

# Simultaneously Modelling Conditional Heteroskedasticity and Scale Change

Yuanhua Feng

University of Konstanz

## Abstract

This paper proposes a semiparametric approach by introducing a smooth scale function into the standard GARCH model so that conditional heteroskedasticity and scale change in a financial time series can be modelled simultaneously. An estimation procedure combining kernel estimation of the scale function and maximum likelihood estimation of the GARCH parameters is proposed. Asymptotic properties of the kernel estimator are investigated in detail. An iterative plug-in algorithm is developed for selecting the bandwidth. Practical performance of the proposal is illustrated by simulation. The proposal is applied to the daily S&P 500 and DAX 100 returns. It is shown that there are simultaneously significant conditional heteroskedasticity and scale change in these series.

*JEL classification:* C22, C14

*Keywords:* Semiparametric GARCH, conditional heteroskedasticity, scale change, nonparametric regression with dependence, bandwidth selection.

## 1 Introduction

This paper considers modelling of heteroskedasticity in an equidistant financial time series, which is one of the most important and interesting themes of financial econometrics. There are at least two components, which result in heteroskedasticity, namely the well known *conditional heteroskedasticity* (CH) and a slowly changing unconditional variance (called scale change). The latter can be modelled by a slowly changing scale (or volatility) function. Two important differences between CH and scale change are: 1. the CH is determined by the past information, whereas the scale function depends only on the time  $t$ ; 2. a process with CH is under common conditions stationary but a process with scale change is no more covariance stationary but (at most) locally stationary.

Well known approaches for modelling CH are the ARCH (autoregressive conditional heteroskedastic, Engle, 1982) and GARCH (generalized ARCH, Bollerslev, 1986) models

together with numerous extensions. To our knowledge, scale change in a discrete financial time series is not yet investigated in detail. Beran and Ocker (2001) fitted SEMIFAR (semiparametric fractional autoregressive) models (Beran, 1999) to some volatility series defined by Ding et al. (1993) and found that sometimes there is a significantly deterministic trend in these series implying that the scale function of these series is no more constant. By checking some financial returns series we found that CH and scale change often appear simultaneously. This motivates us to propose a semiparametric GARCH (SEMIGARCH) model by introducing a smooth scale function  $\sigma(t)$  into the standard GARCH model (Bollerslev, 1986), which provides us a tool for simultaneously modelling CH and scale change.

It is proposed to estimate  $\sigma(t)$  by an approximate kernel smoother of the squared residuals. The parameters of the GARCH model are then estimated using (approximate) maximum likelihood approach. Asymptotic properties of the kernel estimator of  $\sigma(t)$  are investigated. The iterative plug-in idea introduced by Gasser et al. (1991) with some improvements proposed by Beran and Feng (2002a, b) is adapted to select the bandwidth in the current context. Practical performance of the proposal is at first illustrated by a simulation study and by detailed analysis of two simulated data sets. The proposal is then applied to the daily S&P 500 and DAX 100 returns. It is shown that, both the CH and the scale change in these time series are significant. By a fitted GARCH(1, 1) model one often obtains  $\hat{\alpha}_1 + \hat{\beta}_1 \approx 1$ . We found that an uneliminated nonconstant  $\sigma(t)$ , i.e. the covariance nonstationarity is an important reason for this phenomenon.

The paper is organized as follows. Section 2 introduces the model. The semiparametric estimation procedure is described in Section 3. Asymptotic properties of the kernel estimator  $\hat{\sigma}(t)$  are discussed in Section 4. Section 5 proposes the iterative plug-in algorithm. Results of the simulation study are reported in Section 6. The proposal is applied to the log-returns of the daily S&P 500 and DAX 100 indices in Section 7. Section 8 contains some final remarks. Proofs of results are put in the appendix.

## 2 The model

Consider the equidistant time series model

$$Y_i = \mu + \sigma(t_i)\epsilon_i, \tag{1}$$

where  $\mu$  is an unknown constant,  $t_i = t/n$ ,  $\sigma(t) > 0$  is a smooth, bounded scale (or volatility) function and  $\{\epsilon_i\}$  is assumed to be a GARCH( $r, s$ ) process defined by

$$\epsilon_i = \eta_i h_i^{\frac{1}{2}}, \quad h_i = \alpha_0 + \sum_{j=1}^r \alpha_j \epsilon_{i-j}^2 + \sum_{k=1}^s \beta_k h_{i-k} \quad (2)$$

(Bollerslev, 1986). The  $\eta_i$  are independent and identically distributed (i.i.d.)  $N(0, 1)$  random variables,  $\alpha_0 > 0$ ,  $\alpha_1, \dots, \alpha_r \geq 0$  and  $\beta_1, \dots, \beta_s \geq 0$ . Let  $v(t) = \sigma^2(t)$  denote the local variance of  $Y$ . Let  $\alpha = (\alpha_1, \dots, \alpha_r)'$ ,  $\beta = (\beta_1, \dots, \beta_s)'$  and  $\theta = (\alpha_0, \alpha', \beta) = (\alpha_0, \alpha_1, \dots, \alpha_r, \beta_1, \dots, \beta_s)'$  be the unknown parameter vectors. It is assumed that there is a strictly stationary solution of (2) such that  $E(\epsilon_i^8) < \infty$ . This condition is required for the practical implementation of a nonparametric estimator of  $v(t)$ . Necessary and sufficient conditions which guarantee the existence of high order moments of a GARCH process may be found in Ling and Li (1997), Ling (1999) and Ling and McAleer (2002). Note that  $E(\epsilon_i^8) < \infty$  implies in particular that  $\sum_{i=1}^r \alpha_i + \sum_{j=1}^s \beta_j < 1$ . Furthermore, it is convenient to assume that  $\text{var}(\epsilon_i) = E(\epsilon_i^2) = 1$  implying  $\alpha_0 = 1 - \sum_{i=1}^r \alpha_i - \sum_{j=1}^s \beta_j$ .

Model (1) and (2) defines a semiparametric, locally stationary GARCH model by introducing the scale function  $\sigma(t)$  into the standard GARCH model, where  $h_i^{1/2}$  stand for the conditional standard deviations of the standardized process  $\epsilon_i$ . The total standard deviation at  $t_i$  is hence given by  $\sigma(t_i)h_i^{1/2}$ . Our purpose is to estimate  $v(t)$  and  $h_i$  separately. For  $\sigma(t) \equiv \sigma_0$ , model (1) and (2) reduces to the standard GARCH model. If the scale function  $\sigma(t)$  in (1) changes over time, then the assumption of a standard GARCH model is a misspecification. In this case the estimation of the GARCH model will be inconsistent. It can be shown through simulation that, if a non-constant scale function is not eliminated, one will obtain  $\hat{\alpha}_1 + \hat{\beta}_1 \rightarrow 1$  by a fitted GARCH(1, 1) model as  $n \rightarrow \infty$ , even when  $\epsilon_i$  are i.i.d. innovations. Furthermore, in the presence of scale change the estimation of  $v(t)$  is also necessary for the prediction. On the other hand, if  $Y_i$  follows a pure GARCH model but model (1) and (2) is used, then the estimation is still consistent but with some loss in efficiency due to the estimation of  $\sigma(t)$ .

The assumptions of model (1) and (2) are made for simplicity, which can be weakened in different ways. For instance, if the constant mean  $\mu$  in (1) is replaced by a smooth mean function  $g$ , then we obtain the following nonparametric regression with scale change and dependence

$$Y_i = g(t_i) + \sigma(t_i)\epsilon_i, \quad (3)$$

where  $\{\epsilon_i\}$  is a zero mean stationary process. Estimation of the mean functions  $g$  in model (3) with i.i.d.  $\epsilon_i$  was discussed e.g. in Ruppert and Wand (1994), Fan and Gijbels

(1995) and Efromovich (1999). Discussion on the estimation of the scale function in heteroskedastic nonparametric regression may be found e.g. in Efromovich (1999). The focus of this paper is to investigate the estimation of  $\sigma(t)$  under model (1) and (2) in detail. And, we will see that the model we need to estimate  $\sigma(t)$  (or  $v(t)$ ) is a special case of model (3).

### 3 A semiparametric estimation procedure

Model (1) and (2) can be estimated by a semiparametric procedure combining nonparametric estimation of  $v(t)$  and parametric estimation of  $\theta$ . A linear smoother of the squared residuals will estimate  $v(t)$ . Let  $Z_i = (Y_i - \mu)$ , model (1) can be rewritten as follows

$$X_i = v(t_i) + v(t_i)\xi_i, \quad (4)$$

where  $X_i = Z_i^2$  and  $\xi_i = \epsilon_i^2 - 1 \geq -1$  are zero mean stationary time series errors. Model (4) transfers the estimation of the scale function to a general nonparametric regression problem (see Section 4.3 of Efromovich, 1999 for related idea). On the one hand, model (4) is a special case of (3) with  $g(t)$  and  $\sigma(t)$  both being replaced by  $v(t)$ . On the other hand, model (4) also applies to (3) by defining  $Z_i = Y_i - g(t_i)$ . Hence, the extension of our results to model (3) is expected.

In the following a kernel estimator of conditional variance proposed by Feng and Heiler (1998) will be adapted to estimate  $v(t)$ . Let  $y_1, \dots, y_n$ , denote the observations. Let  $\hat{\mu} = \bar{y}$ ,  $\hat{z}_i = y_i - \bar{y}$  and  $\hat{x}_i = \hat{z}_i^2$ . Let  $K(u)$  denote a second order kernel with compact support  $[-1, 1]$ . An *approximate* Nadaraya-Watson estimator of  $v$  at  $t$  is defined by

$$\hat{v}(t) = \frac{\sum_{i=1}^n K(\frac{t_i-t}{b})\hat{x}_i}{\sum_{i=1}^n K(\frac{t_i-t}{b})} =: \sum_{i=1}^n w_i \hat{x}_i, \quad (5)$$

where  $w_i = K(\frac{t_i-t}{b})(\sum_{i=1}^n K(\frac{t_i-t}{b}))^{-1}$  and  $b$  is the bandwidth. And we define  $\hat{\sigma}(t) = \sqrt{\hat{v}}$ . The definition given in (5) does not depend on the dependence structure of the errors, because  $\hat{v}$  is a linear smoother. It is clear that  $\hat{v} > 0$  as far as all observations such that  $|t_i - t| \leq b$  are not identically. The bias of  $\hat{v}$  at a boundary point is of a larger order than that in the interior due to the asymmetry in the observations. This is the so-called boundary effect of the kernel estimator, which can be overcome by using a local linear estimator (see e.g. Härdle et al., 1998). However, as mentioned in Feng and Heiler (1998),

a local linear estimator of  $v$  may sometimes be non-positive. Hence, the kernel estimator is more preferable in the current context.

Furthermore, note that

$$\epsilon_i = z_i/\sigma(t_i). \quad (6)$$

The parameter vector  $\theta$  may be estimated by standard maximum likelihood method (Bollerslev, 1986) with  $\epsilon_i$  being replaced by the standardized residuals

$$\hat{\epsilon}_i = \hat{z}_i/\hat{\sigma}(t_i) = (y_i - \bar{y})/\hat{\sigma}(t_i). \quad (7)$$

$\hat{\theta}$  obtained in this way is also an *approximate* maximum likelihood estimator. Any standard GARCH software can be built in this semiparametric estimation procedure. In this paper the S+GARCH (Martin et al., 1996) will be used. For a given bandwidth, the proposed procedure can already be carried out, e.g. in S-Plus, as follows:

1. Calculate  $\hat{\mu} = \bar{y}$ ;
2. Estimate  $v$  using the S-plus function *ksmooth* with input variables  $\mathbf{t} = (1/n, \dots, 1)'$ ,  $\mathbf{x} = ((y_1 - \bar{y})^2, \dots, (y_n - \bar{y})^2)'$ , a bandwidth  $b$  and a selected built-in kernel function;
3. Obtain  $\hat{h}_i$  by fitting a GARCH model to the series  $((y_1 - \bar{y})/\hat{\sigma}(t_1), \dots, (y_n - \bar{y})/\hat{\sigma}(t_n))$ ;
4. Carry out further predictions with  $\hat{v}(t_i)$  and  $\hat{h}_i$ .

Similar to the results on the approximate maximum likelihood estimators in the SEMI-FAR model (see Beran, 1999 and Beran and Feng, 2002c), it is expected that  $\hat{\theta}$  proposed here is still  $\sqrt{n}$ -consistent as in the parametric case. The simulation results in Section 6 confirm this. However, this will not be investigated here. In the following we will focus on discussing the asymptotic properties of  $\hat{v}$  and developing a data-driven algorithm for the practical implementation of the proposed procedure.

## 4 Asymptotic properties of $\hat{v}$

For the derivation of the asymptotic results the following assumptions are required.

- A1. Model (1) and (2) holds with i.i.d.  $N(0, 1)$   $\eta_i$  and strictly stationary  $\epsilon_i$  such that  $E(\epsilon_i^8) < \infty$ .

A2. The function  $v(t)$  is strictly positive on  $[0, 1]$  and is at least twice continuously differentiable.

A3. The kernel  $K(u)$  is a symmetric density function defined on  $[-1, 1]$ .

A4. The bandwidth  $b$  satisfies  $b \rightarrow 0$  and  $nb \rightarrow \infty$  as  $n \rightarrow \infty$ .

Equation (4) is a nonparametric regression model with a local stationary error process. Results in nonparametric regression with dependence may be found e.g. in Altman (1990) and Hart (1991) for short-range dependent errors and Hall and Hart (1990), Beran (1999) and Beran and Feng (2002c) for long-range dependent errors. The pointwise results obtained in these works may be adapted to the current case without any difficulty. Let  $\gamma_\xi(k)$  denote the autocovariance function of  $\xi_i$ . It is well known that  $\text{var}(\hat{v})$  depends on  $c_f = f(0)$ , where  $f(\lambda) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} \exp(ik\lambda) \gamma_\xi(k)$  is the spectral density of  $\xi_i$ . Let  $r' = \max(r, s)$ . Following equations (6) and (7) in Bollerslev (1986) and observing that  $\alpha_0 = 1 - \sum_{i=1}^r \alpha_i - \sum_{j=1}^s \beta_j$ , we have the ARMA( $r', s$ ) representation of  $\xi_i$ :

$$\xi_i = \sum_{j=1}^{r'} \alpha'_j \xi_{i-j} - \sum_{k=1}^s \beta_k u_{i-k} + u_i, \quad (8)$$

where  $\alpha'_j = \alpha_j + \beta_j$  for  $j \leq \min(r, s)$ ,  $\alpha'_j = \alpha_j$  for  $j, r > s$  and  $\alpha'_j = \beta_j$  for  $j, s > r$  and

$$u_i = \epsilon_i^2 - h_i = (\eta_i^2 - 1)h_i \quad (9)$$

is a sequence of zero mean, uncorrelated random variables with independent  $\eta_i \sim N(0, 1)$ . Equations (8) and (9) allow us to calculate  $c_f$ .

Define  $R(K) = \int K^2(u)du$  and  $I(K) = \int u^2 K(u)du$ . At an interior point  $0 < t < 1$  the following results hold.

**Theorem 1.** *Under assumptions A1 to A4 we have*

*i) The bias of  $\hat{v}(t)$  is given by*

$$E[\hat{v}(t) - v(t)] = \frac{I(K)v''(t)}{2}b^2 + o(b^2). \quad (10)$$

*ii) The variance of  $\hat{v}(t)$  is given by*

$$\text{var}[\hat{v}(t)] = 2\pi c_f R(K) \frac{v^2(t)}{nb} + o\left(\frac{1}{nb}\right). \quad (11)$$

iii) Assume that  $nb^5 \rightarrow d^2$  as  $n \rightarrow \infty$ , for some  $d > 0$ , then

$$(nb)^{1/2}(\hat{v}(t) - v(t)) \xrightarrow{D} N(dD, V(t)), \quad (12)$$

where  $D = I(K)v''(t)/2$  and  $V(t) = 2\pi c_f R(K)v^2(t)$ .

The proof of Theorem 1 is given in the appendix. The asymptotic bias of  $\hat{v}$  is the same as in nonparametric regression with i.i.d. errors. The asymptotic variance of it is similar to that in nonparametric regression with short-range dependent errors, which depends however on the unknown underlying function  $v$  itself. The asymptotic normality of  $\hat{v}(t)$  allows us to test, if there is significant scale change in a time series.

Let  $\phi(z) = 1 - \sum_{i=1}^{r'} \alpha'_i z^i$  and  $\psi(z) = 1 - \sum_{j=1}^s \beta_j z^j$ . Assume further that

A5. The polynomials  $\phi(z)$  and  $\psi(z)$  have no common roots.

Assumption A5 implies in particular that  $\alpha \neq \mathbf{0}$ . For a GARCH(1, 1) model A5 is equivalent to the condition  $\alpha_1 > 0$ . Under assumptions A1 and A5 we have

$$c_f = \frac{E(\epsilon_i^4) |\psi(1)|^2}{3\pi |\phi(1)|^2} = \frac{E(\epsilon_i^4)}{3\pi} \frac{(1 - \sum_{j=1}^s \beta_j)^2}{(1 - \sum_{i=1}^r \alpha_i - \sum_{j=1}^s \beta_j)^2}. \quad (13)$$

If  $\epsilon_i$  follows a GARCH(1, 1) model, then we have

$$\begin{aligned} c_f &= \frac{1}{\pi} \frac{\alpha_0^2 (1 + \alpha_1 + \beta_1)(1 - \beta_1)^2}{(1 - \alpha_1 - \beta_1)^3 (1 - 3\alpha_1^2 - 2\alpha_1\beta_1 - \beta_1^2)} \\ &= \frac{1}{\pi} \frac{(1 + \alpha_1 + \beta_1)(1 - \beta_1)^2}{\alpha_0 (1 - 3\alpha_1^2 - 2\alpha_1\beta_1 - \beta_1^2)}. \end{aligned} \quad (14)$$

The last equation in (14) is due to the standardization of  $\epsilon_i$ . The proof of (13) and (14) is given in the appendix.

The mean integrated squared error (MISE) defined on  $[\Delta, 1 - \Delta]$  will be used as a goodness of fit criterion, where  $\Delta > 0$  is used to avoid the boundary effect of  $\hat{v}$ . Define  $I((v'')^2) = \int_{\Delta}^{1-\Delta} (v''(t))^2 dt$  and  $I(v^2) = \int_{\Delta}^{1-\Delta} v^2(t) dt$ . The following theorem holds.

**Theorem 2.** *Under the assumptions of Theorem 1 we have*

i) *The MISE of  $\hat{v}(t)$  is*

$$\begin{aligned} \text{MISE} &= \int_{\Delta}^{1-\Delta} E[\hat{v}(t) - v(t)]^2 dt \\ &= \frac{I^2(K)I((v'')^2)}{4} b^4 + 2\pi c_f R(K) \frac{I(v^2)}{nb} + o[\max(b^4, \frac{1}{nb})]. \end{aligned} \quad (15)$$

ii) Assume that  $I((v'')^2) \neq 0$ . The asymptotically optimal bandwidth for estimating  $v$ , which minimizes the dominant part of the MISE is given by

$$b_A = C_A n^{-1/5} \quad (16)$$

with

$$C_A = \left( 2\pi c_f \frac{R(K)}{I^2(K)} \frac{I(v^2)}{I((v'')^2)} \right)^{1/5}. \quad (17)$$

The proof of Theorem 2 is straightforward and is omitted. If a bandwidth  $b = O(b_A) = O(n^{-1/5})$  is used, we have  $\hat{v}(t) = v(t)[1 + O_p(n^{-2/5})]$  and  $\text{MISE} = O(n^{-4/5})$ .

## 5 The proposed data-driven algorithm

A plug-in bandwidth selector may be developed by replacing the unknowns  $c_f$ ,  $I(v^2)$  and  $I((v'')^2)$  in (17) with some suitable estimators. At first, it is proposed to estimate  $c_f$  by

$$\hat{c}_f = \frac{\hat{E}(\epsilon_i^4)}{3\pi} \frac{(1 - \sum_{j=1}^s \hat{\beta}_j)^2}{(1 - \sum_{i=1}^r \hat{\alpha}_i - \sum_{j=1}^s \hat{\beta}_j)^2}, \quad (18)$$

where  $\hat{E}(\epsilon_i^4) = \sum_{i=1}^n \hat{\epsilon}_i^4/n$  is a nonparametric estimator of  $E(\epsilon_i^4)$ . Although explicit formulae of  $E(\epsilon_i^4)$  are known (see He and Teräsvirta, 1999a and Karanasos, 1999 for common results and Bollerslev, 1986 and He and Teräsvirta, 1999b for results in some special cases), we prefer to use  $\hat{c}_f$  defined in (18), since the formulae of  $E(\epsilon_i^4)$  are in general too complex. For a GARCH(1, 1) model, another simple estimator,  $\tilde{c}_f$  say, may be defined based on (14) by replacing  $\alpha_0$ ,  $\alpha_1$  and  $\beta_1$  with their estimates. Now  $\hat{c}_f$  and  $\tilde{c}_f$  perform quite similarly. Assume that a bandwidth  $b_\epsilon$  is used for estimating  $E(\epsilon_i^4)$ , which satisfies A4 but is not necessarily the same as  $b$ , then the following holds

**Proposition 1.** *Under the assumptions of Theorem 1 we have*

$$E[\hat{E}(\epsilon_i^4) - E(\epsilon_i^4)] \doteq O(b_\epsilon^2) + O(nb_\epsilon)^{-1} \quad (19)$$

and

$$\text{var}(\hat{E}(\epsilon_i^4)) \doteq 2\pi c_f^\epsilon n^{-1}[1 + o(1)], \quad (20)$$

where  $c_f^\epsilon$  denotes the value of the spectral density of the process  $\epsilon_i^4$  at the origin.



The proof of Proposition 1 is given in the appendix.

**Remark 1.** Equations (19) and (20) show that  $\hat{E}(\epsilon_i^4)$  is  $\sqrt{n}$ -consistent, if  $O(n^{-1/2}) \leq b_\epsilon \leq O(n^{-1/4})$ . The optimal bandwidth in a second order sense, which balances the two terms on the right hand side of (19) is of order  $O(n^{-1/3})$ . In this paper, we propose to use a bandwidth  $b_\epsilon = O(n^{-1/4})$  for estimating  $E(\epsilon_i^4)$  so that the estimator is more stable. Note that  $\hat{E}(\epsilon_i^4)$  is no more  $\sqrt{n}$ -consistent, if a bandwidth  $b_\epsilon = O(b_\Delta) = O(n^{-1/5})$  is used. However, it can be shown that,  $\hat{b}$  is not so sensitive to the bandwidth for estimating  $E(\epsilon_i^4)$ .

The integral  $I(v^2)$  can be estimated by

$$\hat{I}(v^2) = \frac{1}{n} \sum_{i=n_1}^{n_2} \hat{v}(t_i)^2, \quad (21)$$

where  $n_1$  and  $n_2$  denote the integer parts of  $n\Delta$  and  $n(1 - \Delta)$ , respectively,  $\hat{v}$  is the same as defined in (5) but obtained with another bandwidth  $b_v$ , say, which satisfies A4. The following results hold for  $\hat{I}(v^2)$ .

**Proposition 2.** *Under the assumptions of Theorem 1 we have*

$$E[\hat{I}(v^2) - I(v^2)] \doteq O(b_v^2) + O(nb_v)^{-1} \quad (22)$$

and

$$\text{var}(\hat{I}(v^2)) \doteq O(n^{-1}) + O(n^{-2}h^{-1}). \quad (23)$$

The proof of Proposition 2 is given in the appendix.

**Remark 2.** Note that the dominated orders of the bias and variances of  $\hat{E}(\epsilon_i^4)$  and  $\hat{I}(v^2)$  are the same. Hence similar statements given in Remark 1 apply for results given in (22) and (23). This is not surprising, since both  $v^2(t_i)$  and  $\epsilon_i^4$  are related to the fourth moment of the errors.

A well known estimator of  $I((v'')^2)$  is given by

$$\hat{I}((v'')^2) = \frac{1}{n} \sum_{i=n_1}^{n_2} \hat{v}''(t_i)^2 \quad (24)$$

(see e.g. Ruppert et al., 1995), where  $\hat{v}''$  is a kernel estimator of  $v''$  using a fourth order kernel for estimating the second derivative (see e.g. Gasser and Müller, 1984 and Müller, 1988) and again another bandwidth  $b_d$ . Corresponding results as given in Proposition 2 hold for  $\hat{I}((v'')^2)$ , for which the following adapted assumptions are required.

A2'. The function  $v(t)$  is strictly positive on  $[0, 1]$  and is at least four times continuously differentiable.

A3'.  $v''$  is estimated with a symmetric fourth order kernel for estimating the second derivative with compacted support  $[-1, 1]$ .

A4'. The bandwidth  $b_d$  satisfies  $b_d \rightarrow 0$  and  $nb_d^5 \rightarrow \infty$  as  $n \rightarrow \infty$ .

**Proposition 3.** *Under assumptions A1 and A2' to A4' we have*

$$E[\hat{I}((v'')^2) - I((v'')^2)] \doteq O(b_d^2) + O(n^{-1}b_d^{-5}) \quad (25)$$

and

$$\text{var}(\hat{I}((v'')^2)) \doteq O(n^{-1}) + O(n^{-2}b^{-5}). \quad (26)$$

The proof of Proposition 3 is omitted, since it is well known in nonparametric regression (see e.g. Herrmann and Gasser, 1994 and Ruppert et al., 1995 for results with i.i.d. errors and Beran and Feng, 2002a, b for results with dependent errors).

**Remark 3.** The MSE (mean squared error) of  $\hat{I}((v'')^2)$  is dominated by the squared bias. The optimal bandwidth for estimating  $I((v'')^2)$ , which balances the two terms on the right hand side of (25), is of order  $O(n^{-1/7})$ . With a bandwidth  $b_d = O(n^{-1/7})$  we have  $\hat{I}((v'')^2) - I((v'')^2) \doteq O_p(n^{-2/7})$ .

We see, for selecting the bandwidth  $b$  we have to choose at first three pilot bandwidths  $b_\epsilon$ ,  $b_v$  and  $b_d$ . This problem will be solved using the iterative plug-in idea (Gasser et al., 1991) with a so-called exponential inflation method (see Beran and Feng, 2002a, b). Let  $b_{j-1}$  denote the bandwidth for estimating  $v$  in the  $(j-1)$ -th iteration. Then in the  $j$ -th iteration, the bandwidths  $b_{\epsilon,j} = b_{v,j} = b_{j-1}^{5/4}$  and  $b_{d,j} = b_{j-1}^{5/7}$  will be used for estimating  $E(\epsilon^4)$ ,  $I(v^2)$  and  $I((v'')^2)$ , respectively. These inflation methods are chosen so that  $b_{\epsilon,j}$  as well as  $b_{v,j}$  are both of order  $O_p(n^{-1/4})$  and  $b_{d,j}$  is of the optimal order  $O_p(n^{-1/7})$ , when  $b_{j-1}$  is of the optimal order  $O_p(n^{-1/5})$ . The unknown constants in the pilot bandwidths are all omitted. By an iterative plug-in algorithm we also need to choose a starting bandwidth  $b_0$ . In the current context,  $b_0$  should satisfy A4, because we have to estimate  $\theta$  in the first iteration. Theoretically, a bandwidth  $b_0 = O(n^{-1/5})$  is more preferable. Our experience shows that  $b_0 = 0.5n^{-1/5}$  is a good choice. Detailed discusses on this may be found in the next two sections, especially in Section 6.3.

The proposed data-driven algorithm processes as follows:

1. Starting with the bandwidth  $b_0 = c_0 n^{-1/5}$  with e.g.  $c_0 = 0.5$ .
2. In the  $j$ -th iteration
  - a) Calculate  $\hat{v}$  and  $\hat{\theta}$  using the bandwidth  $b_{j-1}$ .
  - b) Calculate  $\hat{E}(\epsilon^4)$  and  $\hat{I}(v^2)$  with  $\hat{v}$  obtained using the bandwidth  $b_{\epsilon,j} = b_{v,j} = b_{j-1}^{5/4}$ .
  - c) Calculate  $\hat{c}_f$  from  $\hat{\theta}$  and  $\hat{E}(\epsilon^4)$ .
  - d) Calculate  $\hat{I}((v'')^2)$  with  $\hat{v}''$  obtained using the bandwidth  $b_{d,j} = b_{j-1}^{5/7}$ .
  - e) Improve  $b_{j-1}$  by

$$b_j = \left( 2\pi\hat{c}_f \frac{R(K)}{I^2(K)} \frac{\hat{I}(v^2)}{\hat{I}((v'')^2)} \right)^{1/5} n^{-1/5}. \quad (27)$$

3. Increase  $j$  by one and repeatedly carry out Step 2 until convergence is reached or until a given maximal number of iterations has been done. Put  $\hat{b} = b_j$ .

The condition  $|b_j - b_{j-1}| < 1/n$  is used as a criterion for the convergence of  $\hat{b}$ , since such a difference is negligible. The maximal number of iterations is put to be twenty. The asymptotic performance of  $\hat{b}$  is quantified by

**Theorem 3.** *Assume that A1, A3, A5, A2' and A3' hold and that  $I((v'')^2) \neq 0$  we have*

$$(\hat{b} - b_A)/b_A \doteq O_p(n^{-2/7}) + O_p(n^{-1/2}). \quad (28)$$

The proof of Theorem 3 is given in the appendix. Note that A4 and A4' are automatically satisfied. The  $O_p(n^{-1/2})$  term in (28) is due to the estimation of  $c_f$  and  $I(v^2)$ , where it is assumed the  $\hat{\alpha}_1$  and  $\hat{\beta}_1$  are  $\sqrt{n}$ -consistent.

The proposed algorithm is coded in an S-Plus function called *SEMIGARCH*. A practical restriction  $1/n \leq b \leq 0.5 - 1/n$  is used in the program for simplicity. Four commonly used kernels, namely the Uniform, the Epanechnikov, the Bisquare and the Triweihgt kernels (see e.g. Müller, 1988) are built in the program. As a standard version we propose the use of the Epanechnikov kernel with  $\Delta = 0.05$  and  $c_0 = 0.5$ , which will be used in the following two sections.

**Remark 4.** Note that  $b_A$  is not well defined, if  $I((v'')^2) = 0$  implying  $v''(t) \equiv 0$ . However, the proposed algorithm also applies to this case. In particular, the SEMIGARCH model does work, even if the underlying model is a standard GARCH model. It can be shown that  $\hat{v}$  in this case is still  $\sqrt{n}$ -consistent but with some loss in the efficiency compared to a parametric estimator.

## 6 The simulation study

### 6.1 Design of the simulation

To show the practical performance of our proposal, a simulation study was carried out. In the simulation study,  $\epsilon_i$  were generated using the *simulate.garch* function in S+GARCH following one of the two GARCH(1, 1) models:

$$\text{Model 1 (M1). } \epsilon_i = \eta_i h_i^{\frac{1}{2}}, \quad h_i = 0.6 + 0.2\epsilon_{i-1}^2 + 0.2h_{i-1} \text{ and}$$

$$\text{Model 2 (M2). } \epsilon_i = \eta_i h_i^{\frac{1}{2}}, \quad h_i = 0.15 + 0.1\epsilon_{i-1}^2 + 0.75h_{i-1}.$$

$y_i$  are generated following model (1) with  $\mu \equiv 0$  and one of the tree scale functions:

$$v_1^{1/2}(t) = \sigma_1(t) = 3.75 + t + (3 \cos(2.75(t - 0.5)\pi) + 22.5 + 2 \tanh(2.75(t - 0.5)\pi))/5,$$

$$v_2^{1/2}(t) = \sigma_2(t) = \sigma_1(t) - 1.2 \text{ and}$$

$$v_3^{1/2}(t) = \sigma_3(t) = 3 + \cos(4(t - 0.25)\pi).$$

$v_1(t)$  and  $v_2(t)$  are quite similar, which are designed following the estimated scale function in the daily DAX 100 returns (see Figure 5(b) in the next section). The scale change with  $v_2$  is stronger than that with  $v_1$ . It is most strong with  $v_3$ . To this end see the bandwidths required for estimating them given in Table 5. The two scale functions  $\sigma_2(t)$  and  $\sigma_3(t)$  may be found in Figures 2(b) and 3(b) in Section 6.3. To confirm the statements in Remark 4, a constant scale function  $v_0(t) = \sigma_0^2(t) \equiv 16$  is also used. The simulation was carried out for three sample sizes  $n = 1000, 2000, 4000$ . For each case 400 replications were done. For each replication, three GARCH (1, 1) models were fitted to  $\epsilon_i, y_i$  and the data-driven  $\hat{\epsilon}_i$ . The estimators of  $\alpha_1$  and  $\beta_1$  are denoted by  $\hat{\alpha}_1^\epsilon, \hat{\beta}_1^\epsilon, \hat{\alpha}_1^y, \hat{\beta}_1^y, \hat{\alpha}_1^{\hat{\epsilon}}, \hat{\beta}_1^{\hat{\epsilon}}$  and  $\hat{\beta}_1^{\hat{\epsilon}}$ , respectively. Here,  $\hat{\alpha}_1^\epsilon, \hat{\beta}_1^\epsilon$  are used as a benchmark.

## 6.2 Results of the simulation study

The sample means, standard deviations and square roots of the MSE's of these estimators in 400 replications are listed in Tables 1 to 3. Note that  $y_i \equiv 4\epsilon_i$  for  $v_0$ . In this case we have  $\hat{\alpha}_1^\epsilon = \hat{\alpha}_1^y$  and  $\hat{\beta}_1^\epsilon = \hat{\beta}_1^y$  for any replication. Hence, results for  $\hat{\alpha}_1^y$  and  $\hat{\beta}_1^y$  with  $v_0$  are omitted.

Consider at first the results on  $\hat{\alpha}_1^\epsilon$  or  $\hat{\beta}_1^\epsilon$ . From Tables 1 to 3 we see that the variances of these estimators converge very fast in all cases. In some cases with small bias, the bias happens to be slightly larger for a larger  $n$  than for a smaller. We think this is due to the randomness. The MSE's of these two estimators seem to be dominated by their variances and converge hence also very fast. For given  $n$ , the MSE's of  $\hat{\alpha}_1^\epsilon$  and  $\hat{\beta}_1^\epsilon$  under M2 are much smaller than those under M1. The difference among the MSE's for the four scale functions in a given case is not clear. By comparing the MSE's for different  $n$  we can find that these estimators seem to be all  $\sqrt{n}$ -consistent. Furthermore, for a given case the MSE of  $\hat{\alpha}_1^\epsilon$  is clearly smaller than that of  $\hat{\beta}_1^\epsilon$ , this means that  $\alpha$  in a SEMIGARCH model is easier to estimate than  $\beta$ .

Results on  $\hat{\alpha}_1^y$  and  $\hat{\beta}_1^y$  show how the estimated parameters perform, if a nonconstant scale function is not eliminated. We see, although the variances of  $\hat{\alpha}_1^y$  and  $\hat{\beta}_1^y$  converge very fast, the MSE's of them do not due to the biases. The MSE's of  $\hat{\alpha}_1^y$  for different  $n$  are about the same. In general, the MSE's of  $\hat{\beta}_1^y$  increases as  $n$  increases, since, as expected, the bias of  $\hat{\beta}_1^y$  increases as  $n$  increases. Observe in particular that  $\hat{\alpha}_1^y + \hat{\beta}_1^y \approx 1$ , even for M1 with  $\alpha_1 + \beta_1 = 0.4$ . For example for  $n = 4000$ , the smallest value of the mean of  $\hat{\alpha}_1^y + \hat{\beta}_1^y$  is 0.883 and the largest 0.997 for M1 with  $v_1$  and  $v_2$ , respectively.

To give a summary of the performance of  $\hat{\alpha}_1^\epsilon$  and  $\hat{\beta}_1^\epsilon$  and to compare them with  $\hat{\alpha}_1^y$  and  $\hat{\beta}_1^y$ , the empirical efficiency (EFF) of an estimator w.r.t. the corresponding one estimated from  $\epsilon_i$  is calculated. For instance,

$$\text{EFF}(\hat{\beta}_1^\epsilon) := \frac{\text{MSE}(\hat{\beta}_1^\epsilon)}{\text{MSE}(\hat{\beta}_1^y)} \times 100\%.$$

These results are listed in Table 4. The difference between two related EFF's, e.g.  $\text{EFF}(\hat{\beta}_1^\epsilon) - \text{EFF}(\hat{\beta}_1^y)$ , in a given case may be thought of as the gain by using the SEMIGARCH model. Table 4 shows that the EFF's of  $\hat{\alpha}_1^\epsilon$  and  $\hat{\beta}_1^\epsilon$  seem to tend to 100%, whereas those of  $\hat{\alpha}_1^y$  and  $\hat{\beta}_1^y$  seem to tend to zero, as  $n \rightarrow \infty$ . Hence, the gains seem to tend to 100%, as  $n \rightarrow \infty$ . However, for  $n = 1000$ , the EFF's of  $\hat{\beta}_1^\epsilon$  in the two cases of M2 with

$v_1$  and  $v_3$  are even smaller than those of  $\hat{\beta}_1^y$ , i.e. the gain in these two cases are slightly negative. This shows that  $n = 1000$  is sometimes not large enough for estimating the scale function. Furthermore, observe that the EFF's of  $\hat{\beta}_1^\epsilon$  under M2 are relatively low. Recall that the MSE's of  $\hat{\beta}_1^\epsilon$  under M2 are smaller than those under M1. This means that the affect due to the estimation of  $v$  is more clear in case when the parameter is easy to estimate.

Now let us consider the quality of  $\hat{b}$ . The sample means, standard deviations and square roots of the MSE's of  $\hat{b}$  together with the true asymptotic optimal bandwidths  $b_A$  are given in Table 5. Note that  $b_A$  and the MSE in cases with  $v_0$  are not defined. Kernel density estimates of  $(\hat{b} - b_A)$  for  $v_1$  to  $v_3$  are shown in Figure 1. We see, the performance of  $\hat{b}$  is satisfactory. In all cases the variance of  $\hat{b}$  decreases as  $n$  increases. It is also true for the bias in most of the cases. Both, the variance and the bias of  $\hat{b}$  depend on the scale function and the model of the errors. For two related cases, the variance of  $\hat{b}$  under M1 is smaller than that under M2. Generally, the stronger the scale change is, the larger the variance of  $\hat{b}$ . The bias of  $\hat{b}$  by  $v_1$  is always negative and it is always positive by  $v_3$ . The bandwidth for  $v_2$  is most easily to choose. The choice of the bandwidth by  $v_3$  is in general easier than that by  $v_1$ , except for the case of M2 with  $n = 1000$ . In this case, the detailed structure of  $v_3$  may sometimes be smoothed away due to the large variation caused by the GARCH model. This shows again that  $n = 1000$  is sometimes not large enough for distinguishing the CH and the scale change.

### 6.3 Detailed analysis of two simulated examples

In the following two simulated data sets are selected to show some details. The first example (called Sim 1) is a typical one of the replications under M2 with the scale function  $\sigma_2(t)$  and  $n = 2000$ . The observations  $y_i$ ,  $i = 1, \dots, 2000$ , are shown in Figure 2(a). For Sim 1 we have  $\hat{b} = 0.160$  by starting with any bandwidth  $3/n \leq b_0 \leq 0.5 - 1/n$ , i.e.  $\hat{b}$  does not depend on  $b_0$ , if  $b_0$  is not too small.  $\sigma_2(t)$  (solid line) together with  $\hat{\sigma}_2(t)$  (dashed line) is shown in Figure 2(b). Figure 2(c) shows the standardized residuals  $\hat{\epsilon}_i$ , which look stationary. The estimated GARCH(1, 1) models are

$$h_i^y = 0.0363 + 0.0540y_{i-1}^2 + 0.9432h_{i-1}^y \quad (29)$$

for  $y_i$  and

$$h_i^\epsilon = 0.2052 + 0.0937\epsilon_{i-1}^2 + 0.6965h_{i-1}^\epsilon \quad (30)$$

for  $\hat{\epsilon}_i$ . For model (29) we have  $\hat{\alpha}_1^y + \hat{\beta}_1^y = 0.9972 \approx 1$  so that the fourth moment of this model does not exist. On the opposite model (30) has finite eighth moment as for the underlying GARCH model. The estimated SEMIGARCH conditional and total standard deviations, i.e.  $(h_i^{\hat{\epsilon}})^{1/2}$  and  $\hat{\sigma}_2(t_i)(h_i^{\hat{\epsilon}})^{1/2}$ , are shown in Figures 2(d) and (e). The true conditional and total standard deviations of  $y_i$ , i.e.  $(h_i)^{1/2}$  and  $\sigma_2(t_i)(h_i)^{1/2}$ , are shown in Figures 2(f) and (g). Figure 2(h) shows the estimated GARCH conditional (in this case also the total) standard deviations  $(h_i^y)^{1/2}$ . The analysis of Sim 1 shows:

1. If a standard GARCH model is used, the scale change will be wrongly estimated as a part of the CH. Furthermore, the total variance tends to be overestimated, when it is large and underestimated, when it is small (compare Figures 2(g) and (h)). This is mainly due to the overestimation of  $\hat{\beta}_1$ .
2. Following the SEMIGARCH model, both, the conditional heteroskedasticity and the scale change are well estimated. The estimated SEMIGARCH total variances are quite close to the true values and are more stable and accurate than those following the standard GARCH model (compare Figures 2(e) and (h)). The errors in  $\hat{\sigma}^2(t_i)h_i^{\hat{\epsilon}}$  are caused by the errors in these two estimates, both of them can be clearly reduced, if more dense observations are available, for instance by analyzing high-frequency financial data. The MSE of the estimated total variances are 0.687 for the SEMIGARCH and 4.979 for the standard GARCH models, the latter is about seven times so large as the former.

Furthermore,  $(h_i^y)^{1/2}$  shown in Figure 2(h) (see also Figures 4(f) and 5(f)) exhibit a clear signal of covariance nonstationarity, a property not shared by the true and the estimated SEMIGARCH conditional standard deviations.

The second simulated data set (called Sim 2) is chosen to show that, sometimes, the selected bandwidth will be wrong, if  $b_0$  is too small or too large. That is a moderate  $b_0$  should be used as proposed in the last section. The data set Sim 2 shown in Figure 3(a) is one of the replications under M1 with  $v_3$  and  $n = 1000$ . For this data set we have,  $\hat{b} = 0.012$  or  $0.12$ , if  $b_0 < 0.020$ . On the other hand, we have  $\hat{b} = 0.499$ , the largest allowed bandwidth in the program, if  $b_0 > 0.262$ . For any starting bandwidth  $b_0 \in [0.021, 0.262]$  a bandwidth  $\hat{b} \doteq 0.120$  will be selected. Now,  $\hat{b}$  does not depend on  $b_0$ . The function  $v_3^{1/2}$  (solid line) together with estimations obtained using  $b_0 = 1/n$  (dots),  $b_0 = 0.126$  (dashes) and  $b_0 = 0.499$  (long dashes) are shown in Figure 3(b). Figure 3(c) shows the data-driven

standardized residuals  $\hat{\epsilon}_i$  obtained using the default  $b_0$ . The relationship between  $\hat{b}$  and  $b_0$  is shown in Figure 3(d).

Note that the proposed default starting bandwidth  $b_0 = 0.5n^{-1/5} = 0.126$  lies in the middle part of the interval  $[0.021, 0.262]$ . In case when it is doubtful, if the selected bandwidth with  $b_0 = 0.5n^{-1/5}$  is the optimal one, we recommend the user to try with some different  $b_0$ 's and choose the most reasonable  $\hat{b}$  from all possible selected bandwidths by means of further analysis (see Feng, 2002).

## 7 Applications

In this section the proposal will be applied to the log-returns of the daily S&P 500 and DAX 100 financial indices from January 03, 1994 to August 23, 2000. For the S&P 500 returns shown in Figure 4(a) we have  $\hat{b} = 0.183$  (for any  $b_0 \geq 0.075$ ). The fitted GARCH models are

$$h_i^y = 5.684 * 10^{-7} + 0.0674y_{i-1}^2 + 0.9302h_{i-1}^y \quad (31)$$

for  $y_i$  and

$$h_i^\epsilon = 0.0649 + 0.0686\epsilon_{i-1}^2 + 0.8676h_{i-1}^\epsilon \quad (32)$$

for  $\hat{\epsilon}_i$ . As before, for model (31) we have  $\hat{\alpha}_1^y + \hat{\beta}_1^y = 0.9976 \approx 1$  so that the fourth moment of this model does not exist. However, model (32) has finite eighth moment. Figure 4(b) shows  $\hat{\sigma}(t)$  together with an about 95% confidence interval for a constant scale function. We see that there is significant scale change. Furthermore, both  $\hat{\alpha}_1$  and  $\hat{\beta}_1$  in model (32) are strongly significant. That is this series has simultaneously significant scale change and CH. Figures 4(c) to (f) show  $\hat{\epsilon}_i$ , the SEMIGARCH conditional standard deviations  $(h_i^\epsilon)^{1/2}$ , the SEMIGARCH total standard deviations  $\hat{\sigma}(t)(h_i^\epsilon)^{1/2}$  as well as the GARCH conditional standard deviations  $(h_i^y)^{1/2}$ . We see again that the estimated total variances following the SEMIGARCH model are more stable.

For the DAX 100 returns we have  $\hat{b} = 0.181$  (for any  $b_0 \geq 0.075$ ). The fitted GARCH models are

$$h_i^y = 2.202 * 10^{-6} + 0.0892y_{i-1}^2 + 0.8957h_{i-1}^y \quad (33)$$

for  $y_i$  and

$$h_i^\epsilon = 0.0651 + 0.0873\epsilon_{i-1}^2 + 0.8481h_{i-1}^\epsilon \quad (34)$$



for  $\hat{\epsilon}_i$ . The condition for the existence of the fourth moment of model (33) is slightly satisfied but the eighth moment of this model does not exist. However, model (34) has finite eighth moment. Same results as given in Figure 4 are shown in Figure 5 for this data set. We see that the S&P 500 and DAX 100 returns series perform quite similarly and the conclusions on the former given above apply to the latter.

## 8 Final remarks

In this paper a SEMIGARCH model is introduced for simultaneously modelling conditional heteroskedasticity and scale change. A data-driven algorithm for the practical implementation of the proposal is developed. Simulation and data examples show that the proposal performs well in practice. There are still many open questions on this topic, e.g. the model selection, the detailed discussion on the properties of  $\hat{\theta}$  and the combination of the SEMIGARCH model with other variants of the GARCH model or with the SEMIFAR model. For the model selection the AIC or BIC criteria can be used. The other questions will be discussed elsewhere.

## Acknowledgements

This work was finished under the advice of Prof. Jan Beran, Department of Mathematics and Statistics, University of Konstanz, Germany, and was financially supported by the *Center of Finance and Econometrics* (CoFE), University of Konstanz. We would like to thank colleagues in CoFE, especially Prof. Winfried Pohlmeier, for their interesting questions to a talk of the author, it were these questions which motivated the author to write this paper. Our special thanks go to Mr. Erik Lüders, CoFE/ZEW, for his helpful suggestions, which lead to improve the quality of this paper.

## Appendix: Proofs of results

In the following  $\hat{z}_i$  and  $\hat{x}_i$  will be replaced by  $z_i$  and  $x_i$ , respectively, since the error in  $\bar{y}$  is negligible.

### Proof of Theorem 1.

*i) The bias.* Note that  $\hat{v}$  is a linear smoother

$$\hat{v}(t) = \sum_{i=1}^n w_i x_i, \quad (\text{A.1})$$

where  $w_i$  are the weights defined by (5). The bias of  $\hat{v}$  is  $E(\hat{v}(t)) - v(t) = \sum_{i=1}^n w_i v(t_i) - v(t)$ , which is just the same as in nonparametric regression with i.i.d. errors. That is, the bias depends neither on the dependence structure nor on the heteroskedasticity of the errors. This leads to the result given in (10).

*ii) The variance.* Let  $\zeta_i = v(t_i)\xi_i$  denote the errors in (4). Note that  $w_i = 0$  for  $|t_i - t| > b$  we have

$$\text{var}(\hat{v}) = \sum_{|t_i - t| \leq b} \sum_{|t_j - t| \leq b} w_i w_j \text{cov}(\zeta_i, \zeta_j). \quad (\text{A.2})$$

For  $|t_i - t| \leq b$  and  $|t_j - t| \leq b$  we have  $\zeta_i = [v(t) + O(b)]\xi_i$  and  $\zeta_j = [v(t) + O(b)]\xi_j$ . This leads to

$$\begin{aligned} \text{cov}(\zeta_i, \zeta_j) &= \text{cov}([v(t) + O(b)]\xi_i, [v(t) + O(b)]\xi_j) \\ &= v^2(t)\gamma_\xi(i - j)[1 + o(1)]. \end{aligned} \quad (\text{A.3})$$

Insert this into (A.2) we have

$$\text{var}(\hat{v}) = v^2(t) \left\{ \sum_{|t_i - t| \leq b} \sum_{|t_j - t| \leq b} w_i w_j \gamma_\xi(i - j) \right\} [1 + o(1)]. \quad (\text{A.4})$$

Results in (11) follow from known results on  $\sum \sum w_i w_j \gamma_\xi(i - j)$  in nonparametric regression with dependent errors (see e.g. Beran, 1999 and Beran and Feng, 2002a).

*iii) Asymptotic normality.* Consider the estimation problem under the model without DV

$$\tilde{X}_i = v(t_i) + v(t)\xi_i = v(t_i) - v(t) + v(t)\epsilon_i^2. \quad (\text{A.5})$$

Define

$$\tilde{v}(t) = \sum_{i=1}^n w_i \tilde{x}_i, \quad (\text{A.6})$$

where  $\tilde{x}_i$  are observations obtained following model (A.5). Following the results in *i*) and *ii*) we see  $(nb)^{1/2}[\hat{v}(t) - \tilde{v}(t)] = o_p(1)$ . Hence  $\hat{v}(t)$  is asymptotically normal, if and only if  $\tilde{v}(t)$  is. The asymptotic normality of  $\tilde{v}(t)$  can be shown using a central limit theorem on nonparametric regression with dependent errors developed by Beran and Feng (2001). Note at first that the error process  $v(t)\epsilon_i^2$  in (A.5) is a squared GARCH process. Under the assumptions of Theorem 1,  $(v(t)\epsilon_i^2)^2 = v^2(t)\epsilon_i^4$  is also second order and strict stationary, whose autocovariances converge to zero as the lag tends to infinite. Furthermore, there exists an extremal index  $\gamma \in (0, 1]$  for the process  $v(t)\epsilon_i^2$  (see Davis et al., 1999). Hence the error process in (A.5) satisfies the condition as given in Theorem 1, Case 2 of Beran and Feng (2001). And it is not difficult to check that the weights  $w_i$  fulfill the conditions of Theorem 4 in Beran and Feng (2001). The asymptotic normality of  $\tilde{v}(t)$  follows from Theorem 4 of Beran and Feng (2001). Theorem 1 is proved.  $\diamond$

**Proof of (13) and (14).** Note that  $\xi_i$  has the ARMA representation

$$\phi(B)\xi_i = \psi(B)u_i, \quad (\text{A.7})$$

where  $\phi(z)$  and  $\psi(z)$  are as defined before. Under A5  $\phi(z)$  and  $\psi(z)$  have no common roots. Under A1 all roots of  $\phi(z)$  and  $\psi(z)$  lie outside the unit circle. Then the spectral density of  $\xi$  is given by

$$f(\lambda) = \frac{\text{var}(u_i) |\psi(e^{-i\lambda})|^2}{2\pi |\phi(e^{-i\lambda})|^2} \quad \text{and} \quad f(0) = \frac{\text{var}(u_i) (\psi(1))^2}{2\pi (\phi(1))^2}. \quad (\text{A.8})$$

Note that  $E(\epsilon_i^4) = 3E(h_i^2)$  (Bollerslev, 1986) and  $\text{var}(u_i) = E(u_i^2) = 2E(h_i^2)$ . The last equation follows from (9). That is  $\text{var}(u_i) = \frac{2}{3}E(\epsilon_i^4)$ . Result in (13) is proved by inserting this formula,  $\psi(1)$  and  $\phi(1)$  into (A.8). Result in (14) is obtained by further inserting explicit formula of  $E(\epsilon_i^4)$  for a GARCH(1, 1) model (Bollerslev, 1986) into (13).  $\diamond$

**A sketched proof of Proposition 1.** Taylor expansion on  $\hat{\epsilon}_i^2$  leads to

$$\begin{aligned} \hat{\epsilon}_i^4 &= (z_i^2/\hat{v}(t_i))^2 \\ &\doteq \left( \frac{z_i^2}{v(t_i)} + O_p(\hat{v}(t_i) - v(t_i)) + O_p(\hat{v}(t_i) - v(t_i))^2 \right)^2 \\ &\doteq \epsilon_i^4 + O_p(\hat{v}(t_i) - v(t_i)) + O_p(\hat{v}(t_i) - v(t_i))^2. \end{aligned} \quad (\text{A.9})$$

We have

$$\begin{aligned} E[\hat{E}(\epsilon_i^4) - E(\epsilon_i^4)] &= O\left(\frac{1}{n} \sum_{i=1}^n E(\hat{v}(t_i) - v(t_i))\right) + O\left(\frac{1}{n} \sum_{i=1}^n E(\hat{v}(t_i) - v(t_i))^2\right) \\ &=: T_1 + T_2. \end{aligned} \quad (\text{A.10})$$

Observe that the bias  $E(\hat{v}(t_i) - v(t_i))$  is of order  $O(b_\epsilon^2)$  in the interior and of order  $O(b_\epsilon)$  at the boundary. We obtain  $T_1 = O(b_\epsilon^2)$ , since the length of the boundary area is  $2b_\epsilon$ . Furthermore,  $T_2 = \text{MISE}_{[0,1]}[1 + o(1)] = O(nb_\epsilon)^{-1} + o(T_1)$ . Results given in (19) are proved.

Observe that  $\hat{\epsilon}_i^4 = \epsilon_i^4[1 + o_p(1)]$ . We have

$$\text{var}(\hat{E}(\epsilon_i^4)) = \text{var}\left(\frac{1}{n} \sum_{i=1}^n \epsilon_i^4\right) [1 + o(1)].$$

Note that  $\epsilon_i^4$  follow a squared ARMA process, which is again a second order stationary process with absolute summable autocovariances under the assumption  $E(\epsilon_i^8) < \infty$ . Hence the spectral density of  $\epsilon_i^4$  exists and

$$n \text{var}\left(\frac{1}{n} \sum_{i=1}^n \epsilon_i^4\right) \rightarrow 2\pi c_f^\epsilon, \quad (\text{A.11})$$

where  $c_f^\epsilon$  is the value of the spectral density of  $\epsilon_i^4$  at the origin (see e.g. Brockwell and Davis, 1991, pp. 218ff). Proposition 1 is proved.  $\diamond$

**A sketched proof of Proposition 2.** Estimation of functionals of the form  $\int \{v^{(\nu)}(t)\}^2 dt$ , where  $v^{(\nu)}$  is the  $\nu$ -th derivative of  $v$ , was investigated in detail by Ruppert et al. (1995) in nonparametric regression with independent errors and Beran and Feng (2002a) in nonparametric regression with dependent errors. Note that  $I(v^2) = \int \{v^2(t)\}^2 dt$  is a special case of such functionals with  $\nu = 0$ . Furthermore, the results in Ruppert et al. (1995) and Beran and Feng (2002a) together show that, the magnitude orders in these results stay unchanged, if short-range dependence and/or a bounded, smooth scale function are introduced into the error process. We obtain the results of Proposition 2 by setting  $k = 0$ ,  $l = 2$  and  $\delta = 0$  in the results in Beran and Feng (2002a), where  $k$  and  $l$  correspond to  $\nu = 0$  and the kernel order used here and  $\delta$  is the long-memory parameter, which is zero in the current context.  $\diamond$

**A sketched proof of Theorem 3.** Note that  $\hat{b} = \hat{C}_A n^{-1/5}$ , where  $C_A$  is as defined in (17). Hence we have

$$(\hat{b} - b_A)/b_A = C_A^{-1}(\hat{C}_A - C_A). \quad (\text{A.12})$$

Taylor expansion shows that

$$\hat{C}_A - C_A \doteq O_p(\hat{c}_f - c_f) + O_p(\hat{I}(v^2) - I(v^2)) + O_p(\hat{I}((v'')^2) - I((v'')^2)). \quad (\text{A.13})$$

Observe that

$$\hat{c}_f - c_f \doteq O_p(\hat{I}(v^2) - I(v^2)) \doteq O_p(n^{-1/2}) \quad (\text{A.14})$$

and

$$\hat{I}((v'')^2) - I((v'')^2) \doteq O_p(n^{-2/7}). \quad (\text{A.15})$$

We obtain the results given in Theorem 3.  $\diamond$

## References

- Altman, N.S. (1990) Kernel smoothing with correlated errors. *J. Am. Statist. Assoc.* 85, 749–759.
- Beran, J. (1999) SEMIFAR models – A semiparametric framework for modelling trends, long range dependence and nonstationarity. Discussion paper, CoFE, No. 99/16, University of Konstanz.
- Beran, J. and Y. Feng (2001) Local polynomial estimation with a FARIMA-GARCH error process. *Bernoulli* 7, 733 – 750.
- Beran, J. and Y. Feng (2002a) Local polynomial fitting with long-memory, short-memory and antipersistent errors. *The Ann. Instit. Statist. Math.* (in press).
- Beran, J. and Y. Feng (2002b) Iterative plug-in algorithms for SEMIFAR models - definition, convergence and asymptotic properties. To appear in *J. Comput. and Graphical Statist.*
- Beran, J. and Y. Feng (2002c) SEMIFAR models - A semiparametric framework for modelling trends, long range dependence and nonstationarity. *Comput. Statist. and Data Anal.* (in press).
- Beran, J. and D. Ocker (2001) Volatility of Stock Market Indices - An Analysis based on SEMIFAR models. *Journal of Business and Economic Statistics* 19, 103-116.
- Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* 31, 307–327.

- Brockwell, P.J. and R.A. Davis (1991) *Time Series: Theory and Methods*. Springer, New York.
- Davis, R.A., T. Mikosch, and B. Basrak (1999) Sample ACF of Multivariate stochastic recurrence equations with application to GARCH. Preprint, University of Groningen.
- Ding, Z., C.W.J. Granger and R.F. Engle (1993) A long memory property of stock market returns and a new model. *J. Empirical Finance* 1, 83-106.
- Efromovich, S. (1999) *Nonparametric Curve estimation: Methods, Theory, and Applications*. Springer, New York.
- Engel, R.F. (1982) Autoregressive conditional heteroskedasticity with estimation of U.K. inflation, *Econometrica* 50, 987–1008.
- Fan, J. and I. Gijbels (1995) Data-driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaptation. *J. Roy. Statist. Soc. Ser. B* 57, 371–394.
- Feng, Y. (2002) An iterative plug-in algorithm for nonparametric modelling of seasonal time series. Discussion Paper, CoFE, No. 02/04, University of Konstanz.
- Feng, Y. and S. Heiler (1998) Locally weighted autoregression. In: *Econometrics in Theory and Practice*. Festschrift für Hans Schneeweiß (Eds. Galata, R., Küchenhoff, H.), pp. 101–117, Physica-Verlag, Heidelberg.
- Gasser, T., A. Kneip and W. Köhler (1991) A flexible and fast method for automatic smoothing. *J. Amer. Statist. Assoc.* 86, 643–652.
- Gasser, T. and H.G. Müller (1984) Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statist.* 11, 171–185.
- Härdle, W., A.B. Tsybakov and L. Yang (1997) Nonparametric vector autoregression. *J. Statist. Plann. Infer.* 68, 221–245.
- Hall, P. and J.D. Hart (1990) Nonparametric regression with long-range dependence. *Stochastic Process. Appl.* 36 339–351.
- Hart, J.D. (1991) Kernel regression estimation with time series errors. *J. R. Statist. Soc. Ser. B* 53, 173–188.

- He, C. T. and Teräsvirta (1999a) Forth moment structure of the GARCH( $p, q$ ) process. *Econometric Theory* 15, 824–846.
- He, C. and T. Teräsvirta (1999b) Properties of autocorrelation function of squared observations for second-order GARCH processes under two sets of parameter constrains. *J. Time Series Analysis* 20, 23–30.
- Herrmann, E. and T. Gasser (1994) Iterative plug-in algorithm for bandwidth selection in kernel regression estimation. Preprint, Darmstadt Institute of Technology and University of Zürich.
- Herrmann, E., T. Gasser, and A. Kneip (1992) Choice of bandwidth for kernel regression when residuals are correlated. *Biometrika* 79, 783–795.
- Karanasos, M. (1999) The second moment and autocovariance function of the squared errors of the GARCH model. *J. Econometrics* 90, 63–76.
- Ling, S. (1999) On probability properties of a double threshold ARMA conditional heteroskedasticity model. *J. App. Prob.* 36, 688–705.
- Ling, S. and W.K. Li (1997) On fractional integrated autoregressive moving-average time series models with conditional heteroskedasticity. *J. Amer. Statist. Assoc.* 92, 1184–1194.
- Ling, S. and M. McAleer (2002) Necessary and sufficient moment conditions for the GARCH( $r, s$ ) and asymmetric power GARCH( $r, s$ ) models. *Econometric Theory* 18, 722–729.
- Martin, R.D., H.Y. Gao, Y. Zhan and Z. Ding (1996) S+GARCH User’s Manual. Math-Soft, Seattle.
- Müller, H.G. (1988) *Nonparametric Analysis of Longitudinal Data*. Springer, Berlin.
- Ruppert, D., S.J. Sheather and M.P. Wand (1995) An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.* 90, 1257–1270.
- Ruppert, D. and M.P. Wand (1994). Multivariate locally weighted least squares regression. *Ann. Statist.* 22 1346–1370.

Table 1: Statistics on the estimated parameters for all cases with  $n = 1000$

Parameter	Statistic	Model 1				Model 2			
		$v_1$	$v_2$	$v_3$	$v_0$	$v_1$	$v_2$	$v_3$	$v_0$
$\hat{\alpha}_1^\epsilon$	Mean	0.198	0.197	0.195	0.196	0.103	0.102	0.104	0.102
	SD	0.048	0.049	0.050	0.052	0.032	0.031	0.030	0.031
	MSE <sup>1/2</sup>	0.048	0.049	0.051	0.052	0.032	0.031	0.030	0.031
$\hat{\beta}_1^\epsilon$	Mean	0.180	0.191	0.178	0.196	0.715	0.723	0.719	0.724
	SD	0.174	0.176	0.189	0.192	0.114	0.103	0.109	0.099
	MSE <sup>1/2</sup>	0.175	0.176	0.190	0.191	0.119	0.106	0.113	0.102
$\hat{\alpha}_1^y$	Mean	0.141	0.076	0.117	—	0.090	0.080	0.099	—
	SD	0.091	0.053	0.072	—	0.036	0.029	0.027	—
	MSE <sup>1/2</sup>	0.109	0.135	0.110	—	0.037	0.035	0.027	—
$\hat{\beta}_1^y$	Mean	0.710	0.909	0.821	—	0.870	0.911	0.877	—
	SD	0.249	0.090	0.159	—	0.070	0.036	0.039	—
	MSE <sup>1/2</sup>	0.568	0.715	0.641	—	0.139	0.165	0.133	—
$\hat{\alpha}_1^{\hat{\epsilon}}$	Mean	0.191	0.191	0.187	0.188	0.100	0.099	0.101	0.098
	SD	0.049	0.049	0.051	0.051	0.032	0.031	0.030	0.032
	MSE <sup>1/2</sup>	0.049	0.050	0.053	0.052	0.032	0.031	0.030	0.032
$\hat{\beta}_1^\epsilon$	Mean	0.159	0.168	0.176	0.166	0.677	0.695	0.701	0.686
	SD	0.177	0.178	0.201	0.185	0.138	0.120	0.135	0.115
	MSE <sup>1/2</sup>	0.181	0.181	0.202	0.188	0.156	0.132	0.143	0.132



Table 2: Statistics on the estimated parameters for all cases with  $n = 2000$

Parameter	Statistic	Model 1				Model 2			
		$v_1$	$v_2$	$v_3$	$v_0$	$v_1$	$v_2$	$v_3$	$v_0$
$\hat{\alpha}_1^\epsilon$	Mean	0.197	0.196	0.194	0.197	0.102	0.099	0.100	0.099
	SD	0.035	0.035	0.034	0.035	0.021	0.022	0.023	0.022
	MSE <sup>1/2</sup>	0.036	0.035	0.035	0.035	0.022	0.022	0.023	0.022
$\hat{\beta}_1^\epsilon$	Mean	0.190	0.190	0.200	0.207	0.732	0.745	0.736	0.737
	SD	0.123	0.116	0.121	0.114	0.065	0.064	0.074	0.076
	MSE <sup>1/2</sup>	0.124	0.116	0.121	0.114	0.068	0.064	0.076	0.077
$\hat{\alpha}_1^y$	Mean	0.132	0.051	0.087	—	0.086	0.074	0.084	—
	SD	0.084	0.037	0.055	—	0.029	0.026	0.025	—
	MSE <sup>1/2</sup>	0.108	0.153	0.125	—	0.032	0.037	0.030	—
$\hat{\beta}_1^y$	Mean	0.734	0.943	0.879	—	0.881	0.919	0.899	—
	SD	0.225	0.048	0.109	—	0.050	0.031	0.035	—
	MSE <sup>1/2</sup>	0.579	0.745	0.687	—	0.140	0.172	0.153	—
$\hat{\alpha}_1^{\hat{\epsilon}}$	Mean	0.193	0.193	0.190	0.192	0.100	0.098	0.098	0.097
	SD	0.035	0.035	0.035	0.035	0.021	0.023	0.022	0.022
	MSE <sup>1/2</sup>	0.036	0.036	0.036	0.036	0.021	0.023	0.022	0.022
$\hat{\beta}_1^{\hat{\epsilon}}$	Mean	0.176	0.182	0.199	0.192	0.716	0.735	0.729	0.714
	SD	0.122	0.120	0.127	0.114	0.070	0.066	0.079	0.084
	MSE <sup>1/2</sup>	0.124	0.121	0.127	0.114	0.078	0.067	0.082	0.091

Table 3: Statistics on the estimated parameters for all cases with  $n = 4000$

Parameter	Statistic	Model 1				Model 2			
		$v_1$	$v_2$	$v_3$	$v_0$	$v_1$	$v_2$	$v_3$	$v_0$
$\hat{\alpha}_1^\epsilon$	Mean	0.197	0.197	0.196	0.195	0.100	0.100	0.099	0.101
	SD	0.024	0.025	0.024	0.024	0.014	0.016	0.015	0.015
	MSE <sup>1/2</sup>	0.024	0.025	0.024	0.024	0.014	0.016	0.015	0.015
$\hat{\beta}_1^\epsilon$	Mean	0.195	0.201	0.202	0.194	0.746	0.745	0.748	0.743
	SD	0.078	0.081	0.081	0.083	0.045	0.051	0.041	0.047
	MSE <sup>1/2</sup>	0.078	0.081	0.081	0.083	0.046	0.051	0.041	0.047
$\hat{\alpha}_1^y$	Mean	0.125	0.038	0.066	—	0.085	0.069	0.078	—
	SD	0.073	0.026	0.043	—	0.021	0.018	0.019	—
	MSE <sup>1/2</sup>	0.105	0.164	0.141	—	0.026	0.036	0.029	—
$\hat{\beta}_1^y$	Mean	0.758	0.959	0.917	—	0.885	0.925	0.907	—
	SD	0.187	0.031	0.068	—	0.035	0.022	0.026	—
	MSE <sup>1/2</sup>	0.589	0.760	0.720	—	0.139	0.176	0.159	—
$\hat{\alpha}_1^{\hat{\epsilon}}$	Mean	0.195	0.195	0.194	0.192	0.100	0.099	0.099	0.100
	SD	0.024	0.025	0.024	0.024	0.014	0.016	0.015	0.016
	MSE <sup>1/2</sup>	0.025	0.025	0.025	0.025	0.014	0.016	0.015	0.016
$\hat{\beta}_1^{\hat{\epsilon}}$	Mean	0.189	0.199	0.202	0.185	0.738	0.740	0.745	0.730
	SD	0.078	0.081	0.081	0.084	0.048	0.053	0.044	0.050
	MSE <sup>1/2</sup>	0.079	0.081	0.081	0.086	0.049	0.054	0.044	0.054

Table 4: Empirical efficiencies (%) of the estimated parameters

Parameter	$n$	Model 1				Model 2			
		$v_1$	$v_2$	$v_3$	$v_0$	$v_1$	$v_2$	$v_3$	$v_0$
$\hat{\alpha}_1^y$	1000	19.8	13.4	21.3	—	73.0	77.9	124.9	—
	2000	10.9	5.1	7.8	—	45.3	36.6	58.2	—
	4000	5.4	2.3	3.0	—	30.2	18.8	26.5	—
$\hat{\beta}_1^y$	1000	9.5	6.0	8.8	—	73.1	41.4	73.0	—
	2000	4.6	2.4	3.1	—	23.4	13.8	24.4	—
	4000	1.8	1.1	1.3	—	10.7	8.5	6.8	—
$\hat{\alpha}_1^{\hat{\epsilon}}$	1000	96.2	97.8	91.3	98.0	96.8	99.8	101.1	95.1
	2000	99.0	94.0	92.5	97.2	100.7	96.5	102.2	99.8
	4000	96.3	97.7	94.0	91.4	97.7	97.8	97.5	97.7
$\hat{\beta}_1^{\hat{\epsilon}}$	1000	93.3	94.2	88.2	103.3	58.5	64.3	62.7	59.9
	2000	99.2	92.7	91.2	99.8	75.2	89.8	85.4	71.7
	4000	97.5	99.8	99.2	94.7	86.5	90.7	86.1	77.6

Table 5: Statistics on the selected bandwidth

$n$	Statistic	Model 1				Model 2			
		$v_1$	$v_2$	$v_3$	$v_0$	$v_1$	$v_2$	$v_3$	$v_0$
1000	$b_A$	0.187	0.166	0.107	—	0.204	0.181	0.116	—
	Mean	0.174	0.167	0.119	0.173	0.184	0.175	0.131	0.191
	SD	0.015	0.011	0.008	0.028	0.024	0.017	0.031	0.037
	MSE <sup>1/2</sup>	0.019	0.011	0.015	—	0.031	0.018	0.034	—
2000	$b_A$	0.163	0.144	0.093	—	0.177	0.151	0.101	—
	Mean	0.153	0.148	0.105	0.141	0.163	0.158	0.113	0.155
	SD	0.011	0.007	0.005	0.018	0.015	0.011	0.008	0.026
	MSE <sup>1/2</sup>	0.015	0.008	0.013	—	0.020	0.014	0.014	—
4000	$b_A$	0.142	0.126	0.081	—	0.154	0.137	0.088	—
	Mean	0.131	0.130	0.091	0.111	0.144	0.140	0.099	0.126
	SD	0.009	0.006	0.003	0.010	0.012	0.008	0.005	0.016
	MSE <sup>1/2</sup>	0.014	0.007	0.010	—	0.015	0.008	0.012	—

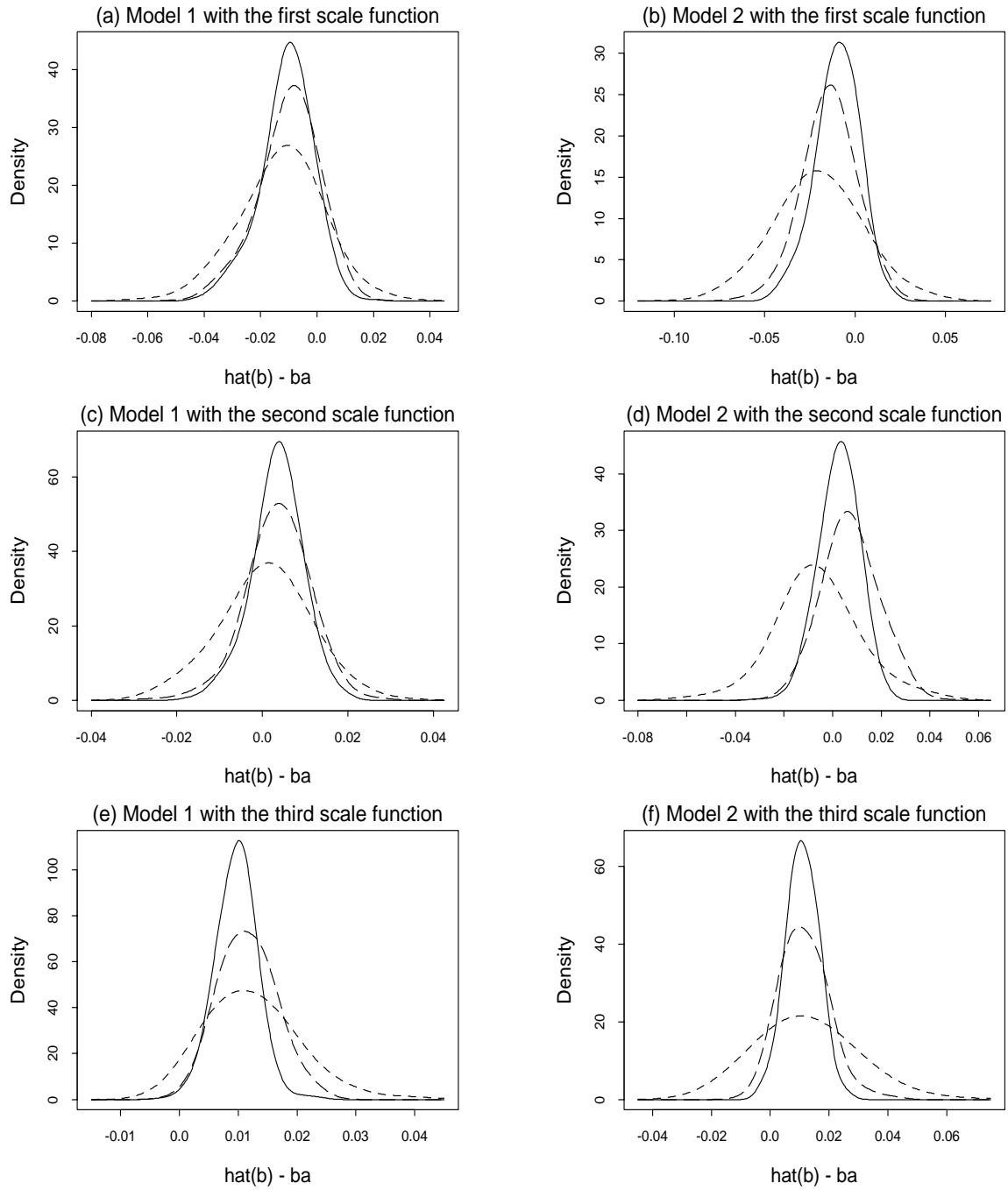


Figure 1: Kernel density estimates of  $\hat{b} - b_A$  (short dashes for  $n = 1000$ , long dashes for  $n = 2000$  and solid line for  $n = 4000$ ).

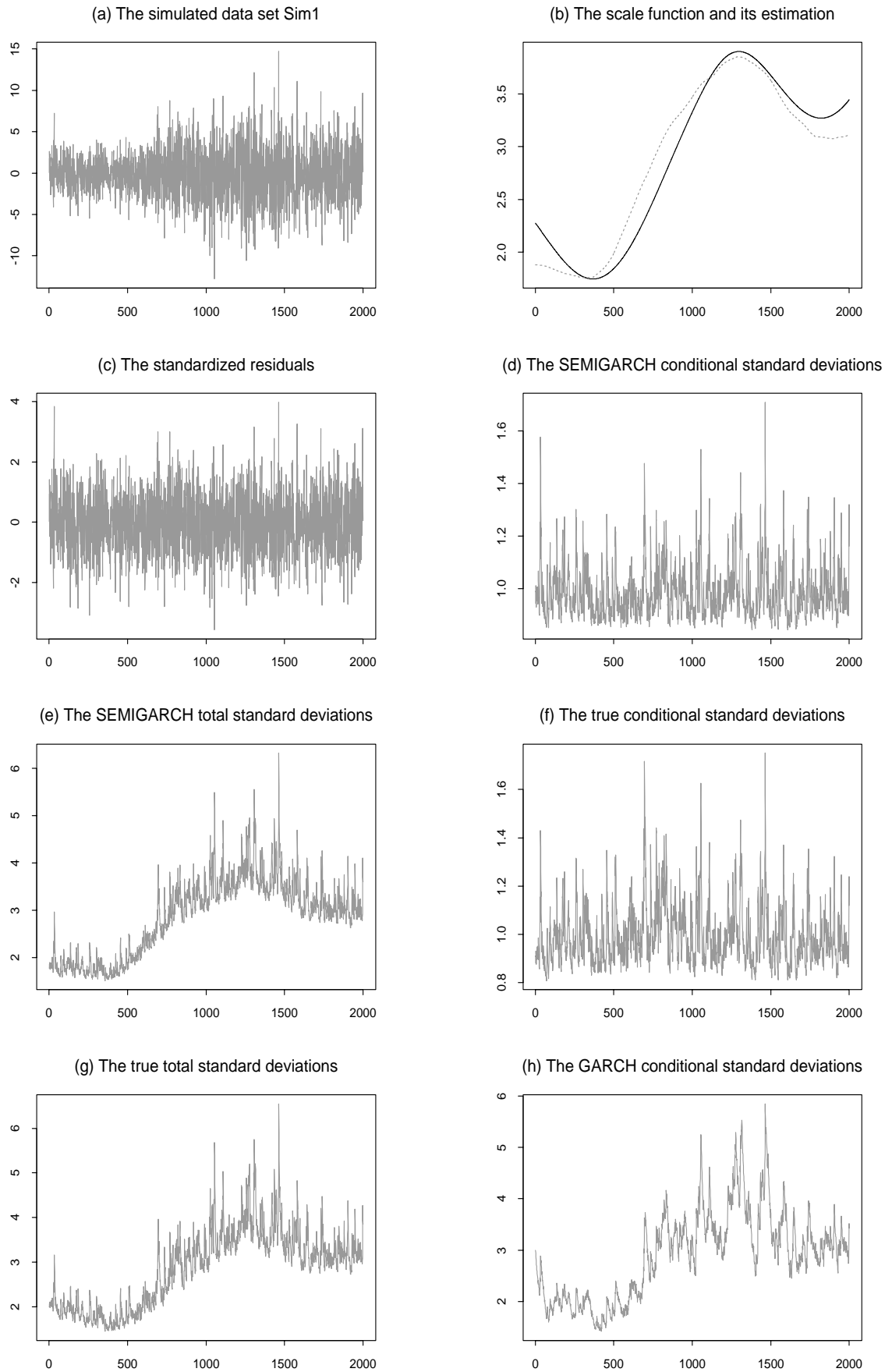


Figure 2: Estimation results for the first simulated data set.

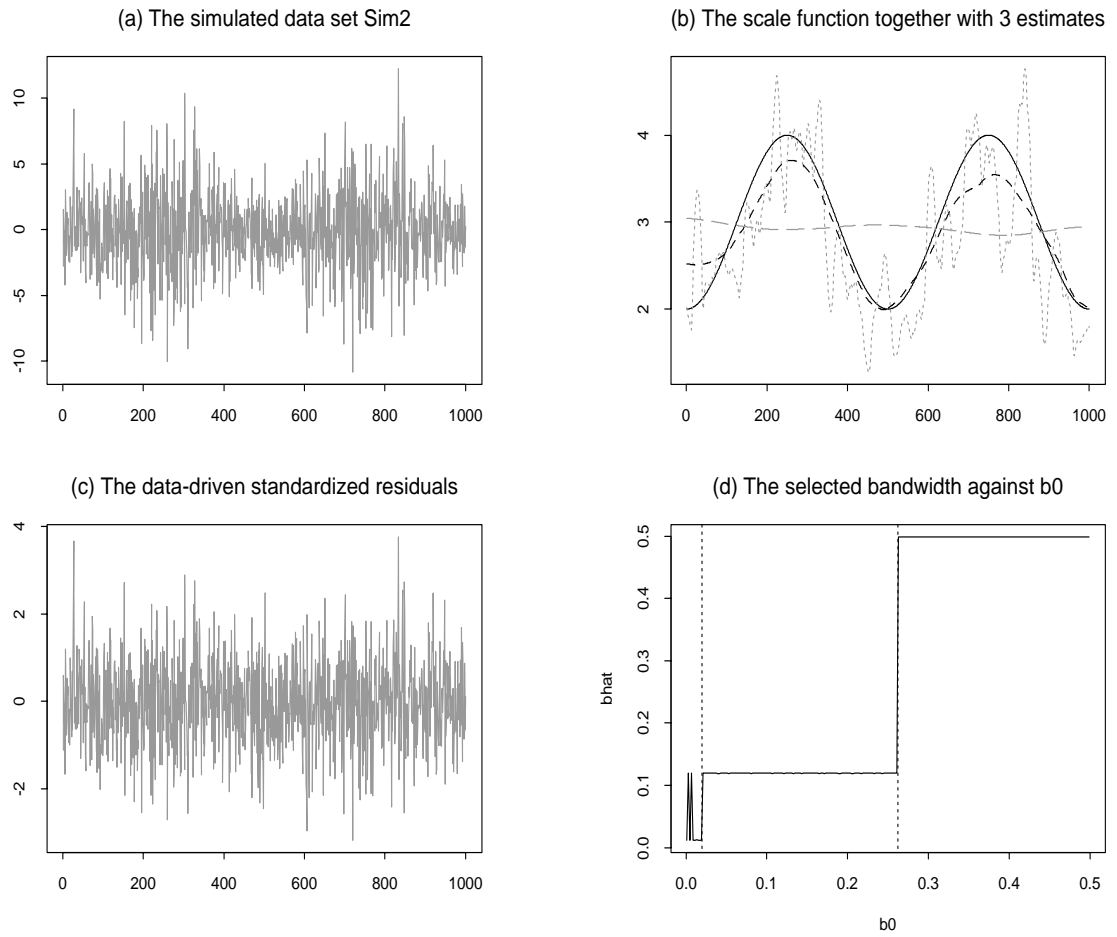


Figure 3: The second simulated data set and some detailed estimation results. Figure (b) shows the scale function  $\sigma_3(t)$  (solid line), the estimation with  $b_0 = n^{-1}$  (dots),  $b_0 = 0.5n^{-1/5}$  (short dashes) and  $b_0 = 0.5 - n^{-1}$  (long dashes).

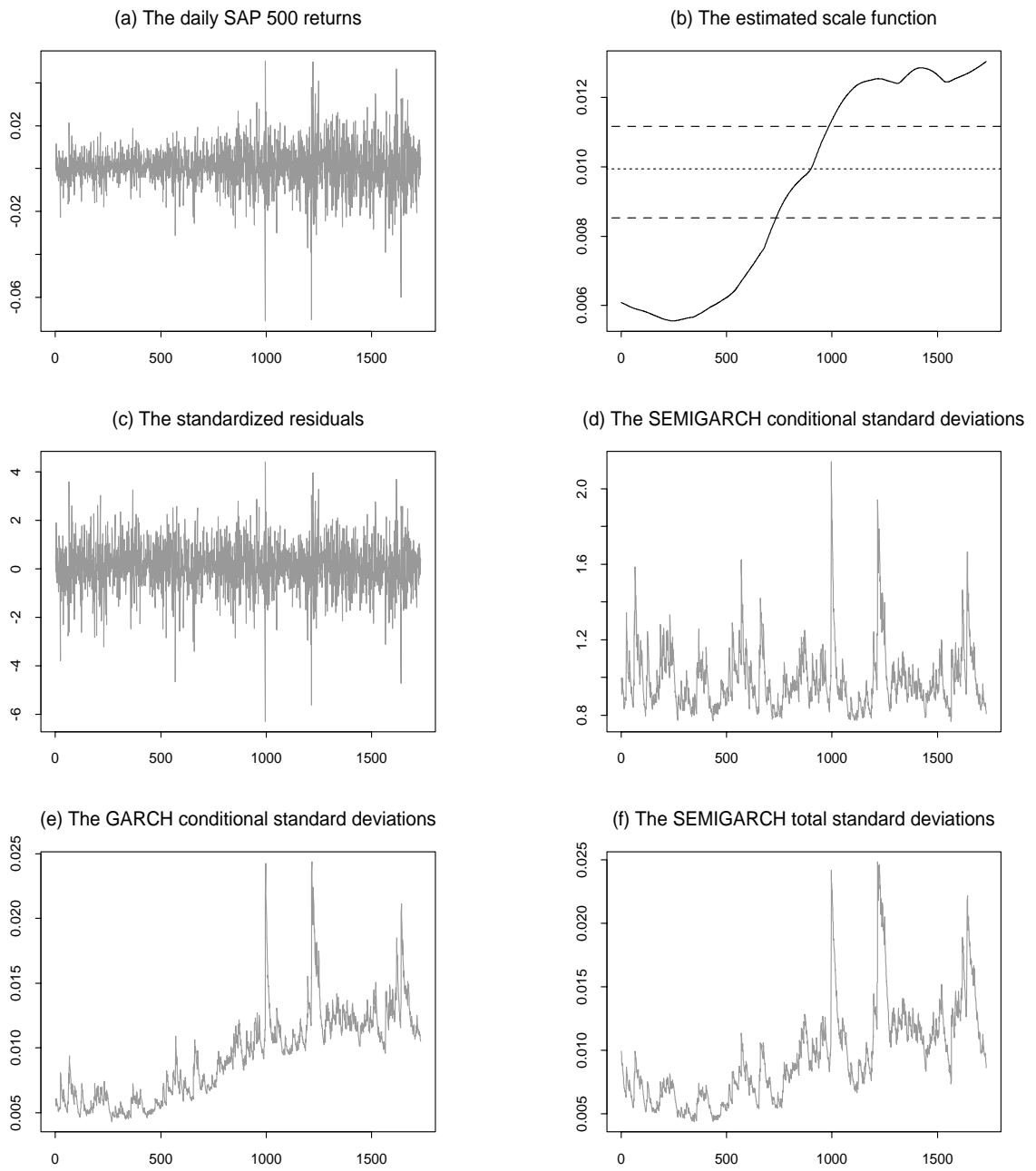


Figure 4: The estimation results for the S&P 500 returns.

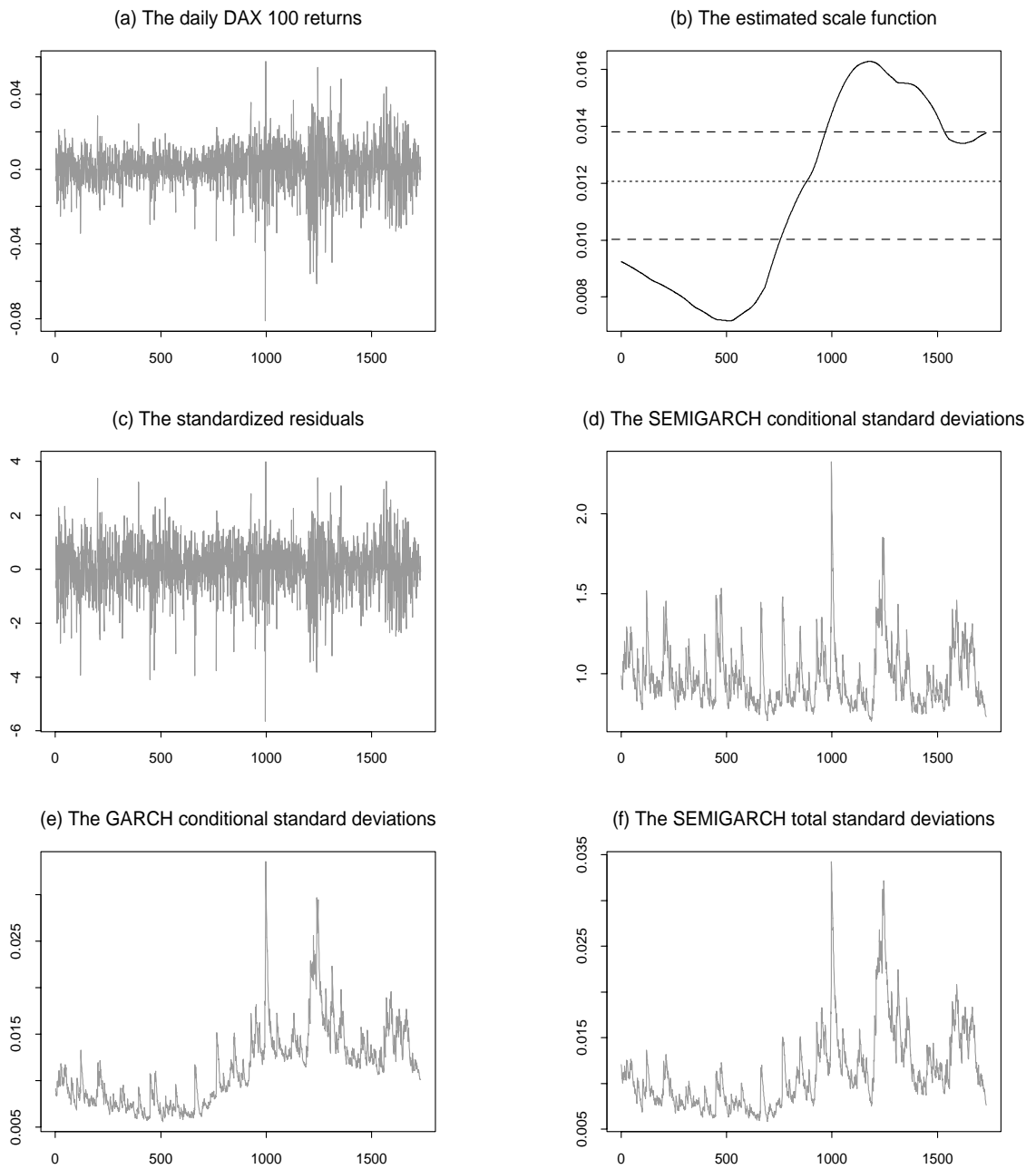


Figure 5: The estimation results for the DAX 100 returns.