Nonparametric GARCH Models

Peter Bühlmann	Alexander J. McNeil *
Seminar für Statistik	Department of Mathematics
Federal Institute of Technology	Federal Institute of Technology
ETH Zentrum	ETH Zentrum
CH-8092 Zurich	CH-8092 Zurich
Tel: +41 1 632 73 38	Tel: +41 1 632 61 62
Fax: +41 1 632 12 28	Fax: +41 1 632 15 23
buhlmann@stat.math.ethz.ch	mcneil@math.ethz.ch

Revised Version
September 2000

Abstract

In this paper we describe a nonparametric GARCH model of first order and propose a simple iterative algorithm for its estimation from data. We provide a theoretical justification for this algorithm and give examples of its application to stationary time series data showing stochastic volatility. We observe that our nonparametric procedure often gives better estimates of the unobserved latent volatility process than parametric GARCH(1,1) modelling, particularly when asymmetries are present in the data. We show how the basic iterative idea may be extended to more complex time series models combining ARMA or GARCH features of possibly higher order.

1 Introduction

Stationary time series data showing fluctuating volatility and, in particular, financial return series have provided the impetus for the study of a whole series of econometric time series models that may be grouped under the general heading of GARCH (generalized, autoregressive, conditionally heteroscedastic models). Examples include the original ARCH model of Engle (1982), the standard GARCH model, integrated GARCH (IGARCH), exponential GARCH (EGARCH) and threshold GARCH (TGARCH), to name but a very few. Two review articles giving details of these models and their many variants are Bollerslev, Chou, and Kroner (1992) and Shephard (1996). In all of these models the hidden variable volatility depends *parametrically* on lagged values of the process and lagged values of volatility. The standard approach to fitting these models to data involves nonlinear maximum likelihood estimation.

In this paper we take a nonparametric approach to GARCH modelling which is less sensitive to model misspecification. We concentrate on a model, motivated by the natural idea in finance, where the hidden volatility depends nonparametrically on one lagged volatility and one lagged value of the process; we term this model a nonparametric GARCH(1,1) process. Nonparametric estimation is a feasible alternative because, in a sense, volatility is only partially hidden. It is possible to devise an *iterative* scheme to estimate the volatility process and thus, perhaps surprisingly, to overcome the problem of *latency*. This scheme makes use of a bivariate smoother and is thus very easy to implement. Using a

*The second author is Swiss Re Research Fellow at ETH Zürich and gratefully acknowledges the financial support of Swiss Re.

software package that provides bivariate smoothing such as S-Plus, we are simply required to program an additional loop.

We argue that such an iterative bivariate smoothing scheme has the same rates of estimation accuracy as one classical bivariate smoothing step, which is usually of the order $n^{-1/3}$ with n denoting the sample size; see Stone (1982). Although the nonparametric GARCH(1,1)-volatility process depends on the infinite past of the process, its estimation can be achieved within reasonable accuracy. In contrast to the parametric case, one should not use an ARCH-approximation scheme in the nonparametric framework: an expansion of the nonparametric GARCH(1,1)- into a nonparametric ARCH(∞)-model and estimation of the latter with a high order nonparametric ARCH(p)-model is prohibitive. The curse of dimensionality would at best lead to an estimation rate of order $n^{-2/(4+p)}$ with p large (Stone 1982). One way to deal with the curse of dimensionality is to consider multiplicative ARCH(p)-models with estimation rate usually of order $n^{-2/5}$ (Yang, Härdle, and Nielsen 1999). But from an interpretational view, we strongly prefer the nonparametric GARCH(1,1)-model where volatility depends only on one-lagged values of the time series and the volatility.

The basic idea of our iterative estimation algorithm allows for many extensions to other time series models with some form of latency. These extensions include nonparametric GARCH models of higher order, nonparametric ARMA models for prediction of conditional expectations and nonparametric GARCH models with a general additive structure for the conditional variance; we give examples in Section 4. To begin with we present in Section 2 the basic first order nonparametric model and its estimation algorithm, for which we provide a justification. In Section 3 we provide examples of the use of the algorithm on both simulated and real data.

2 The Basic Model

In this paper we consider stationary stochastic processes $\{X_t; t \in \mathbb{Z}\}$ adapted to the filtration $\{\mathcal{F}_t; t \in \mathbb{Z}\}$ with $\mathcal{F}_t = \sigma\left(\{X_s; s \le t\}\right)$, and having the form

$$\begin{aligned} X_t &= \sigma_t Z_t, \\ \sigma_t^2 &= f(X_{t-1}, \sigma_{t-1}^2), \end{aligned} \tag{1}$$

where $\{Z_t; t \in \mathbb{Z}\}$ is an iid innovation series with zero mean, unit variance and a finite fourth moment. Z_t is assumed independent of $\{X_s; s < t\}$, and $f : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}_+$ is a strictly positive-valued function; σ_t^2 is then the conditional variance $\mathrm{Var}[X_t \mid \mathcal{F}_{t-1}]$ and σ_t is known as the volatility.

Equations (1) specify a general discrete time stochastic volatility process of first order that includes the parametric ARCH(1) and GARCH(1,1) models, but which also allows for a more complicated dependence of the present volatility on the past. For example, for the parametric GARCH(1,1) model,

$$f(x, \sigma^2) = \alpha_0 + \alpha_1 x^2 + \beta\sigma^2,\ \alpha_0, \alpha_1, \beta > 0,$$

and for a more unusual asymmetric model we might consider

$$f(x, \sigma^2) = \alpha_0 + \alpha_1 x^2 + \beta \mathbf{1}_{\{x \le 0\}} \sigma^2.$$

In this paper we leave the exact form of f unspecified and attempt to estimate it by nonparametric means.

To this end we observe first that the model (1) can be written in terms of an additive noise as follows;

$$\begin{aligned} X_t^2 &= f(X_{t-1}, \sigma_{t-1}^2) + V_t, \\ V_t &= f(X_{t-1}, \sigma_{t-1}^2)\left(Z_t^2 - 1\right), \end{aligned}$$

where V_t is a martingale difference series with $\mathbb{E}[V_t] = \mathbb{E}[V_t \mid \mathcal{F}_{t-1}] = 0$ and $\mathrm{Cov}[V_s, V_t] = \mathrm{Cov}[V_s, V_t \mid \mathcal{F}_{t-1}] = 0$ for $s < t$. It follows that

$$\mathbb{E}[X_t^2 \mid \mathcal{F}_{t-1}] = f(X_{t-1}, \sigma_{t-1}^2),$$

and

$$\mathrm{Var}[X_t^2 \mid \mathcal{F}_{t-1}] = f^2(X_{t-1}, \sigma_{t-1}^2)\left(\mathbb{E}[Z_t^4] - 1\right).$$

This suggests we could estimate f by regressing X_t^2 on the lagged variables X_{t-1} and σ_{t-1}^2 using a nonparametric smoothing technique. The conditional heteroscedasticity of the series X_t^2 suggests a weighted regression might be used. The principal problem is that the volatility σ_{t-1} is an unobserved latent variable. This problem is overcome in the following algorithm.

2.1 An estimation algorithm

Assume we have a data sample $\{X_t; 1 \le t \le n\}$, ideally from a process satisfying (1).

1. Calculate a first estimate of volatility $\{\hat\sigma_{t,0}; 1 \le t \le n\}$ by fitting an ordinary parametric GARCH(1,1) model by standard maximum likelihood. Set $m = 1$.
2. Regress $\{X_t^2; 2 \le t \le n\}$ against $\{X_{t-1}; 2 \le t \le n\}$ and $\{\hat\sigma_{t-1,m-1}^2; 2 \le t \le n\}$ using a nonparametric procedure to obtain an estimate $\hat f_m$ of f. As an added refinement a weighted regression can be performed with regression weights $\{\hat\sigma_{t,m-1}^{-2}; 2 \le t \le n\}$; our experience with simulations suggests this mostly yields slightly improved estimates. (We thereby assume that $\mathbb{E}[Z_t^4] < \infty$ which implies existence of $\mathrm{Var}(X_t^2 | \mathcal{F}_{t-1})$).
3. Calculate $\{\hat\sigma_{t,m}^2 = \hat f_m(X_{t-1}, \hat\sigma_{t-1,m-1}^2); 2 \le t \le n\}$ and impute some sensible value for $\hat\sigma_{1,m}^2$, for example $\hat\sigma_{1,m-1}^2$.
4. Increment m and return to step two if $m < M$.

Our experience, gathered from a number of simulation experiments, suggests that if the nonparametric estimation steps will provide an improvement over the parametric GARCH estimation of volatility, then this improvement will be attained in a small number of steps 1.–4. Thereafter, there is little to pick and choose between volatility estimates $\hat\sigma_{t,m}$ for various values of m. The algorithm can however often be improved by averaging over the final K such estimates to obtain

$$\hat\sigma_{t,*} = (1/K) \sum_{m=M-K+1}^{M} \hat\sigma_{t,m},$$

and then performing a final regression of X_t^2 against X_{t-1} and $\hat\sigma_{t-1,*}^2$ to get final estimates $\hat f$ of f and $\hat\sigma_t^2 = \hat f(X_{t-1}, \hat\sigma_{t-1,*}^2)$; we refer to this as a final smoothing step.

The parameters of the estimation algorithm are n, the sample size; M, the number of basic iterations and K, the number of estimates used in the final smoothing step. There will also be a bandwidth parameter to choose in the nonparametric regression method.

2.2 Justification of the algorithm

The following assumption is crucial for our justification of consistency of the estimation algorithm.

Assumption A1 (contraction with respect to hidden variable).

$$\sup_{x\in\mathbb{R}} |f(x,\sigma^2) - f(x,\tau^2)| \leq D|\sigma^2 - \tau^2| \text{ for some } 0 < D < 1, \text{ for all } \sigma^2, \tau^2 \in \mathbb{R}_+.$$

The population case. A key feature of the algorithm can be understood when simplifying to the case of no estimation error. We recursively define

$$\begin{aligned} f_{t,m}(x,\sigma^2) &= \mathbb{E}[X_t^2 | X_{t-1} = x, \sigma^2_{t-1,m-1} = \sigma^2] \\ \sigma^2_{t,m} &= f_{t,m}(X_{t-1}, \sigma^2_{t-1,m-1}),\ m = 1,2,\ldots\ ;\ t \in \mathbb{Z}, \end{aligned}$$

where $\sigma^2_{t,0}$ are some starting values assumed to be elements of $\mathcal{F}_{t-1}$ for all $t \in \mathbb{Z}$, i.e. they are independent from $\{Z_s;\ s \geq t\}$. The $\sigma^2_{t,m}$ are the population quantities corresponding to the estimates $\hat{\sigma}^2_{t,m}$ of the algorithm.

Assumption A2 (moment assumption).

$$\mathbb{E}|\sigma_t|^4 \leq C_1 < \infty,\ \mathbb{E}|\sigma_{t,0}|^4 \leq C_2 < \infty \text{ for all } t.$$

Theorem 1. *Assume that $\{X_t; t \in \mathbb{Z}\}$ is as in (1), satisfying assumptions A1 and A2. Denote by $\|.\|_2$ the L_2-norm, i.e. $\|Y\|_2 = (\mathbb{E}|Y|^2)^{1/2}$. Then, in the mth iteration step*

$$\|\sigma^2_{t,m} - \sigma^2_t\|_2 \leq D^m(C_1 + C_2),\ t = m+2, m+3, \ldots$$

A proof is given in the Appendix. Theorem 1 shows that in case of no estimation error, iteration from arbitrary starting values $\sigma^2_{t,0}$ $(t \in \mathbb{Z})$ leads exponentially fast to the true squared volatility σ^2_t, due to the contraction property in assumption A1.

An approach dealing with estimation errors. A rigorous mathematical analysis about the estimation error in our procedure seems very difficult, due to the iterative structure of the algorithm. Although Theorem 2 below is mathematically rigorous, its assumptions are hard to justify. Nevertheless, we believe that the theoretical approach presented here gives some useful insight.

We bound the quantity of interest as

$$\|\hat{\sigma}^2_{t,m} - \sigma^2_t\|_2 \leq \|\hat{\sigma}^2_{t,m} - \tilde{\sigma}^2_{t,m}\|_2 + \|\tilde{\sigma}^2_{t,m} - \sigma^2_{t,m}\|_2 + \|\sigma^2_{t,m} - \sigma^2_t\|_2. \tag{2}$$

Thereby,

$$\begin{aligned} &\tilde{f}_{t,m}(x,\sigma^2) = \mathbb{E}[X_t^2 | X_{t-1} = x, \hat{\sigma}^2_{t-1,m-1} = \sigma^2], \\ &\tilde{\sigma}^2_{t,m} = \tilde{f}_{t,m}(X_{t-1}, \hat{\sigma}^2_{t-1,m-1}), \end{aligned}$$

the latter being a true conditional expectation as a function of X_{t-1} and the estimate $\hat{\sigma}^2_{t-1,m-1}$. $\tilde{\sigma}^2_{t,m}$ is an intermediate term between the population quantity $\sigma^2_{t,m}$ above and the estimate $\hat{\sigma}^2_{t,m}$. The terms on the right hand side in (2) can be controlled as follows. The third one is bounded by Theorem 1; the first and second are controlled using assumptions A3 and A4 below.

Assumption A3.

$$\begin{aligned} &\mathbb{E}[\{\tilde{\sigma}^2_{t,m} - \sigma^2_{t,m}\}^2] = \mathbb{E}[\{\tilde{f}_{t,m}(X_{t-1}, \hat{\sigma}^2_{t-1,m-1}) - f_{t,m}(X_{t-1}, \sigma^2_{t-1,m-1})\}^2] \\ \leq\ & G^2 \mathbb{E}[\{\hat{\sigma}^2_{t-1,m-1} - \sigma^2_{t-1,m-1}\}^2] \text{ for some } 0 < G < 1, \end{aligned}$$

for $t = m+2, m+3, \ldots$ and $m = 1, 2, \ldots$

Assumption A4 ($\mathcal{L}_2$ estimation error).

$$\Delta_n^{(\mathcal{L}_2)} = \sup_{m\geq 1} \max_{m+2\leq t\leq n} \|\hat{\sigma}^2_{t,m} - \tilde{\sigma}^2_{t,m}\|_2 \to 0 \text{ as } n \to \infty.$$

Theorem 2. *Assume that* $\{X_t; t \in \mathbb{Z}\}_t$ *is as in (1), satisfying assumptions A1-A4 and* $\max_{2 \le t \le n} \mathbb{E}|\hat{\sigma}^2_{t,0}|^4 \le C_3 < \infty$ *for all* n. *Denote by* $\|.\|_2$ *the* L_2*-norm, i.e.* $\|Y\|_2 = (\mathbb{E}|Y|^2)^{1/2}$. *Then, the estimator* $\hat{\sigma}^2_{t,m}$ *in the* m*th step of the algorithm satisfies*

$$\|\hat{\sigma}^2_{t,m} - \sigma^2_t\|_2 \le \Delta_n^{(\mathcal{L}_2)}/(1-G) + G^M(C_1 + C_3) + D^m(C_1 + C_2)$$
$$\text{for all } t = m+2, m+3, \ldots, n.$$

Particularly, by choosing $m_n = C\{-\log(\Delta_n^{(\mathcal{L}_2)})\}$ *for some* $C \ge \max\{-1/\log(D), -1/\log(G)\}$, *we have*

$$\max_{m_n+2 \le t \le n} \|\hat{\sigma}^2_{t,m_n} - \sigma^2_t\|_2 = O(\Delta_n^{(\mathcal{L}_2)}) = o(1) \text{ as } n \to \infty.$$

A proof is given in the Appendix. Note that similar asymptotic results also hold for the final smoothed estimate $\hat{\sigma}^2_t$.

The last statement of Theorem 2 says that the accuracy of the estimation algorithm is of the same order as the maximal error $\Delta_n^{(\mathcal{L}_2)}$ in one smoothing step. Since the smoothing problem is bivariate, we expect $\Delta_n^{(\mathcal{L}_2)} = O(n^{-2/3})$ (or slightly slower) with a second order kernel, provided some appropriate smoothness conditions on the function f hold and rate-optimal bandwidth choice is used (Stone 1982), although the volatility of the nonparametric GARCH model at time t is a function of the *infinite* past $X_{t-1}, X_{t-2}, \ldots$.

We complete this section with a discussion about assumptions. Assumption A1 is crucial in our reasoning and we conjecture that without a suitable contraction property of the function f with respect to the second, hidden argument, there is no hope in getting a consistent estimate with our iterative scheme. The condition is related to stationarity of the process and holds for stationary parametric GARCH(1,1) models and asymmetric extension thereof, for example the one in section 3.1. Assumption A2 is a mild moment condition. Theorem 1 states a clean mathematical result when there are no estimation errors; it thus points out that it seems indeed possible to solve the problem of nonparametric GARCH estimation with a simple iterative smoothing scheme. Assumptions A3 and A4 are exclusively needed to deal with the case of estimation error. Assumption A3 can be viewed as a requirement on projections: $\mathbb{E}[\{\tilde{\sigma}^2_{t,m} - \sigma^2_{t,m}\}^2] = (\mathbb{E}[X_t^2|X_{t-1}, \hat{\sigma}^2_{t-1,m-1}] - \mathbb{E}[X_t^2|X_{t-1}, \sigma^2_{t-1,m-1}])^2$. The requirement in A3 is thus a contraction for the difference of projections (conditional expectations) with respect to $\mathbb{E}[\{\hat{\sigma}^2_{t-1,m-1} - \sigma^2_{t-1,m-1}\}^2]$, an expected difference between the variables causing the projections to be distinct. Assumption A4 controls the $\mathcal{L}_2$ estimation error when nonparametrically regressing X_t^2 versus X_{t-1} and $\hat{\sigma}^2_{t-1,m-1}$. A rigorous justification for appropriateness of our assumptions A3 and A4 is not given here and still an open area for future research in theoretical statistics.

3 Illustrative Examples

3.1 Simulated Example

We illustrate the method using simulated data from a process satisfying (1). For the volatility surface we take

$$f(x, \sigma^2) = 5 + 0.2x^2 + (0.75 \cdot \mathbf{1}_{\{x>0\}} + 0.1 \cdot \mathbf{1}_{\{x \le 0\}})\sigma^2,$$

which is depicted in Figure 1, and we assume that the innovation distribution is standard normal. Note the asymmetry that has been built into the GARCH effect; the strength of the GARCH effect (that is the magnitude of the σ^2 coefficient) varies with the sign of x. A realisation of $n = 1000$ points from this process is shown in Figure 2.

We apply the algorithm to these data, setting the number of basic iterations to be $M = 8$ and performing a final smooth based on $K = 5$ final iterations. We use the default value of the smoothing parameter in the S-Plus implementation of a bivariate *loess* smoother.

In Figure 3 the estimated surfaces $\hat{f}_m$ calculated at the first 8 iteration and the surface calculated at the final smoothing stage are shown. Comparison with Figure 1 indicates that the nonparametric algorithm is recovering the essential features of the volatility surface. This series of nine pictures illustrates the first envisaged use of the algorithm — as an exploratory graphical tool for investigating the dependence of the present volatility on the immediate past.

To provide a quantitative assessment of the accuracy of the volatility estimates we calculate both a mean squared estimation error and a mean absolute estimation error. Because estimates of volatility at the first few time points may be unreliable we generally omit $r = 10$ values from this calculation. Our measures of estimation error are

$$\mathrm{MSE}(\hat{\sigma}_{\cdot,m}) = \frac{1}{n-r} \sum_{t=r+1}^{n} (\hat{\sigma}_{t,m} - \sigma_t)^2, \quad \mathrm{MAE}(\hat{\sigma}_{\cdot,m}) = \frac{1}{n-r} \sum_{t=r+1}^{n} |\hat{\sigma}_{t,m} - \sigma_t| .$$

In Table 1 we see that the parametric GARCH(1,1) estimate of volatility can be improved upon through additional iterations of the nonparametric algorithm. Even a single iteration reduces both errors by around a third and after 2 or 3 iterations no further improvement is obtained; thereafter the errors oscillate within a band. In this example the final smooth gives the lowest errors of all.

Iteration (m)	$\mathrm{MSE}(\hat{\sigma}_{\cdot,m})$	$\mathrm{MAE}(\hat{\sigma}_{\cdot,m})$
0 (parametric)	0.48	0.58
1	0.31	0.40
2	0.25	0.37
3	0.21	0.33
4	0.19	0.32
5	0.20	0.32
6	0.20	0.32
7	0.20	0.32
8	0.20	0.32
* (final smooth)	0.18	0.31

Table 1: Volatility estimation errors for a realisation of $n = 1000$ points.

In Figure 4 the volatility estimates are compared with the true volatility trajectory for an arbitrary section of 70 observations. In the left-hand picture it is clear that the parametric estimate derived from GARCH(1,1) modelling is unable to follow the true volatility closely through its peaks and troughs; on the other hand, in the right-hand picture we see that the nonparametric GARCH estimate derived from the final smoothing round is fairly faithful to the true volatility.

Essentially it is the strong asymmetry of the simulated process that denigrates the performance of ordinary parametric GARCH as a volatility estimator. For more symmetric processes the difference in performance is less obvious. For *simple* additive surfaces of the form

$$f(x, \sigma^2) = g(x) + h(\sigma^2),$$

where $g : \mathbb{R} \to \mathbb{R}_+$ is a positive-valued function satisfying $g(x) = g(-x)$ (such as $g(x) = \alpha|x|$, $0 < \alpha < 1$) and $h : \mathbb{R}_+ \to \mathbb{R}_+$ is a positive-valued non-decreasing function (such as

$h(\sigma^2) = \beta\sigma$, $0 < \beta < 1$), the performance of ordinary GARCH(1,1) is often as good as or slightly better than that of the nonparametric procedure, even though GARCH(1,1) assumes an erroneous parametric form. For complicated additive functions we note that our procedure can be adapted to fit general nonparametric additive models; for an outline see Section 4.2.

When the true process is exactly GARCH(1,1) then we would expect the parametric procedure to outperform the nonparametric one, and it does so. Our simulation experiments suggest that it is often in situations of asymmetry that the nonparametric procedure has advantages.

3.2 Real Example

For an example with real data we take a 1000-day excerpt from the time series of daily percentage returns on the BMW share price (see Figure 5). Our time series contains the period of high volatility around the 1987 stock market crash.

In Figure 6 we show again the graphical output of the algorithm for the first 8 iterations and the final smoothing step. The pictures have been rotated so that we view them along the line $x = 0$. These estimated volatility surfaces show some evidence of an asymmetric effect, which depends on the sign of the last observation; as long as volatility is low, a large positive return appears to have a much more modest effect on the next day's volatility than does a large negative return. This asymmetric effect of new information is a well-known phenomenon in financial time series and our method clearly picks it up.

Finally in Figure 7 we show the estimated volatilities derived from parametric modelling and the final-smooth stage of nonparametric modelling. Although we do not know the true volatility process for these data a simple comparison of the estimators can be obtained by calculating the mean squared "error" statistic

$$\frac{1}{n-r} \sum_{t=r+1}^{n} (\hat{\sigma}_t^2 - X_t^2)^2$$

for the two volatility estimates. This statistic is based on the additive representation of the squared series as $X_t^2 = \sigma_t^2 + V_t$, as introduced in Section 2; the V_t is a martingale difference series with mean zero, which we regard as the error term. The mean squared error statistic takes the values 106 and 98.3 for the parametric and nonparametric estimators respectively. Thus the error is about 8% smaller for the nonparametric estimate.

4 Extensions

The iterative idea of our estimation algorithm can be extended in a variety of ways and combined with other nonparametric modelling techniques.

4.1 Nonparametric GARCH(p,q)

The estimation algorithm in Section 2.1 and its justification easily extend to the nonparametric GARCH(p, q) model with $0 \leq p, q < \infty$. Equation (1) is now generalised to

$$\begin{aligned} X_t &= \sigma_t Z_t, \\ \sigma_t^2 &= f(X_{t-1}, \ldots, X_{t-p}, \sigma_{t-1}^2, \ldots, \sigma_{t-q}^2). \end{aligned}$$

The mth iteration step in the estimation algorithm is then

$$\hat{\sigma}_{t,m}^2 = \hat{f}_m(X_{t-1}, \ldots, X_{t-p}, \hat{\sigma}_{t-1,m-1}^2, \ldots, \hat{\sigma}_{t-q,m-1}^2),$$

where $\hat{f}_m$ is based on a $p+q$-variate smoother of X_t^2 versus $X_{t-1}, \ldots, X_{t-p}, \hat{\sigma}^2_{t-1,m-1}, \ldots, \hat{\sigma}^2_{t-q,m-1}$. Justification of such an extension can be given as in section 2.2 by replacing assumption A1 with

$$\sup_{x \in \mathbb{R}^p} |f(x_1, \ldots, x_p, \sigma_1^2, \ldots, \sigma_q^2) - f(x_1, \ldots, x_p, \tau_1^2, \ldots, \tau_q^2)| \leq \sum_{i=1}^{q} d_i |\sigma_i^2 - \tau_i^2|,$$

$$\text{for some } 0 < d_1, \ldots, d_q < 1 \text{ with } \sum_{i=1}^{q} d_i = D < 1,$$

for all $\sigma^2,\ \tau^2 \in (\mathbb{R}^+)^q$.

4.2 Nonparametric additive GARCH

The model in (1) or the higher order model of Section 4.1 can be modified to the case where σ_t^2 is a general additive function of lagged values and volatilities. The estimation algorithm is simply adjusted: the smoothing operations are performed according to the additive structure of σ_t^2 by using the backfitting algorithm. For generalized additive models and their estimation see Hastie and Tibshirani (1990)

4.3 Nonparametric ARMA-GARCH of first order

The estimation algorithm also extends to hybrid ARMA-GARCH models; for notational simplicity we focus on the first order case. Consider

$$\begin{aligned} X_t &= \mu_t + \sigma_t Z_t, \\ \mu_t &= g(X_{t-1}, \mu_{t-1}), \\ \sigma_t^2 &= f\left((X_{t-1} - \mu_{t-1}), \sigma_{t-1}^2\right), \end{aligned}$$

where $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a nonparametric function giving the conditional mean of the process; all other elements of the model are as in (1).

Observe that

$$\mathbb{E}[X_t \mid \mathcal{F}_{t-1}] = g(X_{t-1}, \mu_{t-1}),$$

and

$$\mathrm{Var}[X_t \mid \mathcal{F}_{t-1}] = f\left((X_{t-1} - \mu_{t-1}), \sigma_{t-1}^2\right).$$

This suggests that a similar scheme can be applied to estimate both g and f.

We start with estimates $\hat{\mu}_{t,0}$ and $\hat{\sigma}_{t,0}$ of μ_t and σ_t derived by the usual parametric means. We regress X_t against X_{t-1} and $\hat{\mu}_{t-1,0}$, weighting possibly with $\hat{\sigma}_{t,0}^{-1}$, to estimate g and hence to re-estimate μ_t by $\hat{\mu}_{t,1}$. We then regress $(X_t - \hat{\mu}_{t,1})^2$ against $(X_{t-1} - \hat{\mu}_{t-1,1})$ and $\hat{\sigma}^2_{t-1,0}$, weighting possibly with $\hat{\sigma}_{t,0}^{-2}$, to estimate f and hence to re-estimate σ_t by $\hat{\sigma}_{t,1}$. We iterate this procedure to improve estimates of conditional mean and volatility.

5 Discussion

In this paper we have proposed and justified an algorithm for fitting nonparametric GARCH models of first order, and we have suggested extensions to other time series models. The first use we see for our procedure in practical applications is an *exploratory* one: assuming the availability of a bivariate smoother (such as *loess* in S-Plus) it is a simple matter to implement our iterative algorithm to investigate visually the dependence of volatility on lagged values of the time series and the volatility itself.

The second envisaged use for our method is as a volatility estimator in situations where the functional relationship between volatility and the lagged series differs markedly from standard parametric GARCH(1,1), particularly in situations where asymmetries appear to be present. The new estimator could then also be used to compute dynamically changing measures of risk, such as value at risk and expected shortfall conditioned on the observed past; see McNeil and Frey (1999).

A further positive feature of our model is that it is *fully* nonparametric in the sense that not only is the exact functional relationship between volatility and the one-lagged values of the time series and volatility left unspecified, but also no assumptions are made regarding the distributional form of the innovation distribution. (We assume only the existence of a finite fourth moment to justify the use of weighted nonparametric regression.) Much work in empirical finance has actually suggested that GARCH-type models with Gaussian innovations cannot capture the leptokurtosis and conditional heteroscedasticity of typical financial return data.

The proposed algorithm for nonparametric GARCH is thus an attractive flexible fitting method that does neither require specification of the functional form of the volatility nor that of the innovation distribution.

Of course the GARCH framework is not the only way to approach volatility estimation. Another stream of models are the stochastic volatility (SV) models; see Shephard (1996). In contrast to our model (1), the SV model has a fully latent volatility process with no observable structure. It can be formalized as a state space model

$$\begin{aligned} X_t &= \sigma_t Z_t, \\ \{\sigma_t^2\} &= \text{stationary stochastic process of a pre-specified form,} \end{aligned} \tag{3}$$

with Z_t as in (1), and σ_t^2 *conditionally independent* of $\{X_s; s < t\}$ given $\{\sigma_s^2; s < t\}$. Note that this conditional independence structure is not true for our nonparametric GARCH model. This is why we refer to *partially hidden* volatility in (1), since the lagged value X_{t-1} contributes to the partial observability of σ_t^2. The SV model in (3) cannot be fitted with our iterative procedure. If the process $\{\sigma_t^2\}$ and the innovation distribution of Z_t are of finite parametric form, techniques from nonlinear Kalman filtering can be used, e.g. Markov chain Monte Carlo, Gibbs sampling or recursive sampling; see for example Künsch (2000).

Appendix

Proof of Theorem 1. By iteration, $\sigma_{t-1,m-1}^2 \in \mathcal{F}_{t-2}$ is independent from $\{Z_s;\ s \geq t-1\}$. Since Z_t is also independent from $\{X_s;\ s < t\}$, we can write

$$\begin{aligned} \sigma_{t,m}^2 &= \mathbb{E}[X_t^2 | X_{t-1}, \sigma_{t-1,m-1}^2] \\ &= \mathbb{E}\{f(X_{t-1}, \sigma_{t-1}^2) | X_{t-1}, \sigma_{t-1,m-1}^2\},\ m = 1, 2, \dots;\ t \in \mathbb{Z}. \end{aligned}$$

The following argument quantifies the error due to wrong starting values,

$$\begin{aligned} \mathbb{E}\{(\sigma_t^2 - \sigma_{t,m}^2)^2\} &= \mathbb{E}\left([f(X_{t-1}, \sigma_{t-1}^2) - \mathbb{E}\{f(X_{t-1}, \sigma_{t-1}^2) | X_{t-1}, \sigma_{t-1,m-1}^2\}]^2\right) \\ &\leq \mathbb{E}[\{f(X_{t-1}, \sigma_{t-1}^2) - f(X_{t-1}, \sigma_{t-1,m-1}^2)\}^2] \\ &\leq D^{2m} \mathbb{E}[\{\sigma_{t-m}^2 - \sigma_{t-m,0}^2\}^2]. \end{aligned} \tag{4}$$

The first inequality holds since $\mathbb{E}\{f(X_{t-1}, \sigma_{t-1}^2) | X_{t-1}, \sigma_{t-1,m-1}^2\}$ is the best mean square error predictor of $\sigma_t^2 = f(X_{t-1}, \sigma_{t-1}^2)$ as a function of X_{t-1} and $\sigma_{t-1,m-1}^2$, the second inequality holds by iteration and assumption A1. The conclusion then follows from assumption A1, the triangle inequality for $\|\cdot\|_2$ and assumption A2. $\square$

Proof of Theorem 2. Use the decomposition in (2). For the second term on the right hand side we obtain with assumption A3 and the notation in A4,

$$\begin{aligned}\|\tilde{\sigma}^2_{t,m} - \sigma^2_{t,m}\|_2 &\leq G\|\hat{\sigma}^2_{t-1,m-1} - \sigma^2_{t-1,m-1}\|_2 \leq G(\Delta_n^{(\mathcal{L}_2)} + \|\tilde{\sigma}^2_{t-1,m-1} - \sigma^2_{t-1,m-1}\|_2) \\ &\leq \ldots \leq \sum_{j=1}^{m-1} G^j \Delta_n^{(\mathcal{L}_2)} + G^m \|\hat{\sigma}^2_{t-m,0} - \sigma^2_{t-m,0}\|_2.\end{aligned}$$

Using Theorem 1, the bound above implies

$$\|\hat{\sigma}^2_{t,m} - \sigma^2_t\|_2 \leq \sum_{j=0}^{m-1} G^j \Delta_n^{(\mathcal{L}_2)} + G^m \|\hat{\sigma}^2_{t-m,0} - \sigma^2_{t-m,0}\|_2 + D^m(C_1 + C_2),$$
$$t = m+2, m+3, \ldots, n.$$

Using the finite 4th moment of $\hat{\sigma}^2_{t,0}$ and assumption A2, this then yields the first assertion in Theorem 2. The other follows by the choice of m_n yielding $\max\{D^{m_n}, G^{m_n}\} = O(\Delta_n^{(\mathcal{L}_2)})$. □

References

Bollerslev, T., R. Chou, and K. Kroner (1992): "ARCH modeling in finance," *Journal of Econometrics*, 52, 5–59.

Engle, R. (1982): "Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation," *Econometrica*, 50, 987–1008.

Hastie, T., and R. Tibshirani (1990): *Generalized Models*. Chapman & Hall, New York.

Künsch, H. (2000): "State space and hidden Markov," in *Complex Stochastic Systems*, ed. by C. D. Barndorff-Nielsen, O.E., and C. Klüppelberg. CRC Press (to appear).

McNeil, A., and R. Frey (1999): "Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach," preprint, ETH Zürich.

Shephard, N. (1996): "Statistical Aspects of ARCH and Stochastic Volatility," in *Time Series Models in Econometrics, Finance and other Fields*, ed. by D. Cox, D. Hinkley, and O. Barndorff-Nielsen, pp. 1–55, London. Chapman & Hall.

Stone, C. (1982): "Optimal global rates of convergence for nonparametric regression," *Annals of Statistics*, 10, 1040–1053.

Yang, L., W. Härdle, and J. Nielsen (1999): "Nonparametric autoregression with multiplicative volatility and additive mean," *Journal of Time Series Analysis*, 20, 579–604.

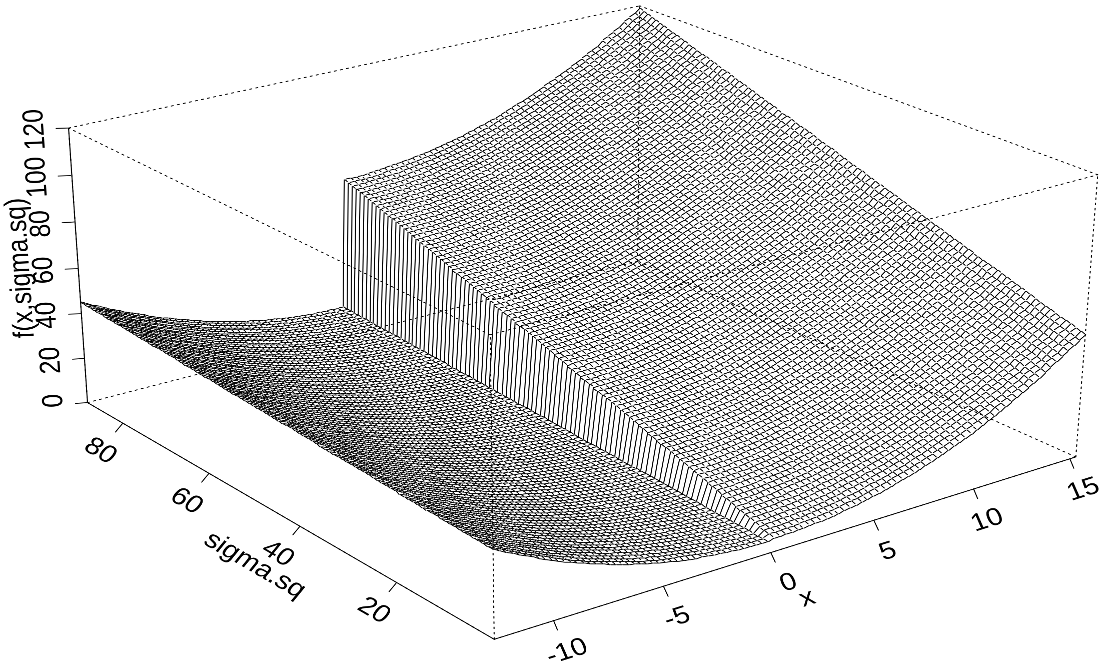

Figure 1: Simulation experiment. The volatility surface $f(x, \sigma^2) = 5 + 0.2x^2 + (0.75\mathbf{1}_{\{x>0\}} + 0.1\mathbf{1}_{\{x\leq 0\}})\sigma^2$; note the asymmetry depending on the sign of x.

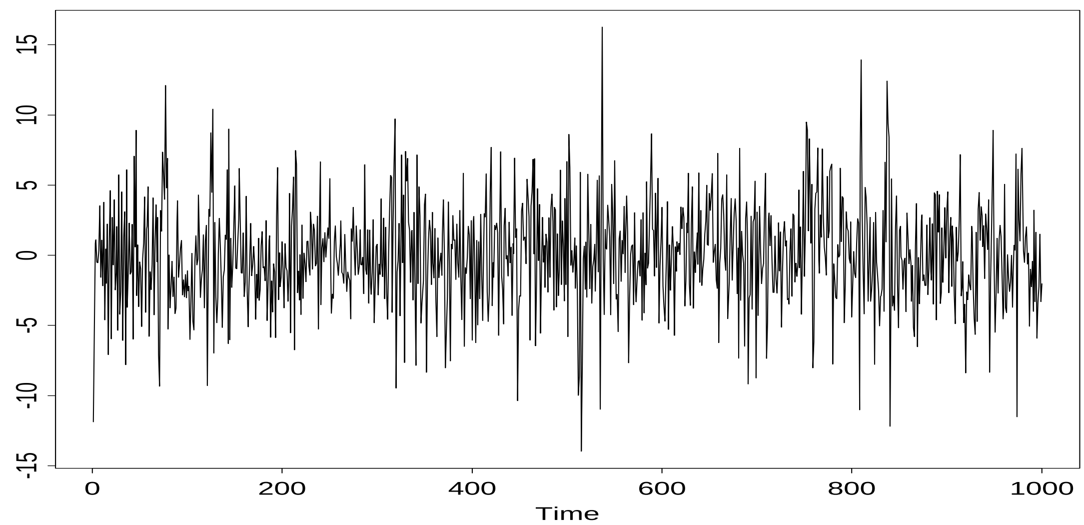

Figure 2: Simulation experiment. The data.

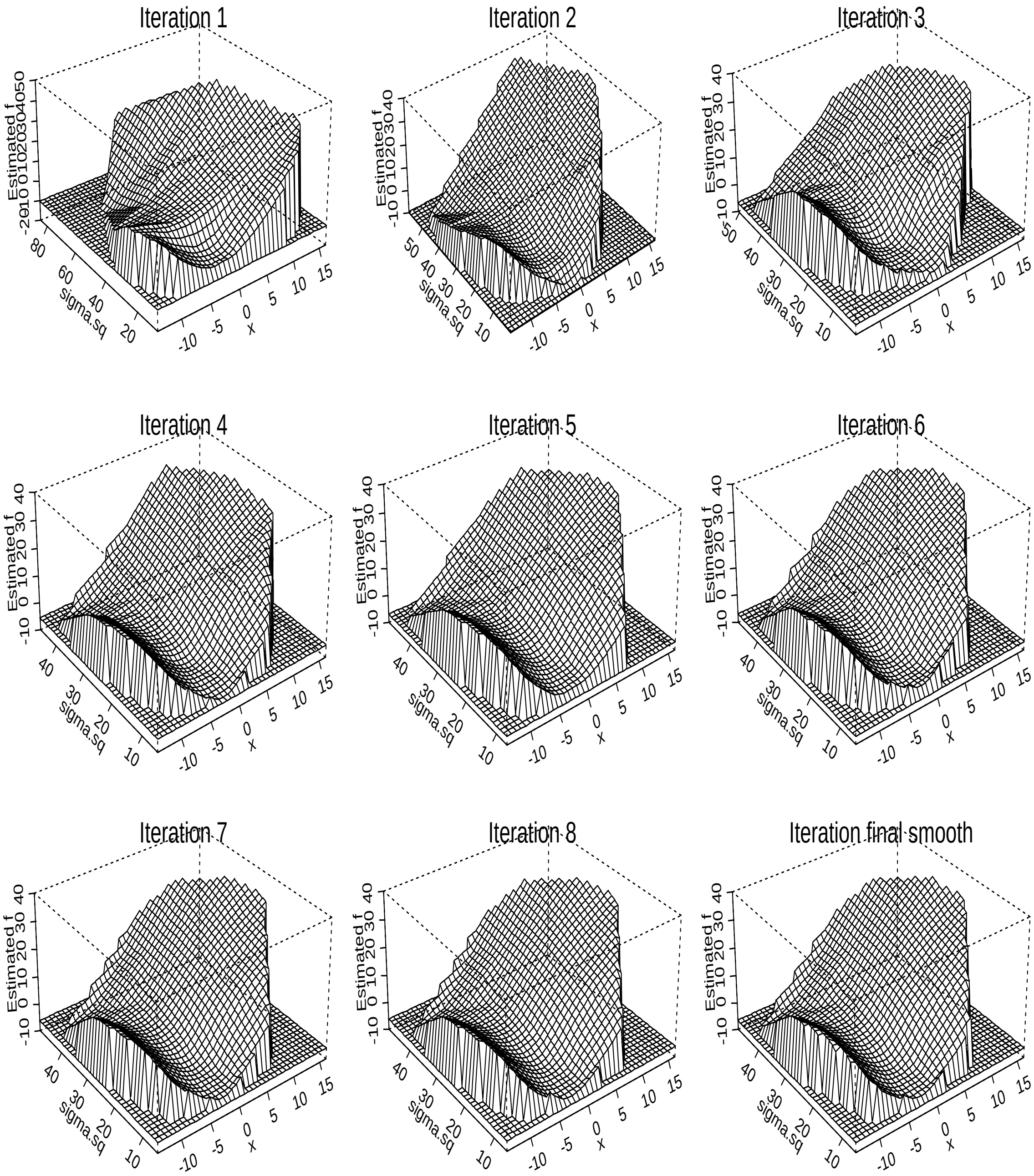

Figure 3: Simulation experiment. The estimated surfaces $\hat{f}_m$ calculated at the first 8 iterations and the surface $\hat{f}$ estimated at the final smoothing stage.

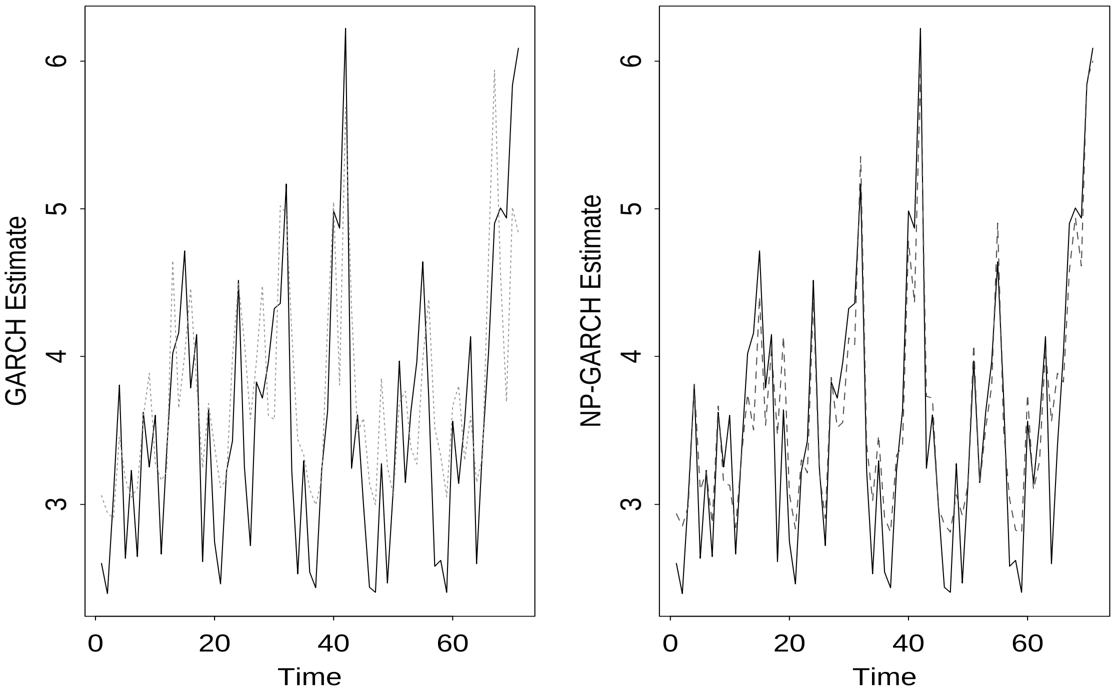

Figure 4: Simulation experiment. For an arbitrary section of 70 observations from the simulated time series, the left-hand plot shows true volatility (solid line) compared with a parametric GARCH(1,1) estimate (dotted line) and the right-hand plot shows true volatility compared with the nonparametric GARCH estimate obtained after a final smooth (dashed line).

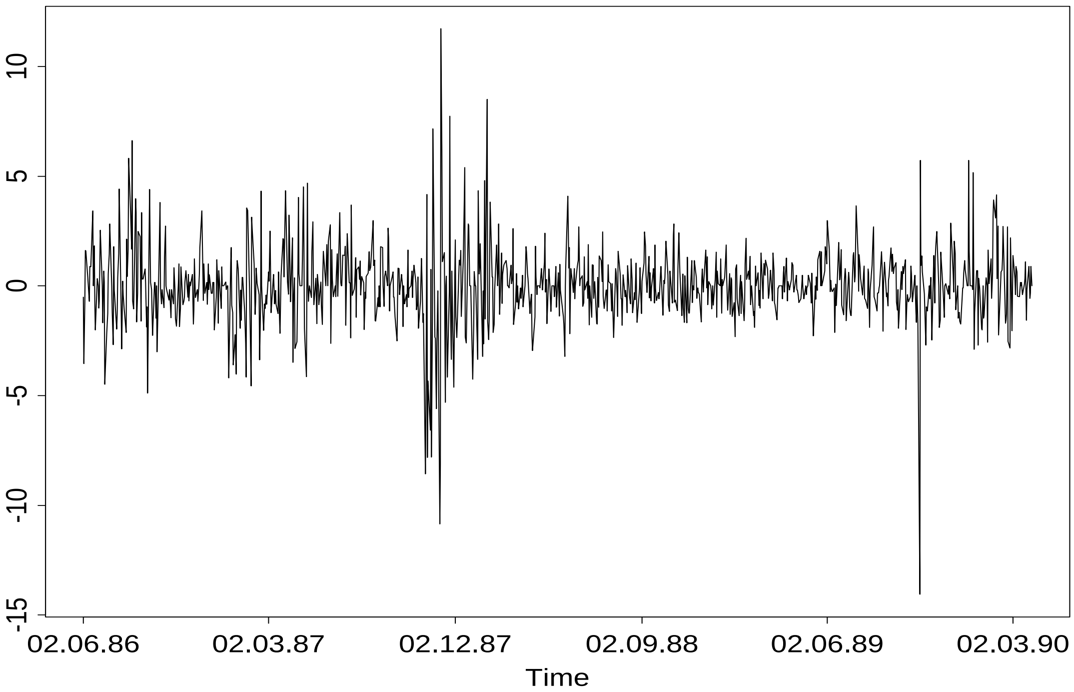

Figure 5: Real data. A 1000-day excerpt from the time series of daily returns on the BMW share price.

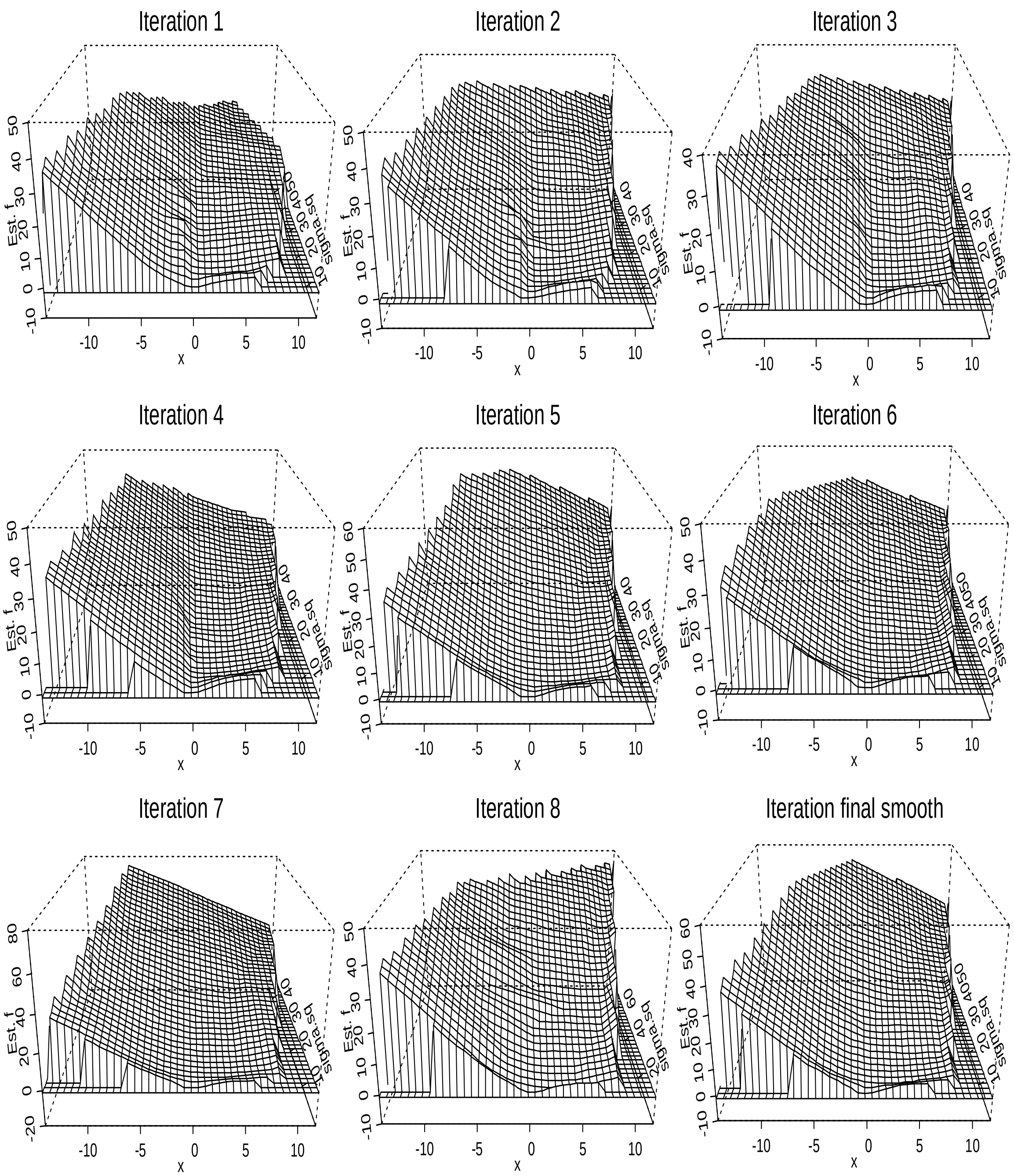

Figure 6: Real data. The estimated surfaces $\hat{f}_m$ calculated at the first 8 iterations and the surface $\hat{f}$ estimated at the final smoothing stage.

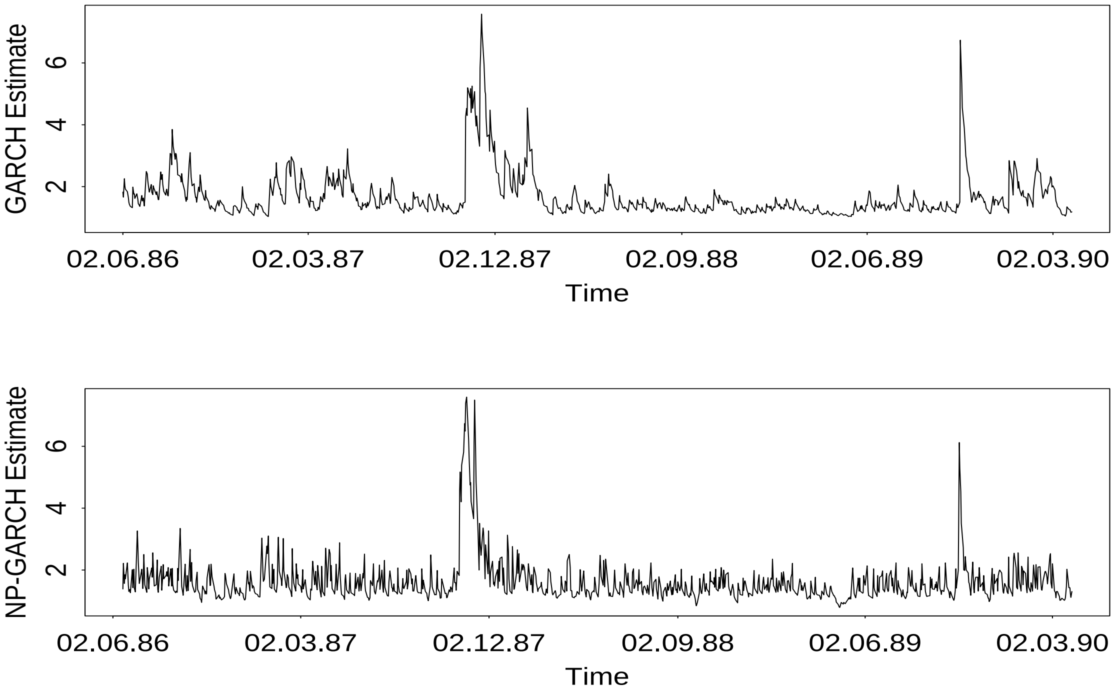

Figure 7: Real data. A comparison of volatilities estimated by ordinary parametric GARCH(1,1) modelling and nonparametric estimation (after the final smooth).