ADAPTIVE ESTIMATION FOR A TIME INHOMOGENEOUS STOCHASTIC-VOLATILITY MODEL

HÄRDLE, W., SPOKOINY, V.G., AND TEYSSIÈRE, G.

Humboldt-Universität zu Berlin,
Spandauer Straße 1, 10178 Berlin, Germany
and
University Würzburg
Sanderring 2, 97070, Würzburg, Germany

ABSTRACT. Let a process $S_1, \ldots, S_T$ obey the conditionally heteroskedastic equation $S_t = v_t \varepsilon_t$ where ε_t is a random noise and v_t is the volatility coefficient which in turn obeys an autoregression type equation $\log v_t^2 = \omega + \alpha S_{t-1}^2 + \eta_t$ with an additional noise η_t. We consider the situation when the parameters ω and α might also depend on the 'time' t, and we study the problem of online estimation of current values of $\omega = \omega(T)$ and $\alpha = \alpha(T)$ from the observations $S_1, \ldots, S_T$. We propose an adaptive method of estimation which does not use any information about time homogeneity of the observed process. We apply this model to two series of FX daily returns on DEM/USD and GBP/USD.

1. Introduction

Standard models of finance theory assume time homogeneous dynamics of the considered asset. Numerous empirical papers for various stock markets have questioned such a homogeneity. A time-varying risk premium has been considered by e.g. Engle et al. (1987), time varying covariances have been investigated by e.g. Bollerslev et al. (1988), Hafner and Herwartz (1988) and Campbell et al. (1997). The provided empirical evidence has shown that time homogeneous dynamic models such as a discrete version of the geometric Brownian motion may be outperformed by models with varying volatility parameters. The dominating models in this respect are the GARCH (Engle, 1982 and Bollerslev, 1986) and stochastic volatility (Taylor, 1986) models. These models are designed to reflect the

Date: October 15, 1999.

1991 *Mathematics Subject Classification.* 62M10; Secondary 62P20.

Key words and phrases. stochastic volatility model, adaptive estimation, local homogeneity .

stylized fact of time inhomogeneous volatility clustering by a suitable presentation of the variance function as a function of lagged values of (external) parameters.

So far these stylized facts were mainly attributed to time-varying behavior of conditional second moments. For the capital asset pricing model, for example, time-varying betas may be obtained by multivariate GARCH and ARCH-M models that naturally link to inhomogeneous risk premium. The increasing applications of GARCH models has led to more adequate specifications, The consistency of a model specification over time, however, is questionable. One might argue that, despite more flexible modelling of covariance patterns more adequate specifications and predictions may be given through identification of interval of homogeneity of a model as time increases.

By the very definition of the time varying covariance models, they are as models also time homogeneous. The model equation contains elements reflecting time inhomogeneous clustering but the mechanism of these volatility patterns stays the same over the observation period. Changes in the volatility parameters are not considered although structural changes are practically relevant. In empirical investigations of financial returns one is very interested in identifying intervals of homogeneity where a certain estimated approximation reflects the market situation well. Identified intervals of homogeneity may lead to certain investment patterns in asset management.

In this paper we present an adaptive estimation technique for intervals of homogeneity in the context of stochastic volatility models. Our approach is adaptive in the sense that we identify intervals of homogeneity without making assumptions on their length or location. The empirical evidence we provide is for the DEM/USD and BPD/USD time series of exchange rate data.

We define an interval of homogeneity for a model as a time interval in which the model parameters do not vary very much relative to the estimation error of these parameters in the considered time interval. This setting seems at first sight to be related with change-point detection. Note, however, that the change-point approach allows only for structural breaks at few separated time moment and assumes the time-homogeneous model structure between changes, see e.g. Antoch, Hušková, and Prášková (1997), Csörgő and Horváth (1997), Kokoszka and Leipus (1998), Spokoiny (1998), Dufour and Ghysels (1996), and Hackl and Westlund (1991) among others.

Our approach is more general since we allow for changing parameters at any time point but we do not consider these changes as time inhomogeneous if they do not appear to be significantly bigger than the expected variation in the considered interval.

The paper is organized as follows. In the next section we formulate the theoretical setup of time-homogeneous stochastic volatility systems. Section 4 introduces time-inhomogeneous models and gives the description of the estimation procedure and their

theoretical properties. In Section 4 we present the adaptive estimation technique. Section 5 discusses theoretical properties of the proposed procedure. Section 6 gives simulation results showing that the procedure works. Estimation results on two samples of FX rates returns at daily frequency are given in Section 7. The proofs are postponed to the Appendix.

2. Time homogeneous stochastic volatility models

Let $S_1, \ldots, S_T$ be the returns of an asset price process obeying the conditional heteroskedasticity equation

$$S_t = v_t \varepsilon_t \tag{2.1}$$

where ε_t is a multiplicative random noise and v_t is the volatility scaling coefficient, $t = 1, 2, \ldots$. Assume that the errors ε_t are i.i.d. random variables satisfying the condition

$$\boldsymbol{E}\ \varepsilon_t^2 = 1$$

The aim is to infer about the structure of the volatility process v_t. We use another representation of this model obtained by transformation to the log-scale, $Y_t = \log S_t^2$. Applying the function $\log(x^2)$ to both sides of (2.1) leads to

$$Y_t = \log v_t^2 + \log \varepsilon_t^2 = z_t + \xi_t \tag{2.2}$$

with

$$\begin{aligned} z_t &= \log v_t^2 + \nu, \\ \xi_t &= \log \varepsilon_t^2 - \nu, \\ \nu &= \boldsymbol{E}\ \log \varepsilon_t^2. \end{aligned}$$

Standard modelling approach for the process (2.1) or the transformed model (2.2), see Gourieroux (1997), is based on certain parametric assumptions about the scaling process v_t. One specific example studied in this paper is given by a *stochastic-volatility model.* In its simplest from with autoregressive term it reads as follows:

$$z_t = \omega + \alpha Y_{t-1} + \eta_t, \qquad t = 2, 3, \ldots \tag{2.3}$$

where η_t is a noise process, and ω, α are unknown coefficients. If we knew the joint distribution of the noise processes ε_t and η_t, then the properties of the volatility process v_t and the asset price process Y_t are completely determined by the coefficients ω and α. This reduces the original problem to the problem of estimating ω, α, and well

developed theory applies in this situation: the maximum likelihood estimate (MLE) is root-n consistent and asymptotically efficient, see e.g. Gourieroux (1997).

In practical applications, however, this approach might be too restrictive and being applied to real time series, it may lead to a considerable modelling bias. One possibility to make this kind of modelling more flexible is to allow for flexible high order autoregression and moving averaging structures as e.g.

$$z_t = \omega + \alpha_1 Y_{t-1} + \cdots + \alpha_p Y_{t-p} + \beta_1 z_{t-1} + \cdots + \beta_q z_{t-q} + \eta_t. \tag{2.4}$$

Then the coefficients $\alpha_1, \ldots, \alpha_p$, $\beta_1, \ldots, \beta_q$ and also the dimensionality parameters p, q could be estimated from data with a reasonable accuracy. This type of modelling will certainly fit better to real data but it is still restricted to the class of time homogeneous processes since the structure of the equation (2.4) is time homogeneous, that is, the involved coefficients do not depend on the time index t.

Dahlhaus (1997), Neumann and von Sachs (1997), Dahlhaus et al. (1999) picked up this point and studied the problem of statistical inference for locally stationary time series. In the case of the simplest model (2.3) their approach allows the coefficients ω and α for time dependence of the form $\alpha = \alpha(t/m)$ and $\omega = \omega(t/m)$. Here $\alpha(\cdot)$ and $\omega(\cdot)$ are smooth unknown functions and the parameter m controls the degree of time homogeneity: the processes S and v are almost stationary inside each time interval of length smaller (in order) than m.

Here we consider the situation when the coefficients ω and α in (2.3) are varying piecewise smooth functions, $\omega = \omega(t)$ and $\alpha = \alpha(t)$. This means that for every point s, there exists a number $m = m(s)$ such that the parameters $\omega(t)$ and $\alpha(t)$ do not change significantly within the interval $[s-m, s]$. The difference between our modelling approach and that of Dahlhaus (1997) is that we also allow the parameter m (controlling the degree of local homogeneity) to vary from point to point. This particularly helps to include the case of spontaneous change in parameters raised by exogeneous (e.g. non stochastic political) perturbations. In our approach the statistical inference focuses on estimation of the "current" values of parameters at the ??? observation T and the primary objective is to determine in a data driven way the largest interval of local homogeneity $[T - m, T]$ corresponding to the time point T. The values $\omega(T)$ $\alpha(T)$ can be estimated from the observations in this interval. The parameter m describing the degree of local homogeneity before the time point T, can be viewed as a "smoothness" parameter. In this sense the problem can be interpreted as choosing a smoothness parameter at the right boundary of the interval of observations. Our aim is to adapt our estimate to this unknown parameter. The proposed method to be presented in Section 4 is led by the ideas of the pointwise adaptive estimation in Spokoiny (1998) for the regression model.

3. Time Heterogeneous Stochastic Volatility Models

A time heterogeneous stochastic-volatility model is described by the following two equations

$$\begin{aligned} S_t &= v_t \varepsilon_t, \\ \log v_t^2 &= \omega(t) + \alpha_1(t) \log S_{t-1}^2 + \cdots + \alpha_p(t) \log S_{t-p}^2 + \eta_t, \qquad t = p+1, p+2, \ldots . \end{aligned}$$

The goal is to estimate the current values $\omega(T)$ and $\alpha_1(T), \ldots, \alpha_p(T)$ from the observations $S_1, \ldots, S_T$ and to determine intervals of homogeneity in which the parameters are not varying too much.

In the sequel we will use the representation (2.2) and treat $Y_t = \log S_t^2$ as our observations. In order to outline the proposed procedure, it is useful to introduce vector notation. Let X_t be the vector in $\mathbb{R}^{p+1}$ with $X_t = (1, Y_{t-1}, \ldots, Y_{t-p})^T$ and $\theta(t) = \bigl(\theta_0(t), \ldots, \theta_p(t)\bigr)^T$ be the vector of parameters where

$$\begin{aligned} \theta_0(t) &= \omega(t) + \mu, \\ \theta_k(t) &= \alpha_k(t), \qquad k = 1, \ldots, p. \end{aligned}$$

The model equation (2.2) reads then as:

$$Y_t = X_t^T \theta(t) + \zeta_t \tag{3.1}$$

with

$$\zeta_t = \xi_t + \eta_t = \log \varepsilon_t^2 - \boldsymbol{E}\ \log \varepsilon_t^2 + \eta_t$$

so that ζ_t is an i.i.d. 'noise' process with $\boldsymbol{E}\ \zeta_t = 0$ and $\sigma^2 = \boldsymbol{E}\ \zeta_t^2 = \boldsymbol{E}\ (\xi_t + \eta_t)^2$.

In order to construct a consistent estimation of $\theta(T)$ we need to specify the assumption of local homogeneity of the process Y_t.

3.1. Local homogeneity assumption

Local time homogeneity means that the parameter vector-function $\theta(t)$ does not differ too much from a constant vector θ within some interval I of the form $I = [t_1, T]$. Here t_1 denotes the left end point and T the right end-point of the interval of homogeneity. The time heterogeneous model (3.1) is approximated within this interval by a homogeneous model

$$Y_t = X_t^T \theta + \zeta_t, \qquad t \in I. \tag{3.2}$$

Given this situation, we construct an estimate of θ from the observations Y_t for $t \in I$ and use this value for estimating $\theta(T)$. We consider the pseudo maximum likelihood

estimate (pseudo MLE)

$$\widetilde{\theta}_I = (\widetilde{\theta}_{0,I}, \dots, \widetilde{\theta}_{p,I})^T = \left(\sum_{t \in I} X_t X_t^T \right)^{-1} \sum_{t \in I} X_t Y_t \tag{3.3}$$

which is the ML estimate for the case of Gaussian errors ζ_t. The interval I here has to be interpreted as the parameter of the method. If no information about the degree of homogeneity of the process Y_t is available, then this approach leads to the problem of adaptive estimation: given a family of estimates $\{\widetilde{\theta}_I\}$ indexed by the parameter I (or, equivalently, by the corresponding left end-point t_1) one has to select in a data-driven way one of them which leads to the minimal possible risk of estimation.

Before presenting our adaptive procedure, we need to study the properties of the estimate $\widetilde{\theta}_I$ in a little more detail.

3.2. **Properties of** $\widetilde{\theta}_I$

Due to our assumption of local homogeneity, the vector $\theta(t)$ is close to a constant vector θ for all $t \in I$. This means that the value

$$\Delta_I = \sup_{t \in I} |\theta(t) - \theta(T)|$$

is small. (Here $|v|$ denotes the Euclidean norm of the vector $v \in \mathbb{R}^{p+1}$.) Define also the random matrices V_I and W_I as:

$$\begin{aligned} V_I &= \sigma^{-2} \sum_{t \in I} X_t X_t^T, \\ W_I &= V_I^{-1}. \end{aligned}$$

The elements of these matrices will be denoted by $v_{ij,I}$ and $w_{ij,I}$ respectively, $i, j = 0, 1, \dots, p$. We also use the notation $\sigma^2_{i,I}$ for the diagonal elements of W_I, $i = 0, \dots, p$.

In the case of a regression model of the form (3.2) with deterministic design $X_1, \dots, X_T$, the estimate $\widetilde{\theta}_I$ is the least squares estimate and W_I is its covariance matrix. In particular, each diagonal element $w_{ii,I} = \sigma^2_{i,I}$ of this matrix is the variance of the estimate $\widetilde{\theta}_{i,I}$, $i = 0, 1, \dots, p$. In our situation the design points X_t's are random and correlated with the errors ζ_t's. By analogy with the regression case, we call the value $\sigma^2_{i,I}$ the *conditional variance* of $\widetilde{\theta}_{i,I}$, $i = 0, \dots, p$.

Since the V_I matrix is random, we introduce a random set where certain regularity conditions are satisfied and in the sequel restrict our consideration to this set. Given positive numbers $b > 0$, $B \geq 1$, $r \geq 1$ and $\delta < 1$, denote by $A_{i,I}$, $i = 0, \dots, p$, the

random set where the following conditions are fulfilled:

$$A_{i,I} = \left\{ \begin{array}{c} b \le w_{ii,I}^{-1} \le bB, \\ w_{ii,I}\|V_I\| \le r, \\ \left| \frac{w_{ij,I}}{w_{ii,I}} \right| \le \rho, \quad \forall j = 0, \ldots, p \end{array} \right\}.$$

Here $\|M\|$ denotes the sup-norm of the matrix M: $\|M\| = \sup_{\lambda \in \mathbb{R}^{d+1}:|\lambda|=1} |M\lambda|$. Note that under usual stationarity assumptions, the normalized matrix $|I|^{-1}V_I = \sigma^{-2}|I|^{-1}\sum_{t \in I} X_t X_t^T$ converges to a deterministic matrix V as $|I|$ grows, see e.g. Anderson and Walker (1964). If in addition this limit matrix is non-degenerated, the inverse matrix $|I|W_I$ converges to V^{-1} which obviously provides for properly selected constants b, B, r and ρ that $\mathbf{P}(A_{i,I})$ is close to one if the interval I contains sufficiently many observations.

Theorem 3.1. *Let $Y_1, \ldots, Y_T$ obey (3.1) and let the errors ζ_t be i.i.d. $\mathcal{N}(0, \sigma^2)$-distributed. Then it holds for the estimate $\widetilde{\theta}_I$*

$$\mathbf{P}\left(|\widetilde{\theta}_{i,I} - \theta_i(T)| > \Delta_I + \lambda\sigma_{i,I}, A_{i,I} \right) \le P_{B,r,\rho,p}(\lambda) \, \exp\left(-\lambda^2/2\right), \qquad i = 0, \ldots, p$$

where $P_{B,r,\rho,p}(\lambda)$ is the following polynom of λ of degree $p+1$

$$P_{B,r,\rho,p}(\lambda) = 4e\,(1 + \log B)\,(1 + 2\rho\sqrt{rp}\lambda)^p\,\lambda. \tag{3.4}$$

Remark 3.1. In Theorem 3.1 we assumed Gaussian errors ζ_t. In principle, this and the following results may be obtained also for the case of i.i.d. errors fulfilling Cramér's condition: $\boldsymbol{E} \exp\kappa|\zeta_i| < \infty$ for some $\kappa > 0$. We restrict ourselves here to the Gaussian case for the sake of exposition simplicity. We also comment on the extensions using Cramér's conditions in Section 5.3.

4. Adaptive choice of the interval of homogeneity

Let us recall the problem setting again. Given observations $Y_1, \ldots, Y_T$ from the time-inhomogeneous model (3.1), we aim to estimate the current value of the vector-parameter $\theta(T)$ using the estimate $\widetilde{\theta}_I$ with a properly selected time interval I of the form $[T-m, T]$ to minimize the corresponding estimation error. Due to Theorem 3.1, the loss of every estimate $\widetilde{\theta}_{i,I}$, $i = 0, \ldots p$, can be bounded with a high probability by the sum of two terms: Δ_I and $\lambda\sigma_{i,I}$, $i = 0, 1, \ldots, p$ (provided that λ is large enough). The first term Δ_I characterizes the variability of the function $\theta(\cdot)$ within the interval I and it is typically small if $|I|$ is small, where $|I|$ denotes the number of points X_t in I, i.e., for

$I = [T-m, T]$, we have $|I| = m+1$. By contrast, the term $\sigma_{i,I}$ is relatively large for small $|I|$ and it decreases with $|I|$. This can be illustrated in the stationary case where the values $\sigma^2_{i,I}$, $i = 0, \dots, p$ – the diagonal elements of the matrix $W_I = \sigma^2 \left(\sum_{t \in I} X_t X_t^T \right)^{-1}$ – decrease in the order of $|I|^{-1}$.

Therefore, the choice of the interval I or equivalently the parameter m relies on minimization of the sum of two terms, one of them (Δ_I) increases with m while the other one decreases. It is well known from nonparametric estimation theory that a reasonable choice of the "smoothing parameter" m is defined by balancing the terms Δ_I and $\sigma_{i,I}$, $i = 0, 1, \dots, p$. One remarkable fact here is that all values $\sigma_{i,I}$ may be directly calculated from the data, since they depend only on the observations $Y_1, \dots, Y_T$ and on the variance σ^2 of the innovations ζ_t. If σ^2 is unknown, it may also be estimated from the data, see Section 4.3 below.

The main problem in selecting m is that the "bias" term Δ_I depends on the unknown target function $\theta(t)$ and is usually unknown a priori. Below we discuss one adaptive (data driven) approach which goes back to the idea of pointwise adaptive estimation, see Lepski (1990), Lepski and Spokoiny (1997) and Spokoiny (1998). The idea of the method can be explained as follows. Suppose a family $\mathcal{I}$ of intervals of the form $I = [T - m, T]$ is fixed. This family is naturally ordered via the values $|I| = m + 1$. With every such interval we associate the estimate $\widetilde{\theta}_I$ of the vector-parameter $\theta(T)$ from (3.3) with the corresponding standard deviations $\sigma_{i,I}$, $i = 0, 1, \dots, p$. We now check successfully the intervals from the family $\mathcal{I}$ starting from the smallest one whether they fulfill the hypothesis of homogeneity. The testing of homogeneity is done by comparison of the estimate $\widetilde{\theta}_I$ based on the observations from I with similar estimates constructed on the base of observations falling in some subintervals of the interval I. If the hypothesis of homogeneity is not rejected we continue with the next larger interval from the given family. Otherwise we stop and utilize for the final estimation the largest non-rejected interval.

Let us present a formal description. Suppose we are given a family $\mathcal{I}$ of interval-candidates. Also, for every $I \in \mathcal{I}$, we also suppose to be given a set $\mathcal{J}(I)$ of testing subintervals J of I (one example of these sets $\mathcal{I}$ and $\mathcal{J}(I)$ is given in the next section). For every $I \in \mathcal{I}$ (resp. every $J \in \mathcal{J}(I)$) we construct the corresponding estimate $\widetilde{\theta}_I$ (resp. $\widetilde{\theta}_J$) from the observations Y_t with $t \in I$ (resp. $t \in J$) according to (3.3).

Given $I \in \mathcal{I}$, we also define $\mathcal{I}_I$ as the set of all intervals $I'' \in \mathcal{I}$ containing I, that is, $\mathcal{I}_I = \{I'' \in \mathcal{I} : I \subseteq I''\}$, and set

$$\sigma^*_{i,I} = \max\{\sigma_{i,I''} : I'' \in \mathcal{I}_I\}.$$

Typically $\sigma_{i,I}$ decreases with $|I|$ so that $\sigma^*_{i,I} = \sigma_{i,I}$. Now, with two constants λ and μ, define the adaptive choice of the interval of homogeneity by the following iterative procedure:

Initialization: Select the smallest interval in $\mathcal{I}$.

Iteration: Select the next interval I in $\mathcal{I}$ and calculate the corresponding estimate $\widetilde{\theta}_I$ and the conditional variances $\sigma^2_{i,I}$, $i \le p$.

Testing homogeneity: Reject I, if there exists one $J \in \mathcal{J}(I)$ and $i \le p$ such that

$$|\widetilde{\theta}_{i,I} - \widetilde{\theta}_{i,J}| > \lambda \sigma_{i,J} + \mu \sigma^*_{i,I};$$

Loop: If I is not rejected, then continue with the iteration step by choosing a larger interval. Otherwise, set $\widehat{I} =$ "the latest non rejected I".

The adaptive estimate $\widehat{\theta}(T)$ of $\theta(T)$ is defined by applying this selected interval $\widehat{I}$:

$$\widehat{\theta}_i(T) = \widetilde{\theta}_{i,\widehat{I}}, \qquad i = 0, \dots, p.$$

It is supposed that the procedure is independently carried out at each time point T, that is $\widehat{I} = \widehat{I}(T)$ and similarly for $\widehat{\theta}$. Some other possibilities allowing to reduce the computational effort are discussed in the next section.

4.1. Choice of the sets $\mathcal{I}$, $\mathcal{J}(I)$ and the parameters λ and μ

The presented algorithm involves the sets $\mathcal{I}$ and $\mathcal{J}(I)$ of considered intervals and two numeric parameters λ and μ. We now discuss how these parameters can be selected starting from the set of intervals $\mathcal{I}$. The simplest proposal is to introduce a regular grid $t_k = m_0 k$ with some natural number m_0 and to consider the intervals $I_k = [t_k, T]$ for all $t_k \le T$. The value m_0 can be selected, e.g., between 10 and 30.

Next, for every such interval $I = [t_k, I]$, we define the set $\mathcal{J}(I)$ of testing intervals J by taking all smaller intervals $I' = [t_{k'}, T]$ with the right end-point T and similarly all smaller intervals $[t_k, t_{k'}]$ with the left end-point t_k, $k < k' \le k_0$:

$$\mathcal{J}(I_k) = \{J = [t_{k'}, T] \text{ or } J = [t_k, t_{k'}] : k < k' < k_0\}.$$

Let N_I stand for the number of subintervals J in $\mathcal{J}(I)$. Clearly, $\mathcal{J}(I_k)$ contains at most $2(k_0 - k)$ elements, that is, $N_{I_k} \le 2(k_0 - k)$.

In the light of Theorem 5.1 below, parameter λ should be chosen to make the value $\left(\sum_{I \in \mathcal{I}} N_I\right) P_{B,r,\rho,p}(\lambda) e^{-\lambda^2/2}$ sufficiently small, which leads to the choice $\lambda \approx \sqrt{(2+\delta)\log M}$ with $M = \sum_{I \in \mathcal{I}} N_I$ and some $\delta > 0$. The simulations we carried out showed that the choice of λ between 3 and 3.5 provides a good quality of estimation.

There is a higher degree of freedom for the choice of μ. The theoretical recommendation is $\mu > \lambda$. Note, however, that if I is essentially larger than J, then $\sigma_{i,I}$ is essentially smaller than $\sigma_{i,J}$ for all $i \le p$ and in such a situation the contribution of

the term $\mu\sigma_{i,I}$ in the critical value $\lambda\,\sigma_{i,J} + \mu\,\sigma_{i,I}$ can be compensated by a slight increase of λ in the first term $\lambda\sigma_{i,J}$. This consideration prompts to take $\mu = 0$ which is in agreement with arguments presented in Lepski and Spokoiny (1997) and Lepski and Levit (1997). Our simulation results are also in agreement with the latter proposal.

4.2. **Computational issues**

To realize the procedure at one point T, one should calculate an estimate $\widetilde{\theta}_J$ for every interval J from $\bigcup_{I\in\mathcal{I}}\mathcal{J}(I)$, that is, the total number of estimates to be calculated is of order $\sum_{I\in\mathcal{I}} N_I$. This may lead to a serious computational effort. To reduce the computation time, we recommend, as in the definition of the sets $\mathcal{I}$ and $\mathcal{J}(I)$ to apply the adaptive procedure only at points $t_k = km_0$ of the same arithmetic grid. For every point T between two neighbor points of the grid, one may keep the left end-point of the latest adaptively selected interval. This means that the procedure is carried over only at successful points $t_1, t_2, \ldots$ and hence, when determining the next adaptive interval of homogeneity at a point $T = t_k$, we have to calculate only the estimates with the right end-point T, keeping the estimates calculated at previous time-moments.

4.3. **Variance estimation**

The previously described procedure requires to know the variance σ^2 of the errors ζ_t. In practical applications this information is typically lacking which leads to the problem of variance estimation from the data. There is a number of proposals in the literature. We refer to Fan and Yao (1998) for an overview. The regression-like representation (3.1) and our local time-homogeneity prompts to apply the residuals-based estimate from Gasser et al. (1986) developed for the nonparametric regression: first, given $m > p$, one set $I = [t-m, t]$ and construct the estimate $\widetilde{\theta}_I$ of the vector $\theta(t)$ due to (3.3) from the data X_s, Y_s, $t - m \le t \le t$. Next the *pseudo-residuals* $\widehat{e}_t$ are defined as $\widehat{e}_t = Y_t - X_t^T\widetilde{\theta}_I$. Finally, the variance estimator $\widehat{\sigma}^2$ is defined by averaging these pseudo-residuals squared:

$$\widehat{\sigma}^2 = \frac{1}{T - t_0}\sum_{t=t_0}^{T}\widehat{e}_t^2.$$

Here t_0 is a starting point which should satisfy $t_0 > m$. It is known in the regression context that such an estimator slightly overestimates the true variance and the pseudo-residuals $\widehat{e}_t$ should be properly weighted to avoid these overestimation, see e.g. Gasser et al. (1986). However, the proposed procedure performs suitably if m is sufficiently large. We apply this estimate with $m = p + 12$ which shows a reasonable performance for the simulated data sets.

5. Theoretical properties

In this section we collect some results describing the quality of the proposed adaptive procedure.

5.1. Accuracy of the adaptive estimate

Let $\widehat{I}$ be the interval selected by our adaptive procedure. We also define the "ideal" choice $I\!I$ as the largest interval for which the variability of the vector-function $\theta(t)$ is not essentially bigger than the conditional standard deviation of the corresponding estimate:

$$I\!I = \operatorname{argmax} \left\{ |I| : I \in \mathcal{I},\, \Delta_I \le D\sigma^*_{i,I},\, i = 0, 1, \ldots, p \right\}. \tag{5.1}$$

For this 'ideal' choice $I = I\!I$ we have the balance between the accuracy of approximation (which is controlled by Δ_I) and the stochastic error characterized by the stochastic variance $\sigma^*_{i,I}$. This, due to Theorem 3.1, allows us to bound with a high probability the losses of the "ideal" estimate $\widetilde{\theta}_{i,I\!I}$ by $(D+\lambda)\sigma^*_{i,I\!I}$ provided that λ is sufficiently large, $i = 0, 1, \ldots, p$. The next assertion claims that the risk of the adaptive estimate is of the same order $\sigma^*_{i,I\!I}$.

Let the random events $A_{i,I}$ be introduced before Theorem 3.1 and

$$A_i = \bigcap_{I \in \mathcal{I}} A_{i,I}.$$

Theorem 5.1. *Let errors $\zeta_1, \ldots, \zeta_T$ be i.i.d. Gaussian r.v.'s with zero mean and the variance σ^2. Then it holds for the adaptive estimate $\widehat{\theta} = \widetilde{\theta}_{\widehat{I}}$ defined in Section 4 with $\mu > \lambda + 2D$:*

$$\begin{aligned} &\mathbf{P}\left(|\widehat{\theta}_i - \theta_i(T)| > (2\lambda + D + \mu)\sigma^*_{i,I\!I}, A_i \right) \\ &\qquad \le \left(1 + \sum_{I \in \mathcal{I}(I\!I)} N_I \right) P_{B,r,\rho,p}(\lambda)\, \exp\left(-\lambda^2/2\right) \end{aligned} \tag{5.2}$$

with $P_{B,r,\rho,p}(\lambda)$ from (3.4).

5.2. Sensitivity to change-points

Here we briefly discuss the behaviour of the proposed adaptive estimate in the situation when the underlying parameter θ is a piecewise constant function of time, that is, this vector spontaneously changes at some time moments $t_{\mathrm{cp},\ell}$, $\ell = 1, 2, \ldots$. To simplify the exposition, we suppose that all $t_{\mathrm{cp},\ell}$'s coincide with points t_k of the grid defining the intervals I from $\mathcal{I}$. This assumption does not restrict generality provided that the grid t_k is dense enough. In this situation, for each time moment T, the corresponding interval of homogeneity coincides with $[t_{\mathrm{cp}}, T]$ where t_{cp} is the latest change-point before T.

Due to the result of Theorem 5.1, the quality of estimating the vector θ is of the same order as for the "ideal" estimate $\widetilde{\theta}_{I\!I}$. We now can show a slightly stronger assertion: the adaptively selected interval $\widehat{I}$ is essentially as $[t_{\rm cp}, T]$ provided that the amplitude of the change is sufficiently large compared to the noise level.

Note first, that the arguments used in the proof of Theorem 5.1 yield that the probability to reject an interval $I \in \mathcal{I}$ which is strictly contained in $[t_{\rm cp}, T]$ is small (provided that λ is large enough). Now it remains to show that, if $I \in \mathcal{I}$ is larger than $[t_{\rm cp}, T]$ in the sense that there is an interval $J \in \mathcal{I}(I)$ lying from the left of $t_{\rm cp}$, then such I will be rejected with a probability close to one.

Let θ denote the parameter vector describing the process Y_t immediately after the change-point $t_{\rm cp}$ and let θ' be the similar parameter vector before $t_{\rm cp}$.

Proposition 5.1. *Let $I\!I = [t_{\rm cp}, T]$ be the interval of homogeneity between the change-point $t_{\rm cp}$ and the point of estimation T. Let next $I = [T_{k'}, T]$ be a larger interval from $\mathcal{I}$ for some $T_{k'} < t_{\rm cp}$. If, for some $i \le p$, it holds with $J = [T_{k'}, t_{\rm cp}]$*

$$|\theta_i - \theta'_i| \ge 2\lambda\sigma_{i,J} + 2(\lambda+\mu)\sigma^*_{i,I\!I}, \tag{5.3}$$

then

$$\mathbf{P}\left(I \text{ is rejected}, A_i\right) \ge 1 - 2P_{B,r,\rho,p}(\lambda)\,\exp\left(-\lambda^2/2\right) \tag{5.4}$$

with $P_{B,r,\rho,p}(\lambda)$ from (3.4).

We now specify the previous results for the most important situation when we estimate the parameters immediately after a change point $t_{\rm cp}$.

Let T be the point of estimation which is close to $t_{\rm cp}$. The corresponding interval of homogeneity is $I\!I = [t_{\rm cp}, T]$. We are interested to describe the delay in detecting the change-point by the adaptive procedure. The fact of detecting the change-point can be treated as rejecting all the intervals I which are essentially larger than $I\!I$, like in Proposition 5.1.

Corollary 5.1. *Let the process Y_t follow the autoregression equation (2.1) with some vector of coefficients θ' before the time moment $t_{\rm cp}$ and with another vector θ after $t_{\rm cp}$ so that*

$$|\theta_i - \theta'_i| \ge b$$

for some $i \le p$ and some $b > 0$. Let then the conditional variances $\sigma^2_{i,I}$ satisfy

$$\sigma^2_{i,I} \le C|I|^{-1}$$

for all $I \in \mathcal{I}$ and for all $i \le p$. If

$$m = Cb^{-2}(4\lambda + 2\mu)^2$$

then the adaptively selected interval $\widehat{I}$ *fulfills for* $T \geq t_{\rm cp} + m$ *with* $I\!I = [t_{\rm cp}, T]$ *:*

$$\mathbf{P}\left(I\!I \subseteq \widehat{I} \subseteq [t_{\rm cp} - m, T]\right) \geq 1 - 2\left(1 + \sum_{I \in I\!I} N_I\right) P_{B,r,\rho,p}(\lambda) \exp\left(-\lambda^2/2\right)$$

with $P_{B,r,\rho,p}(\lambda)$ *from (3.4).*

Remark 5.1. Let $b = \max_{i \leq p} |\theta_i - \theta_i'|$ be the amplitude of the change point. The message of Corollary 5.1 is that the procedure detects this change point (in the sense that it starts to select $\widehat{I} \approx I\!I = [t_m, T]$) when the distance $T - t_{\rm cp}$ between the point of estimation and the change point becomes of order $\lambda^2 b^{-2}$.

5.3. **Extensions**

The proposed approach is based on the local ARCH(p)-assumption (3.1) which generalizes usual ARCH-models to the case of time varying coefficients. It is however well known that an ARCH-assumption is quite restrictive and it is required to take a large order p to get a reasonable fit to real data. A GARCH-type modelling which allows for ARMA-type representation (2.4) for the log-returns $Y_t = \log S_t^2$ is more flexible in mimicking the long-range dependence structure of high frequency data.

Here we shortly discuss how the proposed approach can be applied to the more complicated situation of GARCH-type modelling. One typical example is given by the model

$$\begin{aligned} Y_t &= z_t + \xi_t, \\ z_t &= \omega + \beta_1 z_{t-1} + \eta_t. \end{aligned}$$

Its time heterogeneous analog is

$$\begin{aligned} Y_t &= z_t + \xi_t, \\ z_t &= \omega(t) + \beta_1(t) z_{t-1} + \eta_t. \end{aligned} \tag{5.5}$$

The proposed approach can be extended to such modelling in several ways. One possibility is to construct the pseudo maximum-likelihood estimates of the underlying parameters $\omega(t)$ and $\beta(t)$ and then to analyze the variability in time of these estimates. One technical problem here is that the corresponding pseudo maximum-likelihood estimates are not linear w.r.t. the observations Y_t and their variance depends on the unknown parameter values. This makes more delicate to carry over the adaptive procedure this way.

Another possibility is based on the the idea of equivalent AR-representation. Indeed, by the Wold Theorem (Wold, 1954) every stationary process can be represented as an infinite order autoregressive process. This also applies for the model (5.5) provided that $|\omega| + |\beta| < 1$. As a consequence, the previously proposed procedure applies here as well if the order of autoregression p is taken sufficiently large. A disadvantage of considering a large p is that we need to estimate a large number of parameters at each time point which

makes the procedure computationally extensive and unstable. In addition, this requires to consider larger intervals in the sets $\mathcal{I}$ and $\mathcal{J}(I)$'s which decreases the sensitivity to change-points.

One more possibility which can be recommended for practical application is to perform the same procedure as described having in mind that we estimate only few parameters among others. The underlying idea is that the estimated autoregressive coefficients are strongly correlated and a change of even one of them leads to some variation in all estimates.

The delivered information cannot be used for forecasting the process Y_t but it can be useful for a qualitative analysis of time homogeneity of this process. This can be illustrated by our example from Section 7 where we observe structural breaks simultaneously in all considered parameters.

Another important assumption on the model (3.1) is normality of the innovations ζ_t. This assumption is essentially used to prove the large deviation bound in Theorem 3.1. In practical applications these innovations ζ_t are far from being normally distributed. Usually the multiplicative errors ε_t in the conditional heteroscedastic model (2.1) are assumed normal, so that ζ_t are log-normal. Nevertheless, the result of Theorem 3.1 can be extended to the case of non-normal variables ζ_t under some moment conditions so that the procedure applies in that case as well.

6. Simulations

We test our method by generating processes with controlled change point and check whether the method is able to detect that change point and to estimate the parameters of the process. For both simulations and estimations, the parameter m_0 for the grid size is set to 50.

We consider two DGP: one threshold model, i.e., the transition between the two models is immediate after the change point, and a smooth transition model with gradual transition.

6.1. Threshold model DGP

We consider the following DGP:

$$Y_t = \omega + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \varepsilon_t \quad \varepsilon_t \sim N(0,1) \tag{6.1}$$

where

$$\omega = 0.1 \quad \alpha_1 = 0.15 \quad \alpha_2 = 0.45 \quad \text{for} \quad 1 \le t \le 500 \tag{6.2}$$

$$\omega = 0.2 \quad \alpha_1 = 0.55 \quad \alpha_2 = 0.15 \quad \text{for} \quad 501 \le t \tag{6.3}$$

The three pictures below report the pointwise estimated parameter, the dark line. The pointwise confidence bands are given by the grey lines. These confidence bands are equal to the averaged L_1 errors $S_t(\theta) = N^{-1} \sum_{K=1}^{N} |\widehat{\theta}_t^{(K)} - \theta_t|$.

Thes pictures show that the procedure estimates the parameter with sufficient accuracy and is very sensitive to the change point which is very quickly detected.

FIGURE 1. Constant term ω

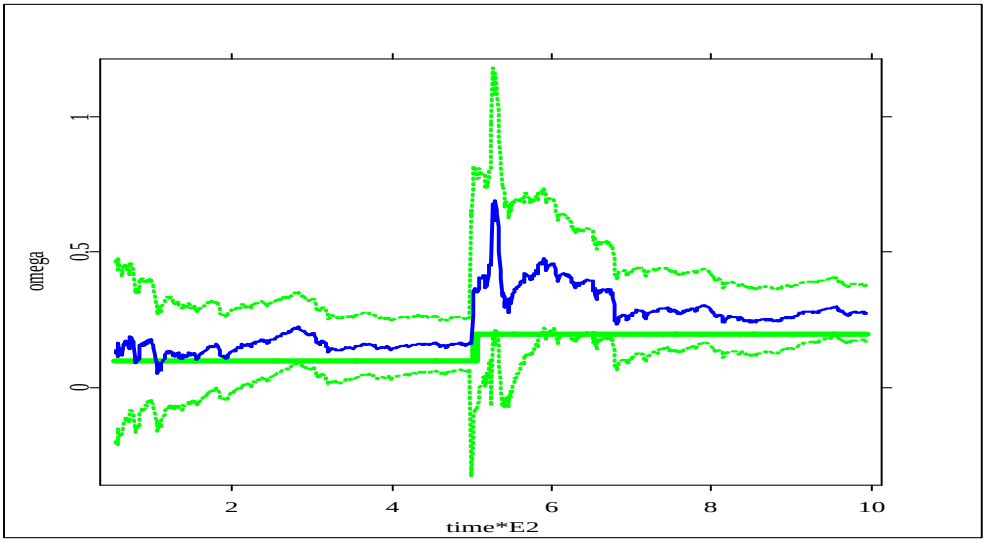


FIGURE 2. α_1

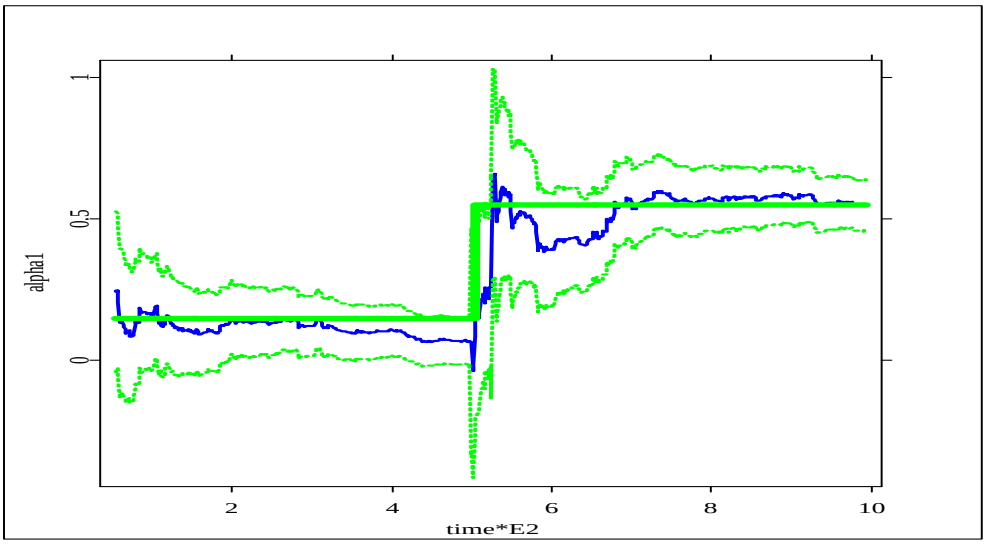
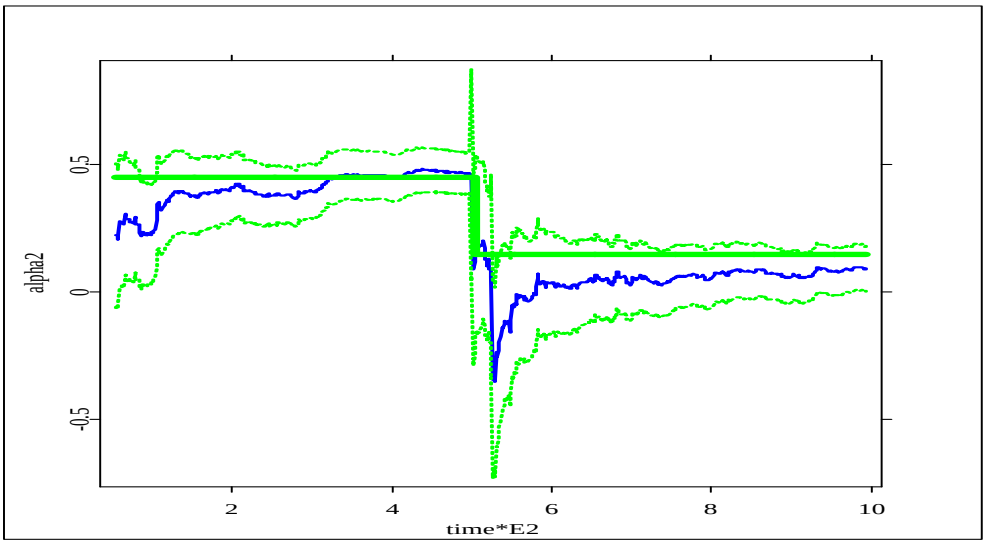


FIGURE 3. α_2



6.2. **Smooth transition model DGP**

We consider the following DGP:

$$Y_t = \omega + \omega^* G(t,c) + (\alpha_1 + \alpha_1^* G(t,c))\, Y_{t-1} + (\alpha_2 + \alpha_2^* G(t,c))\, Y_{t-2} + \varepsilon_t \quad \varepsilon_t \sim N(0,1) \tag{6.4}$$

where

$$G(t,c) = (1 + \exp\{-\gamma(t-c)\})^{-1}, \gamma = 0.025, c = 500, \tag{6.5}$$

$$\omega = 0.1 \quad \alpha_1 = 0.15 \quad \alpha_2 = 0.45 \tag{6.6}$$

$$\omega^* = 0.2 \quad \alpha_1^* = 0.20 \quad \alpha_2^* = -0.25 \tag{6.7}$$

In this transition model, the parameters are time varying between two extremes ω, and $\omega + \omega^*$, α_1 and $\alpha_1 + \alpha_1^*$, α_2 and $\alpha_2 + \alpha_2^*$. The degree of smoothness of the transition is controlled by the parameter γ, when γ tends to ∞, this model reduces to the threshold model presented above.

The three pictures below report the estimated parameters (the dark line), the theoretical curve (the thick line), and the confidence bands estimated from L_1 errors. It clearly apppears that the procedure estimates the parameters with a good accuracy and detects the smooth transition with some delay.

FIGURE 4. Constant term ω

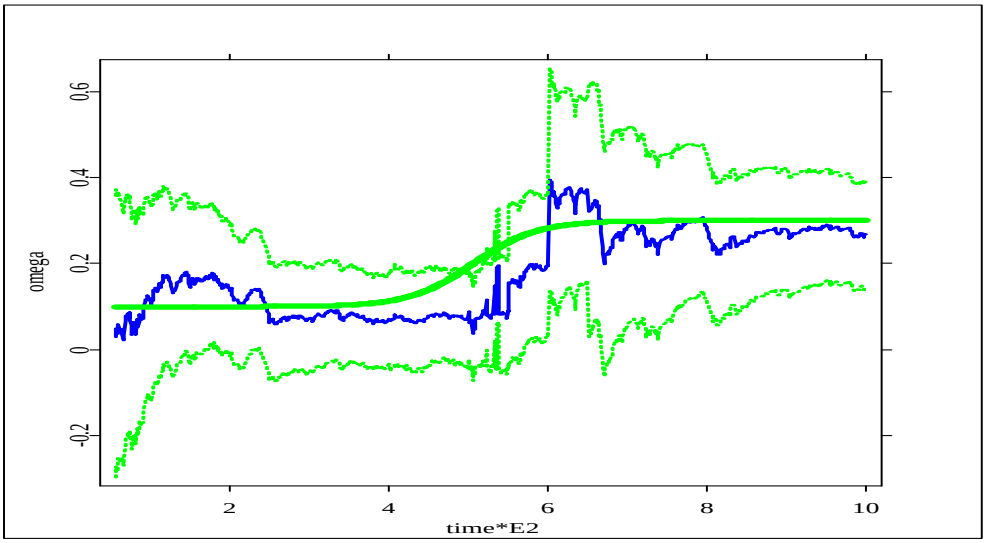


FIGURE 5. α_1

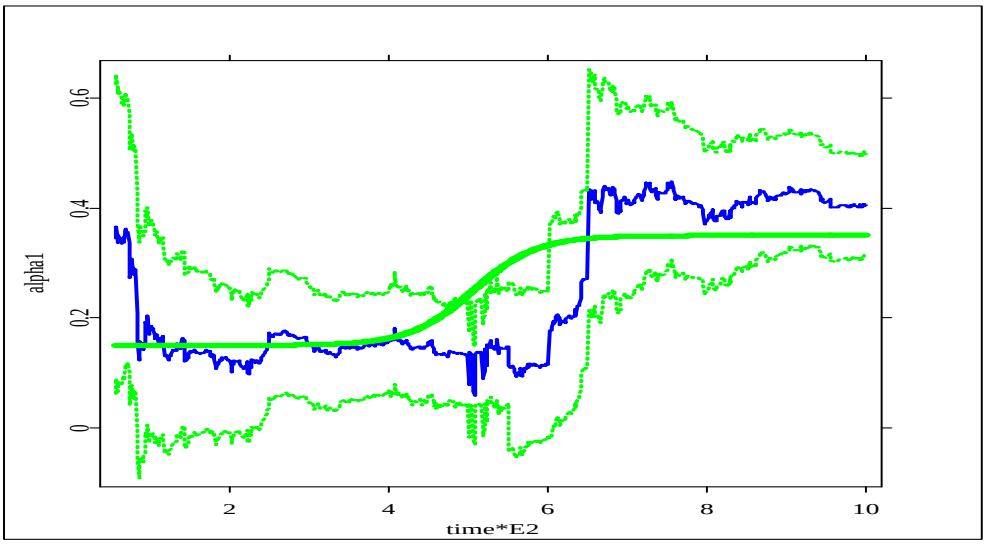
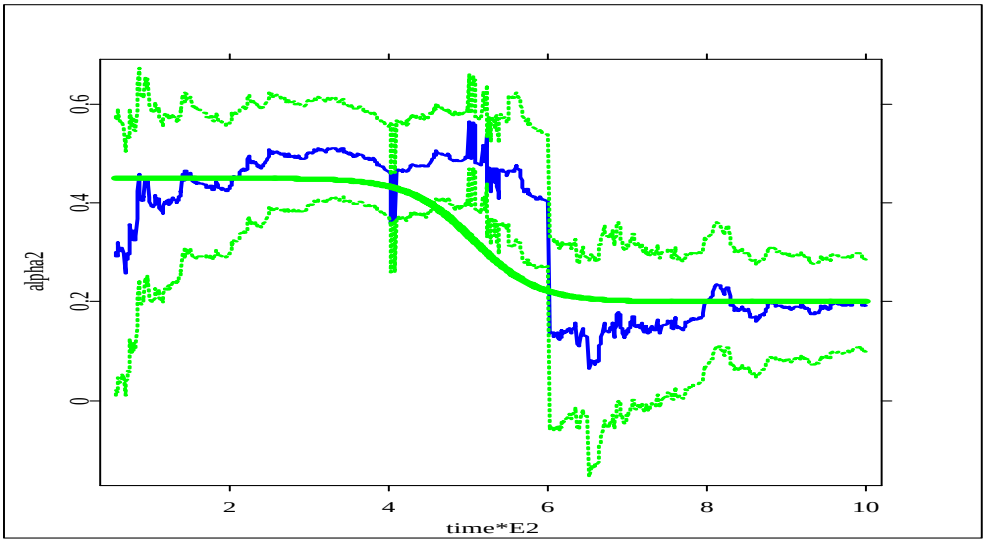


FIGURE 6. α_2



7. Applications

In this section we illustrate the performance of the proposed procedure for real data.

7.1. Practical implementation

Before starting with applications, let us point out some specific features of the finance data:

- real data are characterized by the large noise level which typically exceeds the level of the signal:
- real data contains outliers which come from missing data; in the case of a missing observation one usually utilizes the previous value of the considered process which means zero return and hence, the log-return is not defined or it is very large in the absolute value;
- lastly, the statistical theory of this paper has been developed for normal error terms. Given that the class of stochastic volatility models uses a log transformation, we are working with log-normal error terms.

The problem with outliers can be handled in the following way. Let (Y_t) be the set of log-returns and $\widehat{\sigma}^2$ is the variance estimate, see Section 4.3. Assuming continuity of the trend of the process Y_t, we classify Y_t as outlier if Y_t does not belong to the interval $[Y_{t-1} - 3\widehat{\sigma}, Y_{t-1} + 3\widehat{\sigma}]$, In such case, it is replaced by the forecast from the previous observations.

Although the innovations of the log-return process certainly are not normally distributed we proceed as if they were Gaussian, see Section 5.3.

7.2. The effects of European Monetary System inception on USD-DEM and USD-GBP exchange rates

We consider in this section the adaptive estimation of the stochastic volatility model on the series of DEM/USD and Pound-USD daily Foreign Exchange (FX) rates returns.

We consider the period after April 1979, which coincides with the inception of a new limited variation exchange rate system: the European Monetary System. The effects of the EMS on the volatility of European currencies have been studied by Bollerslev (1990) in a multivariate ARCH framework, and by Teyssière (1997) in a long-memory framework. Recently, Kokoszka and Leipus (1998) have devised a test for structural change in an ARCH framework.

All these studies have concluded that the EMS has affected the volatility structure of the European currencies. Our purpose is to analyze the change in the volatility through our model.

Since both currencies belongs to the same "target zone", it is expected that both volatilities share some common features. The EMS implies that each currency has a fluctuation band of ±2.25%. All EMS participants intervene in the FX market if these bounds are reached. This limited variation mechanism implies of course some interdependencies between the currencies.

We concentrate here on the two most frequently currencies, although Great Britain joined the EMS one month later and withdrew six months afterwards.

FIGURE 7. Constant term ω USD/DEM

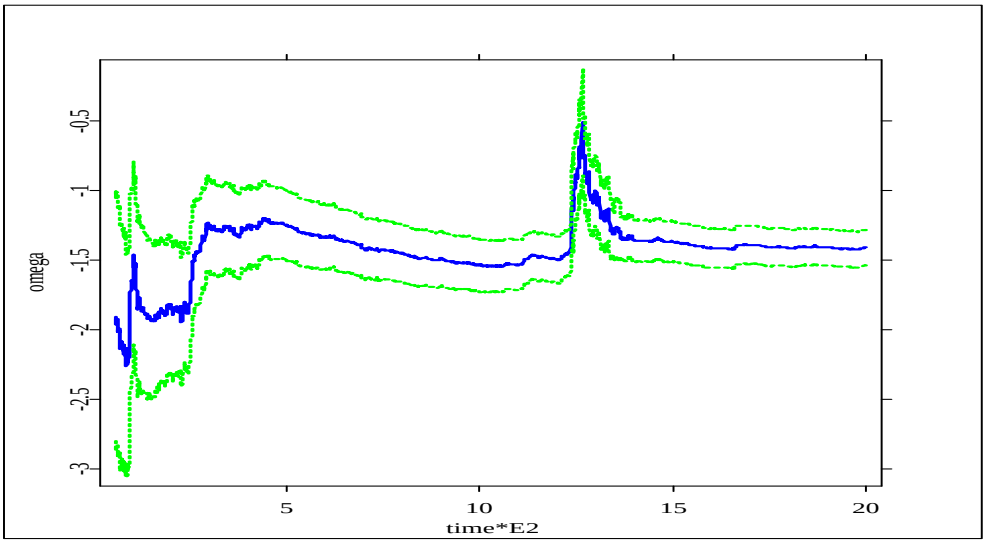


FIGURE 8. Constant term ω USD/GBP

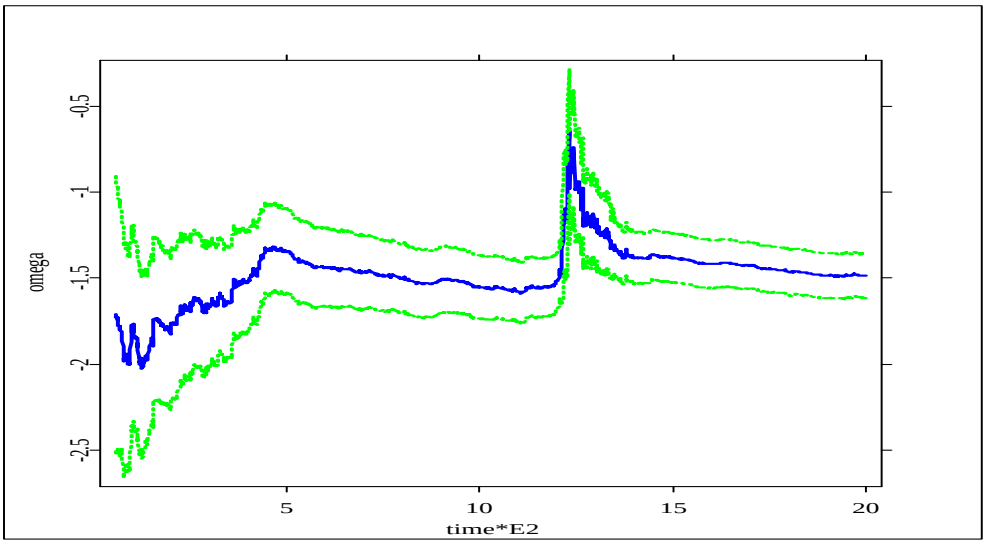


FIGURE 9. α_1 USD/DEM

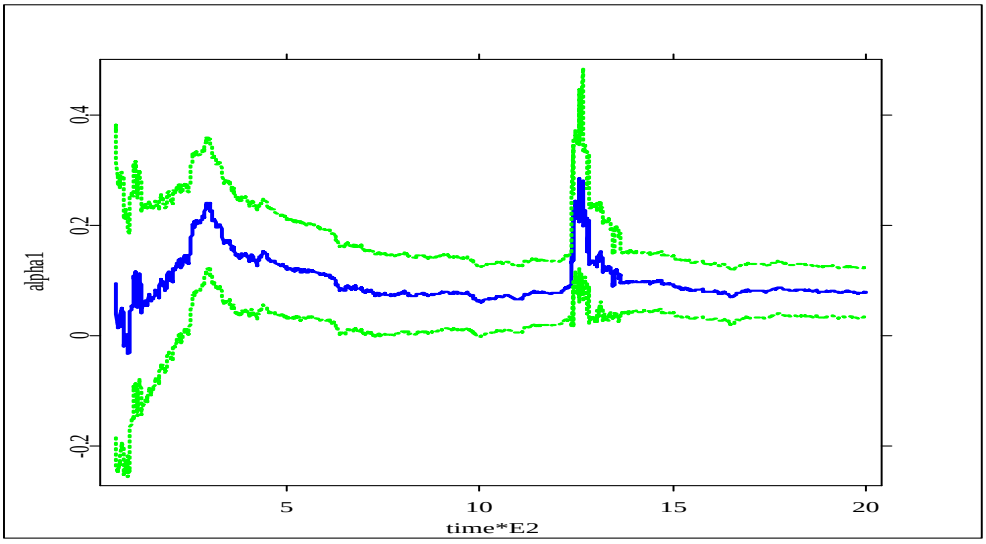


FIGURE 10. α_1 USD/GBP

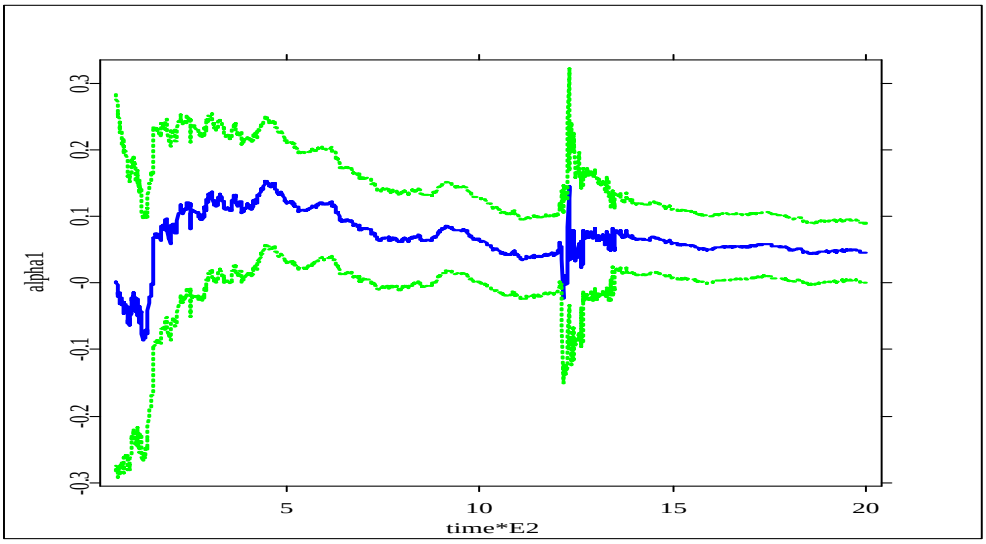


FIGURE 11. α_2 USD/DEM

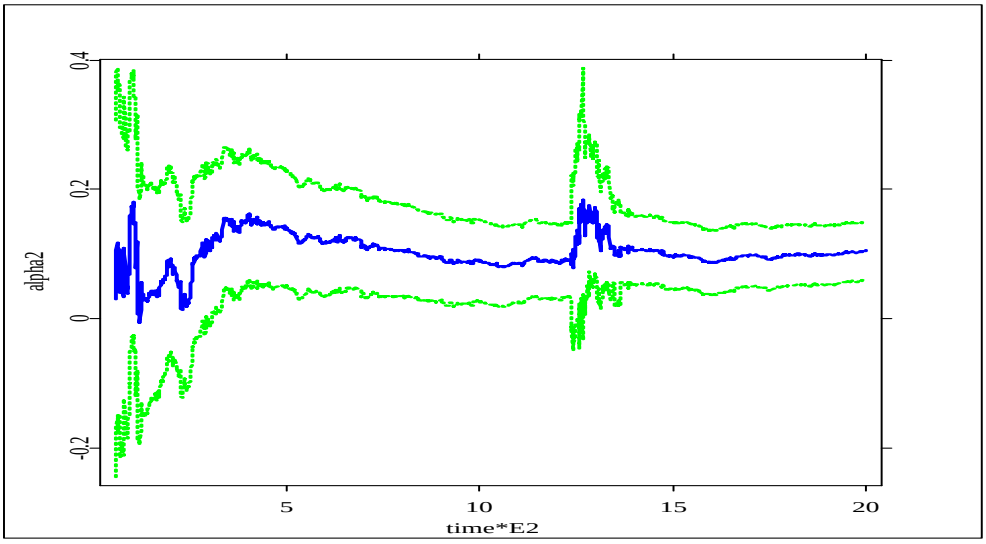
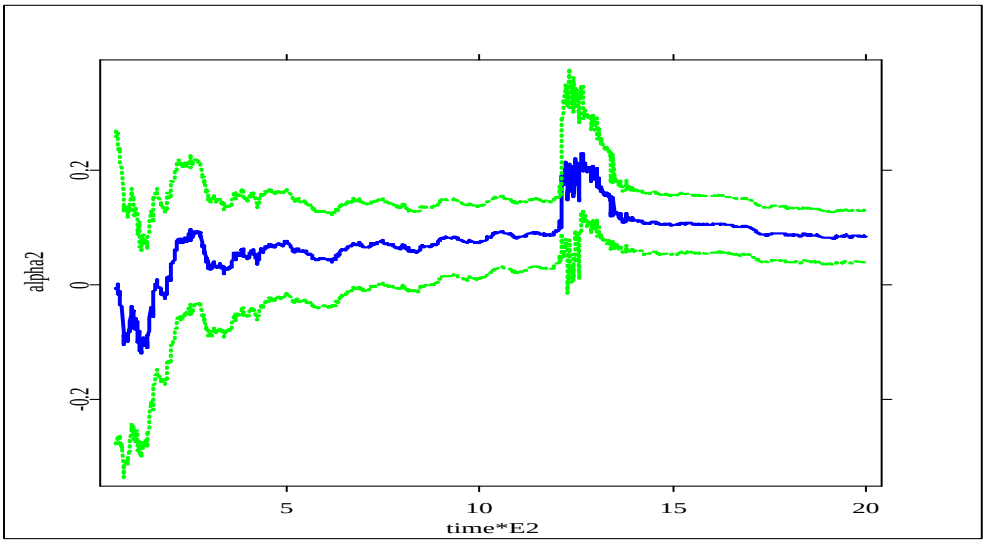


FIGURE 12. α_2 USD/GBP



These picture show that the model is able to detect changes in regime in the volatility equation. What is more interesting is the similarity in the changes of regime in the volatility of both FX rates returns.

Appendix: Proofs

Here we collect the proofs of the assertions from Sections 3.2 and 5.

7.3. **Proof of Theorem 3.1**

The model equation (3.1) and the definition of Δ_I imply

$$\begin{aligned} |\widetilde{\theta}_I - \theta(T)| &= \left| \left(\sum_{t \in I} X_t X_t^T \right)^{-1} \sum_{t \in I} \left(X_t Y_t - \theta(T) X_t X_t^T \right) \right| \\ &= \left| \left(\sum_{t \in I} X_t X_t^T \right)^{-1} \sum_{t \in I} \left(X_t \zeta_t + [\theta(t) - \theta(T)] X_t X_t^T \right) \right| \\ &\leq \left| \left(\sum_{t \in I} X_t X_t^T \right)^{-1} \sum_{t \in I} X_t \zeta_t \right| + \Delta_I. \end{aligned}$$

This bound constitutes the decomposition of the difference $\widetilde{\theta}_I - \theta(T)$ into a bias and a stochastic component. Denote by $\xi_I = (\xi_{0,I}, \ldots, \xi_{p,I})^T$ the stochastic component of $\widetilde{\theta}_I$, i.e.

$$\xi_I = \left(\sum_{t \in I} X_t X_t^T \right)^{-1} \sum_{t \in I} X_t \zeta_t$$

The assertion of Theorem 3.1 follows from a martingale inequality given in from Liptser and Spokoiny (1999), under the conditions of the theorem,

$$\begin{aligned} \mathbf{P}\left(|\xi_{i,I}| > \lambda \sigma_{i,I}, A_{i,I} \right) &\leq 4e \left(1 + \log B\right) \left(1 + 2\rho\sqrt{rp}\lambda\right)^p \lambda \exp\left(-\lambda^2/2\right) \\ &= P_{B,r,\rho,p}(\lambda) \exp\left(-\lambda^2/2\right). \end{aligned}$$

7.4. **Proof of Theorem 5.1**

Let $I\!I$ be the "ideal" interval from (5.1). Obviously

$$\begin{aligned} &\mathbf{P}\left(\left| \widehat{\theta}_i - \theta_i(T) \right| > (2\lambda + \mu + D)\sigma^*_{i,I\!I},\, A_i \right) \\ &\quad \leq \mathbf{P}\left(\left| \widehat{\theta}_i - \theta_i(T) \right| > (2\lambda + \mu + D)\sigma^*_{i,I\!I},\, A_i\,,\, I\!I \subseteq \widehat{I} \right) + \mathbf{P}\left(I\!I \text{ is rejected},\, A_i \right). \end{aligned}$$

We evaluate separately each summand in the right side of this inequality. It holds on the event $\{I\!I \subseteq \widehat{I}\}$ in view of the definition of $\widehat{I}$ and $\sigma^*_{i,I\!I}$

$$|\widehat{\theta}_i - \theta_i(T)| = |\widetilde{\theta}_{i,\widehat{I}} - \widetilde{\theta}_{i,I\!I}| \leq \lambda \sigma_{i,I\!I} + \mu \sigma^*_{i,\widehat{I}} \leq (\lambda + \mu)\sigma^*_{i,I\!I}.$$

On the "ideal" interval $I\!I$ we have $\Delta_{I\!I} \le D\sigma_{i,I\!I}$. An application of Theorem 3.1 then yields

$$\begin{aligned}\mathbf{P}\left(|\widetilde{\theta}_{i,I\!I} - \theta_i(T)| > (\lambda + D)\sigma_{i,I\!I}, A_i\right) &\le \mathbf{P}\left(|\widetilde{\theta}_{i,I\!I} - \theta_i(T)| > \lambda\sigma_{i,I\!I} + \Delta_{I\!I}\,,\, A_i\right) \\ &\le P_{B,r,\rho,p}(\lambda)\exp\left(-\frac{\lambda^2}{2}\right).\end{aligned}$$

Coupled with the previous inequality the latter implies

$$\mathbf{P}\left(|\widehat{\theta}_i - \theta_i(T)| > (2\lambda + \mu + D)\sigma^*_{i,I\!I}, A_i, I\!I \subseteq \widehat{I}\right) \le P_{B,r,\rho,p}(\lambda)\exp\left(-\frac{\lambda^2}{2}\right). \tag{7.1}$$

It remains to evaluate $\mathbf{P}(I\!I \text{ is rejected})$. Due to the definition

$$\{I\!I \text{ is rejected}\} = \bigcup_{I\in\mathcal{J}(I\!I)}\ \bigcup_{J\in\mathcal{J}(I)} \left\{|\widehat{\theta}_{i,I} - \widehat{\theta}_{i,J}| > \lambda\sigma_{i,J} + \mu\sigma^*_{i,I}\right\}. \tag{7.2}$$

Note also that for every $I \in \mathcal{J}(I\!I)$ and every $J \in \mathcal{J}(I)$, it holds

$$\begin{aligned}\Delta_I &\le \Delta_{I\!I} \le D\sigma_{i,I\!I} \le D\sigma^*_{i,I} \\ \Delta_J &\le \Delta_{I\!I} \le D\sigma_{i,I\!I} \le D\sigma^*_{i,I}.\end{aligned}$$

Hence, by Theorem 3.1, using $\mu \ge \lambda + 2D$ and $\sigma_{i,I} \le \sigma^*_{i,I}$

$$\begin{aligned}&\mathbf{P}\left(|\widetilde{\theta}_{i,I} - \widetilde{\theta}_{i,J}| > \lambda\sigma_{i,J} + \mu\sigma^*_{i,I}\,,\, A_i\right) \\ &\quad\le \mathbf{P}\left(|\widetilde{\theta}_{i,I} - \theta_i(T)| > \lambda\sigma_{i,I} + D\sigma^*_{i,I}\,,\, A_i\right) + \mathbf{P}\left(|\widetilde{\theta}_{i,J} - \theta_i(T)| > \lambda\sigma_{i,J} + D\sigma^*_{i,I}\,,\, A_i\right) \\ &\le \mathbf{P}\left(|\widetilde{\theta}_{i,I} - \theta_i(T)| > \lambda\sigma_{i,I} + \Delta_I\,, A_i\right) + \mathbf{P}\left(|\widetilde{\theta}_{i,J} - \theta_i(T)| > \lambda\sigma_{i,J} + \Delta_J\,, A_i\right) \\ &\le 2P_{B,r,\rho,p}(\lambda)\exp\left(-\frac{\lambda^2}{2}\right).\end{aligned}$$

In view of (7.2)

$$\begin{aligned}\mathbf{P}\,(I\!I \text{ is rejected}) &\le \sum_{I\in\mathcal{I}(I\!I)}\ \sum_{J\in\mathcal{J}(I)} \mathbf{P}\left(|\widehat{\theta}_{i,I} - \widehat{\theta}_{i,J}| > \lambda\sigma_{i,J} + \mu\sigma^*_{i,I}\right) \\ &\le \sum_{I\in\mathcal{I}(I\!I)} 2N_I P_{B,r,\rho,p}(\lambda)\exp\left(-\frac{\lambda^2}{2}\right).\end{aligned}$$

This along with (7.1) yields the assertion.

7.5. **Proof of Proposition 5.1**

Clearly the event $\{I \text{ is not rejected}\}$ includes

$$\begin{aligned}|\widetilde{\theta}_{i,I} - \widetilde{\theta}_{i,I\!I}| &\le \lambda\sigma_{i,I\!I} + \mu\sigma^*_{i,I} \le (\lambda + \mu)\sigma^*_{i,I\!I}, \\ |\widetilde{\theta}_{i,I} - \widetilde{\theta}_{i,J}| &\le \lambda\sigma_{i,J} + \mu\sigma^*_{i,I} \le \lambda\sigma_{i,J} + \mu\sigma^*_{i,I\!I}.\end{aligned}$$

Next, by Theorem 5.1, using homogeneity within $I\!I$ and J

$$\mathbf{P}\left(|\widetilde{\theta}_{i,I\!I} - \theta_i| \geq \lambda\sigma_{i,I\!I}, A_i\right) + \mathbf{P}\left(|\widetilde{\theta}_{i,J} - \theta_i'| \geq \lambda\sigma_{i,J}, A_i\right) \leq 2P_{B,r,\rho,p}(\lambda)\exp\left(-\lambda^2/2\right).$$

Since also $\sigma^*_{i,I} \leq \sigma^*_{i,I\!I}$, the inequality (5.3) implies (5.4).

7.6. **Proof of Corollary 5.1**

Let $I\!I = [t_{\rm cp}, T]$. The arguments from the proof of Theorem 5.1 lead to the bound

$$\mathbf{P}\left(I\!I \text{ is rejected}\right) \leq \sum_{I \in \mathcal{I}(I\!I)} 2N_I P_{B,r,\rho,p}(\lambda)\exp\left(-\lambda^2/2\right).$$

Next, set $I = [t_{\rm cp} - m, T]$ and $J = [t_{\rm cp} - m, t_{\rm cp}]$. The conditions of the corollary provide for all $i \leq p$

$$\sigma^2_{i,J} \leq Cm^{-1} \qquad \text{and} \qquad \sigma^2_{i,I\!I} \leq Cm^{-1}$$

so that condition (5.3) is fulfilled for m satisfying $m \geq Cb^{-2}(4\lambda + 2\mu)^2$. An application of Proposition 5.1 yields

$$\mathbf{P}\left(I \text{ is not rejected}\right) \leq 2P_{B,r,\rho,p}(\lambda)\exp\left(-\lambda^2/2\right)$$

and the assertion follows.

References

[1] Anderson, T.W. and Walker, A.M. (1964). On the asymptotic distribution of the autocorrelations of a sample from a linear stochastic process. *Ann. Math. Stat.* **35**, 1128–1135.

[2] Antoch, J., Hušková, M. and Z. Prášková (1997). "Effects of Dependence on Statistics for Determination of Change", *Journal of Statistical Planning and Inference*, **60** , 291–310.

[3] Antoch, J. and M. Hušková (1993). "Change Point Problem", in *Computational Aspects of Model Choice*, J. Antoch editor, Physica Verlag.

[4] Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* **31** 307–327.

[5] Bollerslev, T. (1990). "Modeling the Coherence in Short-run Nominal Exchange Rates: A Multivariate Generalized ARCH Model", *Review of Economics and Statistics*, **72** , 498–505.

[6] Bollerslev, T., Engle, R.F. and D.B. Nelson (1994). "ARCH Models", in *Handbook of Econometrics*, vol **4**, R.F. Engle and D.L. McFadden editors, 2959–3038.

[7] Bollerslev, T., Engle, R.F. and Woolridge, J.M. (1988). A capital asset pricing model with time-varying covariances. *J. of Political Economy* **96** 116–131.

[8] Campbell, J.A., Lo, A. and Mackinlay (1997). The Econometrics of Financial Markets. Prenseton University Press. Princeton, New Jessey.

[9] Csörgő, L. and L. Horváth (1997). "Limit Theorems in Change Point Analysis", Wiley, Chichester.

[10] Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Annals of Statistics*, **25** , no. 1, 1–37.

[11] Dahlhaus, R., Neumann, M. and von-Sachs, R. (1999). "Nonlinear wavelet estimation of time varying autoregressive processes". *Bernoulli* **5** no. 5, 873–907.

[12] Diebold, F.X. and P. Pauly (1987). "Structural Change and the Combination of Forecasts", *Journal of Forecasting*, **6**, 21–41.

[13] Dufour, J.M. and E. Ghysels, editors (1996): "Recent Developments in the Econometrics of Structural Change", Special Issue, *Journal of Econometrics*, **70**, no. 1.

[14] Engle, R.F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. inflation. *Econometrica* **50** 987–1008.

[15] Engle, R.F., Lilien, D.M. and Robins, R.P. (1987). Estimating the Varying Risk Premia in the Term Structure: the GARCH Model. *Econometrica* **55** 391–407.

[16] Fan, J. and Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* **85** no. 3 645–660.

[17] Gasser, T. and Sroka, L. and Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika* **73** 625–633.

[18] Gourieroux, C.(1997). *ARCH Models and Financial applications*. Springer. New York.

[19] Hackl, P. and A.H. Westlund, editors (1991). *Economic Structural Change*, Springer-Verlag.

[20] Hafner, C.M. and Herwatz, H. (1998). "Structural analysis of portfolio risk using beta impulse response functions". *Statistica Neerlandica* **52** no. 3, 336–355.

[21] Kokoszka, P. and R. Leipus (1998). "Change-point Estimation in ARCH Models" *Bernouilli*, forthcoming.

[22] Lepski, O. and Spokoiny, V. (1997) " Optimal Pointwise Adaptive Methods in Nonparametric Estimation." *Annals of Statistics*, **25**, no.6, 2512–2546.

[23] Liptser, R. and Spokoiny, V. (1999). "Deviation probability bound for martingales with applications to statistical estimation." Preprint **471**. Weierstrass-Institute, Berlin.

[24] Neumann, M. and von-Sachs, R. (1997). "Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra." *Annals of Statistics*, **25** no. 1, 38–76.

[25] Spokoiny, V. (1998). " Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice." *Annals of Statistics*, **26**, no. 4, 1356–1378.

[26] Taylor, S.J. (1986). *Modelling Financial Time Series*. Chichster. John Wiley.

[27] Teräsvirta, T. (1998). "Modeling Economic Relationship with Smooth Transition Regressions", in *Handbook of Applied Economic Statistics*, A. Ullah and D.E.A. Giles editors. Marcel Dekker.

[28] Teräsvirta, T. (1994). "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models", *Journal of the American Statistical Association*, **89**, no. 425, 208–218.

[29] Wold, H. (1954). *A Study in the Analysis of Stationary Time Series*. Almqvist and Wiksell Book Co., Uppsala.

Humboldt-Universität zu Berlin, Spandauer Strasse 1, D-10178 Berlin, Germany
E-mail address: haerdle@wiwi.hu-berlin.de

University at Würzburg,, Sanderring 2,, 97070, Würzburg, Germany
E-mail address: spokoiny@wias-berlin.de

Humboldt-Universität zu Berlin, SFB 373, Spandauer Strasse 1, D-10178 Berlin, Germany.
E-mail address: teyssier@wiwi.hu-berlin.de