critical values for models with $\alpha = 0.3, \beta \le 0.8$ and with $\alpha \le 0.3$, $\beta = 0.8$. (The proposed procedure is, however, not overly sensitive to this choice, as we shall see later.)

Finally, let us have a look at the influence of the tuning constants r and ρ in (3.4) on the critical values for several selected models (Table 2). The influence is significant, but can be classified in the following way. Whereas increasing ρ generally leads to an overall decrease of critical values (cf. Theorem 3.1), but primarily for the longer intervals, increasing r leads to an increase of

Table 1. Critical values $\mathfrak{z}_k = \mathfrak{z}(|I_k|)$ of the supremum LR test.

$\mathfrak{z}(\lvert I_k\rvert)$		β									
α	$\lvert I_k\rvert$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.0	10	15.5	15.5	16.4	16.8	17.9	17.3	17.0	17.0	16.9	16.0
	570	5.5	7.2	7.0	7.0	7.5	7.5	7.4	7.3	7.0	6.7
0.1	10	16.3	14.5	15.1	15.9	16.4	15.9	16.1	16.0	16.0	
	570	8.6	9.0	9.1	9.6	9.8	10.7	11.5	12.5	14.0	
0.2	10	16.7	15.2	15.7	16.2	16.9	18.9	20.1	25.1		
	570	9.4	10.6	11.2	11.4	11.4	12.5	13.3	14.2		
0.3	10	18.5	16.4	16.7	16.9	18.1	21.8	26.4			
	570	9.7	10.8	12.0	12.4	12.9	13.5	14.5			
0.4	10	22.1	16.5	18.3	19.3	22.8	30.9				
	570	9.9	12.0	13.0	13.4	13.9	14.7				
0.5	10	26.2	19.1	19.5	25.4	38.1					
	570	10.7	12.6	13.8	14.0	14.6					
0.6	10	33.0	22.8	25.9	32.4						
	570	12.7	12.7	13.9	15.3						
0.7	10	41.1	24.8	29.1							
	570	16.8	14.7	16.1							
0.8	10	66.2	26.4								
	570	31.5	15.8								
0.9	10	88.6									
	570	60.9									

Note: $\omega = 1, r = 1$ and $\rho = 1$.

critical values mainly for the shorter intervals; cf. (3.4). In simulations and real applications, we verified that a fixed choice such as $r = 1$ and $\rho = 1$ performs well. To optimize the performance of the adaptive methods, one can however determine constants r and ρ in a data-dependent way as described in Section 3.3. We use here this strategy for a small grid of $r \in \{0.5, 1.0\}$ and $\rho \in \{0.5, 1.0, 1.5\}$ and find globally optimal r and ρ. We will document, though, that the differences in the average absolute PE (3.6) for various values of r and ρ are relatively small.

5.2. Simulation study

We aim (i) to examine how well the proposed estimation method is able to adapt to long stable (time-homogeneous) periods and to less stable periods with more frequent volatility changes and (ii) to see which adaptively estimated model—local volatility, local ARCH or local GARCH—performs best in different regimes. To this end, we simulated 100 series from two change-point GARCH models with a low GARCH effect (ω, 0.2, 0.1) and a high GARCH effect (ω, 0.2, 0.7). Changes in constant ω are spread over a time span of 1000 days; see Figure 1. There is a long stable period at the beginning (500 days $\approx$ 2 years) and end (250 days $\approx$ 1 year) of time series with several volatility changes between them.

Table 2. Critical values $\mathfrak{z}(|I_k|)$ of the supremum LR test for various values r and ρ.

Model (ω, α, β)		(0.1, 0.0, 0.0)		(0.1, 0.2, 0.0)		(0.1, 0.1, 0.8)	
r	ρ	$\mathfrak{z}(10)$	$\mathfrak{z}(570)$	$\mathfrak{z}(10)$	$\mathfrak{z}(570)$	$\mathfrak{z}(10)$	$\mathfrak{z}(570)$
1.0	0.5	16.3	7.3	17.4	11.2	18.7	17.1
1.0	1.0	15.4	5.5	16.7	9.4	16.0	14.0
1.0	1.5	14.9	4.5	15.9	8.3	15.2	13.4
0.5	0.5	10.7	7.1	11.7	10.1	11.7	10.1
0.5	1.0	8.9	5.5	10.3	8.5	10.3	8.5
0.5	1.5	7.7	4.6	9.3	7.5	9.3	7.5

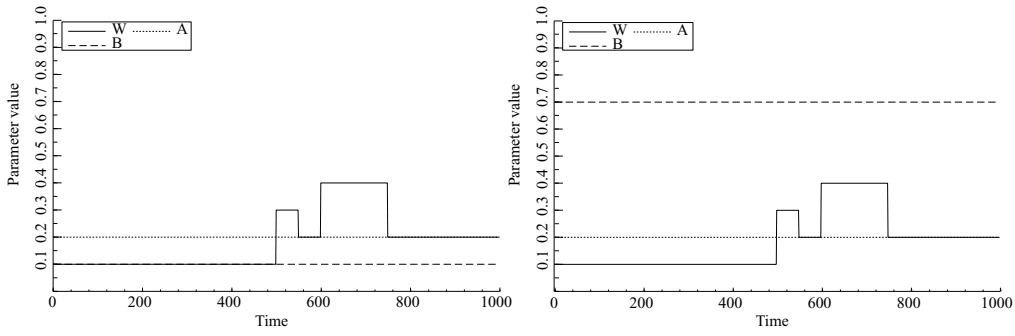

Figure 1. GARCH(1,1) parameters of low (left panel) and high (right panel) GARCH-effect simulations.

5.2.1. Low GARCH effect. Let us now discuss simulation results from the low GARCH-effect model. First, we mention the effect of structural changes in time series on the parameter estimation. Later, we compare the performance of all methods in terms of absolute PE.

Estimating a parametric model from data containing a change point will necessarily lead to various biases in estimation. For example, Hillebrand (2005) demonstrates that a change in volatility level ω within a sample drives the GARCH parameter β very close to 1. This is confirmed when we analyse the parameter estimates for parametric and adaptive GARCH at each time point $t \in [250, 1000]$ as depicted on Figure 2, where the mean (solid line), the 10% and 90% quantiles (dotted lines), and the true values (thick dotted line) of the model parameters are provided. The parametric estimates are consistent before breaks starting at $t = 500$, but the GARCH parameter β becomes inconsistent and converges to 1 once data contain breaks, $t > 500$. The locally adaptive estimates are similar to parametric ones before the breaks and become rather imprecise after the first change point, but they are not too far from the true value on average and stay consistent (in the sense that the confidence interval covers the true values). The low precision of estimation can be attributed to rather short intervals used for estimation (cf. Figure 2 for $t <$ 500).

Next, we would like to compare the performance of parametric and adaptive estimation methods by means of absolute PE: first for the prediction horizon of one day, $\mathscr{H} = \{1\}$, and later for prediction two weeks ahead, $\mathscr{H} = \{1, \ldots, 10\}$. To make the results easier to decipher,

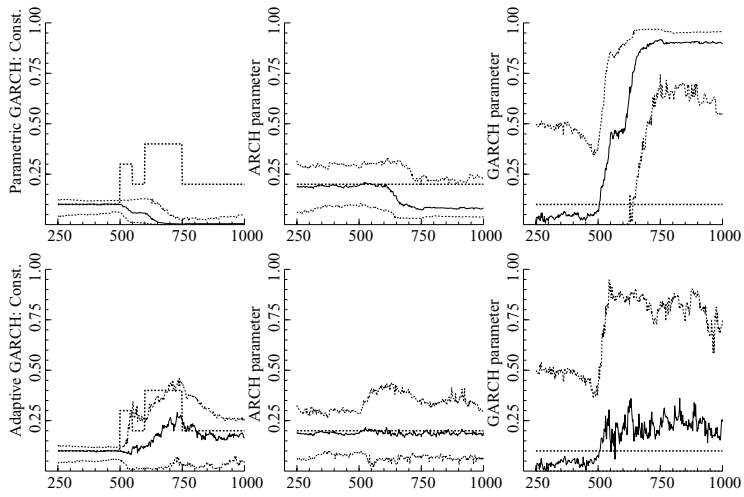

Figure 2. Parameter values estimated by the parametric (top row) and locally adaptive (bottom row) GARCH methods.

we present in what follows PEs averaged over the past month (21 days). The absolute-PE criterion was also used to determine the optimal values of parameters r and ρ (jointly across all simulations and for all $t = 250, \ldots, 1000$). The results differ for different models: $r = 0.5$, $\rho = 0.5$ for local constant, $r = 0.5$, $\rho = 1.0$ for local ARCH, and $r = 0.5$, $\rho = 1.5$ for local GARCH.

Let us now compare the adaptively estimated local constant, local ARCH and local GARCH models with the parametric GARCH, which is the best performing parametric model in this set-up. Forecasting one period ahead, the average PEs for all methods and the median lengths of the selected time-homogeneous intervals for adaptive methods are presented on Figure 3 for $t \in [250, 1000]$. First of all, let us observe in the case of the simplest local constant model that even the (median) estimated interval of homogeneity at the end of the first homogeneous period, $1 \leq t < 500$, can actually be shorter than the true one. The reason is that the probability of some 5 or 10 subsequent observations used as I_0 having their sample variance very different from the underlying one increases with the length of the series.

Next, one can notice that all methods are sensitive to jumps in volatility, especially to the first one at $t = 500$: the parametric ones because they ignore a structural break, the adaptive ones because they use a small amount of data after a structural change. In general, the local GARCH performs rather similarly to the parametric GARCH for $t < 650$ because it uses all historical data. After initial volatility jumps, the local GARCH, however, outperforms the parametric one, $650 < t < 775$. Following the last jump at $t = 750$, where the volatility level returns closer to the initial one, the parametric GARCH is best of all methods for some time, $775 < t < 850$, until the adaptive estimation procedure detects the (last) break, and after it, 'collects' enough observations for estimation. Then the local GARCH and local ARCH become preferable to the parametric model again, $850 < t$. Interestingly, the local ARCH approximation performs almost as well as both GARCH methods and even outperforms them shortly after structural breaks (except for break at $t = 750$), $600 < t < 775$ and $850 < t < 1000$. Finally, the local constant

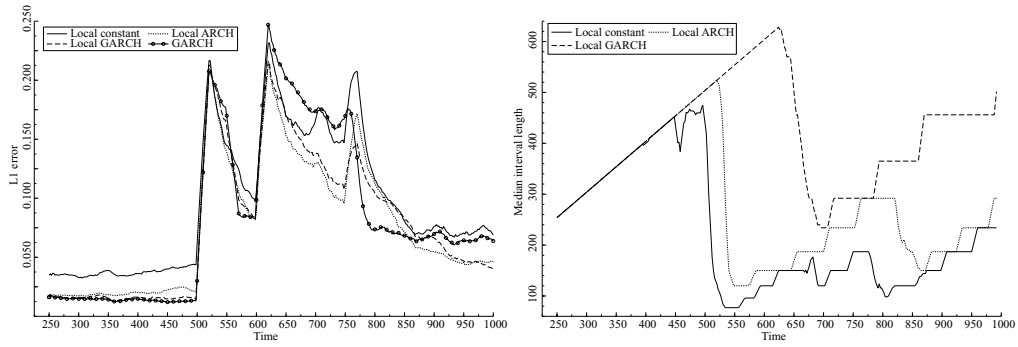

Figure 3. Left-hand panel: Low GARCH-effect simulations—absolute prediction errors one period ahead. Right-hand panel: The median lengths of the adaptively selected intervals.

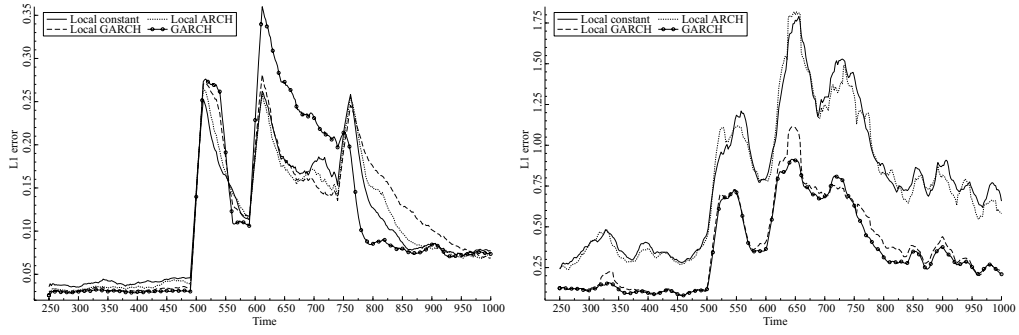

Figure 4. Left-hand panel: Low GARCH-effect simulations—absolute prediction errors 10 periods ahead. Right-hand panel: High GARCH-effect simulations—absolute prediction errors one period ahead.

volatility is lacking behind the other two adaptive methods whenever there is a longer time period without a structural break, but keeps up with them in periods with frequent volatility changes, $500 < t < 650$. All these observations can be documented also by the absolute PE averaged over the whole period $250 \leq t \leq 1000$ (we refer to it as the global PE from now on): the smallest PE is achieved by local ARCH (0.075), then by local GARCH (0.079) and the worst result is from local constant (0.094).

Additionally, all models are compared using the forecasting horizon of 10 days. Most of the results are the same (e.g. parameter estimates) or similar (e.g. absolute PE) to forecasting one period ahead due to the fact that all models rely on at most one past observation. The absolute PEs averaged over one month are summarized for $t \in [250, 1000]$ on Figure 4, which reveals that the difference between local constant volatility, local ARCH and local GARCH models are smaller in this case. As a result, it is interesting to note that: (i) the local constant model becomes a viable alternative to the other methods (it has in fact the smallest global PE 0.107 from all adaptive methods) and (ii) the local ARCH model still outperforms the local GARCH (global

PEs are 0.108 and 0.116, respectively) even though the underlying model is GARCH (with a small value of $\beta = 0.1$ however).

5.2.2. High GARCH effect. Let us now discuss the high GARCH-effect model. One would expect much more prevalent behaviour of both GARCH models, since the underlying GARCH parameter is higher and the changes in the volatility level ω are likely to be small compared to overall volatility fluctuations. Note that the optimal values of tuning constant r and ρ differ from the low GARCH-effect simulations: $r = 0.5$, $\rho = 1.5$ for local constant; $r = 0.5$, $\rho = 1.5$ for local ARCH; and $r = 1.0$, $\rho = 0.5$ for local GARCH.

Comparing the absolute PEs for the one-period-ahead forecast at each time point (Figure 4) indicates that the adaptive and parametric GARCH estimations perform approximately equally well. On the other hand, both the parametric and adaptively estimated ARCH and constant volatility models are lacking significantly. Unreported results confirm, similarly to the low GARCH-effect simulations, that the differences among method are much smaller once a longer prediction horizon of 10 days is used.

6. APPLICATIONS

The proposed adaptive pointwise estimation method will be now applied to real time series consisting of the log-returns of the DAX and S&P 500 stock indices (Sections 6.1 and 6.2). We will again summarize the results concerning both parametric and adaptive methods by the absolute PEs one day ahead averaged over one month. As a benchmark, we employ the parametric GARCH estimated using the last two years of data (500 observations). Since we however do not have the underlying volatility process now, it is approximated by squared returns. Despite being noisy, this approximation is unbiased and provides usually the correct ranking of methods (Andersen and Bollerslev, 1998).

6.1. DAX analysis

Let us now analyse the log-returns of the German stock index DAX from January 1990 till December 2002 depicted at the top of Figure 5. Several periods interesting for comparing the performance of parametric and adaptive pointwise estimates are selected since results for the whole period might be hard to decipher at once.

First, consider the estimation results for years 1991 to 1996. Contrary to later periods, there are structural breaks practically immediately detected by all adaptive methods (July 1991 and June 1992; cf. Stapf and Werner, 2003). For the local GARCH, this differs from less pronounced structural changes discussed later, which are typically detected only with delays of several months. One additional break detected by all methods occurs in October 1994. Note that parameters r and ρ were $r = 0.5$, $\rho = 1.5$ for local constant, $r = 1.0$, $\rho = 1.0$ for local ARCH, and $r = 0.5$, $\rho = 1.5$ for local GARCH.

The results for the period 1991–96 are summarized in the left bottom panel of Figure 5, which depicts the PEs of each adaptive method relative to the PEs of parametric GARCH. First, one can notice that the local constant and local ARCH approximations are preferable till July 1991, where we have less than 500 observations. After the detection of the structural change in June 1991, all adaptive methods are shortly worse than the parametric GARCH due to the limited amount of data used, but then outperform the parametric GARCH till the next structural break in the second half of 1992. A similar behaviour can be observed after the break detected in October 1994,

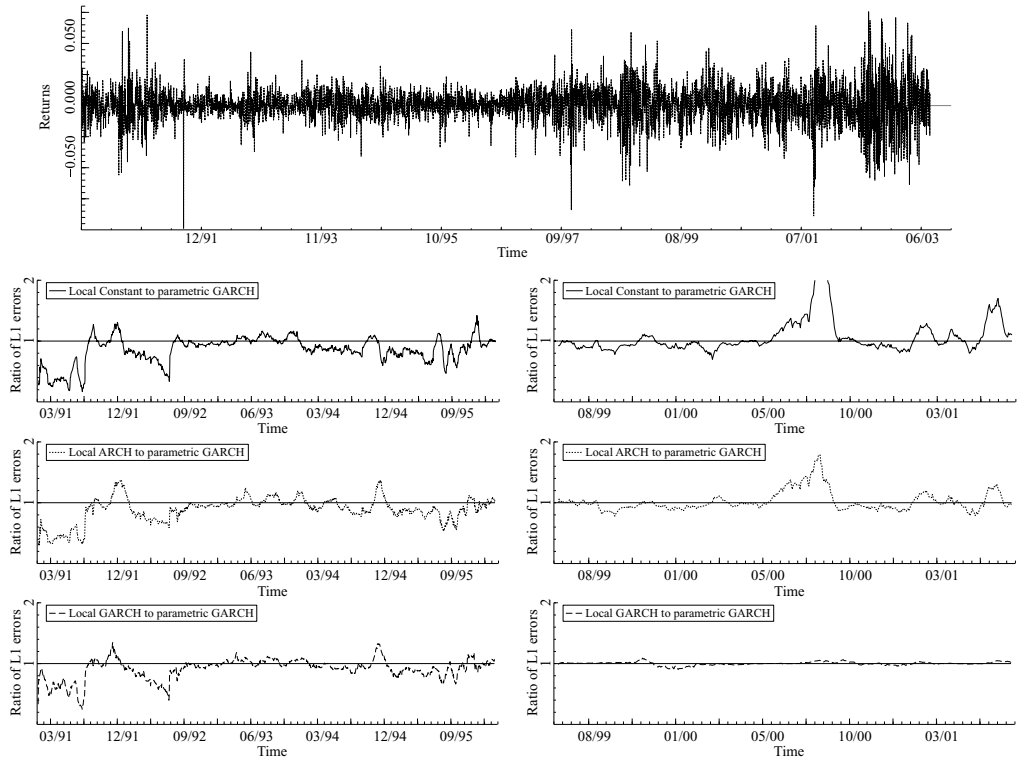

Figure 5. Top panel: The log-returns of DAX series. Bottom panels: The absolute prediction errors of the pointwise adaptive methods relative to the parametric GARCH errors for predictions one period ahead.

where the local constant and local ARCH models actually outperform both the parametric and adaptive GARCH. In the other parts of the data, the performance of all methods is approximately the same, and even though the adaptive GARCH is overall better than the parametric one, the most interesting fact is that the adaptively estimated local constant and local ARCH models perform equally well. In terms of the global PE, the local constant is best (0.829), followed by the local ARCH (0.844) and local GARCH (0.869). This closely corresponds to our findings in simulation study with low GARCH effect in Section 5.2. Note that for other choices of r and ρ, the global PEs are at most 0.835 and 0.851 for the local constant and local ARCH, respectively. This indicates low sensitivity to the choice of these parameters.

Next, we discuss the estimation results for years 1999 to 2001 ($r = 1.0$ for all methods now). After the financial markets were hit by the Asian crisis in 1997 and the Russian crisis in 1998, the market headed to a more stable state in year 1999. The adaptive methods detected the structural breaks in the autumn of 1997 and 1998. The local GARCH detected them, however, with more than a one-year delay—only during 1999. The results in Figure 5 (right bottom panel) confirm that the benefits of the adaptive GARCH are practically negligible compared to the parametric GARCH in such a case. On the other hand, the local constant and ARCH methods perform slightly better than both GARCH methods during the first presented year (July 1999 to June 2000). From July 2000, the situation becomes just the opposite and the performance

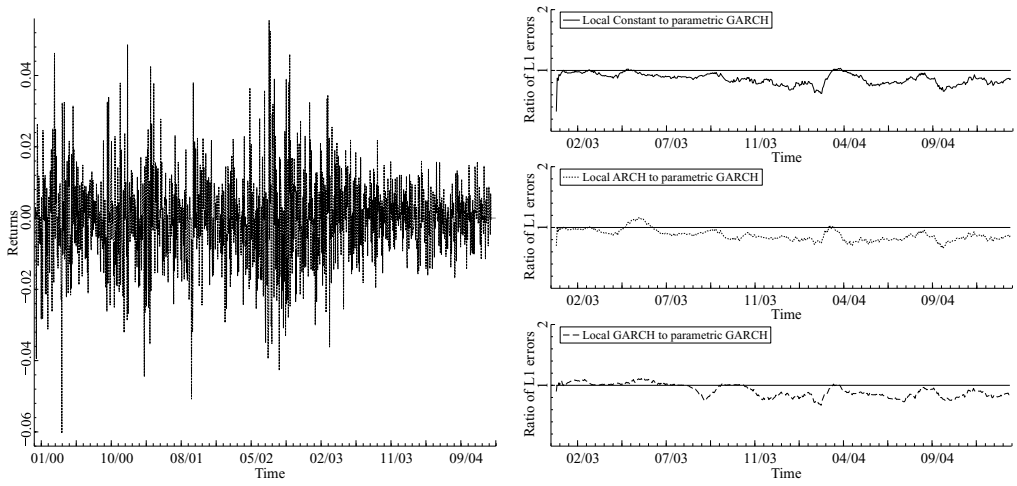

Figure 6. Left-hand panel: The log-returns of S&P 500. Right-hand panel: The absolute prediction errors of the pointwise adaptive methods relative to the parametric GARCH errors for predictions one period ahead.

of the GARCH models is better (parametric and adaptive GARCH estimates are practically the same in this period since the last detected structural change occurred approximately two years ago). Together with previous results, this opens the question of model selection among adaptive procedures as different parametric approximations might be preferred in different time periods. Judging by the global PE, the local ARCH provides slightly better predictions on average than the local constant and local GARCH—despite the 'peak' of the PE ratio in the second half of year 2000 (see Figure 5). This, however, depends on the specific choice of loss Λ in (3.6).

Finally, let us mention that the relatively similar behaviour of the local constant and local ARCH methods is probably due to the use of ARCH(1) model, which is not sufficient to capture more complex time developments. Hence, ARCH(p) might be a more appropriate interim step between the local constant and GARCH models.

6.2. S&P 500

Now we turn our attention to more recent data regarding the S&P 500 stock index considered from January 2000 to December 2004; see Figure 6. This period is marked by many substantial events affecting the financial markets, ranging from September 11, 2001, terrorist attacks and the war in Iraq (2003) to the crash of the technology stock-market bubble (2000–02). For the sake of simplicity, a particular time period is again selected: year 2003 representing a more volatile period (the war in Iraq) and year 2004 being a less volatile period. All adaptive methods detected rather quickly a structural break at the beginning of 2003, and additionally they detected a structural break in the second half of 2003, although the adaptive GARCH did so with a delay of more than eight months. The ratios of monthly PE of all adaptive methods to those of the parametric GARCH from January 2003 to December 2004 are summarized on Figure 6 ($r = 0.5$ and $\rho = 1.5$ for all methods).

In the beginning of year 2003, corresponding with 2002 to a more volatile period (see Figure 6), all adaptive methods perform as well as the parametric GARCH. In the middle of year 2003, the local constant and local ARCH models are able to detect another structural change (possibly less pronounced than the one at the beginning of 2003 because of its late detection by the adaptive GARCH). Around this period, the local ARCH shortly performs worse than the parametric GARCH. From the end of 2003 and in year 2004, all adaptive methods starts to outperform the parametric GARCH, where the reduction of the PEs due to the adaptive estimation amounts to 20% on average. All adaptive pointwise estimates exhibit a short period of instability in the first months of 2004, where their performance temporarily worsens to the level of parametric GARCH. This corresponds to 'uncertainty' of the adaptive methods about the length of the interval of homogeneity. After this short period, the performance of all adaptive methods is comparable, although the local constant performs overall best of all methods (closely followed by local ARCH) judged by the global PE.

Similarly to the low GARCH-effect simulations and to the analysis of DAX in Section 6.1, it seems that the benefit of pointwise adaptive estimation is most pronounced during periods of stability that follow an unstable period (i.e. year 2004) rather than during a presumably rapidly changing environment. The reason is that, despite possible inconsistency of parametric methods under change points, the adaptive methods tend to have a rather large variance when the intervals of time homogeneity become very short.

7. CONCLUSION

We extend the idea of adaptive pointwise estimation to parametric CH models. In the specific case of ARCH and GARCH, which represent particularly difficult cases due to high data demands and dependence of critical values on underlying parameters, we demonstrate the use and feasibility of the proposed procedure: on the one hand, the adaptive procedure, which itself depends on a number of auxiliary parameters, is shown to be rather insensitive to their choice, and on the other hand, it facilitates the global selection of these parameters by means of fit or forecasting criteria. The real-data applications highlight the flexibility of the proposed time-inhomogeneous models since even simple varying-coefficients models such as constant volatility and ARCH(1) can outperform standard parametric methods such as GARCH(1,1). Finally, the relatively small differences among the adaptive estimates based on different parametric approximations indicate that, in the context of adaptive pointwise estimation, it is sufficient to concentrate on simpler and less data-intensive models such as ARCH(p), $0 \leq p \leq 3$, to achieve good forecasts.

ACKNOWLEDGMENTS

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 'Economic Risk'.

REFERENCES

Andersen, T. G. and T. Bollerslev (1998). Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review 39*, 885–905.

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

Andreou, E. and E. Ghysels (2006). Monitoring disruptions in financial markets. *Journal of Econometrics 135*, 77–124.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica 61*, 821–56.

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

Beltratti, A. and C. Morana (2004). Structural change and long-range dependence in volatility of exchange rates: either, neither or both? *Journal of Empirical Finance 11*, 629–58.

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

Cai, Z., J. Fan and Q. Yao (2000). Functional coefficient regression models for nonlinear time series. *Journal of the American Statistical Association 95*, 941–56.

Chen, J. and A. K. Gupta (1997). Testing and locating variance changepoints with application to stock prices. *Journal of the American Statistical Association 92*, 739–47.

Chen, R. and R. J. Tsay (1993). Functional-coefficient autoregressive models. *Journal of the American Statistical Association 88*, 298–308.

Cheng, M.-Y., J. Fan and V. Spokoiny (2003). Dynamic nonparametric filtering with application to volatility estimation. In M. G. Akritas and D. N. Politis (Eds.), *Recent Advances and Trends in Nonparametric Statistics*, 315–33. Amsterdam: Elsevier.

Diebold, F. X. and A. Inoue (2001). Long memory and regime switching. *Journal of Econometrics 105*, 131–59.

Doornik, J. A. (2002). Object-oriented programming in econometrics and statistics using Ox: a comparison with C++, Java and C#. In S. S. Nielsen (Ed.), *Programming Languages and Systems in Computational Economics and Finance*, 115–47. Dordrecht: Kluwer.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50*, 987–1008.

Fan, J. and W. Zhang (2008). Statistical models with varying coefficient models. *Statistics and Its Interface 1*, 179–95.

Francq, C. and J.-M. Zakoian (2007). Quasi-maximum likelihood estimation in GARCH processes when some coefficients are equal to zero. *Stochastic Processes and their Applications 117*, 1265–84.

Glosten, L. R., R. Jagannathan and D. E. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance 48*, 1779–801.

Hansen, B. and S.-W. Lee (1994). Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory 10*, 29–53.

Härdle, W., H. Herwatz and V. Spokoiny (2003). Time inhomogeneous multiple volatility modelling. *Journal of Financial Econometrics 1*, 55–99.

Herwatz, H. and H. E. Reimers (2001). Empirical modeling of the DEM/USD and DEM/JPY foreign exchange rate: structural shifts in GARCH-models and their implications. 2001–83, Discussion Paper SFB 373, Humboldt-Univerzität zu Berlin, Germany.

Hillebrand, E. (2005). Neglecting parameter changes in GARCH models. *Journal of Econometrics 129*, 121–38.

Kokoszka, P. and R. Leipus (2000). Change-point estimation in ARCH models. *Bernoulli 6*, 513–39.

Laurent, S. and J.-P. Peters (2006). *G@RCH 4.2, Estimating and Forecasting ARCH Models*. London: Timberlake Consultants Press.

Mercurio, D. and V. Spokoiny (2004). Statistical inference for time-inhomogeneous volatility models. *Annals of Statistics 32*, 577–602.

Mikosch, T. and C. Starica (1999). Change of structure in financial time series, long range dependence and the GARCH model. Working Paper, Department of Statistics, University of Pennsylvania. See `http://citeseer.ist.psu.edu/mikosch99change.html`.

Mikosch, T. and C. Starica (2004). Changes of structure in financial time series and the GARCH model. *Revstat Statistical Journal 2*, 41–73.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica 59*, 347–70.

Pesaran, M. H. and A. Timmermann (2004). How costly is it to ignore breaks when forecasting the direction of a time series? *International Journal of Forecasting 20*, 411–25.

Sentana, E. (1995). Quadratic ARCH models. *Review of Economic Studies 62*, 639–61.

Spokoiny, V. (1998). Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *Annals of Statistics 26*, 1356–78.

Spokoiny, V. (2009a). Multiscale local change-point detection with applications to value-at-risk. *Annals of Statistics 37*, 1405–36.

Spokoiny, V. (2009b). Parameter estimation in time series analysis. WIAS Preprint No. 1404, Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany.

Stapf, J. and T. Werner (2003). How wacky is DAX? The changing structure of German stock market volatility. Discussion Paper 2003/18, Deutsche Bundesbank, Germany.

Taylor, S. J. (1986). *Modeling Financial Time Series*. Chichester: Wiley.

APPENDIX: PROOFS

Proof of Corollary 2.1: Given the choice of $\mathfrak{z}_\alpha$, it directly follows from (2.5). □

Proof of Theorem 3.1: Consider the event $\mathscr{B}_k = \{\hat{I} = I_{k-1}\}$ for some $k \le K$. This particularly means that I_{k-1} is accepted while $I_k = [T - m_k + 1, T]$ is rejected; i.e. there is $I' = [t', T] \subseteq I_k$ and $\tau \in \mathscr{T}(I_k)$ such that $T_{I_k,\tau} > \mathfrak{z}_k = \mathfrak{z}_{I_k,\mathscr{T}(I_k)}$. For every fixed $\tau \in \mathscr{T}(I_k)$ and $J = I_k \setminus [\tau + 1, T]$, $J^c = [\tau + 1, T]$, it holds by definition of $T_{I_k,\tau}$ that

$$T_{I_k,\tau} \le L_J(\widetilde{\boldsymbol{\theta}}_J) + L_{J^c}(\widetilde{\boldsymbol{\theta}}_{J^c}) - L_I(\boldsymbol{\theta}_0) = L_J(\widetilde{\boldsymbol{\theta}}_J, \boldsymbol{\theta}_0) + L_{J^c}(\widetilde{\boldsymbol{\theta}}_{J^c}, \boldsymbol{\theta}_0).$$

This implies by Theorem 2.1 that $\boldsymbol{P}_{\boldsymbol{\theta}_0}(T_{I_k,\tau} > 2\mathfrak{z}) \le \exp\{\mathfrak{e}(\lambda, \boldsymbol{\theta}_0) - \lambda\mathfrak{z}\}$. Now,

$$\boldsymbol{P}_{\boldsymbol{\theta}_0}(\mathscr{B}_k) \le \sum_{t'=T-m_k+1}^{T-m_0} \sum_{\tau=t'+1}^{T-m_0+1} 2\exp\{\mathfrak{e}(\lambda, \boldsymbol{\theta}_0) - \lambda\mathfrak{z}_k/2\} \le 2\frac{m_k^2}{2}\exp\{\mathfrak{e}(\lambda, \boldsymbol{\theta}_0) - \lambda\mathfrak{z}_k/2\}.$$

Next, by the Cauchy–Schwartz inequality

$$\begin{aligned}\boldsymbol{E}_{\boldsymbol{\theta}_0}|L_{I_K}(\widetilde{\boldsymbol{\theta}}_{I_K}, \hat{\boldsymbol{\theta}})|^r &= \sum_{k=1}^K \boldsymbol{E}_{\boldsymbol{\theta}_0}[|L_{I_K}(\widetilde{\boldsymbol{\theta}}_{I_K}, \widetilde{\boldsymbol{\theta}}_{k-1})|^r \mathbf{1}(\mathscr{B}_k)] \\ &\le \sum_{k=1}^K \boldsymbol{E}_{\boldsymbol{\theta}_0}^{1/2}|L_{I_K}(\widetilde{\boldsymbol{\theta}}_{I_K}, \widetilde{\boldsymbol{\theta}}_{k-1})|^{2r} \boldsymbol{P}_{\boldsymbol{\theta}_0}^{1/2}(\mathscr{B}_k).\end{aligned}$$

Under the conditions of Theorem 2.1, it follows similarly to (2.6) that

$$\boldsymbol{E}_{\boldsymbol{\theta}_0}|L_{I_K}(\widetilde{\boldsymbol{\theta}}_{I_K}, \widetilde{\boldsymbol{\theta}}_{k-1})|^{2r} \le (m_K/m_{k-1})^{2r}\mathfrak{R}_{2r}^*(\boldsymbol{\theta}_0)$$

for some constant $\mathfrak{R}^*_{2r}(\boldsymbol{\theta}_0)$ and $k = 1, \ldots, K$, and therefore,

$$\boldsymbol{E}_{\boldsymbol{\theta}_0}|L_{I_K}(\widetilde{\boldsymbol{\theta}}_{I_K}, \hat{\boldsymbol{\theta}})|^r \le [\mathfrak{R}^*_{2r}(\boldsymbol{\theta}_0)]^{1/2} \sum_{k=1}^{K} m_k (m_K/m_{k-1})^r \exp\{\mathfrak{e}(\lambda, \boldsymbol{\theta}_0)/2 - \lambda \mathfrak{z}_k/4\}$$

and the result follows by simple algebra provided that $a_1\lambda/4 \ge 1$ and $a_2\lambda/4 > 2$. □

LEMMA A.1. *Let $\boldsymbol{P}$ and $\boldsymbol{P}_0$ be two measures such that the Kullback–Leibler divergence $\boldsymbol{E}\log(d\boldsymbol{P}/d\boldsymbol{P}_0)$, satisfies $\boldsymbol{E}\log(d\boldsymbol{P}/d\boldsymbol{P}_0) \le \Delta < \infty$. Then for any random variable ζ with $\boldsymbol{E}_0\zeta < \infty$, it holds that $\boldsymbol{E}\log(1+\zeta) \le \Delta + \boldsymbol{E}_0\zeta$.*

Proof: By simple algebra one can check that for any fixed y the maximum of the function $f(x) = xy - x\log x + x$ is attained at $x = e^y$ leading to the inequality $xy \le x\log x - x + e^y$. Using this inequality and the representation $\boldsymbol{E}\log(1+\zeta) = \boldsymbol{E}_0\{Z\log(1+\zeta)\}$ with $Z = d\boldsymbol{P}/d\boldsymbol{P}_0$ we obtain

$$\begin{aligned}\boldsymbol{E}\log(1+\zeta) &= \boldsymbol{E}_0\{Z\log(1+\zeta)\} \le \boldsymbol{E}_0(Z\log Z - Z) + \boldsymbol{E}_0(1+\zeta)\\ &= \boldsymbol{E}_0(Z\log Z) + \boldsymbol{E}_0\zeta - \boldsymbol{E}_0 Z + 1.\end{aligned}$$

It remains to note that $\boldsymbol{E}_0 Z = 1$ and $\boldsymbol{E}_0(Z\log Z) = \boldsymbol{E}\log Z$. □

Proof of Theorem 4.1: Lemma A.1 applied with $\zeta = \varrho(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})/\boldsymbol{E}_{\boldsymbol{\theta}}\varrho(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})$ yields the result in the view of

$$\begin{aligned}\boldsymbol{E}_{\boldsymbol{\theta}}(Z_{I,\boldsymbol{\theta}}\log Z_{I,\boldsymbol{\theta}}) &= \boldsymbol{E}\log Z_{I,\boldsymbol{\theta}} = \boldsymbol{E}\sum_{t\in I}\log\frac{p[Y_t, g(X_t)]}{p[Y_t, g(X_t(\boldsymbol{\theta}))]}\\ &= \boldsymbol{E}\sum_{t\in I}\boldsymbol{E}\left\{\log\frac{p[Y_t, g(X_t)]}{p[Y_t, g(X_t(\boldsymbol{\theta}))]}\,\middle|\,\mathscr{F}_{t-1}\right\} = \boldsymbol{E}\Delta_{I_k}(\boldsymbol{\theta}).\end{aligned}$$ □

Proof of Corollary 4.1: It is Theorem 4.1 formulated for $\varrho(\boldsymbol{\theta}', \boldsymbol{\theta}) = L_I(\boldsymbol{\theta}', \boldsymbol{\theta})$. □

Proof of Theorem 4.2: The first inequality follows from Corollary 4.1, the second one from condition (3.4) and the property $x \ge \log x$ for $x > 0$. □

Proof of Theorem 4.3: Let $\hat{k} = k > k^*$. This means that I_k is not rejected as homogeneous. Next, we show that for every $k > k^*$ the inequality $T_{I_k,\tau} \le T_{I_k,\mathscr{T}(I_k)} \le \mathfrak{z}_k$ with $\tau = T - m_{k^*} = T - |I_{k^*}|$ implies $L_{I_{k^*}}(\widetilde{\boldsymbol{\theta}}_{I_{k^*}}, \widetilde{\boldsymbol{\theta}}_{I_k}) \le \mathfrak{z}_{k^*}$. Indeed with $J = I_k \backslash I_{k^*}$, this means that, by construction, $\mathfrak{z}_k \le \mathfrak{z}_{k^*}$ for $k > k^*$ and

$$\mathfrak{z}_k \ge T_{I_k,\tau} = L_{I_{k^*}}(\widetilde{\boldsymbol{\theta}}_{I_{k^*}}, \widetilde{\boldsymbol{\theta}}_{I_k}) + L_J(\widetilde{\boldsymbol{\theta}}_J, \widetilde{\boldsymbol{\theta}}_{I_k}) \ge L_{I_{k^*}}(\widetilde{\boldsymbol{\theta}}_{I_{k^*}}, \widetilde{\boldsymbol{\theta}}_{I_k}).$$

It remains to note that

$$|L_{I_{k^*}}(\widetilde{\boldsymbol{\theta}}_{I_{k^*}}, \hat{\boldsymbol{\theta}})|^r \le |L_{I_{k^*}}(\widetilde{\boldsymbol{\theta}}_{I_{k^*}}, \hat{\boldsymbol{\theta}}_{I_{k^*}})|^r \mathbf{1}(\hat{k} < k^*) + \mathfrak{z}^r_{k^*}\mathbf{1}(\hat{k} > k^*),$$

which obviously yields the assertion. □