Adaptive Pointwise Estimation in Time-Inhomogeneous Time-Series Models

Cizek, P.; Haerdle, W.; Spokoiny, V.

Publication date:
2007

No. 2007–35

ADAPTIVE POINTWISE ESTIMATION IN TIME-INHOMOGENEOUS TIME- SERIES MODELS

By P. Čížek, W. Härdle, V. Spokoiny

May 2007

ISSN 0924-7815

Adaptive pointwise estimation in time-inhomogeneous time-series models

P. Čížek,* W. Härdle,† and V. Spokoiny‡

Abstract

This paper offers a new method for estimation and forecasting of the linear and nonlinear time series when the stationarity assumption is violated. Our general local parametric approach particularly applies to general varying-coefficient parametric models, such as AR or 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.

JEL codes: C13, C14, C22

Keywords: adaptive pointwise estimation, autoregressive models, conditional heteroscedasticity models, local time-homogeneity

*Dept. of Econometrics & OR, Tilburg University, P.O.Box 90153, 5000LE Tilburg, The Netherlands.

†Humboldt-Universität zu Berlin, Spandauerstrasse 1, 10178 Berlin, Germany.

‡Weierstrass-Institute, Mohrenstr. 39, 10117 Berlin, Germany.

1 Introduction

A growing amount of econometrical research is devoted to modeling macroeconomic and financial time series and their volatility, which measures dispersion at a point in time (e.g., conditional variance). Although many economies and financial markets have been recently experiencing many shorter and longer periods of instability or uncertainity such as Asian crisis (1997), Russian crisis (1998), start of the European currency (1999), the "dot-Com" technology-bubble crash (2000–2002), or the terrorist attacks (September, 2001) and the war in Iraq (2003), mostly used econometric models are based on the assumption of time homogeneity. This includes linear and nonlinear autoregressive and moving-average models (ARMA) and conditional heteroscedasticity models such as ARCH (Engel, 1982) and GARCH (Bollerslev, 1986), stochastic volatility models (Taylor, 1986), as well as their combinations such as AR-GARCH or ARCH-in-mean (Engle et al., 1987).

On the other hand, the market and institutional changes have long been assumed to cause structural breaks in macroeconomics and financial time series, which was confirmed, for example, in data on GDP (McConnell and Perez-Quiros, 2000), money demand (Wolters et al., 1990), stock prices (Andreou and Ghysels, 2002; Beltratti and Morana, 2004), and exchange rates (Herwatz and Reimers, 2001). There even seems to be causal relationship between shocks and structural changes in macroeconomic and financial indicators (Beltratti and Morana, 2006). Moreover, ignoring these breaks can adversely affect the modeling, estimation, and forecasting as suggested by Diebold and Inoue (2001), Mikosch and Starica (2004), Pesaran and Timmermann (2004), and Hillebrand (2005), for instance. Such findings led to the development of the change-point analysis in the context of linear time-series and conditional-heteroscedasticity models; see for example, Chen and Gupta (1997), Bai and Perron (1998), Kokoszka and Leipus (2000), Andrews (2003), Perron and Zhu (2005), 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 (Fan and Zhang, 1999; Cai et al., 2000; Fan et al., 2003). Without structural assumptions about

the transition of model parameters over time, time-varying coefficient models have to be estimated nonparametrically, for example, under the identification condition that their parameters are smooth functions of time (e.g., Orbe et al., 2005). In this paper, we follow a more general strategy based on the assumption that a time series can be locally, that is 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), that is, for the longest time interval in which the series is described well by a given parametric model. In the context of the local constant approximation, this strategy was employed for volatility modeling by Härdle et al. (2003) and Mercurio and Spokoiny (2004). Our aim is to generalize this approach so that it can identify intervals of homogeneity for any parametric model irrespective 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 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 ARMA and GARCH mimic the majority of stylized facts about macroeconomic and 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 modeling 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 modeling 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 modeling 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, both the traditional structural-change tests (e.g., Bai and Perron, 1998) and the end-of-sample stability tests (e.g., Chu et al., 1996; Andrews, 2003) require that the number of observations before each break point is large (and can grow to infinity) as these tests rely on asymptotic results. This is not very realistic especially in macroeconomic time series (e.g., for yearly data). 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 (2004)'s procedure to the local parametric modeling is a nontrivial problem, which requires new tools and techniques. This is especially true for nonlinear and data-demanding models such as GARCH. The reason is that, at each time point, the procedure starts from a small interval, where 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., 5 or 10 observations). Using standard testing methods, such a procedure might be feasible for simple parametric models, such as the local constant approximation used by Mercurio and Spokoiny (2004), 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 based on Spokoiny and Chen (2007), which relies on a finite-sample theory of testing a growing sequence of historical time intervals on homogeneity against a change-point alternative. We generalize their results to general time-series models, including autoregressive and conditional-heteroscedasticity models, and demonstrate the

feasibility and stability of the proposed method.

The rest of the paper is organized as follows. In Section 2, we discuss the two main ingredients of the method: parametric estimation of autoregressive conditional-heteroscedasticity models and the test of homogeneity for these models. Section 3 introduces the adaptive estimation procedure and discusses the choice of its parameters. Theoretical properties of the method are discussed in Section 4. In the specific case of the GARCH(1,1) model, 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 time-series models

Consider a time series Y_t in discrete time, $t \in N$. A standard way to model such processes is based on one or another parametric specification. One frequently employed example is given by the autoregressive (AR) model:

$$Y_t = \gamma_0 + \sum_{i=1}^{p_m} \gamma_i Y_{t-i} + \eta_t, \tag{2.1}$$

where $\{\eta_t\}_{t\in N}$ is a white noise process and $\boldsymbol{\theta} = (\gamma_0, \ldots, \gamma_{p_m})^\top$ is a parameter vector. (The AR model can also be nonlinear.)

In financial applications, one often applies similar equations for the volatility process which includes a large class of parametric models around the ARCH (Engle, 1982) and GARCH (Bollerslev, 1986) specifications:

$$Y_t = \sigma_t \varepsilon_t, \qquad \sigma_t^2 = \omega + \sum_{i=1}^{p_v} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{q_v} \beta_j \sigma_{t-j}^2, \tag{2.2}$$

where $\{\varepsilon_t\}_{t\in N}$ is again a white noise process. A number of GARCH extensions were proposed to make the model even more flexible; for example, EGARCH (Nelson, 1991), QGARCH (Sentana, 1995), and TGARCH (Glosten et al., 1993) that account for asymmetries in a volatility process. All such conditional-heteroscedast-

icity models can be put into a common class of generalized linear volatility models:

$$\sigma_t = g(X_t), \qquad X_t = \omega + \sum_{i=1}^{p_v} \alpha_i h(Y_{t-i}) + \sum_{j=1}^{q_v} \beta_j X_{t-j}, \tag{2.3}$$

where g and h are known functions, X_t is a (partially) unobserved process (structural variable) that models volatility coefficient σ_t^2 via transformation g, and $\boldsymbol{\theta} = (\omega, \alpha_1, \ldots, \alpha_{p_v}, \beta_1, \ldots, \beta_{q_v})^\top$ is a parameter vector. For example, the GARCH model (2.2) is described by $g(u) = u$ and $h(r) = r^2$.

One can also combine the AR-type modeling (2.1) for the observables Y_t and the GARCH-type modeling (2.3) for the noise: with $Z_t = h(Y_t)$, the model reads

$$\begin{aligned} Y_t &= \gamma_0 + \sum_{i=1}^{p_m} \gamma_t Y_{t-i} + \eta_t \\ \eta_t &= \sigma_t \varepsilon_t = g(X_t)\varepsilon_t, \\ X_t &= \omega + \sum_{i=1}^{p_v} \alpha_i Z_{t-i} + \sum_{j=1}^{q_v} \beta_j X_{t-j}. \end{aligned} \tag{2.4}$$

This model class does not include ARMA and stochastic volatility models, or more generally, the Hidden Markov Chain models. An extension of our approach to such models is possible but involves a number of serious technical problems. However, the class (2.4) is sufficiently large to cover many applications.

We consider a general parametric set-up covering models (2.1)–(2.4), which can be described as follows. Let the observed process Y_t be progressively measurable w.r.t. a filtration $\mathcal{F} = (\mathcal{F}_t)$ and let X_t be a non-observed or partially observed explanatory process which is predictable w.r.t. $\mathcal{F}$, that is, $X_t \sim \mathcal{F}_{t-1}$. In general, both Y_t and X_t can be multivariate. Our basic assumption is that the conditional distribution of Y_t given the "past" $\mathcal{F}_{t-1}$ belongs to some given parametric family $\mathcal{P} = (P_v)$, but the corresponding parameter v depends on the explanatory vector X_t and some parameter $\boldsymbol{\theta}$. We write this relation in the form

$$\mathcal{L}\big(Y_t | \mathcal{F}_{t-1}\big) = P_{g(X_t)} \quad \in \mathcal{P}, \tag{2.5}$$

$$X_t = f_{\boldsymbol{\theta}}(Y^{t-1}, X^{t-1}), \tag{2.6}$$

where Y^{t-1} means the past $Y_{t-1}, Y_{t-2}, \ldots$ of the process Y_t, X^{t-1} is defined analogously, and $f_{\boldsymbol{\theta}}(\cdot)$ is a given parametric class of functions. It is important

to note that, given the parameter $\boldsymbol{\theta}$, the observations $Y_1, \ldots, Y_t$, and the pre-history X^0, Y^0, one can uniquely reconstruct the process $X_t = X_t(\boldsymbol{\theta})$ for $t \geq 1$ by recurrently applying the relation (2.6) for X_t. Under ergodicity conditions, the impact of the pre-history X^0, Y^0 is not significant and we just fix it arbitrary.

In the case of volatility modeling, Y_t means the squared return at time t and $P_\upsilon \in \mathcal{P}$ is the distribution of a squared normal random variable with zero mean and variance υ. The parametric assumption (2.6) means that the time-varying volatility $\sigma_t^2 = g(X_t)$ is a parametric function of the process X_t.

Model (2.5)-(2.6) is time homogeneous in the sense that the process Y_t follows the same structural equation at each time point. In other words, parameter $\boldsymbol{\theta}$ and hence the structural dependence in Y_t is constant over time. Even though models like (2.4) can often fit data well over a longer period of time, the assumption of homogeneity is too restrictive in practical applications: to guarantee sufficient amount of data for sufficiently precise estimation, these models are often applied for time spans of many years both in the case of macroeconomic data (e.g., due to low sampling frequency) and financial data (e.g., due to high data demands of GARCH models). On the contrary, the strategy pursued in this paper 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 model (2.5)-(2.6). 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, which then enable predicting Y_t and X_t.

Now, Section 2.1 describes in details the method of parameter estimation in the model (2.5)-(2.6). Its accuracy in finite samples is discussed in Section 2.2.

2.1 Parameter estimation

This section discusses the parameter estimation for model (2.5)-(2.6) using observations Y_t from some time interval $I = [t_0, t_1]$. To give a specific example, we can focus on the problem of volatility modeling with P_υ being the distribution of $\upsilon\varepsilon^2$, where ε is standard normal. The process Y_t and the structural process X_t

for volatility $\sigma_t^2 = g(X_t)$ are described by equations (2.5)-(2.6). We write these processes in the form $Y_t = Y_t(\boldsymbol{\theta})$ and $X_t = X_t(\boldsymbol{\theta})$ for $t \in I$. In the parametric case, $\boldsymbol{\theta}^*$ denotes the underlying value of $\boldsymbol{\theta}$.

For estimating $\boldsymbol{\theta}^*$, we apply the quasi maximum likelihood (quasi-MLE) approach under the assumption of Gaussian errors ε_t, which guarantees efficiency under the normality of innovations and consistency under rather general moment conditions (Hansen, 1982; Hansen and Lee, 1994). The quasi log-likelihood for the model (2.5)-(2.6) on an interval I can be represented in the form

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

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

$$\widetilde{\boldsymbol{\theta}}_I = \operatorname*{argmax}_{\boldsymbol{\theta} \in \Theta} L_I(\boldsymbol{\theta}) = \operatorname*{argmax}_{\boldsymbol{\theta} \in \Theta} \sum_{t \in I} \ell\{Y_t(\boldsymbol{\theta}), g(X_t(\boldsymbol{\theta}))\}, \tag{2.7}$$

and denote by $L_I(\widetilde{\boldsymbol{\theta}}_I)$ the corresponding maximum. Symbol Θ denotes the parametric space containing only parameter values ensuring ergodicity of process Y_t.

2.2 Accuracy of parametric situation

This section discusses some specific properties of the proposed (quasi-)MLE estimate $\widetilde{\boldsymbol{\theta}}$ that apply for an arbitrarily long interval I (i.e., for an arbitrarily small or large sample size). This feature is very important for the proposed adaptive procedure because it starts the search for a structural break in parameter values at very small intervals. For a given interval I, we now discuss the accuracy of estimation by $\widetilde{\boldsymbol{\theta}}_I$ in the time-homogeneous situation when the process Y_t follows a parametric model with the parameter vector $\boldsymbol{\theta}^*$ for all $t \in I$.

The main result concerning the quality of estimation of $\boldsymbol{\theta}$ concerns the value of maximum $L_I(\widetilde{\boldsymbol{\theta}}_I) = \max_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta})$ rather than the point of maximum $\widetilde{\boldsymbol{\theta}}$. More precisely, we consider difference $L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta}^*) = L_I(\widetilde{\boldsymbol{\theta}}_I) - L_I(\boldsymbol{\theta}^*)$. 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 $\boldsymbol{\theta}^*$. Later, we comment on

how the accuracy of estimation of the parameter $\boldsymbol{\theta}^*$ by $\widetilde{\boldsymbol{\theta}}_I$ relates to the value $L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta}^*)$. We will also see that the bound for $L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta}^*)$ yields the confidence set for the parameter $\boldsymbol{\theta}^*$, which is then used for the employed change-point test.

Theorem 2.1 (Spokoiny, 2007)**.** *Assume that the process Y_t follows the model (2.5)–(2.6) with the parameter $\boldsymbol{\theta}^* \in \Theta$. Let the parameter set Θ be compact and the function $f_{\boldsymbol{\theta}}(x)$ be continuously differentiable w.r.t. $\boldsymbol{\theta}$. Let the process $X_t(\boldsymbol{\theta})$ belong to a compact set $\mathcal{X}$ for all $t \in N$ and $\boldsymbol{\theta} \in \Theta$ and $g(x)$ is smooth and separated away from zero on $\mathcal{X}$. Then there are some $\lambda > 0$ and values $\mathfrak{e}(\lambda, \boldsymbol{\theta}^*)$ and $\mathfrak{e}(\lambda)$ depending on only $f_{\boldsymbol{\theta}}$, g, and $\mathcal{X}$ such that*

$$\log \boldsymbol{E}_{\boldsymbol{\theta}^*} \exp\bigl\{\lambda L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta}^*)\bigr\} \le \mathfrak{e}(\lambda, \boldsymbol{\theta}^*) \le \mathfrak{e}(\lambda).$$

Moreover, for any $\mathfrak{z} > 0$, it holds that

$$\boldsymbol{P}_{\boldsymbol{\theta}^*}\bigl(L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta}^*) > \mathfrak{z}\bigr) \le \exp\bigl\{\mathfrak{e}(\lambda) - \lambda\mathfrak{z}\bigr\}, \tag{2.8}$$

and for any $r > 0$ and every $\boldsymbol{\theta}^* \in \Theta$, there is a constant $\mathfrak{R}_r(\boldsymbol{\theta}^*)$ depending on $\mathcal{X}$ and r only such that

$$\boldsymbol{E}_{\boldsymbol{\theta}^*}\bigl|L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta}^*)\bigr|^r \le \mathfrak{R}_r(\boldsymbol{\theta}^*) \le \mathfrak{R}_r\,.$$

Remark 2.1. The condition that $X_t(\boldsymbol{\theta})$ is bounded can be easily relaxed to the condition that $X_t(\boldsymbol{\theta})$ is ergodic and thus belongs to a bounded set with a high probability but the corresponding formulation is more involved. A special case of this result can be found in Spokoiny and Chen (2007).

One very attractive feature of Theorem 2.1, formulated in the following corollary, is that it enables constructing the exact non-asymptotic confidence sets and testing the parametric hypothesis on the base 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 $\mathfrak{z}_\alpha$ fulfill $\mathfrak{e}(\lambda, \boldsymbol{\theta}^*) - \lambda\mathfrak{z}_\alpha < \log\alpha$ for some $\alpha < 1$. Then the random set $\mathcal{E}_I(\mathfrak{z}_\alpha) = \{\boldsymbol{\theta} : L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta}) \le \mathfrak{z}_\alpha\}$ is an α-confidence set for $\boldsymbol{\theta}^*$ in the sense that $\boldsymbol{P}_{\boldsymbol{\theta}^*}\bigl(\boldsymbol{\theta}^* \notin \mathcal{E}_I(\mathfrak{z}_\alpha)\bigr) \le \alpha$.*

Table 1: Simulated mean values and standard deviations (in brackets, multiplied by factor $|I|^{1/2}$) of parameter estimates and likelihood differences $L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta}^*)$. The parameters of a constant volatility model and two GARCH(1,1) models with normally distributed error terms are estimated for intervals I with 25 to 1000 observations using 10000 simulated samples.

		Interval length $\|I\|$				
Model	Parameter	25	75	150	400	1000
Constant vol.	ω ($\omega^* = 1.0$)	0.994	0.995	0.997	0.999	0.999
		(1.432)	(1.423)	(1.413)	(1.414)	(1.431)
	$L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta}^*)$	0.523	0.545	0.576	0.690	1.008
		(0.732)	(0.760)	(0.783)	(0.893)	(1.001)
GARCH(1,1)	ω ($\omega^* = 1.0$)	0.676	0.765	0.833	0.905	0.968
		(2.754)	(3.788)	(4.620)	(5.770)	(5.957)
	α_1 ($\alpha_1^* = 0.2$)	0.179	0.180	0.184	0.190	0.196
		(1.242)	(1.397)	(1.474)	(1.580)	(1.541)
	β_1 ($\beta_1^* = 0.1$)	0.355	0.291	0.236	0.175	0.126
		(2.027)	(2.943)	(3.554)	(4.476)	(4.417)
	$L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta}^*)$	1.006	1.186	1.297	1.513	1.839
		(0.889)	(1.034)	(1.110)	(1.280)	(1.374)
GARCH(1,1)	ω ($\omega^* = 1.0$)	3.217	2.451	1.758	1.236	1.081
		(18.06)	(20.70)	(18.79)	(13.79)	(10.52)
	α_1 ($\alpha_1^* = 0.2$)	0.166	0.199	0.202	0.200	0.200
		(1.209)	(1.309)	(1.288)	(1.211)	(1.200)
	β_1 ($\beta_1^* = 0.7$)	0.463	0.528	0.605	0.672	0.691
		(2.023)	(2.561)	(2.640)	(2.179)	(1.829)
	$L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta}^*)$	1.331	1.468	1.430	1.499	1.637
		(0.969)	(1.301)	(1.451)	(1.574)	(1.629)

The benefits of considering the distance $L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta}^*)$ and the result of Theorem 2.1 can be illustrated by a small simulation study. Consider the constant volatility and GARCH(1,1) models with two different sets of parameters. In Table 1, estimation results are presented in terms of the mean and the normalized standard deviation $|I|^{1/2}\big(\widetilde{\boldsymbol{\theta}}_I - \boldsymbol{\theta}^*\big)$ of the estimate $\widetilde{\boldsymbol{\theta}}_I$ and in terms of the fitted log-likelihood $L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta}^*)$. We see that for the constant volatility model, the results are nicely in agreement with the root-n consistency of the estimate $\widetilde{\boldsymbol{\theta}}_I$, whereas for the GARCH(1,1) models, the results for small sample sizes are awful in the sense that the standard deviation of estimates is of order of the range of the parameter set. At the same time, the fitted log-likelihood demonstrates a very stable behaviour for all models and sample sizes.

As already mentioned, Theorem 2.1 gives a non-asymptotic and fixed upper bound for the risk of estimation which applies to an arbitrary sample size $|I|$. To understand the relation of this result to the classical rate result, we 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

$$L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta}^*) = 0.5\big(\widetilde{\boldsymbol{\theta}}_I - \boldsymbol{\theta}^*\big)^\top \nabla^2 L(\boldsymbol{\theta}'_I)\big(\widetilde{\boldsymbol{\theta}}_I - \boldsymbol{\theta}^*\big), \tag{2.9}$$

where $\boldsymbol{\theta}'_I$ is some point on the line connecting $\boldsymbol{\theta}^*$ and $\widetilde{\boldsymbol{\theta}}_I$. Under the usual ergodicity assumptions and for sufficiently large $|I|$, the normalized matrix $|I|^{-1}\nabla^2 L(\boldsymbol{\theta})$ is close to some matrix $V^2(\boldsymbol{\theta})$, which depends only on the stationary distribution of $X_t(\boldsymbol{\theta})$ and is continuous in $\boldsymbol{\theta}$. Then the result of the theorem approximately means that $\|V(\boldsymbol{\theta}^*)\big(\widetilde{\boldsymbol{\theta}}_I - \boldsymbol{\theta}^*\big)\|^2 \le \mathfrak{z}/|I|$ for some fixed constant $\mathfrak{z}$. In other words, the large deviation result of Theorem 2.1 yields the root-$|I|$ consistency of the MLE estimate $\widetilde{\boldsymbol{\theta}}_I$.

3 Adaptive nonparametric estimation

An obvious feature of the model (2.5)-(2.6) 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 model. A natural generalization leads to models whose coefficients may vary with time. For example, Cai et al. (2000) considered the following varying-coefficient model

$$Y_t = \gamma_0(t) + \sum_{i=1}^{p} \gamma_i(t) Y_{t-i} + \varepsilon_t, \tag{3.1}$$

where $\gamma_0(t), \ldots, \gamma_p(t)$ are functions of time and have to be estimated from the observations Y_t. In general, one can assume that the structural process X_t satisfies the relation (2.6) but the vector of coefficients $\boldsymbol{\theta}$ may vary with the time t, $\boldsymbol{\theta} = \boldsymbol{\theta}(t)$. The estimation of the coefficients as functions of time is a hard problem and it is possible only under some additional assumptions on these functions. Typical examples are given by (i) varying coefficients are smooth functions of time (Cai et al., 2000) and (ii) varying coefficients are piecewise constant functions (Bai and Perron, 1998; 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 $\mathcal{L}\big(Y_t|\mathcal{F}_{t-1}\big) \sim P_{g(X_t)}$, see (2.5), and at each point T, there exists a historical interval $I(T) = [t, T]$ in which the process X_t "nearly" follows the parametric specification (2.6) (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 parameter $\boldsymbol{\theta} = \boldsymbol{\theta}(T)$. The estimate $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}(T)$ leads to the value $\widehat{X}_T = f_{\widehat{\boldsymbol{\theta}}}(Y^{T-1}, X^{T-1})$ of the process X_t at T and can be used for further modeling, for example, for forecasting the next value Y_{T+1}. Moreover, this assumption includes the above mentioned "smooth transition" and "switching regime" models as special cases: parameters $\widehat{\boldsymbol{\theta}}(T)$ vary over time as $I(T)$ changes, and at the same time, discontinuities and jumps in $\widehat{\boldsymbol{\theta}}(T)$ as a function of time are possible.

The idea of choosing the historical *interval of homogeneity* $I(T)$ is to find the longest interval I with the right-end point T, where data do not contradict the hypothesis of model homogeneity. Starting at each time T with a very short interval $I = [t, T]$, the search is done 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. This method strongly differs from that of Bai and Perron (1998) and Mikosch and Starica (1999) who try to detect all change points in a given time series. Our approach is local in the sense that it focuses on the local change-point analysis near the point of estimation or forecasting and tries to find only one change closest to the reference point.

In the rest of this section, we first rigorously describe the adaptive pointwise estimation procedure (Section 3.1), which relies on the test of time-homogeneity against a change-point alternative discussed in Section 3.2. Next, the choice of parameters and implementation of the adaptive procedure is described in Section 3.3. Theoretical properties of the method are studied in Section 4.

3.1 Adaptive choice of the interval of homogeneity

This section proposes an adaptive method for the pointwise selection of the longest historical interval of homogeneity for the parametric model (2.6). For a given point T, we aim to estimate the unknown parameter $\boldsymbol{\theta} = \boldsymbol{\theta}(T)$ from historical data Y_t, $t \leq T$. This procedure repeats for every current time point T as new data arrives.

The choice of the longest homogeneous interval at a given point T is done by the multiscale *local change point (LCP)* detection procedure. The procedure is based on a family of nested interval-candidates $I_0 \subset I_1 \subset I_2 \subset \ldots \subset I_K = I_{K+1}$ of the form $I_k = [T - m_k, T]$, where T is the right-end point and m_k is the interval length growing with k. Every interval leads to the estimate $\widetilde{\boldsymbol{\theta}}_k = \widetilde{\boldsymbol{\theta}}_{I_k}$ of the parameter $\boldsymbol{\theta}$ from the interval I_k. The adaptive procedure selects one interval $\widehat{I}$ out of the given family and thus the corresponding estimate $\widehat{\boldsymbol{\theta}} = \widetilde{\boldsymbol{\theta}}_{\widehat{I}}$.

The idea of the proposed method is to sequentially screen each interval $\mathcal{T}_k = I_k \setminus I_{k-1} = [T - m_k, T - m_{k-1}[, k = 1, \ldots, K,$ and check each point $\tau \in \mathcal{T}_k$ for a possible change-point location (see Section 3.2 for details of the test). The interval I_k is accepted if no change point is detected within $\mathcal{T}_1, \ldots, \mathcal{T}_k$ (interval I_0 has to be so short that homogeneity can always be assumed). If the hypothesis of homogeneity is rejected for an interval-candidate I_k, the procedure stops and

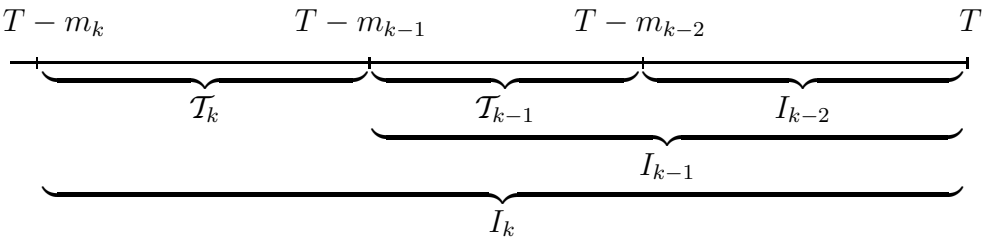

Figure 1: Choice of the intervals I_k and $\mathcal{T}_k$

selects the latest accepted interval. The formal description reads as follows.

Start the procedure with $k = 1$. Then (see Figure 1 for illustration)

1. test the hypothesis $H_{0,k}$ of no structural change within $\mathcal{T}_k$ using data from testing interval I_{k+1};

2. if no change points were found in $\mathcal{T}_k$, interval I_k is accepted. Take the next interval $\mathcal{T}_{k+1}$ and repeat the previous step with $k := k+1$ until homogeneity is rejected or the largest possible interval $I_K = [T - m_K, T]$ is accepted;

3. if $H_{0,k}$ is rejected for $\mathcal{T}_k$, the estimated interval of homogeneity is the last accepted interval $\widehat{I} = I_{k-1}$;

4. if the largest possible interval I_K is accepted, we take $\widehat{I} = I_K$.

In the description of the adaptive procedure above, $\widehat{I}_k = I_k$ is the latest accepted interval after the first k steps of the procedure, provided that the procedure has not stop yet. The corresponding quasi-MLE estimate on I_k is then $\widehat{\boldsymbol{\theta}}_k = \widetilde{\boldsymbol{\theta}}_{I_k}$. The final adaptively selected interval $\widehat{I}(T)$ at T is the latest accepted interval from the whole procedure, that is, $\widehat{I}(T) = \widehat{I}$. The corresponding adaptive pointwise estimate $\widehat{\boldsymbol{\theta}}(T)$ is then defined as $\widehat{\boldsymbol{\theta}}(T) = \widetilde{\boldsymbol{\theta}}_{\widehat{I}(T)}$.

3.2 Test of homogeneity against a change-point alternative

The main ingredient of the adaptive estimation procedure is the test of local homogeneity for a tested interval $\mathcal{T}$, which we now describe.

Let $\mathcal{T}$ be a *tested* historical interval. This means that we want to check every point of this interval for a possible change in the dependency structure. The test statistic is based on the observations Y_t from a larger *testing* interval I of the form $I = [T - m, T]$. We consider the supremum likelihood ratio (LR) test introduced by Andrews (1993). The null hypothesis for I means that the observations $\{Y_t\}_{t \in I}$ follow the parametric model (2.6) with a parameter $\boldsymbol{\theta} = \boldsymbol{\theta}^*$, which yields the parametric estimate $\widetilde{\boldsymbol{\theta}}_I$ due to (2.7) and the corresponding fitted log-likelihood $L_I(\widetilde{\boldsymbol{\theta}}_I)$. The change-point alternative given by the tested set $\mathcal{T}$ can be described as follows. Every point $\tau \in \mathcal{T} \subset I$ splits the interval I in two subintervals $J = [\tau+1, T]$ and $J^c = I \setminus J = [T-m, \tau]$. A change-point alternative with a location at $\tau \in \mathcal{T}(I)$ means that Y_t follows the parametric model with parameter $\boldsymbol{\theta}_J$ for $t \in J$ and $\boldsymbol{\theta}_{J^c}$ for $t \in J^c$ with $\boldsymbol{\theta}_J \neq \boldsymbol{\theta}_{J^c}$. Under such an alternative, data $\{Y_t\}_{t \in I}$ are associated with the log-likelihood $L_J(\widetilde{\boldsymbol{\theta}}_J) + L_{J^c}(\widetilde{\boldsymbol{\theta}}_{J^c})$.

To test against a single change-point alternative with a known fixed $\tau \in \mathcal{T}$, the LR test statistic can be used:

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

Considering an unknown change point $\tau \in \mathcal{T}$, the test of homogeneity for intervals I can be defined as the maximum (supremum) of the LR statistics over all $\tau \in \mathcal{T}$:

$$T_{I,\mathcal{T}} = \max_{\tau \in \mathcal{T}} T_{I,\tau}. \tag{3.2}$$

A change point is detected within I if the test statistic $T_{I,\mathcal{T}}$ exceeds a critical value $\mathfrak{z}$, which may generally depend on the intervals I and $\mathcal{T}$.

In the adaptive procedure described in Section 3.1 at every step $k \geq 1$, this test is applied to the tested interval $\mathcal{T}_k$ and testing interval I_{k+1}. The hypothesis of homogeneity is rejected if the test statistic $T_{I_{k+1},\mathcal{T}_k}$ exceeds the critical value $\mathfrak{z}_k$, which depends on the step k.

3.3 Parameters and implementation details of the method

To run the proposed procedure, one has to fix some of its parameters and ingredients. This particularly concerns the choice of intervals I_k. However, the most

important ingredient of the method is a collection of the critical values $\mathfrak{z}_k$, which is used for testing the presence of a change point in the interval $\mathcal{T}_k$ at every step k. Their choice and related computational issues are discussed in Sections 3.3.2–3.3.4.

3.3.1 Set of intervals

This section presents our way of selecting the sets I_k for $k = 0, \ldots, K$. Note however that our proposal is just an example. The procedure and the theoretical results in Section 4 apply under rather general conditions on these sets. In what follows, similarly to Spokoiny and Chen (2007), we fix some geometric grid $\{m_k = [m_0 a^k], k = 0, \ldots, K\}$ with an initial length $m_0 \in N$ and a multiplier $a > 1$ to define intervals $I_k = [t_k, T] = [T - m_k, T]$, $k = 0, \ldots, K$.

Our experiments show that results are rather insensitive to the choice of the parameters a and m_0. Results presented in Section 5 employ a multiplier $a = 1.25$ and the initial length $m_0 = 10$. For simpler parametric models (2.6), such as local constant volatility or AR(p), $m_0 = 5$ could be a reasonable choice.

3.3.2 Choice of critical values $\mathfrak{z}_k$

The proposed estimation method can be viewed as a hierarchic multiple testing procedure. At every step k, the hypothesis of homogeneity is tested against a change-point alternative with possible locations in the interval $\mathcal{T}_k$. The corresponding test statistic $T_{I_{k+1},\mathcal{T}_k}$ is compared with the critical value $\mathfrak{z}_k$. The parameters $\mathfrak{z}_k$ are selected to provide the prescribed error level under the null hypothesis, that is, in the time-homogeneous parametric situation. Because the proposed adaptive choice of the interval of homogeneity is based on the supremum LR test applied sequentially in rather small samples, the well-known asymptotic properties of the supremum LR statistics $T_{I,\mathcal{T}}$ defined in (3.2) for a single interval I (Andrews, 1993) are not applicable. For a practically relevant choice of critical values, we therefore apply, instead of asymptotic bounds, the finite-sample theoretical concepts and results presented in this section.

In what follows we assume that the considered set of intervals $I_0, \ldots, I_K$ is

fixed. The parameters $\mathfrak{z}_k$ are then selected so that they provide the below prescribed features of the procedure under the parametric (time-homogeneous) model (2.6) with some fixed parameter vector $\boldsymbol{\theta}^*$.

Let $\widehat{I}$ be the selected interval and $\widehat{\boldsymbol{\theta}}$ be the corresponding adaptive estimate for data generated from a time-homogeneous parametric model. Both the interval $\widehat{I}$ and estimate $\widehat{\boldsymbol{\theta}}$ depend implicitly on the critical values $\mathfrak{z}_k$. Under the null hypothesis, the desirable feature of the adaptive procedure is that, with a high probability, it does not reject any interval I_k as being time-homogeneous and thus selects the largest possible interval I_K. Equivalently, the selected interval $\widehat{I}_k$ after the first k steps and the corresponding adaptive estimate $\widehat{\boldsymbol{\theta}}_k$ should coincide with a high probability with their non-adaptive counterparts I_k and $\widetilde{\boldsymbol{\theta}}_k = \widetilde{\boldsymbol{\theta}}_{I_k}$. Motivated by the results of Section 2.2 and following Spokoiny and Chen (2007), this condition can be stated in the form

$$E_{\boldsymbol{\theta}^*} \left| L_{I_k}(\widetilde{\boldsymbol{\theta}}_k, \widehat{\boldsymbol{\theta}}_k) \right|^r \le \rho \mathfrak{R}_r(\boldsymbol{\theta}^*), \qquad k = 1, \ldots, K, \tag{3.3}$$

where r and ρ are a given positive constant and $\mathfrak{R}_r(\boldsymbol{\theta}^*)$ is the log-likelihood risk of the parametric estimation: $\mathfrak{R}_r(\boldsymbol{\theta}^*) = \max_{k \le K} \boldsymbol{E}_{\boldsymbol{\theta}^*} |L_{I_k}(\widetilde{\boldsymbol{\theta}}_k, \boldsymbol{\theta}^*)|^r$. In total, (3.3) states K conditions on the choice of K parameters $\mathfrak{z}_k$ that implicitly enter the definition of $\widehat{\boldsymbol{\theta}}$. There are two ways to determine the values of $\mathfrak{z}_k$ so that they satisfy (3.3). First, one can fix the values $\mathfrak{z}_k$ sequentially starting from $\mathfrak{z}_1$ by the procedure described in the next Section 3.3.3. Second, an alternative way is to apply the critical values which linearly depend on $\log(|I_k|)$, where $|I_k|$ denotes the length of interval I_k. This second choice is justified by the following result.

Theorem 3.1. *Suppose that* $|I_k| = m_0 a^k$ *for some* $a > 1, m_0 \in N$, *and* $k = 0, \ldots, K$, *and assume* $r > 0$ *and* $\rho > 0$. *Then there are constants* a_1, a_2, *and* a_3 *such that conditions (3.3) hold for the choice*

$$\mathfrak{z}_k = a_1 \log \rho^{-1} + a_2 \log(|I_K|/|I_k|) + a_3 \log(|I_k|). \tag{3.4}$$

Note however that the bound of Theorem 3.1 on the critical values $\mathfrak{z}_k$ is very rough and leads to a very conservative procedure. Below we discuss one practically relevant choice of critical values using Monte-Carlo simulations.

3.3.3 Sequential choice of the critical values by simulations

Let us now present a general way of selecting the critical values $\mathfrak{z}_k$ using Monte Carlo simulations from a parametric model so that they satisfy condition (3.3). To specify the contribution of $\mathfrak{z}_1$ to the final risk of the method, we set all the remaining values $\mathfrak{z}_2, \ldots, \mathfrak{z}_K$ equal to infinity: $\mathfrak{z}_2 = \ldots = \mathfrak{z}_K = \infty$. For every particular $\mathfrak{z}_1$, the whole set of critical values $\mathfrak{z}_k$ is thus fixed and one can run the procedure leading to the estimates $\widehat{\boldsymbol{\theta}}_k(\mathfrak{z}_1)$ for $k = 2, \ldots, K$. The value $\mathfrak{z}_1$ is selected as the minimal one for which

$$\boldsymbol{E}_{\boldsymbol{\theta}^*} \big| L_{I_k}\big(\widetilde{\boldsymbol{\theta}}_k, \widehat{\boldsymbol{\theta}}_k(\mathfrak{z}_1)\big)\big|^r \le \frac{\rho \mathfrak{R}_r(\boldsymbol{\theta}^*)}{K}, \qquad k = 1, \ldots, K. \tag{3.5}$$

Such a value exists because the choice $\mathfrak{z}_1 = \infty$ leads to $\widehat{\boldsymbol{\theta}}_k(\mathfrak{z}_1) = \widetilde{\boldsymbol{\theta}}_k$ for all k.

Next, with $\mathfrak{z}_1$ fixed, we can continue this way for all $\mathfrak{z}_k$, $1 < k \le K$. Suppose $\mathfrak{z}_1, \ldots, \mathfrak{z}_{k-1}$ have been already fixed. We set $\mathfrak{z}_{k+1} = \ldots = \mathfrak{z}_K = \infty$ and adjust $\mathfrak{z}_k$. Every particular choice of $\mathfrak{z}_k$ leads to estimates $\widehat{\boldsymbol{\theta}}_l(\mathfrak{z}_1, \ldots, \mathfrak{z}_k)$ for $l = k+1, \ldots, K$ coming out of the procedure with the parameters $\mathfrak{z}_1, \ldots, \mathfrak{z}_k, \infty, \ldots, \infty$. We select $\mathfrak{z}_k$ as the minimal value which fulfills

$$\boldsymbol{E}_{\boldsymbol{\theta}^*} \big| L_{I_l}\big(\widetilde{\boldsymbol{\theta}}_l, \widehat{\boldsymbol{\theta}}_l(\mathfrak{z}_1, \ldots, \mathfrak{z}_k)\big)\big|^r \le \frac{k\rho \mathfrak{R}_r(\boldsymbol{\theta}^*)}{K}, \qquad l = k, \ldots, K. \tag{3.6}$$

By simple induction arguments one can see that such a value exists since the choice $\mathfrak{z}_k = \infty$ provides a stronger inequality (3.6) for $k := k - 1$. Clearly, the final procedure with the parameters defined in the described way fulfills (3.3).

It is also worth mentioning that the numerical complexity of the proposed procedure is not very high. One can first generate M samples using the data generating process (2.5)-(2.6) with parameter $\boldsymbol{\theta}^*$ and compute and store estimates $\widetilde{\boldsymbol{\theta}}_k^{(m)}$, test statistics $T_{I_k}^{(m)}$, and values $L_{I_l}(\widetilde{\boldsymbol{\theta}}_k^{(m)})$ for every realization $m = 1, \ldots, M$ and all $k \le l = 1, \ldots, K$. Now, for a given set of critical values $\{\mathfrak{z}_k\}_{k=1}^K$, running the procedure and computing the estimates $\widehat{\boldsymbol{\theta}}_k^{(m)}$ and the loss $L_{I_k}(\widetilde{\boldsymbol{\theta}}_k, \widehat{\boldsymbol{\theta}}_k)$ requires only a fixed number of operations proportional to K. One can thus roughly bound the numerical complexity of selecting $\mathfrak{z}_k$'s by CMK^2 for some fixed constant C.

3.3.4 Selecting parameters r and ρ by minimizing the forecast error

The choice of critical values using inequality (3.3) 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. 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 $\widehat{\boldsymbol{\theta}}^{(r,\rho)}(T)$ and to different forecasts $\widehat{Y}^{(r,\rho)}_{T+h|T}$ of the future value 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:

$$(\widehat{r},\widehat{\rho}) = \operatorname*{argmin}_{r,\rho} PE_{\Lambda,\mathcal{H}}(r,\rho) = \operatorname*{argmin}_{r,\rho} \sum_{T} \sum_{h\in\mathcal{H}} \Lambda\{Y_{T+h}, \widehat{Y}^{(r,\rho)}_{T+h|T}\}, \tag{3.7}$$

where Λ is a loss function and $\mathcal{H}$ is the forecasting horizon set. For example, one can take $\Lambda_r(v,v') = |v-v'|^r$ for $r \in [1/2, 2]$. Another reasonable choice is $\Lambda^{\mathcal{K}}_c(v,v') = \big|\mathcal{K}(v,v')\big|^c$, where $\mathcal{K}(v,v')$ is the Kullback-Leibler divergence for measures P_v and $P_{v'}$ from the family $\mathcal{P}$. In the case of the volatility family, $\mathcal{K}(v,v') = -0.5\big\{\log(v/v') + 1 - v/v'\big\}$ and $c = 1$ or $c = 1/2$. For daily data, the forecasting horizon could be one day, $\mathcal{H} = \{1\}$, or two weeks, $\mathcal{H} = \{1,\ldots,10\}$.

4 Theoretic properties

In this section we collect results describing the quality of the proposed adaptive procedure. Note first that the definition of the procedure ensures the prescribed performance in the parametric situation, see (3.3). We however claimed that the pointwise adaptive estimation applies even if the process Y_t is locally only approximated by a parametric model. Therefore, we now define locally "nearly parametric" process, fow 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 modeling bias condition

This section discusses the concept of "nearly parametric" case. To treat this case in a rigorous mathematical way, we have to quantify the distance between the true latent process X_t, which drives the observed data Y_t by (2.5), and the parametric process $X_t(\boldsymbol{\theta})$ described by the parametric model (2.6) for some $\boldsymbol{\theta} \in \Theta$. To simplify presentation, we restrict ourselves to the case of the MLE estimation for model (2.5)-(2.6). In context of volatility modeling, this means that we consider the case of Gaussian innovations. Chen and Spokoiny (2007) explain how the study can be extended to the case of sub-Gaussian and heavy tailed-innovations.

Following Spokoiny and Chen (2007), we introduce for every interval I and every parameter $\boldsymbol{\theta} \in \Theta$ the random quantity

$$\Delta_I(\boldsymbol{\theta}) = \sum_{t \in I} \mathcal{K}\{g(X_t), g(X_t(\boldsymbol{\theta}))\},$$

where $\mathcal{K}(\upsilon, \upsilon')$ denotes the Kullback-Leibler distance between two distributions: $\mathcal{K}(\upsilon, \upsilon') = \boldsymbol{E}_\upsilon\{\log p(y, \upsilon) - \log p(y, \upsilon')\}$, where $p(y, \upsilon)$ represents the density of a measure $P_\upsilon \in \mathcal{P}$. In the case of volatility modeling, $\mathcal{K}(\upsilon, \upsilon') = -0.5\{\log(\upsilon/\upsilon') + 1 - \upsilon/\upsilon'\}$.

In the parametric case with $X_t = X_t(\boldsymbol{\theta}^*)$, we clearly have $\Delta_I(\boldsymbol{\theta}^*) = 0$. To characterize the "nearly parametric case," we introduce *small modeling bias* (SMB) condition, which simply means that, for some $\boldsymbol{\theta} \in \Theta$, $\Delta_I(\boldsymbol{\theta})$ 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 by the parametric one with the parameter $\boldsymbol{\theta}$. In this situation, the parameter $\boldsymbol{\theta}$ of the approximating model can be considered as the target of estimation and $\widetilde{\boldsymbol{\theta}}_I$ is the estimate of $\boldsymbol{\theta}$.

The following theorem then claims that the results on the accuracy of estimation given in Theorem 2.1 can be extended from the parametric case to the general nonparametric situation under the SMB condition. Let $\varrho(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta})$ be any loss function for an estimate $\widehat{\boldsymbol{\theta}}$ constructed from the observations $\{Y_t\}_{t \in I}$. Define also the corresponding risk under the parametric measure $\boldsymbol{P}_{\boldsymbol{\theta}}$: $\mathcal{R}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \boldsymbol{E}_{\boldsymbol{\theta}} \varrho(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta})$.

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

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

Then it holds for any estimate $\widehat{\boldsymbol{\theta}}$ *measurable with respect to* $\mathcal{F}_I$ *that*

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

If we now apply this general result to the quasi-MLE estimation with the loss function $L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta})$ and combine it with the bound for the parametric risk from Theorem 2.1, we yield the following corollary.

Corollary 4.1. *Assume that the SMB condition (4.1) holds for some interval* I *and* $\boldsymbol{\theta} \in \Theta$. *Then*

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

This result shows that the estimation loss $\bigl|L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta})\bigr|^r$ 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 nonparametric 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.

Furthermore, to understand the meaning of the SMB condition and the derived results in the context of varying-coefficient models, let us suppose that process Y_t follows the varying-coefficient model (3.1) with a time-dependent parameter vector $\boldsymbol{\theta}(t) \in \Theta$. Then the SMB condition can be easily reformulated in terms of the process $\boldsymbol{\theta}(t)$: for $\boldsymbol{\theta} \in \Theta$, the process $X_t(\boldsymbol{\theta})$ defined recurrently by the formula (2.6) is smooth with respect to the vector $\boldsymbol{\theta}$. In addition, the Kullback-Leibler divergence $\mathcal{K}(\upsilon, \upsilon')$ is typically proportional to $|\upsilon - \upsilon'|^2$. This implies that the value $\Delta_I(\boldsymbol{\theta})$ is of the same order as the squared L_2-distance between the "true" time-varying function $\boldsymbol{\theta}(t)$ and its constant approximation $\boldsymbol{\theta}$ on the interval I: $\Delta_I(\boldsymbol{\theta}) \asymp \sum_{t\in I}\bigl\|\boldsymbol{\theta}(t) - \boldsymbol{\theta}\bigr\|^2$. The SMB condition for the varying-coefficient models then means that $\boldsymbol{\theta}(t)$ is close to $\boldsymbol{\theta}$ within the whole interval I.

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 still neglect deviations of the underlying process (e.g., varying-coefficient model with a parameter $\boldsymbol{\theta}(t)$) from a parametric model with a fixed parameter $\boldsymbol{\theta}$. Therefore, we say that the choice k^* is "optimal" if there exists $\boldsymbol{\theta} \in \Theta$ such that

$$\boldsymbol{E}\Delta_{I_k}(\boldsymbol{\theta}) \leq \Delta, \qquad k \leq k^*, \tag{4.2}$$

for a fixed $\Delta > 0$ and that (4.2) does not hold for $k > k^*$.

By construction, the pointwise adaptive procedure described in Section 3 provides the prescribed performance if the underlying process follows a parametric model (2.4). Now, condition (3.3) combined with of Theorem 4.1 implies similar performance in the first k^* steps of the adaptive estimation procedure.

Theorem 4.2. *Let $\boldsymbol{\theta}$ and k^* be such that (4.2) holds for some $\Delta \geq 0$. Then*

$$\begin{aligned} \boldsymbol{E}\log\left(1 + \frac{\left|L_{I_{k^*}}(\widetilde{\boldsymbol{\theta}}_{k^*}, \boldsymbol{\theta})\right|^r}{\mathfrak{R}_r(\boldsymbol{\theta})}\right) &\leq 1 + \Delta \\ \boldsymbol{E}\log\left(1 + \frac{\left|L_{I_{k^*}}(\widetilde{\boldsymbol{\theta}}_{k^*}, \widehat{\boldsymbol{\theta}}_{k^*})\right|^r}{\rho\mathfrak{R}_r(\boldsymbol{\theta})}\right) &\leq 1 + \Delta, . \end{aligned}$$

Theorem 4.2 documents that the distance between oracle estimate $\widetilde{\boldsymbol{\theta}}_{k^*}$ and $\boldsymbol{\theta}$, measured by $L_{I_{k^*}}(\widetilde{\boldsymbol{\theta}}_{k^*}, \boldsymbol{\theta})$, is of the same magnitude as the distance $L_{I_{k^*}}(\widetilde{\boldsymbol{\theta}}_{k^*}, \widehat{\boldsymbol{\theta}}_{k^*})$ between the adaptive estimate $\widehat{\boldsymbol{\theta}}_{k^*}$ and the oracle $\widetilde{\boldsymbol{\theta}}_{k^*}$ itself. Thus, within the propagation phase under the condition (4.2), the adaptive pointwise estimation 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 $\Delta_{I_k}(\boldsymbol{\theta})$ can be large and the bound for the risk becomes meaningless due to the factor e^{Δ}. To establish the result about the quality of the final estimate, we thus have to show that the quality of estimation cannot be significantly destroyed at the steps $k > k^*$; that is, if $\widehat{\boldsymbol{\theta}}_k$ becomes equal to $\widetilde{\boldsymbol{\theta}}_l$ for some $l > k^*$. For this, we utilize the following "stability" result.

Theorem 4.3. *Suppose that interval I_k is accepted as time-homogeneous for some $k, k^* \le k \le K$. Then it holds for every $k^* \le l < k$ that*

$$\left|L_{I_l}(\widetilde{\boldsymbol{\theta}}_l, \widetilde{\boldsymbol{\theta}}_{l+1})\right| = \left|L_{I_l}(\widetilde{\boldsymbol{\theta}}_l) - L_{I_l}(\widetilde{\boldsymbol{\theta}}_{l+1})\right| \le \mathfrak{z}_{l+1}.$$

To understand the meaning of this "stability" result, we utilize the quadratic expansion of the fitted log-likelihood $L_I(\widetilde{\boldsymbol{\theta}}_I, \boldsymbol{\theta})$. Quadratic expansion (2.9) implies that $\|\widetilde{\boldsymbol{\theta}}_{I_l} - \widetilde{\boldsymbol{\theta}}_{I_{l+1}}\|^2 \le C\mathfrak{z}_{l+1}/|I_l|$ for any $k^* \le l < k$. By simple arguments, see Spokoiny and Chen (2007), it follows that the difference between the adaptive estimate $\widehat{\boldsymbol{\theta}}_k$ and the "oracle" estimate $\widetilde{\boldsymbol{\theta}}_{k^*}$ is of order $|I_{k^*}|^{-1/2}$ up to the multiplicative factor $\mathfrak{z}_{k^*}$. As the "oracle" estimate itself has the accuracy $|I_{k^*}|^{-1/2}$, this means nearly "oracle" quality of estimation. Polzehl and Spokoiny (2006) have shown that this "oracle" property implies the rate optimal estimation quality for the case of smoothly varying model coefficients. Spokoiny and Chen (2007) argued that this result guarantees the smallest (in rate) possible delay in detecting a change of the structure from the observed data.

5 Simulation study

In the last two sections, we present simulation study and real data applications documenting the performance of the proposed adaptive estimation procedure. Despite the generality of the proposed method, we concentrate on the volatility estimation using parametric and pointwise adaptive estimation of ARCH(1) and GARCH(1,1) models. The reason for this choice is to verify the practical applicability of the method in a complex settings. Specifically, the estimation of GARCH models requires generally hundreds of observations for reasonable quality of estimation (cf. Table 1), 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 3.3.2 depend on the underlying parameter values in the case of (G)ARCH (contrary to the case of homoscedastic autoregressive models, for instance).

Hence, we limit ourselves to adaptively estimated varying-coefficient constant

volatility, ARCH(1), and GARCH(1,1) models (for the sake of brevity, referred to also as the local constant, local ARCH, and local GARCH approximations). We first study the finite-sample critical values for the test of homogeneity by means of Monte Carlo simulations (Section 5.1). Later, we demonstrate the performance of the proposed pointwise adaptive estimation procedure in simulated samples and real data (Sections 5.2 and 6, respectively). Additionally, note that, throughout this section, we identify the GARCH(1,1) models by triplets (ω, α, β) : for example, $(1, 0.1, 0.3)$ -model. Constant volatility and ARCH(1), are then indicated by $\alpha = \beta = 0$ and $\beta = 0$, respectively. Finally, GARCH estimation is done using GARCH 3.0 package (Laurent and Peters, 2006) and Ox 3.30 (Doornik, 2002).

5.1 Finite-sample critical values for test of homogeneity

A practical application of the proposed adaptive procedure requires critical values for the test of local homogeneity of a time series. For given r and ρ, the average risk (3.3) between the adaptive and oracle estimates can be bounded for critical values that linearly depend on the logarithm of interval length $|I_k|$:

$$\mathfrak{z}_k = b_0 + b_1 k = c_0 + c_1 \log(|I_k|) \tag{5.1}$$

(see Theorem 3.1). Since such critical values are generally decreasing with the interval length, the linear approximation cannot be used for an arbitrarily long interval. Hence, we recommend to simulate critical values up to a certain interval length, for example $|I| = 1000$, and to use the critical values obtained for the latest interval considered also for longer intervals if needed.

Unfortunately, the critical values depend on the parameters of the underlying (G)ARCH model (in contrast to the case of local constant approximation). We simulated the critical values for ARCH(1) and GARCH(1,1) models with different values of underlying parameters; see Table 2 for critical values corresponding to $r = 1$ and $\rho = 1$. The adaptive estimation was performed sequentially on intervals with length ranging from $|I_0| = 10$ to $|I_K| = 570$ observations using a geometric grid with the initial interval length $m_0 = 10$ and multiplier $a = 1.25$, see Section 3.1. Note however that the results are not sensitive to the choice of a.

Table 2: Critical values $\mathfrak{z}_k = \mathfrak{z}(|I_k|)$ of the sequential supremum LR test defined by line (5.1) for various ARCH(1) and GARCH(1,1) models; $r = 1, \rho = 1$.

Model (ω, α, β)	$\mathfrak{z}(10)$	Slope	$\mathfrak{z}(570)$
$(0.1, 0.0, 0.0)$	15.4	-0.55	5.5
$(0.1, 0.2, 0.0)$	16.6	-0.40	9.4
$(0.1, 0.4, 0.0)$	23.4	-0.74	10.1
$(0.1, 0.6, 0.0)$	30.8	-1.05	11.9
$(0.1, 0.8, 0.0)$	73.6	-3.37	16.4
$(0.1, 0.1, 0.8)$	19.5	-0.29	14.3
$(0.1, 0.2, 0.7)$	26.3	-0.68	14.1
$(0.1, 0.3, 0.6)$	25.1	-0.58	14.6
$(0.1, 0.4, 0.5)$	28.9	-0.74	15.6
$(0.1, 0.5, 0.4)$	29.8	-0.83	14.9
$(0.1, 0.6, 0.3)$	34.4	-1.05	15.5
$(0.1, 0.7, 0.2)$	27.1	-0.66	15.2
$(0.1, 0.8, 0.1)$	29.2	-0.75	15.7
$(0.1, 0.05, 0.90)$	16.1	-0.14	13.6
$(0.1, 0.10, 0.85)$	19.4	-0.23	15.8
$(0.1, 0.20, 0.75)$	36.2	-1.15	15.5

Generally, the critical values seem to increase with the values of the ARCH parameter or the sum of the ARCH and GARCH parameters. To deal with the dependence of the critical values on the underlying model parameters, we propose to choose the largest (most conservative) critical values corresponding to any estimated parameter in the analyzed data. For example, if the largest estimated parameters of GARCH(1,1) are $\widehat{\alpha} = 0.3$ and $\widehat{\beta} = 0.8$, one should use $\mathfrak{z}(10) = 25.1$ and $\mathfrak{z}(570) = 14.6$. The proposed adaptive search procedure is however not overly sensitive to this choice as we shall see later.

Table 3: Critical values $\mathfrak{z}(|I|)$ of the sequential supremum LR test defined by line (5.1) for various ARCH(1) and GARCH(1,1) models and various values r and ρ.

Model (ω, α, β)	r	ρ	$\mathfrak{z}(10)$	Slope	$\mathfrak{z}(570)$
$(0.1, 0.0, 0.0)$	1.0	0.5	16.3	-0.50	7.3
	1.0	1.0	15.4	-0.54	5.5
	1.0	1.5	14.9	-0.58	4.5
	0.5	0.5	10.7	-0.20	7.1
	0.5	1.0	8.9	-0.19	5.5
	0.5	1.5	7.7	-0.17	4.6
$(0.1, 0.2, 0.0)$	1.0	0.5	16.0	-0.27	11.2
	1.0	1.0	16.5	-0.39	9.5
	1.0	1.5	16.4	-0.45	8.3
	0.5	0.5	11.7	-0.09	10.1
	0.5	1.0	10.3	-0.09	8.5
	0.5	1.5	9.3	-0.10	7.5
$(0.1, 0.1, 0.8)$	1.0	0.5	18.7	-0.09	17.1
	1.0	1.0	19.4	-0.28	14.4
	1.0	1.5	18.6	-0.29	13.4
	0.5	0.5	11.7	-0.09	10.1
	0.5	1.0	10.3	-0.10	8.5
	0.5	1.5	9.3	-0.10	7.5

Finally, let us have a look at the influence of the tuning constants r and ρ in (3.3) on the critical values for several selected models (Table 3). 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 critical values primarily for the shorter intervals (cf. (3.3)). In simulations and real applications, we verified that a fixed choice such as $r = 1$ and $\rho = 1$ performs well. To optimize

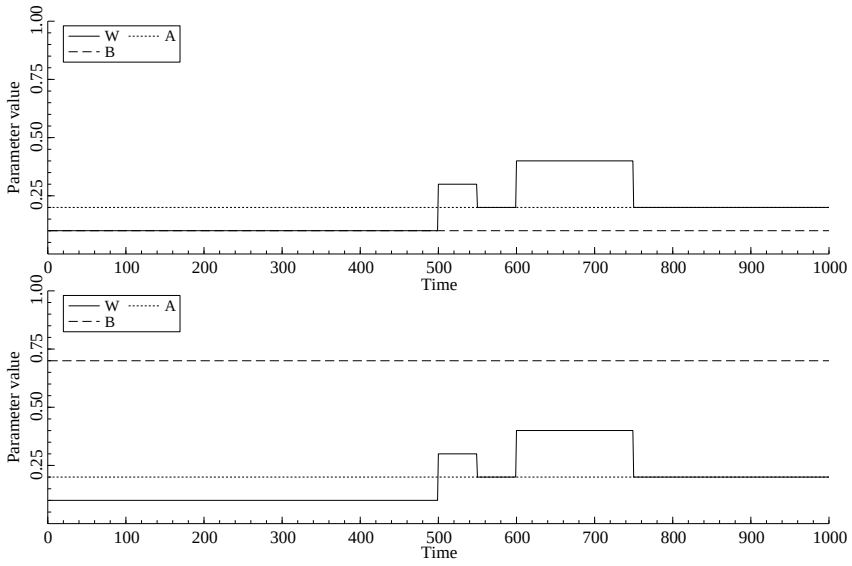

Figure 2: GARCH(1,1) parameters of low (upper panel) and high (lower panel) GARCH-effect simulations for $t = 1, \ldots, 1000$.

the performance of the adaptive methods, one can however determine constants r and ρ in a data-dependent way as described in Section 3.3.2. In the rest of this section and in Section 6, we use 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 the optimal choice of r and ρ; we will document though that the differences in the average prediction errors (3.7) for various values of r and ρ are relatively small.

5.2 Simulation study

The main concern of this simulation study is two-fold: (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 $(\omega, 0.2, 0.1)$ and a high GARCH-effect $(\omega, 0.2, 0.7)$. Changes in constant ω are spread over a time span of 1000 days, see Figure 2. There is a long stable period at the beginning

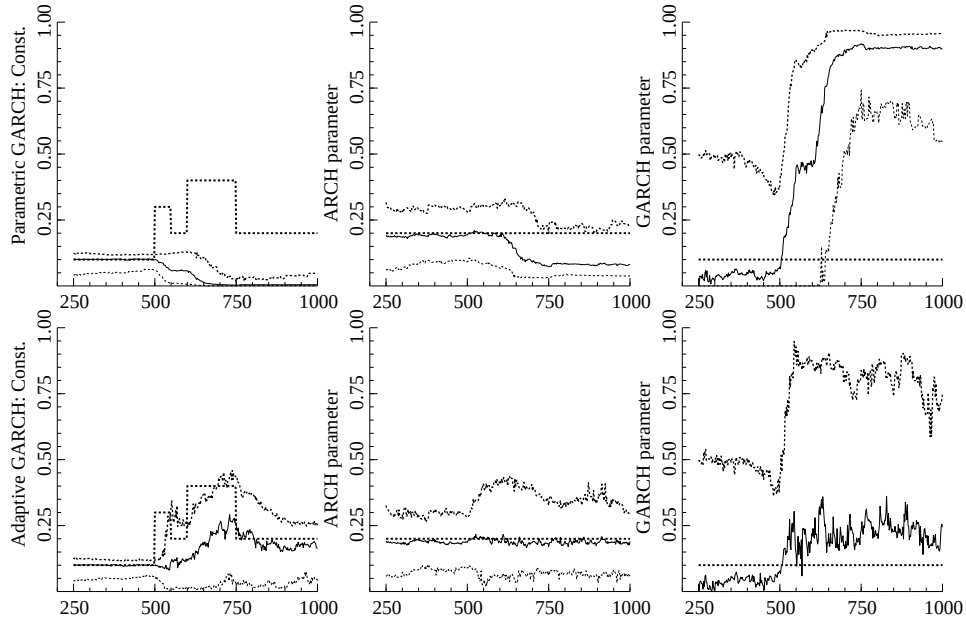

Figure 3: Parameters estimated by the parametric (upper row) and locally adaptive (lower row) GARCH methods, $t = 250, \ldots, 1000$. Thick dotted line represents the true parameter value, solid line is the mean estimate, and the upper and lower dotted lines represent 10% and 90% quantiles of parameter estimates.

(500 days $\approx$ 2 years) and end (250 days $\approx$ 1 year) of time series with several volatility changes between them.

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 prediction error (PE), that is, in terms of $PE_{\Lambda_1,\mathcal{H}}(r,\rho)$ from (3.7).

Estimating a single parametric model from data containing a change point will necessarily lead to various biases in estimation. For example, Hillebrand (2005) and Mikosch and Starica (2003) demonstrate that a change in volatility level ω within a sample drives the GARCH parameter β very close to 1. Therefore, it is interesting to see the estimated parameters from the adaptive estimation proce-

dure, which should try to detect and avoid change points. The parameter estimates both for parametric and adaptive GARCH at each time point $t \in [250, 1000]$ are depicted on Figure 3. Let us first observe that, whereas the parametric estimates are consistent before breaks starting at $t = 500$, 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. Table 1 and Figure 3 for $t < 500$).

Next, we would like to compare the performance of parametric and adaptive estimation methods by means of absolute prediction error $PE_{\Lambda_1,\mathcal{H}}$ in (3.7), where $\Lambda_1(v, v') = |v - v'|$; first for the prediction horizon of one day, $\mathcal{H} = \{1\}$, and later for prediction two weeks ahead, $\mathcal{H} = \{1, \ldots, 10\}$. To make the results easier to decipher, 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 for all simulations). 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. Hence, all methods "prefer" in this case flatter critical-value lines corresponding to $r = 0.5$.

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 setup. The average PEs for all methods forecasting one period ahead are presented on Figure 4. First of all, 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 $t < 625$. After initial volatility jumps, the local GARCH however outperforms the parametric one, $625 < t < 775$. Following the last jump at $t = 750$, where the volatility

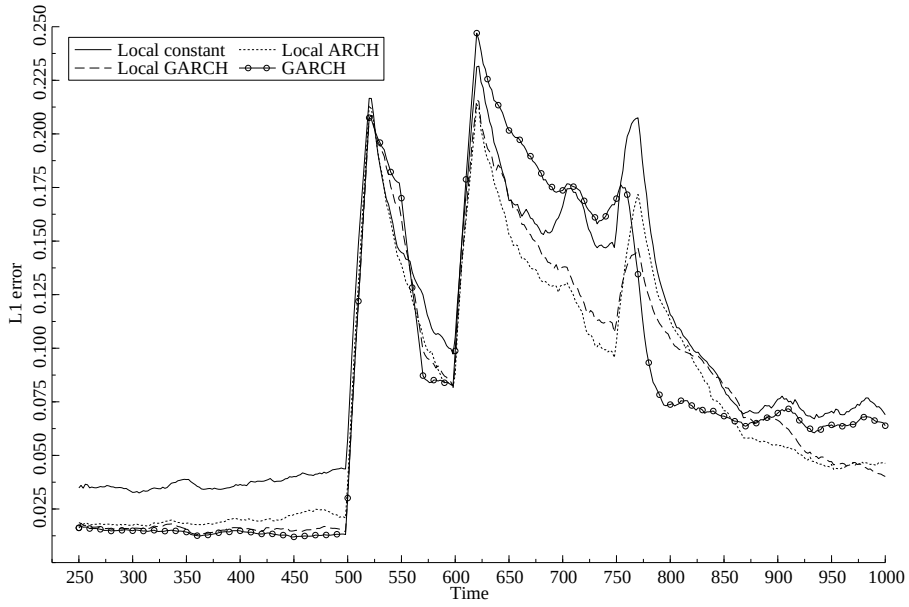

Figure 4: Low GARCH-effect simulations: absolute prediction errors one period ahead averaged over last month for the parametric GARCH and adaptive local constant, local ARCH, and local GARCH models; $t \in [250, 1000]$.

level returns closer to the original level (before $t < 500$), the parametric GARCH is best of all methods for some time, $775 < t < 850$, until the adaptive estimation procedure "collects" enough observations for estimation after the last change in volatility. Then the local GARCH (and also local ARCH) become preferable to the parametric model again, $850 < t$. Next, it is interesting to note that the local ARCH approximation performs almost as well as the GARCH methods and even outperforms them after several structural breaks, $600 < t < 775$ and $850 < t < 1000$ (the only exception to this is the last break at $t = 750$). Finally, the local constant estimation 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 more frequent volatility changes, $500 < t < 650$. All these observations can be documented also by the absolute PE averaged over the whole period, $t = 250, \ldots, 1000$ (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).

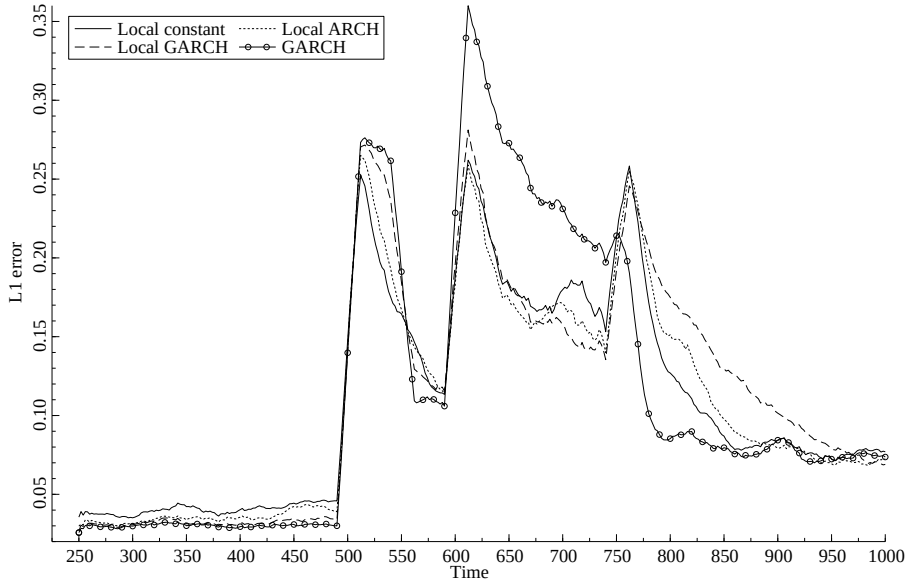

Figure 5: Low GARCH-effect simulations: absolute prediction errors ten periods ahead averaged over last month for the parametric GARCH and adaptive local constant, local ARCH, and local GARCH models; $t \in [250, 1000]$.

Additionally, all models are compared using the forecasting horizon of ten 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 on Figure 5, 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 even though the underlying model is GARCH (the global PEs are 0.108 and 0.116, respectively). The GARCH parameter $\beta = 0.1$ is however relatively small.

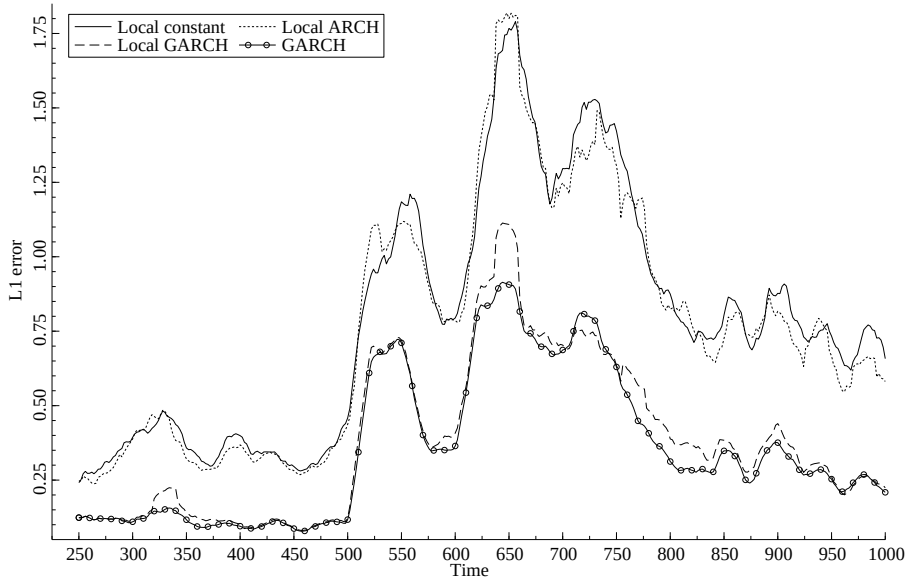

Figure 6: High GARCH-effect simulations: absolute prediction errors one period ahead averaged over last month for the parametric GARCH and adaptive local constant, local ARCH, and local GARCH models; $t \in [250, 1000]$.

5.2.2 High GARCH-effect

Let us now discuss the high GARCH-effect model. One would expect much more prevalent behavior 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" choice of critical values corresponded to the different values of tuning constant r and ρ than in the case of 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 one-period-ahead forecast at each time point (Figure 6) indicates that the adaptive and parametric GARCH estimation 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

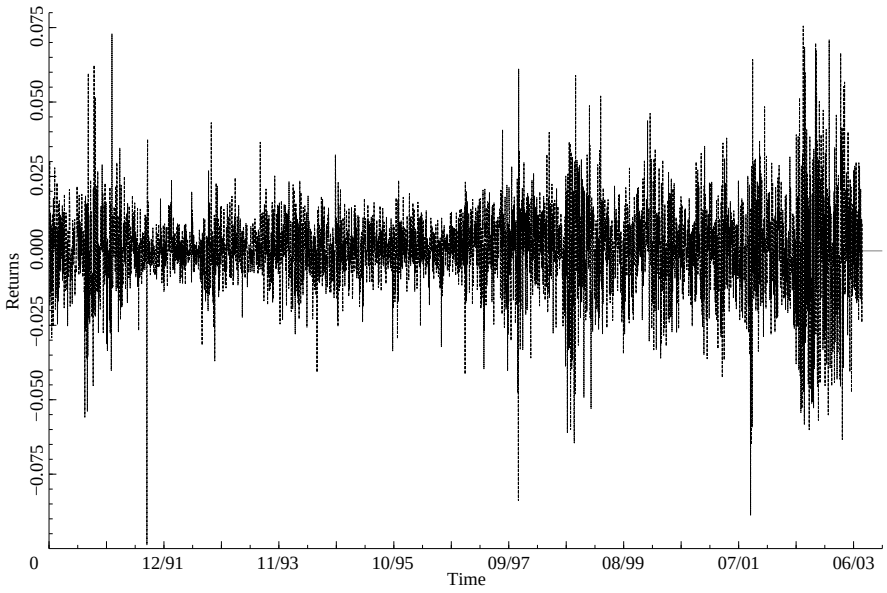

Figure 7: The log-returns of DAX series from January, 1990 till December, 2002.

of ten 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). Similarly to Section 5.2, we summarize the results concerning both parametric and adaptive methods by looking at absolute PEs one-day ahead averaged over one month throughout this section. As a benchmark, we use now the parametric GARCH estimated using last two years of data (500 observations). Since we however do not have the underlying volatility process in this case, we approximate the underlying volatility by squared returns. Despite being noisy, this approximation is unbiased and provides usually the correct ranking of methods (Andersen and Bollerslev, 1998).

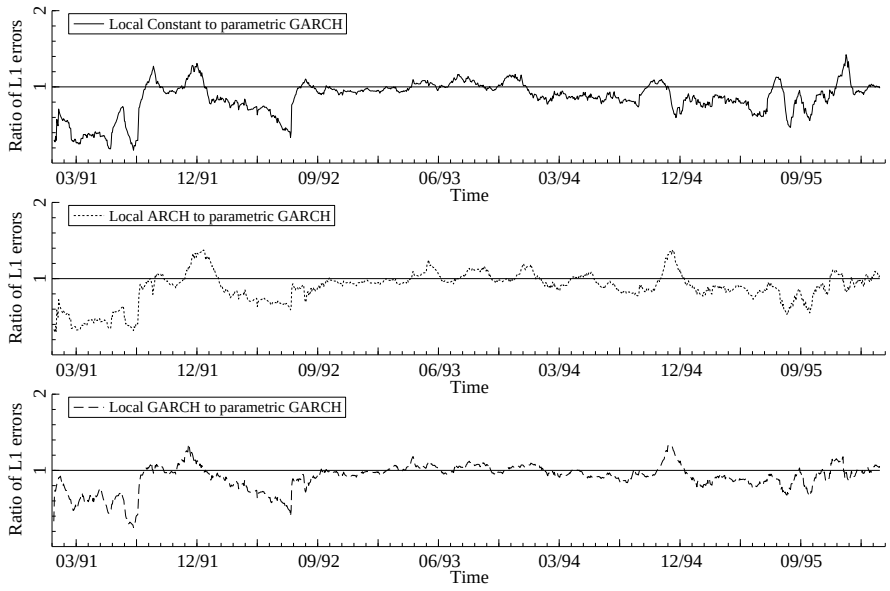

Figure 8: The ratio of the absolute prediction errors of the three pointwise adaptive methods to the parametric GARCH for predictions one period ahead averaged over one month. The DAX index is considered from January, 1992 to March, 1997.

6.1 DAX analysis

Let us now analyze the log-returns of the German stock index DAX from January, 1990 till December, 2002. The whole time series is depicted on Figure 7. In the following analysis, we select several periods interesting for comparing the performance of parametric and adaptive pointwise estimates since results for the whole period might be hard to decipher at once.

First, let us have a look at 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). In the case of the local GARCH approximation, this differs from less pronounced structural changes discussed later, which are typically detected only with several months delays. One additional break detected by all methods occurs in October 1994. Let us 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 this period are summarized in Figure 8, 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 optimal at the beginning of the period, where we have less than 500 observations. After the detection of the structural change in June 1991, from July 1991 on, all adaptive methods are shortly worse than the parametric GARCH due to limited amount of data used, but then outperform the parametric GARCH till the next structural break in the second half of 1992. A similar behavior can be observed after the break detected in October 1994, 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. Further, note that even the worst choice of r and ρ results in the global prediction errors 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 would like to 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 Russian crisis in 1998, market headed to a more stable state in year 1999. In this case, the adaptive methods detected the structural breaks in the fall of 1997 and 1998. The local GARCH detected them however with more than one-year delay, that is, only in the course of year 1999. The results presented in Figure 9 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 of the GARCH

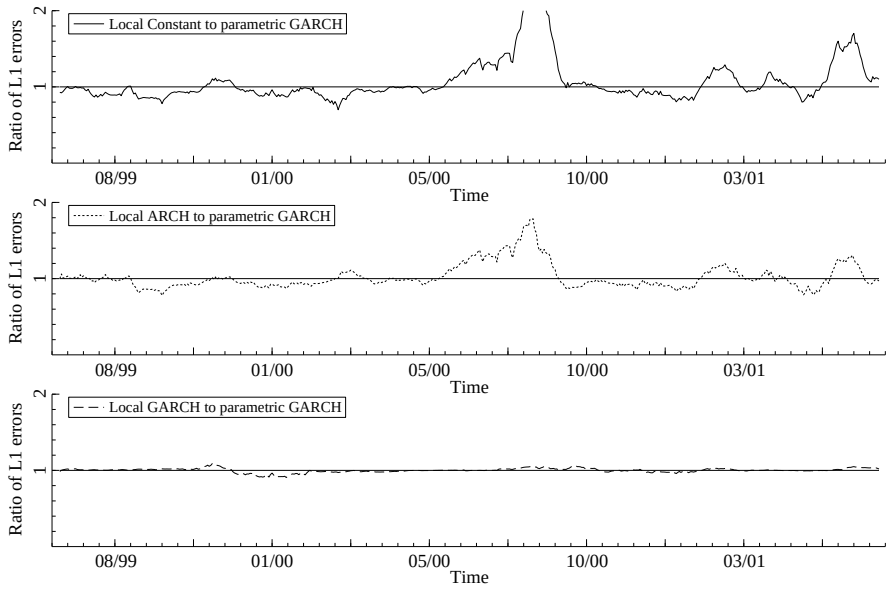

Figure 9: The ratio of the absolute prediction errors of the three pointwise adaptive methods to the parametric GARCH for predictions one period ahead averaged over one month. The DAX index is considered from July, 1999 to June, 2001.

models is better (parametric and adaptive GARCH estimates are practically the same in this period since the last detected structural change occured approximately two years ago). Together with previous results, this opens the question of model selection among adaptive procedures as different base models (ARCH or GARCH) might be preferred in different time periods. Nevertheless, judging by the global PE, the local ARCH surprisingly provides slightly better predictions on average than the local GARCH approximation – despite the "peak" of the PE ratio in the second half of year 2000 (see Figure 9).

Finally, let us mention that the relatively similar behavior 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.

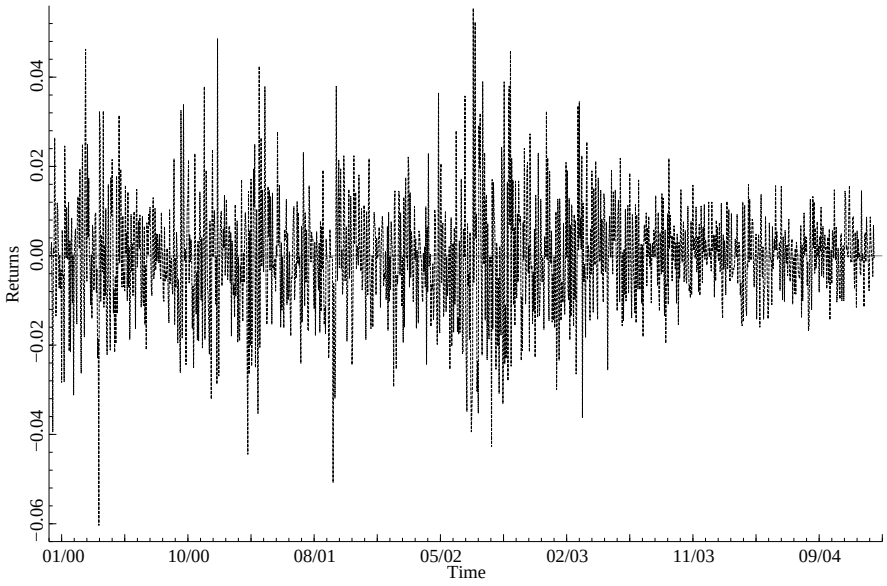

Figure 10: The log-returns of S&P 500 from January, 2000 till December, 2004.

6.2 S&P 500

Now we turn our attention to more recent data regarding the S&P 500 stock index considered from January, 1990, to December, 2004, see Figure 10. 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 (2000–2002). For the sake of simplicity, a particular time period is again selected. The estimation results for years 2003 and 2004, where the first one represent a more volatile period (war on terrorism in Iraq) and the latter one is a less volatile period. All adaptive methods detected rather quickly a structural break at the beginning of 2003; additionally, all methods also detected a structural break in the second half of 2003, although the adaptive GARCH did so with a delay of more than 8 months. The ratios of monthly PE of all adaptive methods to the parametric GARCH are summarized on Figure 11 ($r = 0.5$ and $\rho = 1.5$ for all methods).

In the beginning of 2003, which together with previous year 2002 corresponds to a more volatile period (see Figure 10), all adaptive methods perform as well

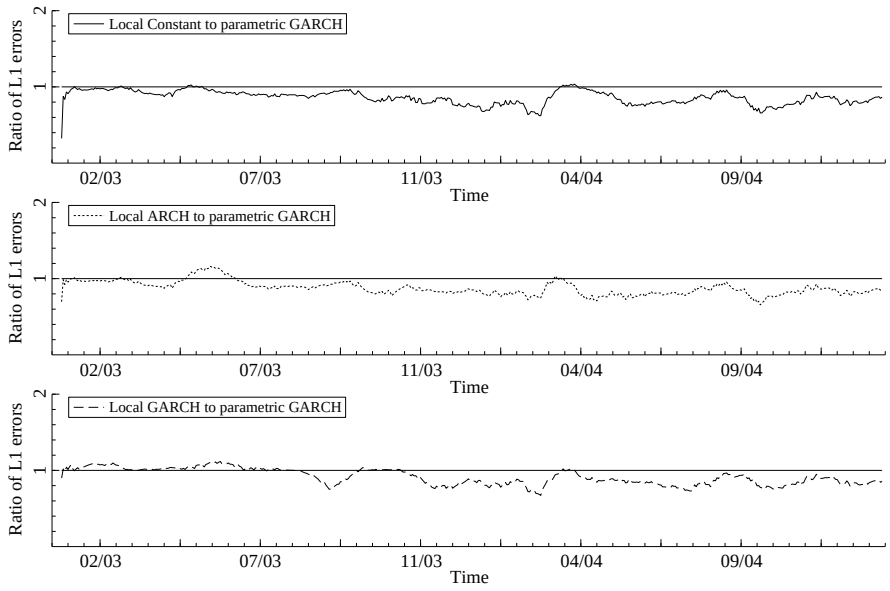

Figure 11: The ratio of the absolute prediction errors of the three pointwise adaptive methods to the parametric GARCH for predictions one period ahead averaged over one month horizon. The S&P 500 index is considered from January, 2003 to December, 2004.

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 pointwise adaptive 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., in year 2004 here) 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 rather large variance when the intervals of time homogeneity become very short.

7 Conclusion

In this paper, we extend the idea of adaptive pointwise estimation to more complex parametric models that belong to the class of autoregressive moving-average conditional-heteroscedasticity models. In the specific case of ARCH and GARCH, which represent particularly difficult case 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 the choice of these parameters, and on the other hand, it facilitates the global selection of these parameters by means of fit or forecasting criteria. Further, the real-data applications highlight the flexibility of the proposed time-inhomogeneous models since even simple varying-coefficients models such as ARCH(1) can outperform standard parametric methods such as GARCH(1,1).

A Proofs

Proof of Corollary 2.1. Given the choice of $\mathfrak{z}_\alpha$, it directly follows from (2.8). □

Proof of Theorem 3.1. We present only the idea of the proof. The detailed proof of a similar assertion for the local constant volatility modeling can be found in Spokoiny and Chen (2007). The main fact behind the proof is that selecting the critical values of order $\log|I_K|$ allows to make the probability of rejecting the null very small.

It follows from the definition that for every $\tau \in \mathcal{T}_k \subset I_{k+1}$, the test statistic $T_{I_{k+1},\tau}$ satisfies $T_{I_{k+1},\tau} \le L_J(\widetilde{\boldsymbol{\theta}}_J, \boldsymbol{\theta}^*) + L_{J^c}(\widetilde{\boldsymbol{\theta}}_{J^c}, \boldsymbol{\theta}^*)$ with $J = [\tau, T]$ and $J^c = I_{k+1} \setminus J$. By Theorem 2.1, there are some positive constants C and λ such that $\boldsymbol{P}_{\boldsymbol{\theta}^*}\big(L_J(\widetilde{\boldsymbol{\theta}}_J, \boldsymbol{\theta}^*) > \mathfrak{z}\big) \le Ce^{-\lambda\mathfrak{z}}$ for any $\mathfrak{z} > 0$; an analogous inequality holds also for $L_{J^c}(\widetilde{\boldsymbol{\theta}}_{J^c}, \boldsymbol{\theta}^*)$. Therefore,

$$\boldsymbol{P}_{\boldsymbol{\theta}^*}\big(T_{I_{k+1},\tau} > 2\mathfrak{z}\big) \le 2Ce^{-\lambda\mathfrak{z}},$$

which obviously implies

$$\boldsymbol{P}_{\boldsymbol{\theta}^*}\big(T_{I_{k+1},\mathcal{T}_k} > 2\mathfrak{z}\big) \le 2C|\mathcal{T}_k|e^{-\lambda\mathfrak{z}}.$$

Selecting $\mathfrak{z}_k$ in the form (3.4) then ensures that the probability of rejecting the null hypothesis at step k is smaller than $Const.(e^{-\lambda a_1} + |I_K/I_k|^{-\lambda a_2} + e^{-\lambda a_3})$ for some sufficiently large a_1, a_2, a_3. □

Proof of Theorem 4.1. The proof is based on the following general result. □

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*

$$\boldsymbol{E}\log\big(1+\zeta\big) \le \Delta + \boldsymbol{E}_0\zeta.$$

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 $\boldsymbol{E}\log\big(1+\zeta\big) = \boldsymbol{E}_0\big\{Z\log\big(1+\zeta\big)\big\}$ with $Z = d\boldsymbol{P}/d\boldsymbol{P}_0$ we obtain

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

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

We now apply this lemma with $\zeta = \varrho(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta})/\mathcal{R}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta})$ and utilize that $\boldsymbol{E}_0\zeta = \boldsymbol{E}_\theta \varrho(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta})/\mathcal{R}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = 1$. This yields

$$\begin{aligned}\boldsymbol{E}_{\boldsymbol{\theta}}\big(Z_{I,\boldsymbol{\theta}} \log Z_{I,\boldsymbol{\theta}}\big) &= \boldsymbol{E} \log Z_{I,\boldsymbol{\theta}} = \boldsymbol{E} \sum_{t\in I} \log \frac{p[Y_t, g(X_t)]}{p[Y_t, g(X_t(\boldsymbol{\theta}))]} \\ &= \boldsymbol{E} \sum_{t\in I} \boldsymbol{E}\Big\{ \log \frac{p[Y_t, g(X_t)]}{p[Y_t, g(X_t(\boldsymbol{\theta}))]} \Big| \mathcal{F}_{t-1} \Big\} = \boldsymbol{E}\Delta_I(\boldsymbol{\theta}). \qquad \square\end{aligned}$$

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

Proof of Theorem 4.2. The first inequality follows from Corollary 4.1, the second one from condition (3.3) and the property $x \geq \log x$ for $x > 0$. $\square$

Proof of Theorem 4.3. By assumption $\widehat{k} > l$, the interval I_{l+1} is accepted by our multiple testing procedure. This particularly means that the sup-LR statistic $T_{I_{l+1}}$ does not exceed $\mathfrak{z}_{l+1}$, and therefore for $\tau = t_l$, it holds $T_{I_{l+1},t_l} \leq \mathfrak{z}_{l+1}$. With $J_l = I_{l+1} \setminus I_l$, this means that

$$\begin{aligned}T_{I_{l+1},t_l} &= L_{I_l}(\widetilde{\boldsymbol{\theta}}_{I_l}) + L_{J_l}(\widetilde{\boldsymbol{\theta}}_{J_l}) - L_{I_{l+1}}(\widetilde{\boldsymbol{\theta}}_{I_{l+1}}) \\ &= L_{I_l}(\widetilde{\boldsymbol{\theta}}_{I_l}, \widetilde{\boldsymbol{\theta}}_{I_{l+1}}) + L_{J_l}(\widetilde{\boldsymbol{\theta}}_{J_l}, \widetilde{\boldsymbol{\theta}}_{I_{l+1}}) \geq L_{I_l}(\widetilde{\boldsymbol{\theta}}_{I_l}, \widetilde{\boldsymbol{\theta}}_{I_{l+1}}).\end{aligned}$$

Therefore, $T_{I_{l+1}} \leq \mathfrak{z}_{l+1}$ implies $L_{I_l}(\widetilde{\boldsymbol{\theta}}_{I_l}, \widetilde{\boldsymbol{\theta}}_{I_{l+1}}) \leq \mathfrak{z}_{l+1}$ as required. $\square$

References

[1] Andersen, T. G., Bollerslev, T., 1998. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. International Economic Review 39(4), 885–905.

[2] Andreou, E., Ghysels, E., 2002. Detecting multiple breaks in financial market volatility dynamics. Journal of Applied Econometrics 17(5), 579–600.

[3] Andreou, E., Ghysels, E., 2006. Monitoring disruptions in financial markets. Journal of Econometrics 135, 77–124.

[4] Andrews, D. W. K., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61(4), 821–856.

[5] Andrews, D. W. K., 2003. End-of-Sample Instability Tests. Econometrica 71(6), 1661–1694.

[6] Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66, 47–78.

[7] Beltratti, A., Morana, C., 2004. Structural change and long-range dependence in volatility of exchange rates: either, neither or both? Journal of Empirical Finance 11(5), 629–658.

[8] Beltratti, A., Morana, C., 2006. Breaks and persistency: macroeconomic causes of stock market volatility. Journal of Econometrics 131, 151–177.

[9] Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31(3), 307–327.

[10] Cai, Z., Fan, J., Li, R., 2000. Efficient estimation and inferences for varying-coefficient models. Journal of the American Statistical Association 95, 888–902.

[11] Charles, A., Darn, O., 2006. Large shocks and the September 11th terrorist attacks on international stock markets. Economic Modelling 23, 683–698.

[12] Chen, J., Gupta, A. K., 1997. Testing and locating variance changepoints with application to stock prices. Journal of the American Statistical Association 92, 739–747.

[13] Chen, Y., Spokoiny, V., 2007. Robust risk management: accounting for non-stationarity and heavy tails. WIAS Discussion Paper 1207, Berlin, Germany.

[14] Cheng, M.-Y., Fan, J., Spokoiny, V., 2003. Dynamic nonparametric filtering with application to volatility estimation. In Akritas, M. G., Politis, D. N. (Eds.) Recent Advances and Trends in Nonparametric Statistics. Elsevier, North Holland, 315–333.

[15] Chu, C.-S. J., Stinchcombe, M., White, H., 1996. Monitoring Structural Change. Econometrica 64, 1045–1065.

[16] Diebold, F. X., Inoue, A., 2001. Long memory and regime switching. Journal of Econometrics 105(1), 131–159.

[17] Doornik, J. A., 2002. Object-oriented Programming in Econometrics and Statistics using Ox: A Comparison with C++, Java and C#. In Nielsen, S. S. (Ed.), Programming Languages and Systems in Computational Economics and Finance. Kluwer Academic Publishers, Dordretch, 115–147.

[18] Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50(4), 987–1008.

[19] Engle, R. F., Lillian, D. M., and Robins, R. P., 1987. Estimating time-varying risk premium in the term structure: ARCH-M Model. Econometrica 55, 391–407.

[20] Fan, J., Zhang, W., 1999. Additive and varying coefficient models – statistical estimation in varying coefficient models. The Annals of Statistics 27(5), 1491–1518.

[21] Fan, J., Yao, Q., Cai, Z., 2003. Adaptive varying-coefficient linear models. Journal of the Royal Statistical Society, Series B 65(1), 57–80.

[22] Glosten, L. R., Jagannathan, R., Runkle, D. E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. The Journal of Finance 48(5), 1779–1801.

[23] Hansen, L. P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

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

[25] Härdle, W., Herwatz, H., Spokoiny, V., 2003. Time inhomogeneous multiple volatility modelling. Journal of Financial Econometrics 1, 55–99.

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

[27] Hillebrand, E., 2005. Neglecting parameter changes in GARCH models. Journal of Econometrics 129(1–2), 121–138.

[28] Kokoszka, P., Leipus, R., 2000. Change-point estimation in ARCH models. Bernoulli 6.

[29] Laurent, S., Peters, J.-P., 2006. G@RCH 4.2, estimating and forecasting ARCH models. Timberlake Consultants Press, London.

[30] McConnell, M. M., Perez-Quiros, G., 2000. Output fluctuations in the United States: what has changed since the early 1980's? The American Economic Review 90(5), 1464–1476.

[31] Mercurio, D., Spokoiny, V., 2004. Statistical inference for time-inhomogeneous volatility models. The Annals of Statistics 32(2), 577–602.

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

[33] Mikosch, T., Starica, C., 2003. Non-stationarities in financial time series, the long-range dependence, and IGARCH effects. The Review of Economics and Statistics 86, 378–390.

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

[35] Nelson, D. B., 1991. Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59(2), 347–370.

[36] Orbe, S., Ferreira, E., Rodriguez-Poo, J., 2005. Nonparametric estimation of time varying parameters under shape restrictions. Journal of Econometrics 126(1), 53–77.

[37] Perron, P., Zhu, X., 2005. Structural breaks with deterministic and stochastic trends. Journal of Econometrics 129, 65–119.

[38] Pesaran, M. H., Timmermann, A., 2004. How costly is it to ignore breaks when forecasting the direction of a time series? International Journal of Forecasting 20(3), 411–425.

[39] Polzehl, J., Spokoiny, V., 2006. Propagation-separation approach for local likelihood estimation. Probability Theory and Related Fields 135, 335–362.

[40] Sentana, E., 1995. Quadratic ARCH Models. The Review of Economic Studies 62(4), 639–661.

[41] Sowell, F., 1996. Optimal tests for parameter instability in the generalized method of moments framework. Econometrica 64, 1085–1107.

[42] Spokoiny, V., 1998. Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. The Annals of Statistics 26(4), 1356–1378.

[43] Spokoiny, V., Chen, Y., 2007. Multiscale local change-point detection with with applications to Value-at-Risk. To appear in The Annals of Statistics. See `http://www.wias-berlin.de/people/spokoiny/listpubm.html`.

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

[45] Taylor, S. J., 1986. Modeling financial time series. Wiley, Chichester.

[46] Wolters, J., Terasvirta, T., Lütkepohl, H., 1990. Modeling the demand for M3 in the unified Germany. The Review of Economics and Statistics 80(3), 399–409.