Modelling Conditional Heteroscedasticity in Nonstationary Series

Cizek, P.

Publication date:
2010

CentER

Discussion Paper

No. 2010–84

MODELLING CONDITIONAL HETEROSCEDASTICITY IN NONSTATIONARY SERIES

By Pavel Čížek

August 2010

ISSN 0924-7815

TILBURG UNIVERSITY

1 Modelling conditional heteroscedasticity in nonstationary series

P. Čížek, Dept. of Econometrics & OR,
Tilburg University, The Netherlands
August 2010

To accommodate the inhomogenous character of financial time series over longer time periods, standard parametric models can be extended by allowing their coefficients to vary over time. Focusing on conditional heteroscedasticity models, we discuss various strategies to identify and estimate varying-coefficients models and compare all methods by means of a real-data application.

JEL codes: C14, C22, C53
Keywords: adaptive estimation, conditional heteroscedasticity, varying-coefficient models, time series

1.1 Introduction

A vast amount of econometrical and statistical research deals with modeling financial time series and their volatility, which measures the dispersion of a series at a point in time (i.e., conditional variance). Although financial markets have been experiencing many shorter and longer periods of instability or uncertainty in last decades such as Asian crisis (1997), start of the European currency (1999), the "dot-Com" technology-bubble crash (2000–2002) or the terrorist attacks (September, 2001), the war in Iraq (2003) and the current global recession (2008–2009), mostly used econometric models are based on the assumption of stationarity and time homogeneity; in other words, structure and parameters of a model are supposed to be constant over time. This includes linear and nonlinear autoregressive (AR) and moving-average models and conditional heteroscedasticity (CH) models such as ARCH (Engel, 1982) and GARCH (Bollerslev, 1986), stochastic volatility models (Taylor, 1986), as well as their combinations.

On the other hand, the market and institutional changes have long been assumed to cause structural breaks in financial time series, which was for example confirmed in data on stock prices (e.g., Andreou and Ghysels, 2002, Beltratti and Morana, 2004, and Eizaguirre et al., 2010) and exchange rates (e.g., Herwatz and Reimers, 2001, or Morales-Zumaqueroa and Sosvilla-Rivero, 2010). Moreover, ignoring these breaks can adversely affect the modeling, estimation, and forecasting of volatility as suggested, for example, by Diebold and Inoue (2001), Mikosch and Starica (2004), Pesaran and Timmermann (2004), and Hillebrand (2005). Such findings led to the development of the change-point analysis in the context of CH models; see for example Chen and Gupta (1997), Kokoszka and Leipus (2000), Andreou and Ghysels (2006), and Chen et al. (2010). Although these methods to detect structural changes in the time series are useful in uncovering major change points, their power is often rapidly decreasing with the number of change points in a given time series. Combined with the fact that the distance between the end of the sample and a change point has to increase with the sample size and with the assumption of time-homogeneity between any two breaks, they cannot relax the assumption of time-homogeneity of CH models to a larger extent. This is particularly visible when forecasting a time series as it is usually not possible to detect structural changes close to the end of the current time series (an exception being e.g. Andrews, 2003, in the linear AR models).

An alternative approach, which we concentrate upon in this chapter, lies in relaxing the assumption of time-homogeneity and allowing some or all model parameters to vary over time (e.g., as in Chen and Tsay, 1993, Cai et al., 2000, and Fan and Zhang, 2008). Without structural assumptions about the transition of model parameters over time, time-varying coefficient models have to be estimated nonparametrically under some additional identification conditions. A classical identification assumption in the context of varying coefficient models is that the parameters of interest are smooth functions of time (e.g., Cai et al., 2000, Xu and Phillips, 2008, and Fryzlewicz et al., 2008). Models with parameters smoothly varying over time are very flexible, but their main assumption precludes sudden changes in the parameter values. Thus, smoothly-varying coefficient models cannot account for classical structural breaks.

A different strategy, which allows for nonstationarity of a time series, is 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 the parametric model. In the context of the local constant approximation, this kind of strategy was employed for the volatility modeling by Härdle et al. (2003), Mercurio and Spokoiny (2004), Starica and Granger (2005), and Spokoiny (2009), for instance. A generalization to ARCH and GARCH models can be found in Čížek et al. (2009) and, in a slightly different context, in Polzehl and Spokoiny (2004). The main advantage of the approach using the local-approximation assumption to search for the longest time-homogeneous interval in a given series is that it unifies the change point analysis and smoothly-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 (2004), for instance, but it does not require prior information such as the number of changes and it does not need a large number of observations before each break point (because no asymptotic results are used for the selection of time-homogenous intervals). 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, but in addition to that, the parameter values can suddenly jump in contrast to models assuming smooth development of the parameters over time (Fan and Zhang, 2008).

To understand the benefits of various varying coefficient models, we will discuss here the conditional heteroscedasticity models (Section 1.2) and their time-varying alternatives: smoothly-varying CH models (Section 1.3), pointwise adaptive estimation of CH models (Section 1.4), and adaptive weights smoothing of CH models (Section 1.5). A real-world comparison will be facilitated by means the analysis of the S&P 500 stock index. In particular, daily data on the log-returns of the index are used from years 1997 to 2005. The reason for choosing this data set is that it is a difficult one for modelling using time-varying coefficients: someone forecasting the stock index by varying coefficient models has a hard time to outperform the standard GARCH model, whereas this is possibly an easy task with other types of data (e.g., for the exchange rate series as shown by Fryzlewicz et al., 2008).

1.2 Parametric conditional heteroscedasticity models

Consider a time series Y_t in discrete time, $t \in N$, which represents the log-returns of an observed asset-price process S_t: $Y_t = \log(S_t/S_{t-1})$. Modelling Y_t using the conditional heteroscedasticity assumption means that $Y_t = \sigma_t \varepsilon_t, t \in N$, where ε_t is a white noise process and σ_t is a predictable volatility (conditional variance) process. Identification and estimation of the volatility process σ_t typically relies on some parametric CH specification such as the ARCH (Engle,

1982) and GARCH (Bollerslev, 1986) models:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{1.1}$$

where $p \in N$, $q \in N$, and $\theta = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)^\top$ is the parameter vector; the ARCH and GARCH models correspond then to $q = 0$ and $q > 0$, respectively. An attractive feature of this model is that, even with very few coefficients, one can model most stylized facts of financial time series like volatility clustering or excessive kurtosis, for instance. A number of (G)ARCH extensions were proposed to make the model even more flexible; for example, EGARCH (Nelson, 1991), QGARCH (Sentana, 1995), and TGARCH (Glosten et al., 1993) that account for asymmetries in a volatility process.

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

$$Y_t = \sigma_t \varepsilon_t = \sqrt{g(X_t)} \varepsilon_t, \tag{1.2}$$

$$X_t = \omega + \sum_{i=1}^{p} \alpha_i h(Y_{t-i}) + \sum_{j=1}^{q} \beta_j X_{t-j}\,, \tag{1.3}$$

where g and h are known functions and X_t is a (partially) unobserved process (structural variable) that models the volatility coefficient σ_t^2 via transformation g: $\sigma_t^2 = g(X_t)$. The GARCH model (1.1) is described by $g(u) = u$ and $h(r) = r^2$, for instance. Despite its generality, the generalized linear volatility model is time homogeneous in the sense that the process Y_t follows the same structural equation at each time point. In other words, the parameter θ and hence the structural dependence in Y_t is constant over time. Even though models like (1.2)–(1.3) can often fit data well over a longer period of time, the assumption of homogeneity is too restrictive in practical applications: to guarantee a sufficient amount of data for reasonably precise estimation, these models are often applied over time spans of many years.

1.2.1 Quasi-maximum likelihood estimation

The parameters in model (1.2)–(1.3) are typically estimated by the quasi maximum likelihood (quasi-MLE) approach, which employs the estimating equations generated under the assumption of Gaussian errors ε_t. This guarantees efficiency under the normality of innovations and consistency under rather general moment conditions (Hansen and Lee, 1994, and Francq and Zakoian, 2007). Using the observations Y_t from some time interval $I = [t_0, t_1]$, the log-likelihood

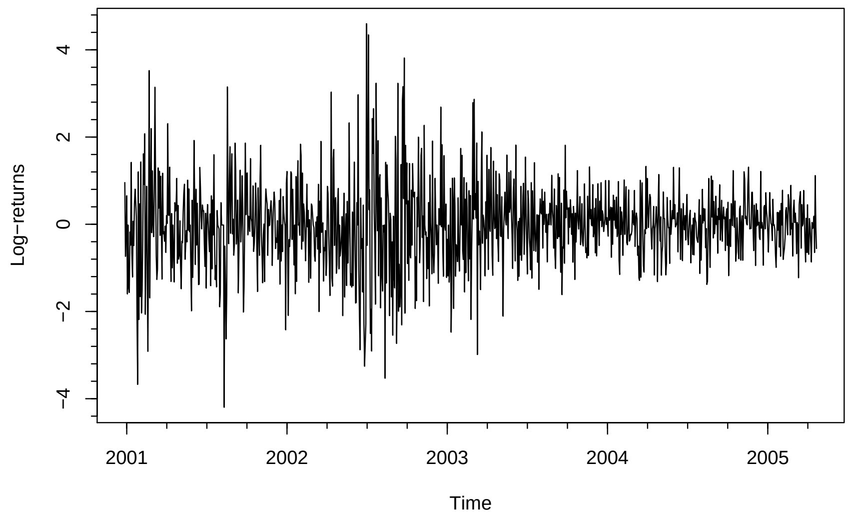

Figure 1.1: The standardized log-returns of the S&P500 index in years 2001–2004.

STF2tvch01.r

for the model (1.2)–(1.3) on an interval I can be represented in the form

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

with the log-likelihood function $\ell(y, v) = -0.5 \left\{\log(v) + y^2/v\right\}$ because the conditional distribution of Y_t has a zero mean and variance $\sigma^2 = g[X_t(\theta)]$. We define the quasi-MLE estimate $\tilde{\theta}_I$ of the underlying parameter value θ_0 by maximizing the log-likelihood $L_I(\theta)$:

$$\tilde{\theta}_I = \operatorname*{argmax}_{\theta \in \Theta} L_I(\theta) = \operatorname*{argmax}_{\theta \in \Theta} \sum_{t \in I} \ell\{Y_t, g[X_t(\theta)]\}. \tag{1.4}$$

Consider now the class of GARCH models, that is, $g(u) = u$ and $h(r) = r^2$. The commonly used models in this class are the ARCH(p) and GARCH(1,1) models. There are several reasons why, if the partial autocorrelation structure of Y_t^2 does not indicate an ARCH process, GARCH(1,1) with only one lag in both components is typically used. On the one hand, one needs several hundreds of observations to obtain significant and reasonably precise estimates of GARCH parameters (e.g., see Čížek et al., 2009). On the other hand, even GARCH(1,1) provides very good one-period ahead forecasts, which often outperform more complicated parametric models (Andersen and Bollerslev, 1998). This is especially true for the aggregated series such as the stock indices.

1.2.2 Estimation results

Let us thus use the GARCH(1,1) model to estimate and predict the volatility of the stock index S&P 500. Although the data span from 1997 to 2005, we mostly concentrate on predictions within years 2001–2004. 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); see Figure 1.1 for the log-returns of the S&P 500 index in years 2001–2004. In this and other examples, where predictions are made and evaluated, we suppose that t_1 represents a current day and our aim is to forecast the volatility tomorrow, that is, $\hat{\sigma}^2_{t_1+1}$ at time t_1+1 using the currently known data from times $\{t_0,\ldots,t_1\}$. Thus, $\hat{\sigma}^2_{t_1+1}$ always represents an out-of-sample forecast here. To evaluate this out-of-sample forecasting performance over a given period $\{t_s,\ldots,t_e\}$, the mean absolute prediction error is used:

$$MAPE(t_s,t_e) = \frac{1}{t_e - t_s + 1}\sum_{t=t_s}^{t_e} |Y^2_{t+1} - \hat{\sigma}^2_{t+1}|, \tag{1.5}$$

where the squared future returns Y^2_{t+1} are used as a noisy, but unbiased approximation of the underlying volatility (Andersen and Bollerslev, 1998). Note that we will report MAPE within each year $2001,\ldots,2004$ in tables and running monthly averages of MAPE in graphs.

Predicting the volatility one day ahead using all available data points $\{1,\ldots,t_1\}$ for the GARCH(1,1) estimation results in the forecast on Figure 1.2. (Note that the data were for convenience rescaled so that the unconditional variance equals 1.) In the light of the (ex post) knowledge of possible structural changes in the series, one can however wonder whether using all available historical data is a good estimation strategy. To this end, we run the GARCH estimation using various historical windows, that is, the data $\{t_1-W,\ldots,t_1\}$ for $W = 125, 250, 500$, and 1000 representing periods of one half to four years. The MAPEs of all estimates are summarized for every year in Table 1.1. Obviously, the estimation using all available historical data is best in years 2001 and 2002, whereas the optimal forecasts in years 2003 and 2004 are achieved using data from the past two years and the past six months, respectively.

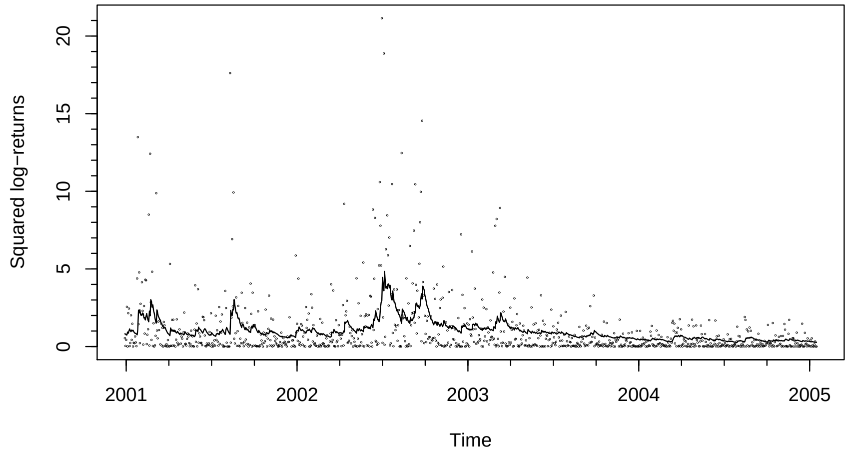

Figure 1.2: The volatility forecasts of GARCH(1,1) for S&P 500 using all historical data.

STF2tvch02.r

Thus, the optimal historical interval to use differs accross time, most likely due to structural changes in the observed series. Unfortunately, there is no clear rule how to select the optimal historical window with well defined statistical properties, at least not in this basic setup. One can only select the overall best method based on the historical performance. To judge it, Table 1.1 contains also MAPE across all four years 2001–2004 (see row 'Total'). Since the total mean reflects more the most volatile years than the other ones, we also present the weighted mean with weights indirectly proportional to the unconditional variance of raw data within each year (see row 'Weighted'). In the case of GARCH(1,1), the best total performance can be attributed to the forecasts based on the past two years of data.

Alternatively, various attempts to account for the nonstationary nature of financial time series are discussed in the following Sections 1.3–1.5. To facilitate comparison across methods, we will mostly use the GARCH(1,1) results using all available data as a benchmark since it provides the overall best forecasting performance in years 2001–2003.

Table 1.1: Mean absolute forecast errors in volatility by GARCH(1,1) using last 125, 250, 500, 1000, and all observations.

	Estimation window				
Year	$W = 125$	$W = 250$	$W = 500$	$W = 1000$	All
2001	1.207	1.215	1.188	1.167	1.156
2002	1.807	1.773	1.739	1.728	1.714
2003	0.808	0.815	0.804	0.823	0.818
2004	0.355	0.360	0.367	0.397	0.435
Total	1.044	1.041	1.025	1.029	1.031
Weighted	1.046	1.050	1.041	1.064	1.087

`STF2tvch03.r`

1.3 Time-varying coefficient models

An obvious feature of the generalized linear volatility model (1.2)–(1.3) is that the parametric structure of the process is assumed constant over the whole sample and cannot thus incorporate changes and structural breaks at unknown times in the model. A natural generalization leads to models whose coefficients may change over time (Fan and Zhang, 2008). In this context, a standard assumption is that the structural process X_t satisfies the relation (1.3) at any time, but the vector of coefficients θ may vary with the time t, $\theta = \theta(t)$. The estimation of the coefficients as general functions of time is possible only under some additional assumptions on these functions. Typical assumptions are (i) time-varying coefficients are smooth functions of time (Cai et al., 2000, and Fryzlewicz et al., 2008) and (ii) time-varying coefficients are piecewise constant functions (Bai and Perron, 1998, and Mikosch and Starica, 2004). Due to the limitations of the latter approach such as the asymptotically increasing length of the intervals, where a parametric model (1.2)–(1.3) holds, we concentrate on the smoothly-varying coefficient models here.

Following Cai et al. (2000), for instance, one can define the following time-varying equivalent of the model (1.2)–(1.3) by

$$Y_t = \sigma_t \varepsilon_t = \sqrt{g(X_t)}\varepsilon_t, \tag{1.6}$$

$$X_t = \omega(t) + \sum_{i=1}^{p} \alpha_i(t) h(Y_{t-i}) + \sum_{j=1}^{q} \beta_j(t) X_{t-j}\,, \tag{1.7}$$

where $\omega(t)$, $\alpha_i(t)$, and $\beta_j(t)$ are smooth functions of time and have to be estimated from the observations Y_t. This very general model can be estimated for example by means of the kernel quasi-MLE method, see (1.4):

$$\hat{\theta}_I(t_1) = \underset{\theta \in \Theta}{\operatorname{argmax}} \sum_{t \in I} W\left(\frac{t - t_1}{b|I|}\right) \ell\{Y_t, g[X_t(\theta)]\}, \tag{1.8}$$

where $\theta(t) = (\omega(t), \alpha_1(t), \ldots, \alpha_p(t), \beta_1(t), \ldots, \beta_q(t))$, b denotes the bandwidth parameter, $|I|$ is the length of the interval I, and $W : [-1/2, 1/2] \to R$ is a symmetric weighting function, which integrates to 1 (e.g., $W(z) = 1$ on $[-1/2, 1/2]$ in the simplest case). The general model (1.6)–(1.7) is however not often used in practice due to data demands. Estimating parameters as functions of time means that $\theta(t)$ has to be estimated locally using possibly rather small numbers of observations from time periods close to t. This is however very difficult to achieve even with the basic GARCH(1,1) (see also Section 1.4 for more details). Hence, the main interest concerning the smoothly time-varying CH models lies in the time-varying ARCH (tvARCH) models.

1.3.1 Time-varying ARCH models

Formally, the tvARCH(p) model is defined by the structural equation (Dahlhaus and Subba Rao, 2006)

$$X_t = \omega(t/|I|) + \sum_{i=1}^{p} \alpha_i(t/|I|) Y_{t-i}^2, \tag{1.9}$$

so that the parameter vector $\theta(t) = (\omega(t), \alpha_1(t), \ldots, \alpha_p(t))$ consists of real-valued functions on $[0, 1]$ and all observations are assigned time $t/|I|$ within $[0, 1]$ irrespective of the sample size $|I|$. This model is able to characterize the data with slowly decaying sample autocorrelations of the squared returns Y_t^2, which are normally atributed to structural breaks or long memory in the series.

To estimate the parameters of tvARCH(p), the kernel quasi-MLE method defined in (1.8) can be used, but it has a number of disadvantages in the context of the small-sample or local estimation. When the sample size is small, the likelihood function tends to be flat around its minimum, which leads to a large variance of estimates. Moreover, the kernel estimation requires solving a large number of these quasi-MLE problems, which is a computationally intensive task (especially taking into account also the selection of the bandwidth b discussed later). An alternative available in the class of ARCH models is the least squares (LS) estimation studied in the context of tvARCH by Fryzlewicz et al. (2008) because of its good small sample properties, closed form solution, and fast computation. The kernel-LS estimator for the tvARCH(p) process minimizes at a given time t_1 the following expression:

$$\hat{\theta}_I(t_1) = \underset{\theta\in\Theta}{\operatorname{argmax}} \sum_{\{t,t-p\}\subset I} W\left(\frac{t-t_1}{b|I|}\right) \frac{(Y_t^2 - \omega - \sum_{j=1}^{p}\alpha_j Y_{t-j}^2)^2}{\kappa(t_1, Y_{t-1}, \ldots, Y_{t-p})^2}, \quad (1.10)$$

where W is a symmetric weighting function as in (1.8), b represents the bandwidth, and κ is also a positive weighting function. Although one does not need weighting by κ in principle, its use is recommended for heavy-tailed data since using $\kappa(t_1, Y_{t-1}, \ldots, Y_{t-p})$ proportional to $Y_{t-1}^2 + \ldots + Y_{t-p}^2$ reduces the number of moments required for the consistency and asymptotic normality of the kernel-LS estimator.

Since the (1.10) has a closed form solution, an operational procedure – the two-step kernel-LS estimator – can be defined for a given interval I and a time point $t_1 \in I$ as follows (Fryzlewicz et al., 2008). Denoting $\mathcal{Y}_{t-1} = (1, Y_{t-1}^2, \ldots, Y_{t-p}^2)^\top$,

1. estimate variance of Y_t^2 at t_1, $\hat{\mu}_I(t_1) = \sum_{t\in I} W\left(\frac{t-t_1}{b|I|}\right) Y_t^2/b|I|$;
2. compute $\hat{\kappa}(t_1, \mathcal{Y}_{t-1}) = \hat{\mu}_I(t_1) + Y_{t-1}^2 + \ldots + Y_{t-p}^2$,

$$\hat{R}_I(t_1) = \sum_{\{t,t-p\}\subset I} W\left(\frac{t-t_1}{b|I|}\right) \frac{\mathcal{Y}_{t-1}\mathcal{Y}_{t-1}^\top}{\hat{\kappa}(t_1, \mathcal{Y}_{t-1})^2},$$

$$\hat{r}_I(t_1) = \sum_{\{t,t-p\}\subset I} W\left(\frac{t-t_1}{b|I|}\right) \frac{\mathcal{Y}_{t-1}Y_t^2}{\hat{\kappa}(t_1, \mathcal{Y}_{t-1})^2};$$

3. and set kernel-LS estimator to $\hat{\theta}_I(t_1) = \hat{R}_I^{-1}(t_1)\hat{r}_I(t_1)$.

The only missing component needed for estimation is the bandwidth b. This can be determined either by the standard leave-one-out cross-validation (Fryzlewicz et al., 2008) or by some global forecasting criterion (Cheng et al., 2003) if the aim is to use tvARCH(p) for predicting volatility as is the case here. Specifically, suppose the current time is t_1 and we want to forecast the volatility $\hat{\sigma}_{t_1+1}^2$ at time $t_1 + 1$. For a given bandwidth b and a historical interval I, one can simply estimate $\hat{\theta}_I(t_1)$ to identify the ARCH(p) model valid at time t_1 and then predict $\hat{\sigma}_{t_1+1}^2(b)$ as in the case of the standard ARCH(p) model with parameters equal to $\hat{\theta}_I(t_1)$. Because the bandwidth is unknown, we can choose it by evaluating the recent out-of-sample forecasts $\hat{\sigma}_{t+1}^2$ at times $t \in J = [\tau, t_1]$ and minimizing the prediction error:

$$\hat{b} = \underset{b>0}{\operatorname{argmin}}\, PE_{\lambda,\mathcal{H}}(b) = \underset{b>0}{\operatorname{argmin}} \sum_{t\in J}\sum_{h\in\mathcal{H}} |Y_{t+h}^2 - \hat{\sigma}_{t+h}^2(b)|^\lambda, \quad (1.11)$$

where $\lambda > 0$ determines the form of the loss function and $\mathcal{H}$ is the forecasting horizon set. In the case of MAPE in (1.5), $\lambda = 1$ and the forecasting horizon is one day day, $\mathcal{H} = \{1\}$. The historical period J can contain last three or six months of data, for instance.

1.3.2 Estimation results

Let us now have a look at the estimation and prediction of S&P 500 using the tvARCH(p) models. First, consider $p = 1$, for which the estimation results are summarized in Figures 1.3 and 1.4. By looking at the forecasted volatility and prediction errors of tvARCH(1) relative to GARCH(1,1) (Figure 1.4), we see that it performs similarly or slightly worse in year 2001 and 2002, where the GARCH using all data was the best method, and outperforms GARCH(1,1) in years 2003 and 2004, where a very short window was optimal for GARCH(1,1) (see Section 1.2). Before comparing numerically the performance of the estimators, the parameter estimates and the used bandwidth are of interest (Figure 1.3). One can see that the bandwidth $b|I|$ ranges from very small values such as two weeks (10 days) in periods after possible structural changes or possible outliers (cf. Figure 1.1) to almost half a year (130 days). The bandwidth choice exhibits a repetitive pattern: after every possible change point or outlier in the stock-index returns, the selected bandwidth suddenly drops and then gradually increases until the next nonstationarity is encountered. The parameter estimates change more or less smoothly as functions of time except for the periods just after structural changes: due to a very low bandwidth and a low precision of estimation, one can then observe large fluctuations in the parameter estimates. Nevertheless, the ARCH coefficient has small values below 0.2 most of the time and there are prolonged periods with the ARCH coefficient being zero.

The estimation is now performed for the tvARCH(p) models with $p \in \{0, 1, 3, 5\}$ and the corresponding MAPEs for years 2001–2004 are summarized in Table 1.2. Naturally, the results vary with the complexity of the ARCH model, but the differences are generally rather small. On the one hand, one could thus model volatility locally as a constant without losing much of a prediction power since tvARCH(0) performs overall as good as GARCH(1,1) in terms of MAPE; considering the weighted total MAPE, tvARCH(0) actually outperforms GARCH(1,1) using any estimation window (cf. Table 1.1). On the other hand, more complex models such as tvARCH(3) and tvARCH(5) seem to perform even better because (i) they do not perform worse than simpler models despite more parameters that have to be estimated locally (e.g., in year 2004, the bandwidth is rather small for all methods and yet tvARCH(5) matches tvARCH(1)) and (ii) they are by definition approximations of GARCH(1,1)

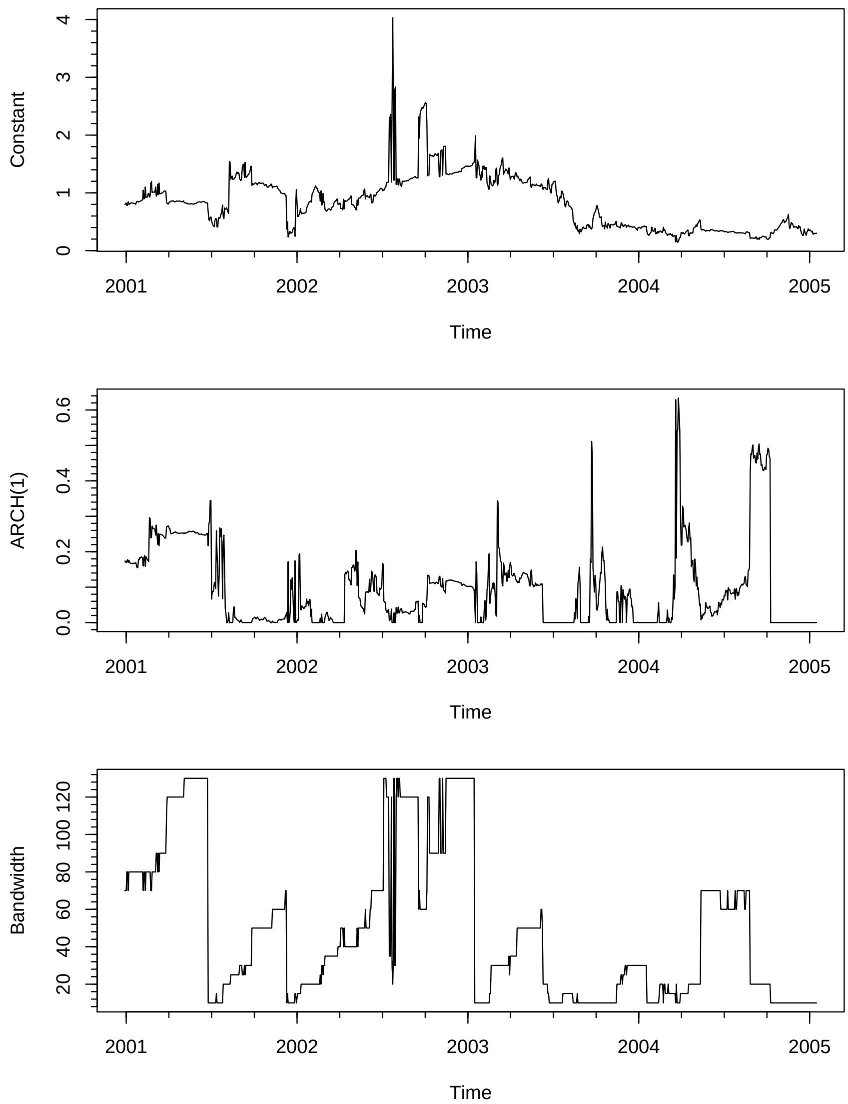

Figure 1.3: The estimated parameters (top panels) and the used bandwidth (bottom panel) of tvARCH(1) as functions of time for S&P 500.
STF2tvch04.r

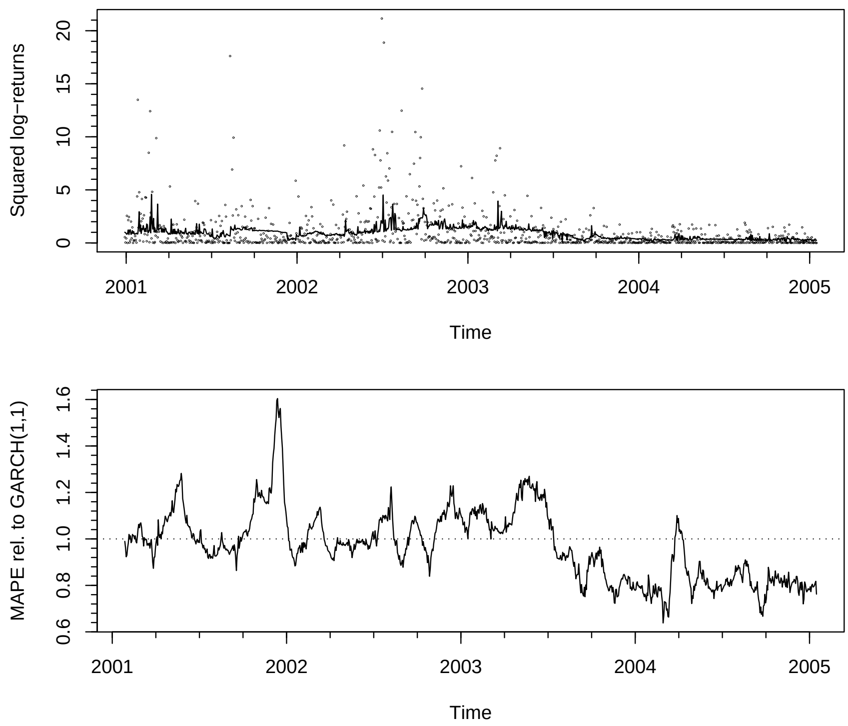

Figure 1.4: The volatility forecast (top panel) by tvARCH(1) and the mean absolute prediction errors (bottom panel) of tvARCH(1) relative to GARCH(1,1) for S&P 500.

STF2tvch04.r

and can thus match its performance in the periods where GARCH(1,1) is optimal. Finally, note that the total performance of tvARCH(3) and tvARCH(5) across the whole period is better than the best GARCH(1,1) model using two years of data (i.e., last 500 observations) by all criteria. The relatively small differences between tvARCH and GARCH can be explained by many pronounced structural changes in the series (see Figure 1.3): while structural changes obviously affect adversely the globally-specified GARCH(1,1) model, they indirectly worsen the performance of the time-varying models as well since the intervals used for estimation shortly after a structural change are short and the estima-

Table 1.2: Mean one-day-ahead forecast errors in volatility by GARCH(1,1) and tvARCH(p) for $p = 0, 1, 3$, and 5.

Year	tvARCH(p) $p = 0$	$p = 1$	$p = 3$	$p = 5$	Global GARCH(1,1)
2001	1.192	1.195	1.185	1.189	1.156
2002	1.791	1.748	1.698	1.665	1.714
2003	0.778	0.843	0.806	0.786	0.818
2004	0.358	0.353	0.358	0.355	0.435
Total	1.030	1.034	1.012	0.998	1.031
Weighted	1.033	1.046	1.029	1.016	1.087

STF2tvch05.r

tion is thus less precise.

1.4 Pointwise adaptive estimation

The main limitation the smoothly-varying CH models discussed in the previous Section 1.3 is the assumption of the continuity of $\theta(t)$, which precludes structural changes. Although we have seen in Section 1.3.2 that tvARCH(p) can actually deal with nonstationary data rather well, this was sometimes achieved by estimating within a window of length $b|I| = 10$. This is formally possible only if the sample size is fixed as the asymptotic theory requires $b|I| \to \infty$ as the interval length increases. These limitations motivate an alternative approach to the estimation of varying coefficient models (1.6)–(1.7), which is based on a finite-sample theory and thus does not suffer from the limitations of the classical change point detection or smoothly-varying coefficient models.

This alternative strategy is based on the assumption that a time series can be locally, that is, over short periods of time, approximated by a parametric CH model such as (1.2)–(1.3). The aim is to find – by means of a finite-sample theory of testing – the longest historical time interval, where such a parameteric approximation is valid. This methodology was proposed by Mercurio and Spokoiny (2004) in the context of the local constant approximation of the volatility process and generalized by Čížek et al. (2009) to the local ARCH and GARCH approximations. Formally, we assume that the observed data Y_t are described by an unobserved process X_t as in (1.2), and at each point

t_1, there exists a historical interval $I(t_1) = [t_0, t_1]$ in which the process X_t "nearly" follows the parametric specification (1.3). This assumption enables us to apply locally the well developed parametric estimation for data $\{Y_t\}_{t \in I(t_1)}$ to estimate the underlying parameter vector $\theta(t_1)$ by $\hat{\theta}(t_1) = \tilde{\theta}_{I(t_1)}$. Obviously, this modelling strategy results in a model with time-varying coefficients $\theta(t_1)$, but contrary to smoothly-varying coefficient models in Section 1.3, it allows for discontinuous jumps in the parameter values of model (1.6)–(1.7): the lengths of intervals $I(t_1)$ and $I(t_1 - 1)$ are namely not assumed to be related in any way.

To estimate $\hat{\theta}(t_1)$, we have to find the historical interval of homogeneity $I(t_1)$, that is, the longest interval $I = [t_0, t_1]$, where data do not contradict the specified parametric model with some fixed parameter values. Starting at each time t_1 with a very short interval $I = [t_1 - m + 1, t_1]$, where m is a small fixed integer independent of the sample size, 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 procedure is repeated until a change point is found or I contains all past observations.

To test the null hypothesis that the observations $\{Y_t\}_{t \in I}$ follow the parametric model (1.2)–(1.3) with a fixed parameter θ_0, one can use, for example, the supremum likelihood-ratio (supLR) test as proposed in Andrews (1993). Since the alternative is that the process Y_t follows different parameteric models within I, the supLR test statistics equals

$$
\begin{aligned}
T_{I,\mathcal{T}(I)} &= \sup_{\tau \in \mathcal{T}(I)} 2[\max_{\theta_J, \theta_{J^c} \in \Theta} \{L_J(\theta_J) + L_{J^c}(\theta_{J^c})\} - \max_{\theta \in \Theta} L_I(\theta)] \quad (1.12) \\
&= \sup_{\tau \in \mathcal{T}(I)} 2[L_J(\tilde{\theta}_J) + L_{J^c}(\tilde{\theta}_{J^c}) - L_I(\tilde{\theta}_I)], \quad (1.13)
\end{aligned}
$$

where intervals $J = [t_0, \tau]$ and $J^c = [\tau + 1, t_1]$ and the supremum is taken over a set $\mathcal{T}(I) = \{\tau : t_0 + m' \le \tau \le t_1 - m''\}$ for some fixed integers $m', m'' > 0$. Although this procedure has well established asymptotic properties for one interval I and $\min\{m', m''\} \to \infty$ as $|I| \to \infty$ (Andrews, 1993), the challenge of the search for the longest interval of homogeneity lies in performing sequentially multiple supLR tests and using some small fixed m' and m'', which do not increase with the interval length.

This sequential search for the longest time-homogeneous region and the appropriate choice of critical values for the test statistics $T_{I,\mathcal{T}(I)}$ are discussed in Sections 1.4.1 and 1.4.2, respectively.

1.4.1 Search for the longest interval of homogeneity

The main steps in the pointwise adaptive estimation procedure are now described. At each point t_1, the aim is to estimate the unknown parameter vector $\theta(t_1)$ from historical data Y_t, $t \leq t_1$. First, the procedure selects a historical interval $\hat{I}(t_1) = [t_0, t_1]$ in which the data do not contradict the parametric model (1.2)–(1.3). Afterwards, the quasi-MLE estimation is applied to data within the selected historical interval $\hat{I}(t_1)$ to obtain the estimate $\hat{\theta}(t_1) = \tilde{\theta}_{\hat{I}(t_1)}$.

To perform the search, suppose that a growing set $I_0 \subset I_1 \subset \ldots \subset I_K$ of historical interval candidates $I_k = [t_1 - m_k + 1, t_1]$ with the right-end point t_1 is fixed, where the smallest interval $I_0 = [t_1 - m_0 + 1, t_1]$ is so short that it can be accepted automatically as time-homogeneous. Every larger interval I_k will be successively tested for time-homogeneity using the test statistic $T_{I_k,\mathcal{T}(I_k)}$ defined in (1.12) and critical values $\mathfrak{z}_k$ constructed for each interval in Section 1.4.2. The interval $\hat{I}(t_1)$ will then simply be chosen as the longest interval, where the null hypothesis of time-homogeneity is accepted.

To perform and describe the procedure, it is thus necessary to select (i) a specific parametric model (1.2)–(1.3) (e.g., constant volatility, ARCH(p), GARCH(1,1)); (ii) the set $\mathcal{I} = (I_0, \ldots, I_K)$ of interval candidates (e.g., $I_k = [t_1 - m_k + 1, t_1]$ using a geometric grid $m_k = [m_0 a^k], a > 1$, and $m_0 \in N$); and (iii) the critical values $\mathfrak{z}_1, \ldots, \mathfrak{z}_K$ as described later in Section 1.4.2. The complete description of the procedure as introduced in Čížek et al. (2009) follows.

1. Set $k = 1$, $\hat{I} = I_0$, and $\hat{\theta} = \tilde{\theta}_{I_0}$.
2. Test the hypothesis $H_{0,k}$ of no change point within the interval I_k using test statistics $T_{I_k,\mathcal{T}(I_k)}$ defined in (1.12) and the critical values $\mathfrak{z}_k$. If a change point is detected, that is, $T_{I_k,\mathcal{T}(I_k)} > \mathfrak{z}_k$, go to point 4. Otherwise continue with point 3.
3. Set $\hat{\theta} = \tilde{\theta}_{I_k}$ and $\hat{\theta}_{I_k} = \tilde{\theta}_{I_k}$. Further, set $k := k+1$. If $k \leq K$, repeat point 2. Otherwise go to point 4.
4. Define $\hat{I} = I_{k-1}$ = "the last accepted interval" and $\hat{\theta} = \tilde{\theta}_{\hat{I}}$ = "the final estimate." Additionally, set $\hat{\theta}_{I_k} = \ldots = \hat{\theta}_{I_K} = \hat{\theta}$ if $k \leq K$.

The result of the procedure is the longest interval of homogeneity $\hat{I}$ and the corresponding pointwise adaptive estimate $\hat{\theta}$. Additionally, the estimate $\hat{\theta}_{I_k}$ after k steps of the procedure is defined. Even though this method is rather insensitive to the choice of its parameters, the set-up of the procedure cannot be completely arbitrary. Specifically, the tested intervals depend on the multiplier a, which is often chosen equal to 1.25 in this context, and the length m_0 of the initial interval I_0. The choice of m_0 should take into account the selected parametric model because I_0 is always assumed to be time-homogeneous and m_0 thus has to reflect flexibility of the parametric model. For example, while $m_0 = 20$ might be reasonable for GARCH(1,1) model, $m_0 = 5$ could be a reasonable choice for the locally constant approximation of a volatility process.

1.4.2 Choice of critical values

The choice of the longest time-homogeneous interval $\hat{I}$ is a multiple testing procedure. The critical values $\mathfrak{z}_k$ are thus selected in the classical way to achieve a prescribed performance under the null hypothesis, that is, in the pure parametric situation. If data come from a parametric model (1.2)–(1.3), the correct choice of $\hat{I}$ is the largest considered interval I_K and the choice of any shorter interval can be interpreted as a "false alarm." The critical values are then selected to minimize the "loss" due to a false alarm.

Čížek et al. (2009) propose to measure the loss associated with a false alarm by the value $L_{I_K}(\tilde{\theta}_{I_K}, \hat{\theta}) = L_{I_K}(\tilde{\theta}_{I_K}) - L_{I_K}(\hat{\theta})$, that is, by the increase of the log-likelihood caused by estimating θ by $\hat{\theta}$ rather than by the optimal estimate $\tilde{\theta}_{I_K}$ under the null hypothesis. Given $r > 0$ and the upper bound on the log-likelihood risk $E_{\theta_0}|L_{I_K}(\tilde{\theta}_{I_K}, \theta_0)|^r \le \mathfrak{R}_r(\theta_0)$, which is data-independent (Theorem 2.1 of Čížek et al., 2009), one can require the loss due to a false alarm to be bounded in the following way:

$$E_{\theta_0}\left|L_{I_K}(\tilde{\theta}_{I_K}, \hat{\theta})\right|^r \le \rho\mathfrak{R}_r(\theta_0), \tag{1.14}$$

where $\rho > 0$ is a constant similar in meaning to the level of a test. Since one performs a test sequentially K times, one can decompose the upper bound on the log-likelihood loss in (1.14) to K equal parts (one per each step) and require

$$E_{\theta_0}\left|L_{I_k}(\tilde{\theta}_{I_k}, \hat{\theta}_{I_k})\right|^r \le \rho_k\mathfrak{R}_r(\theta_0), \tag{1.15}$$

where $\rho_k = \rho k/K \le \rho$ for $k = 1, \ldots, K$. These K inequalities then define the K critical values $\mathfrak{z}_k$, $k = 1, \ldots, K$. To simplify this rather general choice, Čížek et al. (2009) show that $\mathfrak{z}_k$ can be chosen as a linear function of $|I_k|$ and describe its (straightforward) construction by simulation for a given r and ρ in details. This facilitates a proper and general finite-sample testing procedure (see Čížek, 2009, for theoretical properties) with only one possible caveat: the critical values can depend on the true parameter values of the underlying process because (1.15) is specific to θ_0.

Similarly to smoothly-varying coefficient models discussed in Section 1.3, the pointwise adaptive estimation also depends on some auxiliary parameters. A simple strategy is to use conservative values for these parameters and the corresponding set of critical values (e.g., $r = 1$ and $\rho = 1$). The choice of r and ρ can however be data-dependent by using the same strategy as for the bandwidth choice of the tvARCH model – minimizing some global forecasting error (Cheng et al., 2003). Since different values of r and ρ lead to different sets of critical values and hence to different estimates $\hat{\theta}^{(r,\rho)}(t_1)$ and out-of sample forecasts $\hat{\sigma}^2_{t_1+h}(r,\rho)$, a data-driven choice of r and ρ can be done by minimizing the following prediction error:

$$(\hat{r},\hat{\rho}) = \operatorname*{argmin}_{r>0,\rho>0} PE_{\lambda,\mathcal{H}}(r,\rho) = \operatorname*{argmin}_{r>0,\rho>0} \sum_{t\in J}\sum_{h\in\mathcal{H}} |Y^2_{t+h} - \hat{\sigma}^2_{t+h}(r,\rho)|^{\lambda}, \quad (1.16)$$

where $\lambda > 0$, $\mathcal{H}$ is the forecasting horizon such as $\mathcal{H} = \{1\}$, and $J = [\tau, t_1]$ represents a historical interval used for the comparison (e.g., J contains last three or six months of data).

1.4.3 Estimation results

The pointwise adaptive estimation method will be now applied to the log-returns of the S&P 500 stock index. We consider the method applied using three models: the volatility process is locally approximated by means of the constant volatility, ARCH(1), and GARCH(1,1) models. The applicability is more complicated than in the case of the tvARCH model since the critical values constructed from (1.15) depend on the values of the ARCH and GARCH parameters (see Čížek et al., 2009, for details and specific critical values). Since these methods are also computationally demanding, the values of the tuning parameters r and ρ are not selected by the criterion (1.16), but they are instead fixed to $r = 0.5$ and $\rho = 1.5$ in all cases. This has naturally a negative influence on the performance of the pointwise adaptive estimation, but a closely related adaptive weights smoothing presented in the next Section 1.5 will demonstrate the possibilities of the methodology under a data-driven choice of tuning parameters.

To understand the nature of estimation results, let us first look at the parameter estimates and the chosen interval lengths for the modelling based on the local ARCH(1) assumption; see Figure 1.5. The chosen intervals of homogeneity resemble the choice of bandwidth $b|I|$ of tvARCH(1) in years 2001–2003, but differ substantially in the second half of 2003 and year 2004: the adaptively chosen intervals of homogeneity are much longer than the bandwidth choice of tvARCH(1). There are also some differences in the parameter estimates since the local ARCH(1) approximation results much more often in the ARCH parameter being equal to zero and thus in the local constant volatility model. Let us note at this point that the locally adaptive GARCH(1,1) estimates are

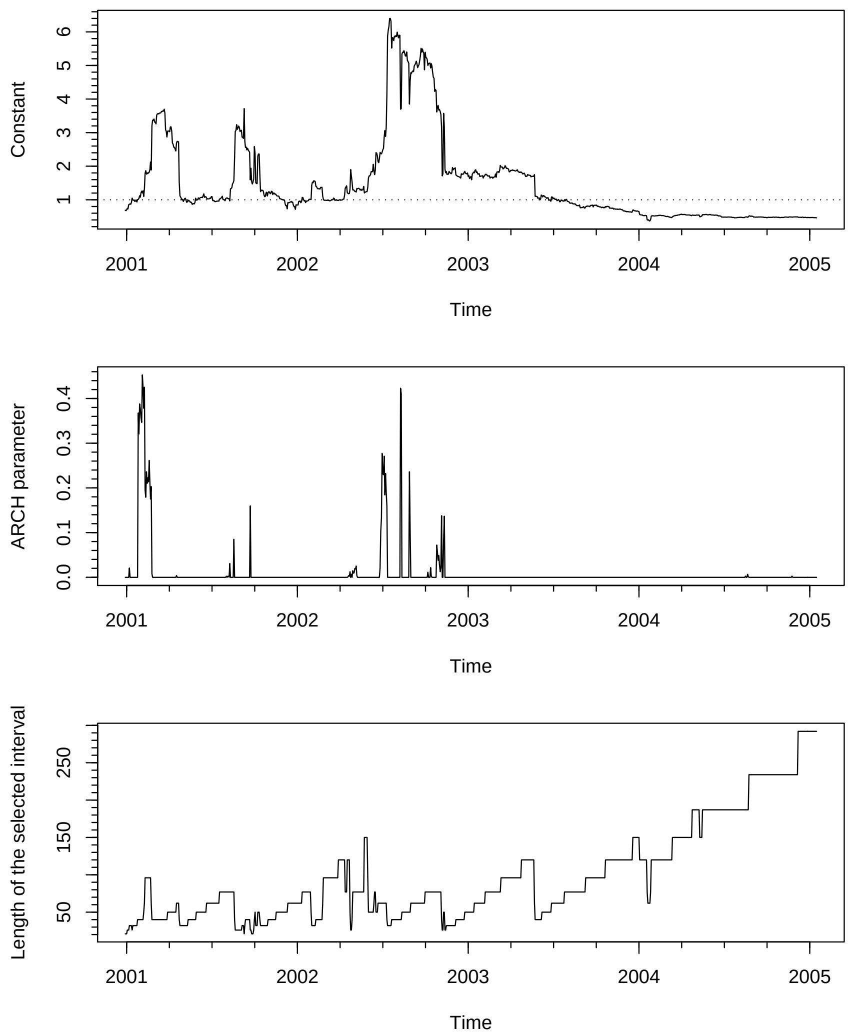

Figure 1.5: The estimated parameters of the pointwise adaptive ARCH(1) model (top panels) and the estimated lengths of the time-homogenous intervals (bottom panel) for S&P 500.

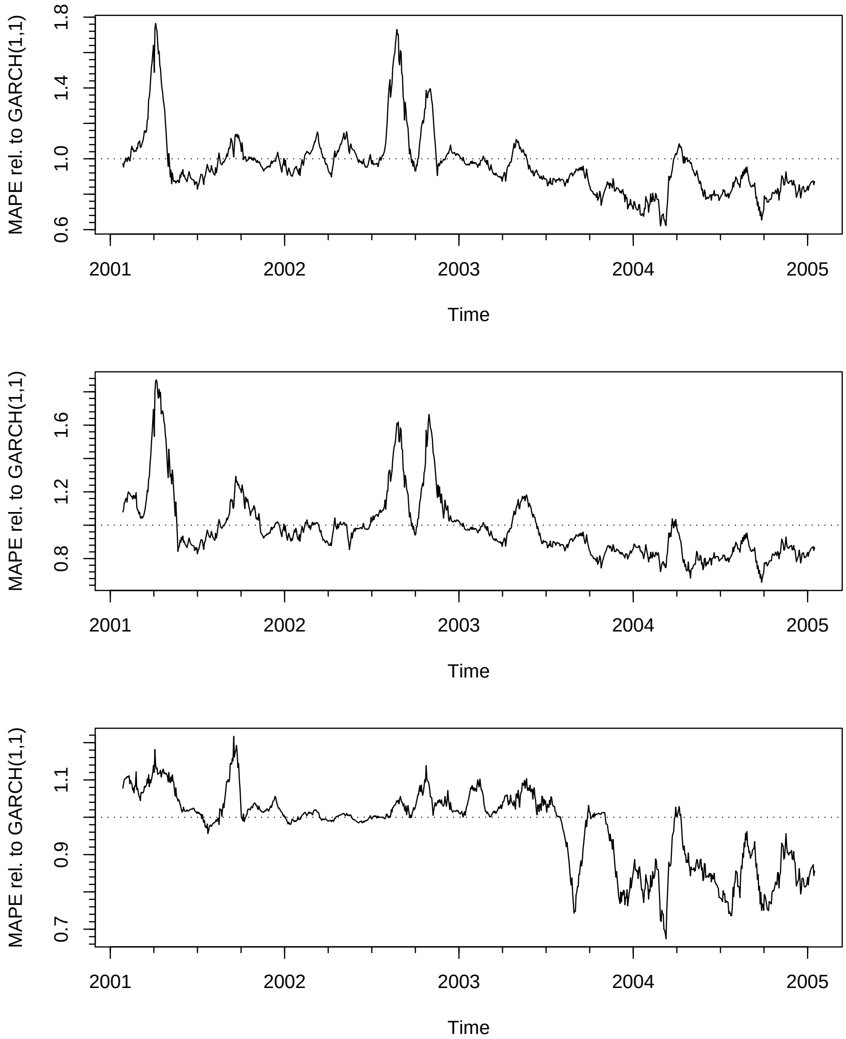

Figure 1.6: The mean average prediction errors of the pointwise adaptive constant (top panel), ARCH(1) (middle panel), and GARCH(1,1) (bottom panel) models relative to the forecast errors of GARCH(1,1) for S&P 500.

Table 1.3: Mean absolute forecast errors in volatility by the pointwise adaptive estimation based on the local approximation of volatility by constant, ARCH(1), and GARCH(1,1).

Year	Local approximation by			Global
	Constant	ARCH(1)	GARCH(1,1)	GARCH(1,1)
2001	1.228	1.298	1.251	1.156
2002	1.853	1.891	1.722	1.714
2003	0.745	0.760	0.818	0.818
2004	0.361	0.351	0.367	0.435
Total	1.046	1.075	1.041	1.031
Weighted	1.040	1.058	1.057	1.087

of different nature in this series: the ARCH parameter mostly fluctuates between 0.05 and 0.2 and the GARCH parameter attains large values ranging from 0.8 to 0.95 most of the time.

The performance of all pointwise adaptive estimates is characterized in terms of MAPE relative to the forecasting errors of the GARCH(1,1) on Figure 1.6. Similarly to tvARCH(1), the pointwise adaptive estimation seems to perform slightly worse than GARCH(1,1) in years 2001 and 2002, while outperforming GARCH(1,1) in years 2003 and 2004. One can notice that the large peaks characterizing the worst performance of the pointwise adaptive estimation correspond to the times following structural changes, that is, the times with very short intervals of time-homogeneity (cf. Figure 1.5). The biggest performance gain of the pointwise adaptive estimation can be therefore observed in year 2004, which seems free of further structural shocks to the market and thus more stable (see Figure 1.5 again). Furthermore, the local GARCH(1,1) modelling matches the GARCH(1,1) forecast performance much more closely than the other two approximations of the volatility process. At this point, the experience with tvARCH(p) would hint that using an ARCH(p) approximation would be beneficial, but it is practically infeasible due to the dependence of the critical values on parameters.

To formally judge the performance of all methods, MAPEs per year are summarized in Table 1.3. These figures confirm the better predictive performance of GARCH(1,1) in years 2001 and 2002, although the difference between local and global GARCH(1,1) estimation is rather small in 2002. In year 2003 and 2004, the pointwise adaptive methods outperform GARCH(1,1), in the case of the local constant and ARCH(1) estimation even irrespective of the number of past observations used for GARCH(1,1) estimation (cf. Table 1.1). Further, it is interesting to note that, even though the GARCH(1,1)-based estimators exhibit overall better performance in terms of the total MAPE, the local constant approximation is actually preferable to the local GARCH(1,1) modelling in all years but 2002 and it outperforms both the local and global GARCH(1,1) forecasts in terms of the weighted total MAPE (again irrespective of the historical-window size). Similarly to the tvARCH results, the locally applied ARCH(1) model is not a good option as it reduces to the locally constant volatility model most of the time. Finally, another advantage of the local constant approximation stems from the critical values independent of parameter values.

1.5 Adaptive weights smoothing

The last approach to the local volatility modelling discussed here combines in a sense features of the smoothly-varying coefficient models and of the pointwise adaptive estimation. The adaptive weights smoothing (AWS) idea proposed by Polzehl and Spokoiny (2000) starts from an initial nonparametric fit such as the kernel quasi-MLE, that is, it first estimates a given time-varying model (1.6)–(1.7) nonparametrically using observations within small neighborhoods of a time period t_1. Later, AWS however tries to expand these local neighborhoods, and similarly to the pointwise adaptive estimation, to find at each time point the largest interval I containing t_1, where $\theta(t), t \in I$, in (1.6)–(1.7) is constant, that is, where a parametric model with fixed parameter values is applicable. The main difference with respect to tvARCH and similar models is that the initial neighborhoods do not have to increase with the sample size, which makes it more flexible and well defined at times shortly after structural breaks. Compared to the pointwise adaptive estimation, AWS does not search for the historical intervals of homogeneity at each time point, but within the whole data set. Additionally, AWS does not necessarily require a time point to be surely be within or outside of a time-homogeneous region: a point can be in such a region only with a certain probability.

We describe here the AWS procedure only in the special case when the unobserved volatility process is locally approximated by a constant. There are several reasons for this. Despite the existing extension of AWS to varying coefficient models (Polzehl and Spokoiny, 2003), the implementation exists only for models with one or two explanatory variables. This limits practical applications to volatility approximations by ARCH(p) models with a small p, which are typically outperformed by the local constant modelling (e.g., see Sections 1.3 and 1.4). The local constant modelling also seems to be similar or preferable to local GARCH(1,1) approximations of the volatility process (e.g., see Section 1.4 and Polzehl and Spokoiny, 2004), which is a likely result of the high data demand of the GARCH(1,1) model.

1.5.1 The AWS algorithm

For the sake of simplicity, let us now thus describe the basic AWS procedure as introduced by Polzehl and Spokoiny (2000) adapted for the volatility modelling; an extention to the time-varying GARCH(1,1) models is discussed in Polzehl and Spokoiny (2004). Similarly to the pointwise adaptive estimation, AWS searches the largest neighborhood of a given time point t and thus requires an increasing sequence of neighborhoods $U_0(t) \subset U_1(t) \subset \ldots \subset U_K(t)$, where typically $U_k(t) = \{\tau : |\tau - t| \leq d_k\}$ for an increasing sequence of $\{d_k\}_{k=0}^{K}$. To describe AWS, one additionally needs again a symmetric integrable kernel function W and an alternative notation $g(t) \equiv g(X_t)$, see (1.6).

1. Initialization: set $k = 0$, define for all $t, t' \in I$ weights $w_k(t, t') = 1$, and at each $t \in I$, compute the initial estimates of the volatility function

$$\hat{g}_k(t) = \frac{\sum_{t' \in U_k(t)} w_k(t, t') Y_{t'}^2}{\sum_{t' \in U_k(t)} w_k(t, t')}$$

   and its variance

$$\hat{s}_k^2(t) = \frac{1}{|I|} \sum_{t' \in I} \{Y_{t'} - \hat{g}_k(t')\}^2 \frac{\sum_{t' \in U_k(t)} w_k^2(t, t')}{\left[\sum_{t' \in U_k(t)} w_k(t, t')\right]^2}.$$

2. Adaptation: set $k = k + 1$, compute for all $t, t' \in I$ weights

$$w_k(t, t') = W\left\{\frac{\hat{g}_{k-1}(t) - \hat{g}_{k-1}(t')}{\phi \hat{s}_{k-1}(t)}\right\},$$

   and at each $t \in I$, compute new estimates of the volatility function

$$\hat{g}_k(t) = \frac{\sum_{t' \in U_k(t)} w_k(t, t') Y_{t'}^2}{\sum_{t' \in U_k(t)} w_k(t, t')}$$

   and its variance

$$\hat{s}_k^2(t) = \frac{1}{|I|} \sum_{t' \in I} \{Y_{t'} - \hat{g}_k(t')\}^2 \frac{\sum_{t' \in U_k(t)} w_k^2(t, t')}{\left[\sum_{t' \in U_k(t)} w_k(t, t')\right]^2}.$$

3. Verification: if $|\hat{g}_k(t) - \hat{g}_{k-s}(t)| > \eta \hat{s}_{k-s}(t)$ for any $s \leq k$, set $\hat{g}_k(t) = \hat{g}_{k-1}(t)$.
4. Stopping rule: stop if $k = K$ or if $\hat{g}_k(t) = \hat{g}_{k-1}(t)$ for all $t \in I$.

The step 1 represents the initial nonparametric estimation within neighborhoods $U_0(t)$. The adaptation step 2 defines the weights used for nonparametric estimation of $\hat{g}_k(t)$ within a larger neighborhood $U_k(t), k > 0$. Most importantly, the weights $w_k(t, t')$ are not proportional to the distance $|t - t'|$ as for classical nonparametric estimates, but they depend on the difference between volatility values at time t and t'. Hence, more distant observations at t' from a larger neighborhood $U_k(t)$ are used only if the previously estimated volatility at times t and t' are close relative to the variance of these values and a tuning parameter ϕ. The third step contains another tuning parameter η, which prevents large changes of the estimated values when neighborhoods are enlarged. Note that, in more recent versions of AWS, η instead defines a convex combination of the estimates $\hat{g}_k(t)$ and $\hat{g}_{k-s}(t)$ (e.g., Polzehl and Spokoiny, 2003) and is thus usually fixed. Finally, the algorithm stops if the largest neighborhood is reached or if there are no changes in the estimated volatility values anymore.

The procedure depends again on a tuning parameter, which is denoted ϕ this time. Similarly to the pointwise adaptive estimation, one can either use a fixed conservative value or make the choice of ϕ data-dependent by minimizing some global forecasting error. Analogously to (1.16), one can select ϕ by

$$\hat{\phi} = \operatorname*{argmin}_{\phi > 0} PE_{\lambda,\mathcal{H}}(\phi) = \operatorname*{argmin}_{\phi > 0} \sum_{t \in J} \sum_{h \in \mathcal{H}} |Y_{t+h}^2 - \hat{\sigma}_{t+h}^2(\phi)|^{\lambda}, \tag{1.17}$$

where $\lambda > 0$, $\mathcal{H}$ is the forecasting horizon, and $J = [\tau, t_1]$ represents a historical interval used for the comparison. Due to the fact that the volatility is locally approximated only by a constant here, J can be selected shorter than in the previous cases (e.g., one or two months).

1.5.2 Estimation results

The data analyzed by AWS will be again the S&P500 index in years 2001–2004. The forecasts using ϕ determined by (1.17) are presented on Figure 1.7, including the ratio of MAPE relative to the MAPE of GARCH(1,1). The yearly averages of absolute errors are reported in Table 1.4. The character of predictions differs substantially from the ones performed by GARCH(1,1), see Figure 1.2. Similarly to the pointwise adaptive estimation, AWS performs worse than GARCH(1,1) in the first two years, while outperforming it in the last two years of data. What is much more interesting is that – with the exception of

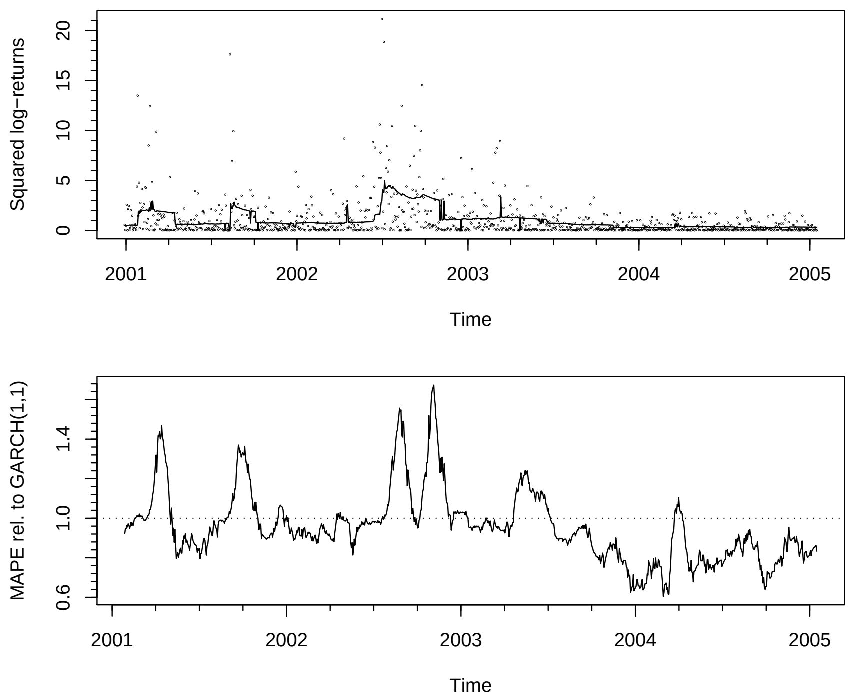

Figure 1.7: The volatility forecasts and mean absolute prediction errors of the adaptive weights smoothing relative to GARCH(1,1) for S&P500.
STF2tvch06.r

year 2002 – AWS performs better than GARCH(1,1) estimated using the 500-observation window. Additionally, AWS outperms any GARCH(1,1) estimate in years 2003 and 2004 and also in terms of the weighted total MAPE (cf. Table 1.1). This means that the local approximation of volatility by a constant practically performs better or as well as GARCH(1,1) once one primitively "protects" the model against structural breaks and uses just one or two years of data instead of the whole available history.

Table 1.4: Mean absolute forecast errors in volatility by the adaptive weights smoothing.

Year	AWS	GARCH(1,1)
2001	1.185	1.156
2002	1.856	1.714
2003	0.782	0.818
2004	0.352	0.435
Total	1.044	1.031
Weighted	1.037	1.087

STF2tvch07.r

1.6 Conclusion

In this chapter, several alternatives to the standard parametric conditional heteroscedasticity modelling were presented: the semiparametric and adaptive nonparametric estimators of models with time-varying coefficients, which can account for nonstationarities of the underlying volatility process. In most cases, adaptive or semiparametric estimation combined with a simpler rather than more complicated models resulted in forecasts of the same or better quality than the standard GARCH(1,1) model. Although this observation is admitedly limited to the data used for demonstration, the results in Fryzlewicz et al. (2008), Polzehl and Spokoiny (2004), or Čížek et al. (2009) indicate that it is valid more generally. The only exception to this rule is tvARCH(p) with a larger number of lags, which behaves similarly to the local constant or ARCH(1) approximations of the volatility in instable periods, but at the same time, is able to capture the features of GARCH(1,1), where this parametric model is predicting optimally. Unfortunately, it is currently not possible to study analogs of tvARCH(5) in the context of the pointwise adaptive estimation or adaptive weights smoothing due to practical limitations of these adaptive methods.

Bibliography

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

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

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

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

Andrews, D. W. K. (2003). End-of-sample instability tests, *Econometrica* **71**, 1661–1694.

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

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

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* **31**, 307–327.

Cai, Z., Fan, J. and Yao, Q. (2000). Functional coefficient regression models for nonlinear time series, *Journal of the American Statistical Association* **95**, 941–956.

Chen, G., Choi, Y. K., and Zhou, Y. (2010). Nonparametric estimation of structural change points in volatility models for time series, *Journal of Econometrics* **126**, 79–114.

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

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

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

Čížek, P., Härdle, W. and Spokoiny, V. (2009). Adaptive pointwise estimation in time-inhomogeneous conditional heteroscedasticity models, *The Econometrics Journal* **12**, 248–271.

Dahlhaus, R. and Subba Rao, S. (2006). Statistical inference for time-varying ARCH processes, *The Annals of Statistics* **34**, 1075–1114.

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

Eizaguirre, J. C., Biscarri, J. G., and Hidalgo, F. P. G. (2010). Structural term changes in volatility and stock market development: evidence for Spain, *Journal of Banking & Finance* **28**, 1745–1773.

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

Fan, J. and Zhang, W. (2008). Statistical models with varying coefficient models, *Statistics and Its Interface* **1**, 179–195.

Francq, C. and Zakoian, J.-M. (2007). Quasi-maximum likelihood estimation in GARCH processes when some coefficients are equal to zero, *Stochastic Processes and their Applications* **117**, 1265–1284.

Fryzlewicz, P., Sapatinas, T., and Subba Rao, S. (2008). Normalised least-squares estimation in time-varying ARCH models, *Annals of Statistics* **36**, 742–786.

Glosten, L. R., Jagannathan, R. and 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**, 1779–1801.

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

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

Herwatz, H. and 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.

Hillebrand, E. (2005). Neglecting parameter changes in GARCH models, *Journal of Econometrics* **129**, 121–138.

Kokoszka, P. and Leipus, R. (2000). Change-point estimation in ARCH models, *Bernoulli* **6**, 513–539.

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

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

Morales-Zumaquero, A., and Sosvilla-Rivero, S. (2010). Structural breaks in volatility: evidence for the OECD and non-OECD real exchange rates, *Journal of International Money and Finance* **29**, 139–168.

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

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

Polzehl, J. and Spokoiny, V. (2000). Adaptive weights smoothing with applications to image restoration, *Journal of the Royal Statistical Society, Ser. B* **62**, 335–354.

Polzehl, J. and Spokoiny, V. (2003). Varying coefficient regression modelling by adaptive weights smoothing, Preprint No. 818, WIAS, Berlin, Germany.

Sentana, E. (1995). Quadratic ARCH Models, *The Review of Economic Studies* **62**, 639–661.

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

Spokoiny, V. (2009). Multiscale local change-point detection with applications to Value-at-Risk, *Annals of Statistics* **37**, 1405–1436.

Starica, C., and Granger, C. (2005) Nonstationarities in stock returns, *The Review of Economics and Statistics* **87**, 503–522.

Taylor, S. J. (1986). *Modeling financial time series*, Chichester: Wiley.

Xu, K.-L., and Phillips, P. C. B. (2008). Adaptive estimation of autoregressive models with time-varying variances, *Journal of Econometrics* **142**, 265–280.