Econometric Theory, **20**, 2004, 563–596. Printed in the United States of America.
DOI: 10.1017/S0266466604203061

SIMULTANEOUSLY MODELING CONDITIONAL HETEROSKEDASTICITY AND SCALE CHANGE

YUANHUA FENG
University of Konstanz

This paper proposes a semiparametric approach by introducing a smooth scale function into the standard generalized autoregressive conditional heteroskedastic (GARCH) model so that conditional heteroskedasticity (CH) and scale change in financial returns can be modeled 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 estimators are investigated in detail. It is shown that asymptotically normal, $\sqrt{n}$-consistent parameter estimation is available. A data-driven algorithm is developed for practical implementation. Finite sample performance of the proposal is studied through simulation. The proposal is applied to model CH and scale change in the daily S&P 500 and DAX 100 returns. It is shown that both series have simultaneously significant scale change and CH.

1. INTRODUCTION

Modeling of heteroskedasticity in financial returns is one of the most important and interesting themes of financial econometrics. Well-known conditional heteroskedastic (CH) models are the autoregressive conditional heteroskedastic ARCH (Engle, 1982) and (generalized ARCH) GARCH (Bollerslev, 1986) together with numerous extensions. Most GARCH variants are however stationary models and are hence time homoskedastic with constant unconditional variance. In practice it is realized that financial returns are often not only conditional but also time heteroskedastic with time varying unconditional variance. This is shown by, e.g., Beran and Ocker (2001) by fitting a trend function to some

We are very grateful to the co-editor and two referees for their helpful comments and suggestions, which led to a substantial improvement of this paper. The paper was finished under the advice of Professor 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 thank colleagues in CoFE, especially Professor Winfried Pohlmeier, for their interesting questions at a talk of the author. It was these questions that motivated the author to write this paper. Our special thanks go to Dr. Erik Lüders, Department of Finance and Insurance, Laval University, and Stern School of Business, New York University, for his helpful suggestions.
Address correspondence to: Yuanhua Feng, Department of Mathematics and Statistics, University of Konstanz, D-78457, Konstanz, Germany; e-mail: yuanhua.feng@uni-konstanz.de.

 0266-4666/04 $12.00

volatility series defined by Ding, Granger, and Engle (1993). Nonstationarity in financial returns is investigated in detail by, e.g., Mikosch and Stărică (2004). They show that the phenomenon $\hat{\alpha}_1 + \hat{\beta}_1 \approx 1$ by a fitted GARCH(1, 1) model often implies nonstationarity.

In recent years different approaches for simultaneously modeling conditional and time heteroskedasticity have been introduced in the literature by defining the volatility as a function not only of the past values but also of the time, e.g., GARCH model with change points (the piecewise GARCH model of Mikosch and Stărică, 2004) and local time homogeneous model with change points (Mercurio and Spokoiny, 2002). A general continuous time model to perform this may be found in Fan, Jiang, Zhang, and Zhou (2002). One can also obtain a similar model for discrete time series by introducing past information into the mean and volatility functions in the indexed stochastic model proposed by Yao and Morgan (1999). Another proposal in this context is the time heteroskedastic stochastic volatility model (Härdle, Spokoiny, and Teyssière, 2000).

In this paper another approach, called a semiparametric GARCH (SEMIGARCH) model is proposed by introducing a scale function $\sigma(t)$ into the parametric GARCH model. This proposal is motivated by the observation that one important reason for the time heteroskedasticity is a slowly changing scale function in volatility. The advantages of this approach are as follows. 1. The volatility is decomposed into two multiplicative components corresponding to the location and the past information, respectively. 2. The GARCH parameters are estimated globally, and hence asymptotically normal, $\sqrt{n}$-consistent estimators are available. 3. The SEMIGARCH model can also be used for predicting the future volatility. A semiparametric estimation procedure combining kernel estimation of the scale function and maximum likelihood estimation of the GARCH parameters is proposed. Asymptotic properties of the estimators are investigated in detail. A data-driven algorithm is developed for practical implementation. Finite sample performance of the proposal is examined through a simulation study. The proposal is applied to model CH and scale change in the daily S&P 500 and DAX 100 returns. It is shown that both series have simultaneously significant scale change and CH.

This approach provides an interesting alternative for modeling financial volatility. Whether or not it is better than another approach depends on the case considered. The idea proposed in this paper can be used to obtain semiparametric generalizations of other GARCH variants. Change points can also be introduced into the SEMIGARCH model.

The paper is organized as follows. Section 2 introduces the model. Section 3 describes the semiparametric estimation procedure. Asymptotic properties of the proposals are investigated in Section 4. Section 5 proposes a data-driven algorithm for practical implementation. 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 discussion. Proofs of results are in the Appendix.

2. THE MODEL

Consider the equidistant time series model

$$Y_i = \mu + \sigma(t_i)\epsilon_i, \quad i = 1,2,\ldots,n, \tag{1}$$

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

$$\epsilon_i = \eta_i h_i^{1/2}, \qquad h_i = \alpha_0 + \sum_{j=1}^{r} \alpha_j \epsilon_{i-j}^2 + \sum_{k=1}^{s} \beta_k h_{i-k} \tag{2}$$

(Bollerslev, 1986), where η_i are independent and identically distributed (i.i.d.) $N(0,1)$ random variables, $\alpha_0 > 0$ and $\alpha_1,\ldots,\alpha_r,\beta_1,\ldots,\beta_s \geq 0$. Let $v(t) = \sigma^2(t)$ denote the local variance of Y_i. The rescaled time index $t_i = i/n$ is introduced to guarantee that the information increases as n increases and the availability of a consistent estimator of v. Now, model (1) defines indeed a sequence of processes.

Let $\theta = (\alpha_0,\alpha_1,\ldots,\alpha_r,\beta_1,\ldots,\beta_s)'$ be the unknown parameter vector. It is assumed that $\sum_{i=1}^{r}\alpha_i + \sum_{j=1}^{s}\beta_j < 1$, which ensures the existence of a unique strictly stationary solution of (2). The practical implementation of a nonparametric estimator $\hat{v}(t)$ requires the moment condition $E(\epsilon_i^8) < \infty$. However, as pointed out by an anonymous referee, the condition of $E(\epsilon_i^4) < \infty$ is sufficient for the derivation of the asymptotic results. Necessary and sufficient conditions that 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). It is further assumed $\mathrm{var}(\epsilon_i) = E(\epsilon_i^2) = 1$, implying $\alpha_0 = 1 - \sum_{i=1}^{r}\alpha_i - \sum_{j=1}^{s}\beta_j$, to avoid identifiability problems.

The process defined by (1) and (2) is locally stationary in the sense of Dahlhaus (1997), which is a special case of Example 1 given there. Such a model provides a semiparametric extension of the standard GARCH model (Bollerslev, 1986) by introducing the scale function $\sigma(t)$ into it, where $h_i^{1/2}$ stand for the conditional standard deviations of the standardized process ϵ_i. The total standard deviation at t_i is hence given by $\sigma(t_i)h_i^{1/2}$. For $\sigma(t) \equiv \sigma_0$, model (1) and (2) reduces to the standard GARCH model. Our purpose is to estimate $v(t)$ and h_i separately. If the scale function $\sigma(t)$ in (1) changes over time, then the assumption of a 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 nonconstant scale function is not eliminated, one will obtain $\hat{\alpha}_1 + \hat{\beta}_1 \to 1$ by a fitted GARCH(1,1) model as $n \to \infty$, even when ϵ_i are i.i.d. 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 GARCH model but model (1) and (2) is used, then the estimation is still $\sqrt{n}$-consistent but with some loss in efficiency due to the estimation of $\sigma(t)$.

The assumptions of model (1) and (2) can be weakened in different ways. For instance, if the constant mean μ in (1) is replaced by a smooth mean function g, then we obtain the following nonparametric regression with heteroskedastic and dependent errors:

$$Y_i = g(t_i) + \sigma(t_i)\epsilon_i, \tag{3}$$

where $\{\epsilon_i\}$ is a zero mean stationary process. Estimation of the mean function g in model (3) with i.i.d. ϵ_i is discussed in, e.g., 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 in, e.g., Efromovich (1999). This paper focuses on investigating the estimation of $\sigma(t)$ and θ under model (1) and (2).

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 θ. A linear smoother of the squared residuals will estimate $v(t)$. Let $Z_i = (Y_i - \mu)$. Then model (1) can be rewritten as follows:

$$X_i = v(t_i) + v(t_i)\xi_i, \tag{4}$$

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 (for a related idea, see Efromovich, 1999, Sect. 4.3). 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.

The kernel estimator of conditional variance proposed by Feng and Heiler (1998) will be adapted to estimate $v(t)$. Let $y_1, \ldots, y_n$, 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 $[-1,1]$. The Nadaraya–Watson estimator of v at t based on $\hat{x}_i$ is defined by

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

where $w_i = K((t_i - t)/b)[\sum_{i=1}^{n} K((t_i - t)/b)]^{-1}$ and b is the bandwidth. And we define $\hat{\sigma}(t) = \sqrt{\hat{v}}$. It is assumed that $b \to 0$, $nb \to \infty$ as $n \to \infty$, which

https://doi.org/10.1017/S0266466604203061 Published online by Cambridge University Press

together with other regular conditions ensures the consistency of $\hat{\sigma}(t)$. The estimator defined in (5) does not depend on the dependence structure of the errors because $\hat{v}$ is a linear smoother. It is clear that $\hat{v} > 0$ if all observations for which $|t_i - t| \leq b$ are not identical. The bias of $\hat{v}$ at a boundary point is of a larger order than in the interior because of 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, Tsybakov, and Yang, 1998). However, as mentioned in Feng and Heiler (1998), a local linear estimator of v may sometimes be nonpositive. Hence, the kernel estimator is more preferable in the current context.

Following Bollerslev (1986), the conditional Gaussian log-likelihood in a parametric GARCH model takes the form (ignoring constants)

$$L(\theta) = \frac{1}{n}\sum_{i=1}^{n} l_i, \quad \text{where } l_i = -\frac{1}{2}\ln(h_i(\epsilon;\theta)) - \frac{\epsilon_i^2}{2h_i(\epsilon;\theta)}. \tag{6}$$

The maximizer of $L(\theta)$, denoted by $\tilde{\theta}$, is not available, because ϵ_i are unobservable in the current context. Hence we define the approximate log-likelihood by

$$\hat{L}(\theta) = \frac{1}{n}\sum_{i=1}^{n} l_i, \quad \text{where } l_i = -\frac{1}{2}\ln(h_i(\hat{\epsilon};\theta)) - \frac{\hat{\epsilon}_i^2}{2h_i(\hat{\epsilon};\theta)}, \tag{7}$$

where $\hat{\epsilon}_i$ are the standardized residuals given by

$$\hat{\epsilon}_i = \hat{z}_i/\hat{\sigma}(t_i) = (y_i - \bar{y})/\hat{\sigma}(t_i). \tag{8}$$

The symbols $h_i(\epsilon;\theta)$ and $h_i(\hat{\epsilon};\theta)$ are used to indicate that, for a given value of θ, $h_i(\epsilon;\theta)$ in $L(\theta)$ depends on ϵ_i and $h_i(\hat{\epsilon};\theta)$ in $\hat{L}(\theta)$ on $\hat{\epsilon}_i$. Similar to the parameter estimation in the SEMIFAR (semiparametric fractional autoregressive) model (Beran, 1999), θ will be estimated by $\hat{\theta}$, the maximizer of $\hat{L}(\theta)$. Any standard GARCH packet can be used for estimating $\hat{\theta}$ from $\hat{\epsilon}_i$. In this paper the S+ GARCH will be used. $\hat{\theta}$ obtained in this way is an approximate maximum likelihood estimator (MLE), which may perform differently from $\tilde{\theta}$ (provided $\tilde{\theta}$ were available).

4. MAIN RESULTS

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

A1. Model (1) and (2) holds with i.i.d. $N(0,1)$ η_i and strictly stationary ϵ_i such that $E(\epsilon_i^4) < \infty$. Furthermore, it is assumed that $\sum_{i=1}^{r} \alpha_i > 0$.

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

https://doi.org/10.1017/S0266466604203061 Published online by Cambridge University Press

A3. The kernel $K(u)$ is a symmetric density with compact support $[-1,1]$.

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

Assumptions A2–A4 are regular conditions in nonparametric regression. A1 summarizes conditions required on the GARCH model. For a GARCH(1,1) model, these conditions are stronger than those used by, e.g., Lee and Hansen (1994) and Lumsdaine (1996). Now, the condition $E(\epsilon_i^4) < \infty$ implies in particular $\alpha_1 + \beta_1 < 1$, and hence $E[\ln(\alpha_1 \eta_i^2 + \beta_1)] < 0$, one of the conditions used by Lee and Hansen (1994) and Lumsdaine (1996). In this paper the innovations η_i are assumed to be i.i.d. $N(0,1)$ random variables as in, e.g., Bollerslev (1986) and Ling and Li (1997) for simplicity, which implies Assumption 2 in Lumsdaine (1996). If non-Gaussian innovations are considered, suitable moment conditions have to be used, which might depend on the orders of the GARCH model. For instance, for a GARCH(1,1) model, Lumsdaine (1996) introduces the moment condition $E(\eta_i^{32}) < \infty$ together with further regular conditions on the distribution of η_i (Assumption 2 therein). Furthermore, it can be shown that, under A1, other assumptions in Lee and Hansen (1994) hold. The additional assumption $\sum_{i=1}^r \alpha_i > 0$ in A1 is introduced to avoid the naive case with $\alpha_i \equiv 0$ for all $i = 1,\dots,r$.

4.1. Asymptotic Properties of $\hat{v}$

Equation (4) is a nonparametric regression model with dependent and heteroskedastic errors. Pointwise results in nonparametric regression with dependent errors as given in, e.g., Altman (1990) and Hart (1991) can be adapted to $\hat{v}$ defined in (5) without any difficulty. Let $\gamma_\xi(k)$ denote the autocovariance function of ξ_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 ξ_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 ξ_i:

$$\xi_i = \sum_{j=1}^{r'} \alpha_j' \xi_{i-j} - \sum_{k=1}^{s} \beta_k u_{i-k} + u_i, \tag{9}$$

where $\alpha_j' = \alpha_j + \beta_j$ for $j \le \min(r,s)$, $\alpha_j' = \alpha_j$ for $j > s$, if $r > s$, and $\alpha_j' = \beta_j$ for $j > r$, if $s > r$, and

$$u_i = \epsilon_i^2 - h_i = (\eta_i^2 - 1)h_i \tag{10}$$

is a sequence of zero mean, uncorrelated random variables with independent $\eta_i \sim N(0,1)$. Equations (9) and (10) 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–A4 we have the following results.*

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

$$E[\hat{v}(t) - v(t)] = \frac{I(K)v''(t)}{2} b^2 + o(b^2). \tag{11}$$

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

$$\operatorname{var}[\hat{v}(t)] = 2\pi c_f R(K) \frac{v^2(t)}{nb} + o\left(\frac{1}{nb}\right). \tag{12}$$

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

$$(nb)^{1/2}(\hat{v}(t) - v(t)) \xrightarrow{\mathcal{D}} N(dD, V(t)), \tag{13}$$

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 $\hat{v}$ it is similar to that in nonparametric regression with short-range dependence, which depends, however, on the unknown underlying function v itself.

Let $\phi(z) = 1 - \sum_{i=1}^{r'} \alpha_i' z^i$ and $\psi(z) = 1 - \sum_{j=1}^{s} \beta_j z^j$. Under A1 we have

$$c_f = \frac{E(\epsilon_i^4)}{3\pi} \frac{|\psi(1)|^2}{|\phi(1)|^2} = \frac{E(\epsilon_i^4)}{3\pi} \frac{\left(1 - \sum_{j=1}^{s} \beta_j\right)^2}{\left(1 - \sum_{i=1}^{r} \alpha_i - \sum_{j=1}^{s} \beta_j\right)^2}. \tag{14}$$

If ϵ_i follows a GARCH(1,1) model, we have

$$\begin{aligned} c_f &= \frac{1}{\pi} \frac{\alpha_0^2(1 + \alpha_1 + \beta_1)(1 - \beta_1)^2}{(1 - \alpha_1 - \beta_1)^3(1 - 3\alpha_1^2 - 2\alpha_1\beta_1 - \beta_1^2)} \\ &= \frac{1}{\pi} \frac{(1 + \alpha_1 + \beta_1)(1 - \beta_1)^2}{\alpha_0(1 - 3\alpha_1^2 - 2\alpha_1\beta_1 - \beta_1^2)}. \end{aligned} \tag{15}$$

The last equation in (15) is due to the standardization of ϵ_i. The proof of (14) and (15) 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.

https://doi.org/10.1017/S0266466604203061 Published online by Cambridge University Press

THEOREM 2. *Under the assumptions of Theorem 1 we have the following results.*

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

$$\text{MISE} = \int_{\Delta}^{1-\Delta} E[\hat{v}(t) - v(t)]^2\, dt$$

$$= \frac{I^2(K)I((v'')^2)}{4} b^4 + 2\pi c_f R(K) \frac{I(v^2)}{nb} + o[b^4 + (nb)^{-1}]. \tag{16}$$

(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*

$$b_{\mathrm{A}} = C_{\mathrm{A}} n^{-1/5} \tag{17}$$

with

$$C_{\mathrm{A}} = \left(2\pi c_f \frac{R(K)}{I^2(K)} \frac{I(v^2)}{I((v'')^2)}\right)^{1/5}. \tag{18}$$

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

4.2. Asymptotic Properties of $\hat{\theta}$

Asymptotic properties of $\tilde{\theta}$ defined in Section 3 are investigated by Ling and Li (1997) under the general fractionally autoregressive integrated moving average–GARCH (FARIMA-GARCH) framework. More detailed asymptotic results in the special case of a GARCH(1,1) model may be found in Lee and Hansen (1994) and Lumsdaine (1996). Asymptotic properties of $\hat{\theta}$ will be studied by comparing its performance with that of $\tilde{\theta}$ based on the results in Ling and Li (1997). At first we will introduce a general lemma. Let $\theta_0 = (\theta_1^0, \dots, \theta_m^0)'$ be the true value of a m-dimensional parameter vector θ and be in the interior of the compact set Θ. Assume that there exists a consistent MLE $\tilde{\theta}$ satisfying the equation $\partial L(\theta)/\partial\theta = 0$, where $L(\theta)$ is a standard likelihood or log likelihood function. Furthermore, assume that $L(\theta)$ is three times differentiable, $L''(\theta)$ converges in probability to a positive definite matrix, and all third-order partial derivatives of $L(\theta)$ have bounded expectations in Θ. Let $\hat{L}(\theta)$ be a consistent estimate of $L(\theta)$. Then we have the following result.

LEMMA 1. *Assume* $\hat{L}(\theta) \xrightarrow{p} L(\theta)$ *for* θ *in a neighborhood of* θ_0. *Under the preceding regular conditions on* $L(\theta)$ *there exists a consistent* MLE $\hat{\theta}$ *satisfying* $\partial\hat{L}(\theta)/\partial\theta = 0$ *and*

$$(\hat{\theta} - \tilde{\theta}) \doteq -[\hat{L}''(\tilde{\theta})]^{-1}\hat{L}'(\tilde{\theta})$$
$$= O_p[\hat{L}'(\tilde{\theta})]. \tag{19}$$

The proof of Lemma 1 is straightforward and is omitted. Lemma 1 ensures the existence of an approximate MLE and provides a tool to quantify the distance between it and an infeasible MLE. Note that $\tilde{\theta}$ is in general $\sqrt{n}$-consistent and asymptotically normal. Hence, $\hat{\theta}$ will have the same properties if $\hat{L}'(\tilde{\theta}) = o_p(n^{-1/2})$.

Now, denote by $\theta_0 = (\alpha_0^0, \alpha_1^0, \ldots, \alpha_r^0, \beta_1^0, \ldots, \beta_s^0)'$ the true value of the unknown parameter vector θ. Assumption A1 ensures that θ_0 is in the interior of a compact parameter set Θ. Let $\hat{\theta}$ and $\tilde{\theta}$ be as defined in Section 3. Let

$$\Omega_\theta = E\left[\frac{1}{2h_i^2(\epsilon;\theta)} \frac{\partial h_i(\epsilon;\theta)}{\partial\theta} \frac{\partial h_i(\epsilon;\theta)}{\partial\theta'}\right] \tag{20}$$

and Ω_0, the value of Ω_θ at $\theta = \theta_0$, denote the information matrix. Then, following Lemma 1 and Theorems 3.1 and 3.2 in Ling and Li (1997), we have the following result.

THEOREM 3. *Assume that A1–A4 hold.*

(i) There exists a MLE $\hat{\theta}$ *satisfying* $\partial\hat{L}(\theta)/\partial\theta = 0$ *and* $\hat{\theta} \xrightarrow{p} \theta_0$ *as* $n \to \infty$.
(ii) Let $B_\theta = E(\hat{\theta} - \tilde{\theta})$. *Then* $\sqrt{n}(\hat{\theta} - B_\theta - \theta_0) \xrightarrow{\mathcal{D}} N(0, \Omega_0^{-1})$.
(iii) The bias vector B_θ *defined in (ii) is of the order of magnitude* $O[b^2 + (nb)^{-1}]$.

We see that $\hat{\theta}$ is $\sqrt{n}$-consistent and asymptotically normal up to a bias term B_θ. The proof of Theorem 3 is given in the Appendix and shows that the $O(b^2)$ term in B_θ is due to $E[\hat{v}(t_i) - v(t_i)]$ and the $O[(nb)^{-1}]$ term is due to $\operatorname{cov}[\epsilon_i^2, \hat{v}(t_i)]$. If $O(n^{-1/2}) < b < O(n^{-1/4})$, B_θ is negligible, and we have $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{D}} N(0, \Omega_0^{-1})$. Similar observations have been made in other semiparametric contexts, e.g., within the context of partially linear models. There, for a certain choice of bandwidth the nonparametric part has no effect on the rate of convergence of the parametric estimator (see Härdle, Liang, and Gao, 2000). If $\hat{\theta}$ is estimated using $b = O(b_A)$, then $B_\theta = O(n^{-2/5})$. If Y_i follow a GARCH model and $b > O(n^{-1/2})$, then $\hat{\theta}$ is $\sqrt{n}$-consistent and asymptotically normal because now $\hat{v}$ is unbiased.

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 (18) with some suitable estimators. At first, it is proposed to estimate c_f by

$$\hat{c}_f = \frac{\hat{E}(\epsilon_i^4)}{3\pi} \frac{\left(1 - \sum_{j=1}^{s} \hat{\beta}_j\right)^2}{\left(1 - \sum_{i=1}^{r} \hat{\alpha}_i - \sum_{j=1}^{s} \hat{\beta}_j\right)^2}, \tag{21}$$

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 formulas of $E(\epsilon_i^4)$ are known (for common results, see He and Teräsvirta, 1999a; Karanasos, 1999; for results in some special cases, see Bollerslev, 1986; He and Teräsvirta, 1999b), we prefer to use $\hat{c}_f$ defined in (21) because the formulas 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 (15) by replacing α_0, α_1, and β_1 with their estimates. Now $\hat{c}_f$ and $\tilde{c}_f$ perform quite similarly. Assume that a bandwidth b_ϵ is used for estimating $E(\epsilon_i^4)$, which satisfies A4 but is not necessarily the same as b. Furthermore, make the following assumption.

A1′. The same as A1 but with $E(\epsilon_i^8) < \infty$.

Then the following proposition holds.

PROPOSITION 1. *Under Assumptions A1′ and A2–A4 we have*

$$E[\hat{E}(\epsilon_i^4) - E(\epsilon_i^4)] = O(b_\epsilon^2) + O([nb_\epsilon]^{-1}) \tag{22}$$

and

$$\mathrm{var}(\hat{E}(\epsilon_i^4)) = 2\pi c_f^\epsilon n^{-1}[1 + o(1)], \tag{23}$$

where c_f^ϵ denotes the value of the spectral density of the process ϵ_i^4 at the origin.

The proof of Proposition 1 is given in the Appendix.

Remark 1. Equations (22) and (23) show that $\hat{E}(\epsilon_i^4)$ is $\sqrt{n}$-consistent, if $O(n^{-1/2}) \le b_\epsilon \le O(n^{-1/4})$. The optimal bandwidth in a second-order sense, which balances the two terms on the right-hand side of (22), 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 longer $\sqrt{n}$-consistent if a bandwidth $b_\epsilon = O(b_{\mathrm{A}}) = O(n^{-1/5})$ is used. The finally selected bandwidth is not so sensitive to the bandwidth for estimating $E(\epsilon_i^4)$.

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

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

https://doi.org/10.1017/S0266466604203061 Published online by Cambridge University Press

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

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

$$E[\hat{I}(v^2) - I(v^2)] = O(b_v^2) + O([nb_v]^{-1}) \tag{25}$$

and

$$\operatorname{var}(\hat{I}(v^2)) = O(n^{-1}) + O([n^{-2}b_v^{-1}]). \tag{26}$$

The proof of Proposition 2 is given in the Appendix.

Remark 2. Note that the dominated orders of the biases and variances of $\hat{E}(\epsilon_i^4)$ and $\hat{I}(v^2)$ are the same. Hence similar statements as given in Remark 1 apply for results given in (25) and (26). This is not surprising because both $v^2(t_i)$ and ϵ_i^4 are related to the fourth moment of the errors.

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

$$\hat{I}((v'')^2) = \frac{1}{n}\sum_{i=n_1}^{n_2} \hat{v}''(t_i)^2 \tag{27}$$

(see, e.g., Ruppert, Sheathec, and Wand, 1995), where $\hat{v}''$ is a kernel estimator of v'' using a fourth-order kernel K_2 for estimating the second derivative (see, e.g., 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 compact support $[-1,1]$.

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

PROPOSITION 3. *Under Assumptions A1′–A4′ we have*

$$E[\hat{I}((v'')^2) - I((v'')^2)] \doteq O(b_d^2) + O(n^{-1}b_d^{-5}) \tag{28}$$

and

$$\operatorname{var}(\hat{I}((v'')^2)) \doteq O(n^{-1}) + O(n^{-2}b_d^{-5}). \tag{29}$$

The proof of Proposition 3 is omitted because it is well known in nonparametric regression (for results with i.i.d. errors, see, e.g., Ruppert et al., 1995; for results with dependent errors, see, e.g., Beran and Feng, 2002a, 2000b).

https://doi.org/10.1017/S0266466604203061 Published online by Cambridge University Press

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 (28), 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 that for selecting the bandwidth b we have to choose at first three pilot bandwidths b_ϵ, b_v, and b_d. This problem will be solved using the iterative plug-in idea (Gasser, Kneip, and Köhler, 1991) with a so-called exponential inflation method (see Beran and Feng, 2002a, 2002b). Let b_{j-1} denote the bandwidth for estimating v in the $(j-1)$th iteration. Then in the jth 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 $b_{\epsilon,j}$ and $b_{v,j}$ 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 $O_p(n^{-1/5})$. By an iterative plug-in algorithm the unknown constants in the pilot bandwidths can be simply omitted. Furthermore, we also need to choose a starting bandwidth b_0. In the current context, b_0 should satisfy A4 because we have to estimate θ 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 discussions on this topic may be found in the next two sections, especially in Section 6.3.

The proposed data-driven algorithm processes as follows:

1. Start with the bandwidth $b_0 = c_0 n^{-1/5}$ with, e.g., $c_0 = 0.5$.
2. In the jth 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

$$b_j = \left(2\pi\hat{c}_f \frac{R(K)}{I^2(K)} \frac{\hat{I}(v^2)}{\hat{I}((v'')^2)}\right)^{1/5} n^{-1/5}; \tag{30}$$

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 completed. Put $\hat{b} = b_j$.

The condition $|b_j - b_{j-1}| < 1/n$ is used as a convergence criterion of $\hat{b}$, because such a difference is negligible. The maximal number of iterations is put to be 20. In this algorithm, $\hat{\theta}$ is estimated using b_{j-1} as for $\hat{v}$ because we do not have a proper bandwidth selector for estimating θ. The asymptotic performance of $\hat{b}$ is quantified by the following theorem

THEOREM 4. *Assume that A3 and A1′–A3′ hold and that* $I((v'')^2) \neq 0$. *Then we have*

$$(\hat{b} - b_A)/b_A \doteq O_p(n^{-2/7}) + O(n^{-2/5}). \tag{31}$$

The proof of Theorem 4 is given in the Appendix. Note that A4 and A4′ are automatically satisfied. The second $O(n^{-2/5})$ term on the right-hand side of (31) is due to the error in $\hat{c}_f$ caused by the bias in $\hat{\theta}$, which is indeed negligible compared with the first term.

The proposed algorithm is coded in an S-Plus function called *SEMIGARCH*. A practical restriction $1/n \le b \le 0.5 - 1/n$ is used in the program for simplicity. Four commonly used kernels, namely, the uniform, the Epanechnikov, the bisquare, and the triweight kernels (see, e.g., Müller, 1988), are built into 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 next two sections.

Remark 4. Note that b_{A} is not well defined if $I((v'')^2) = 0$ implying $v''(t) \equiv 0$. However, the SEMIGARCH model also applies to this case. In particular, the proposed algorithm does work if Y_i follow a GARCH model. Now it can be shown that, theoretically, $b_j \to O_p(1)$ as $j \to \infty$. Following the context after Theorem 3, $\hat{\theta}$ has the same asymptotic properties as by a GARCH model because $\hat{b} \gg O_p(n^{-1/2})$. And $\hat{v}(t)$ is $\sqrt{n}$-consistent with some loss in the efficiency compared with a parametric estimator, provided that no maximal number of iterations is given, because $(nb_j)^{-1} \to O_p(n^{-1})$ now.

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, ϵ_i were generated using the *simulate.garch* function in S+GARCH following one of the two GARCH(1,1) models.

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

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

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

$$
\begin{aligned}
v_1^{1/2}(t) &= \sigma_1(t) = 3.75 + t \\
&\qquad + (3\cos(2.75(t-0.5)\pi) + 22.5 + 2\tanh(2.75(t-0.5)\pi))/5, \\
v_2^{1/2}(t) &= \sigma_2(t) = \sigma_1(t) - 1.2, \quad \text{and} \\
v_3^{1/2}(t) &= \sigma_3(t) = 3 + \cos(4(t-0.25)\pi).
\end{aligned}
$$

The terms $v_1(t)$ and $v_2(t)$ are quite similar, and they are designed following the estimated scale function in the daily DAX 100 returns. The scale change with v_2 is stronger than that with v_1. It is strongest with v_3. To this end see the bandwidths required for estimating them given in Table 2. The scale function $\sigma_2(t)$ may be found in Figure 2b, which follows. To confirm the statements in

https://doi.org/10.1017/S0266466604203061 Published online by Cambridge University Press

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 $n = 1{,}000,\ 2{,}000,\ 4{,}000$. For each case 400 replications were done. For each replication, three GARCH (1,1) models were fitted to ϵ_i, the data-driven $\hat{\epsilon}_i$ and y_i. The estimators of α_1 and β_1 are denoted by $\hat{\alpha}_1^{\epsilon}$, $\hat{\beta}_1^{\epsilon}$, $\hat{\alpha}_1^{\hat{\epsilon}}$, $\hat{\beta}_1^{\hat{\epsilon}}$, $\hat{\alpha}_1^{y}$, and $\hat{\beta}_1^{y}$, respectively. For v_0 we have $\hat{\alpha}_1^{\epsilon} \equiv \hat{\alpha}_1^{y}$ and $\hat{\beta}_1^{\epsilon} \equiv \hat{\beta}_1^{y}$. Here, $\hat{\alpha}_1^{\epsilon}$ and $\hat{\beta}_1^{\epsilon}$ are used as a benchmark. Note in particular that the estimated parameters may sometimes be negative using the S+GARCH.

6.2. Results of the Simulation Study

To give a summary of the performance of $\hat{\alpha}_1^{\hat{\epsilon}}$ and $\hat{\beta}_1^{\hat{\epsilon}}$, 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 ϵ_i is calculated. For instance,

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

These results are listed in Table 1. The difference between two related EFFs, e.g., $\mathrm{EFF}(\hat{\beta}_1^{\hat{\epsilon}}) - \mathrm{EFF}(\hat{\beta}_1^{y})$, in a given case may be thought of as the gain by using the SEMIGARCH model. Table 1 shows that the EFFs of $\hat{\alpha}_1^{\hat{\epsilon}}$ and $\hat{\beta}_1^{\hat{\epsilon}}$ seem to tend to 100%, whereas those of $\hat{\alpha}_1^{y}$ and $\hat{\beta}_1^{y}$ seem to tend to zero, as $n \to \infty$. Hence, the gains seem to tend to 100% as $n \to \infty$. The EFFs of $\hat{\beta}_1^{\hat{\epsilon}}$

TABLE 1. Empirical efficiencies (%) of the estimated parameters

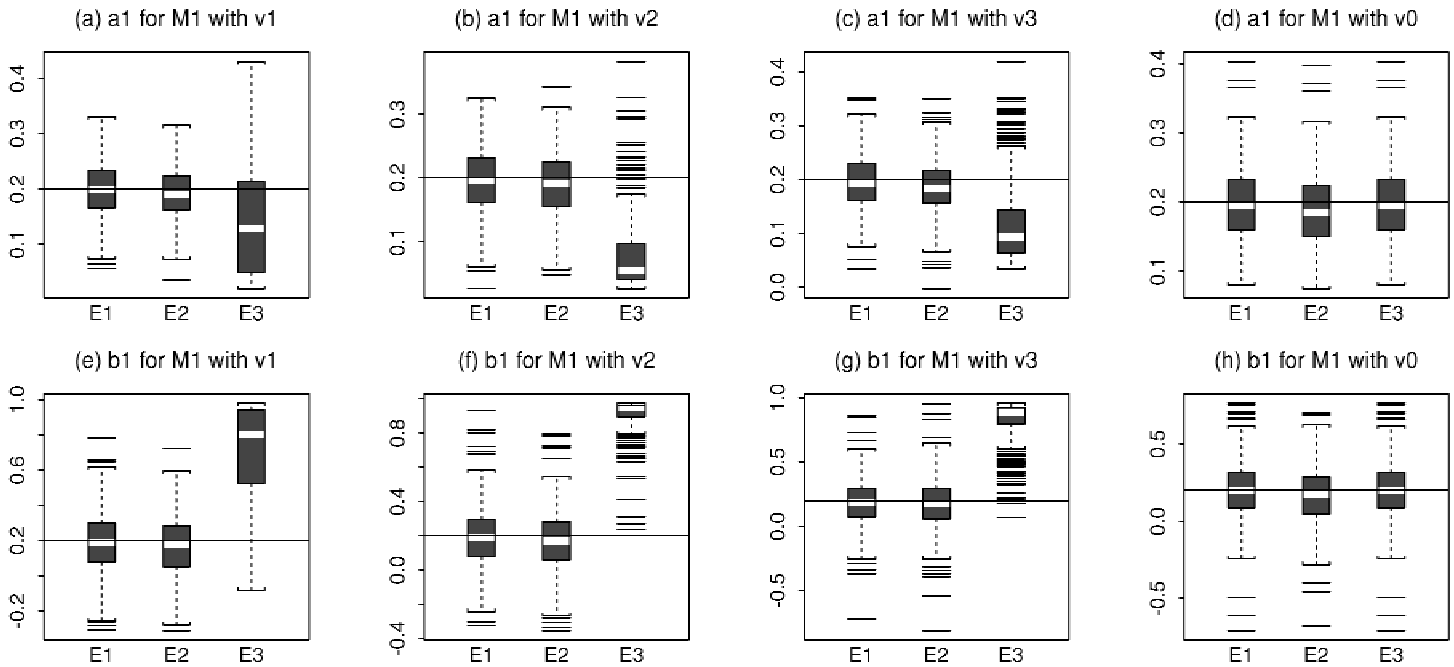
		Model 1				Model 2			
Parameter	n	v_1	v_2	v_3	v_0	v_1	v_2	v_3	v_0
$\hat{\alpha}_1^{\hat{\epsilon}}$	1,000	96.2	97.8	91.3	98.0	96.8	99.8	101.1	95.1
	2,000	99.0	94.0	92.5	97.2	100.7	96.5	102.2	99.8
	4,000	96.3	97.7	94.0	91.4	97.7	97.8	97.5	97.7
$\hat{\beta}_1^{\hat{\epsilon}}$	1,000	93.3	94.2	88.2	103.3	58.5	64.3	62.7	59.9
	2,000	99.2	92.7	91.2	99.8	75.2	89.8	85.4	71.7
	4,000	97.5	99.8	99.2	94.7	86.5	90.7	86.1	77.6
$\hat{\alpha}_1^{y}$	1,000	19.8	13.4	21.3	—	73.0	77.9	124.9	—
	2,000	10.9	5.1	7.8	—	45.3	36.6	58.2	—
	4,000	5.4	2.3	3.0	—	30.2	18.8	26.5	—
$\hat{\beta}_1^{y}$	1,000	9.5	6.0	8.8	—	73.1	41.4	73.0	—
	2,000	4.6	2.4	3.1	—	23.4	13.8	24.4	—
	4,000	1.8	1.1	1.3	—	10.7	8.5	6.8	—

under M2 are relatively low. In particular, for $n = 1{,}000$, the EFFs of $\hat{\beta}_1^{\hat{\epsilon}}$ 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 cases is slightly negative. This shows that $n = 1{,}000$ is sometimes not large enough for estimating the scale function when β_1 is large.

Box plots of the 400 replications of $\hat{\alpha}_1^{\epsilon}$, $\hat{\beta}_1^{\epsilon}$, $\hat{\alpha}_1^{\hat{\epsilon}}$, $\hat{\beta}_1^{\hat{\epsilon}}$, $\hat{\alpha}_1^y$, and $\hat{\beta}_1^y$ for $n =$ 1,000 are shown in Figures 1a–1f, where the symbols E1, E2, and E3 denote estimators obtained from ϵ_i, $\hat{\epsilon}_i$, and y_i, respectively. Those for $n = 2{,}000$ and $n = 4{,}000$ are omitted to save space. The simulation results show that $\hat{\alpha}_1^{\hat{\epsilon}}$ and $\hat{\beta}_1^{\hat{\epsilon}}$ perform in general quite well. One clear problem arises with $\hat{\beta}_1^{\hat{\epsilon}}$ under M2 with $n = 1{,}000$. Now, both the variance and the bias of $\hat{\beta}_1^{\hat{\epsilon}}$ are strongly affected by some extremely small estimates (see Figures 1m–1p). This is due to the non-robustness of the bandwidth selection. Hence, it is worthwhile to develop a robust procedure to improve the poor performance of $\hat{\beta}_1^{\hat{\epsilon}}$ for small n. The quality of $\hat{\alpha}_1^{\epsilon}$ and $\hat{\beta}_1^{\epsilon}$ is clearly improved as n increases. In particular, the estimation becomes more and more stable. Detailed statistics (in the first version of this paper) show that the standard deviations of $\hat{\alpha}_1^{\hat{\epsilon}}$ and $\hat{\beta}_1^{\hat{\epsilon}}$ seem to converge at the same rate as those of $\hat{\alpha}_1^{\epsilon}$ and $\hat{\beta}_1^{\epsilon}$, but their biases converge a little more slowly. This confirms the results of Theorem 3. The simulation results show clearly that, in the case with scale change, $\hat{\alpha}_1^y$ and $\hat{\beta}_1^y$ are inconsistent as a result of their biases. The situation become worse as n increases. In particular, we can see that $\hat{\beta}_1^y$ will tend to one as $n \to \infty$, no matter how large β_1^0 is. However, if there is no scale change the estimators $\hat{\alpha}_1^y$ and $\hat{\beta}_1^y$ should of course be used. It is hence helpful to test whether or not the estimated scale function is significant. For the data examples given in the next section it is proposed to carry out such a test based on simulation.

Now let us consider the quality of $\hat{b}$. The sample means, standard deviations, and square roots of the MSEs of $\hat{b}$ together with the true asymptotic optimal bandwidths b_{A} are given in Table 2. Note that b_{A} and the MSE in cases with v_0 are not defined. Kernel density estimates of $(\hat{b} - b_{\mathrm{A}})$ (omitted to save space) show that 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, 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 easiest 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 $n = 1{,}000$. In this case, the detailed structure of v_3 may sometimes be smoothed away because of the large variation caused by the GARCH model. This shows again that $n =$ 1,000 is sometimes not large enough for distinguishing the CH and the scale change.

Remark 5. As suggested by a referee, the performance of the proposed procedure for the cases with highly persistent GARCH effect is investigated through


(a) a1 for M1 with v1
(b) a1 for M1 with v2
(c) a1 for M1 with v3
(d) a1 for M1 with v0
(e) b1 for M1 with v1
(f) b1 for M1 with v2
(g) b1 for M1 with v3
(h) b1 for M1 with v0
E1
E2
E3

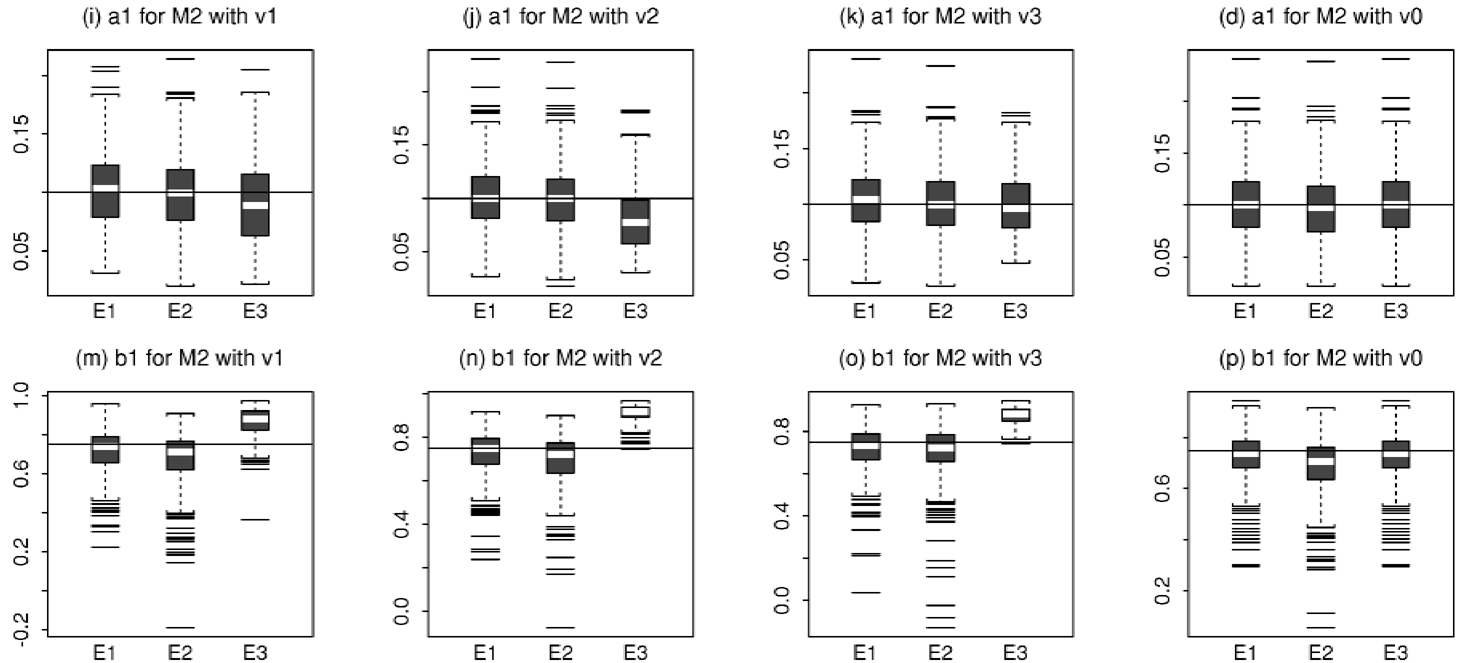


FIGURE 1. Box plots of $\hat{\alpha}_1^{\epsilon}$ and $\hat{\beta}_1^{\epsilon}$ (E1), $\hat{\alpha}_1^{\hat{\epsilon}}$ and $\hat{\beta}_1^{\hat{\epsilon}}$ (E2), and $\hat{\alpha}_1^{y}$ and $\hat{\beta}_1^{y}$ (E3), respectively, with $n = 1{,}000$, where the horizontal lines show the true values.

TABLE 2. Statistics on the selected bandwidth

		Model 1				Model 2			
n	Statistic	v_1	v_2	v_3	v_0	v_1	v_2	v_3	v_0
1,000	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
	$\mathrm{MSE}^{1/2}$	0.019	0.011	0.015	—	0.031	0.018	0.034	—
2,000	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
	$\mathrm{MSE}^{1/2}$	0.015	0.008	0.013	—	0.020	0.014	0.014	—
4,000	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
	$\mathrm{MSE}^{1/2}$	0.014	0.007	0.010	—	0.015	0.008	0.012	—

an additional simulation under a third model, model 3 (M3), with $\alpha_1 = 0.07$ and $\beta_1 = 0.87$ and without trend. As expected, the proposed procedure does not work well for $n = 1{,}000$ because the variance of $\hat{\alpha}_1^{\hat{\epsilon}}$ and in particular that of $\hat{\beta}_1^{\hat{\epsilon}}$ are too large as a result of some extreme estimates. This shows again that a robust estimation procedure should be developed. For $n \geq 2{,}000$, the procedure works well. The empirical efficiencies are a little lower than those for M2. Detailed results of this additional simulation are omitted to save space.

Remark 6. In this paper, the bandwidth is selected by minimizing the dominant part of the MISE of $\hat{v}$. In a semiparametric context, the performance of the bandwidth selection and the resulting parameter estimation may be improved if a plug-in algorithm that takes the MSE of $\hat{\theta}$ into account is developed. For this purpose a more detailed formula of the MSE of $\hat{\theta}$ is required, and one has to develop a suitable procedure to estimate the MSE. This is still an important open question and will be discussed elsewhere.

6.3. Detailed Analysis of Two Simulated Examples

In the following discussion, two simulated data sets are selected to show some details. The first example (called Sim 1) is a typical example of the replications under M2 with the scale function $\sigma_2(t)$ and $n = 2{,}000$. The observations y_i, $i = 1, \ldots, 2{,}000$, are shown in Figure 2a. 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. The $\sigma_2(t)$ (solid line) and $\hat{\sigma}_2(t)$ (dashed line) are shown

https://doi.org/10.1017/S0266466604203061 Published online by Cambridge University Press

in Figure 2b. Figure 2c shows the standardized residuals $\hat{\epsilon}_i$, which look stationary. The estimated GARCH(1,1) models are

$$h_i^y = 0.0363 + 0.0540 y_{i-1}^2 + 0.9432 h_{i-1}^y \tag{32}$$

for y_i and

$$h_i^{\hat{\epsilon}} = 0.2052 + 0.0937 \hat{\epsilon}_{i-1}^2 + 0.6965 h_{i-1}^{\hat{\epsilon}} \tag{33}$$

for $\hat{\epsilon}_i$. For model (32) 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 (33) has finite moments until at least twelfth order 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 2d and 2e. 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 2f and 2g. Figure 2h shows the estimated GARCH conditional (in this case also the total) standard deviations $(h_i^y)^{1/2}$. The analysis of Sim 1 shows the following results.

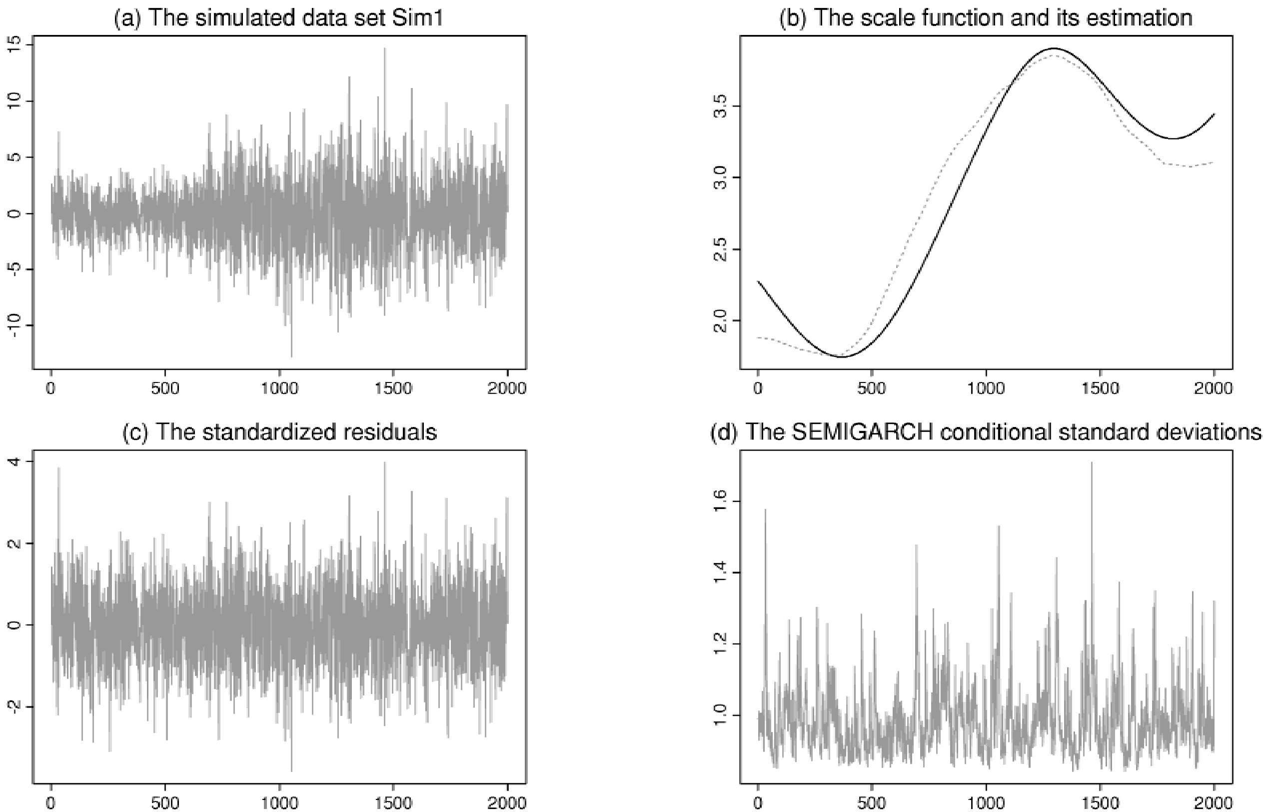
(1) If a standard GARCH model is used, the scale change will be wrongly estimated as a part of the CH. Furthermore, the total variance tends to be overestimated when it is large and underestimated when it is small (compare Figures 2g and 2h). This phenomenon is mainly due to the overestimation of $\hat{\beta}_1$ and will be called the (volatility) inflation effect of the GARCH model in the presence of scale change.

(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 2e and 2h). The errors in $\hat{\sigma}^2(t_i)h_i^{\hat{\epsilon}}$ are caused by the errors in these two estimates, and both of them can be clearly reduced if more dense observations are available, e.g., 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 as large as the former.

Furthermore, $(h_i^y)^{1/2}$ shown in Figure 2h (see also Figure 3f) 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 one of the replications under M1 with v_3 and $n = 1{,}000$, which 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 Section 5. For this data set we have either $\hat{b} = 0.012$ or $\hat{b} = 0.12$ 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. 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

https://doi.org/10.1017/S0266466604203061 Published online by Cambridge University Press


(a) The simulated data set Sim1
(b) The scale function and its estimation
(c) The standardized residuals
(d) The SEMIGARCH conditional standard deviations
0
500
1000
1500
2000
15
10
5
0
-5
-10
3.5
3.0
2.5
2.0
4
2
0
-2
1.6
1.4
1.2
1.0

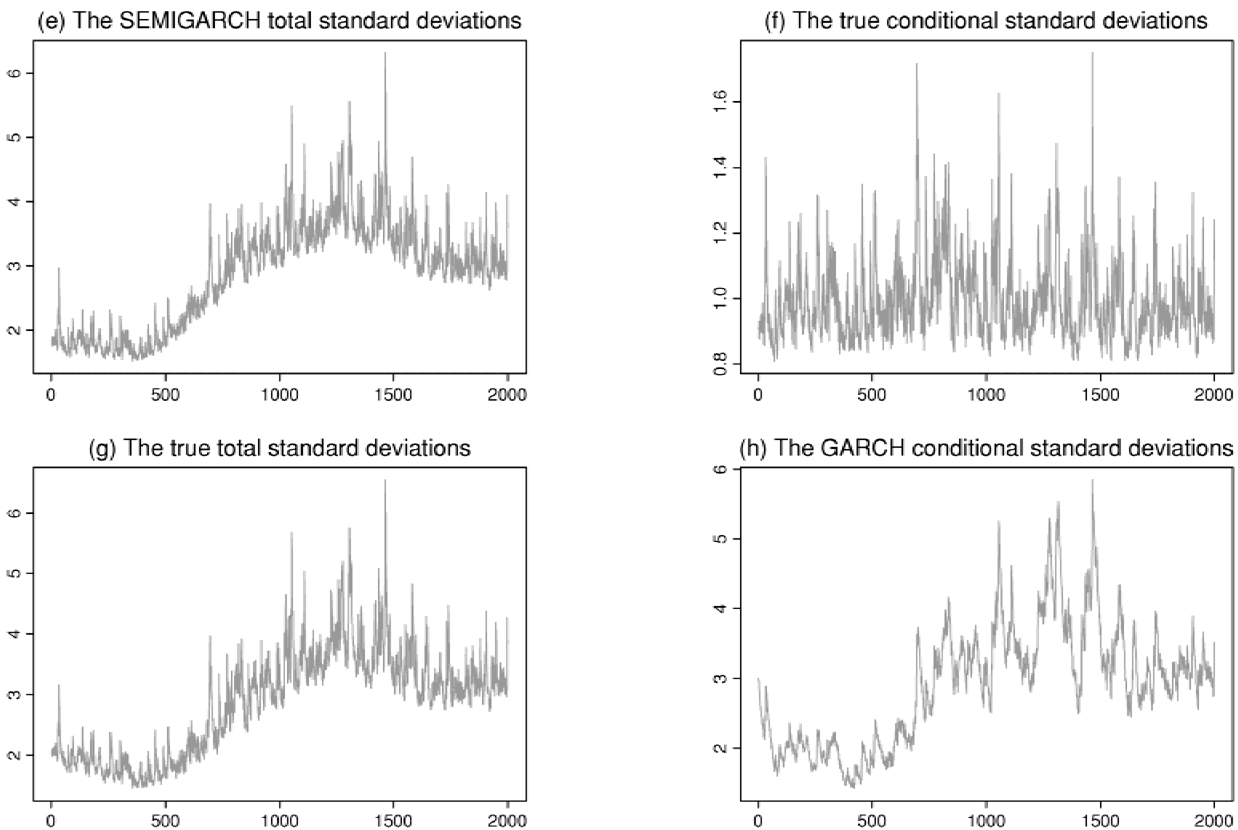


FIGURE 2. Estimation results for the first simulated data set.

case when it is doubtful, if the selected bandwidth with $b_0 = 0.5n^{-1/5}$ is the optimal one, we recommend that the user 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 indexes from January 3, 1994, to August 23, 2000. For the S&P 500 returns shown in Figure 3a we have $\hat{b} = 0.183$ (for any $b_0 \geq 0.075$). The fitted GARCH models are

$$h_i^y = 5.684 \times 10^{-7} + 0.0674 y_{i-1}^2 + 0.9302 h_{i-1}^y \tag{34}$$

for y_i and

$$h_i^{\hat{\epsilon}} = 0.0649 + 0.0686 \hat{\epsilon}_{i-1}^2 + 0.8676 h_{i-1}^{\hat{\epsilon}} \tag{35}$$

for $\hat{\epsilon}_i$. As before, for model (34) 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. Model (35) has finite moments until twelfth order. To test whether the estimated trend is significantly nonconstant, 400 replications were generated following model (35) with the corresponding sample variance and without trend. The scale function was then estimated with the bandwidth $b = 0.183$ from each replication. Symmetric Monte Carlo confidence bounds that covered 95% or 99% of all estimated trends were calculated and are shown in Figure 3b together with the sample standard deviation (0.0099) and the estimated scale function $\hat{\sigma}(t)$. We see that there is significant scale change in this data set. Furthermore, both $\hat{\alpha}_1$ and $\hat{\beta}_1$ in model (35) are strongly significant. That is, this series has simultaneously significant scale change and CH. Figures 3c–3f show $\hat{\epsilon}_i$, the SEMIGARCH conditional standard deviations $(h_i^{\hat{\epsilon}})^{1/2}$, the SEMIGARCH total standard deviations $\hat{\sigma}(t)(h_i^{\hat{\epsilon}})^{1/2}$, and the GARCH conditional standard deviations $(h_i^y)^{1/2}$. Comparing Figures 3e and 3f we see again that the estimated total variances following the SEMIGARCH model are more stable and those following the GARCH model are inflated.

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

$$h_i^y = 2.202 \times 10^{-6} + 0.0892 y_{i-1}^2 + 0.8957 h_{i-1}^y \tag{36}$$

for y_i and

$$h_i^{\hat{\epsilon}} = 0.0651 + 0.0873 \hat{\epsilon}_{i-1}^2 + 0.8481 h_{i-1}^{\hat{\epsilon}} \tag{37}$$

for $\hat{\epsilon}_i$. The condition for the existence of the fourth moment of model (36) is slightly satisfied, but the eighth moment of this model does not exist. Again,

https://doi.org/10.1017/S0266466604203061 Published online by Cambridge University Press

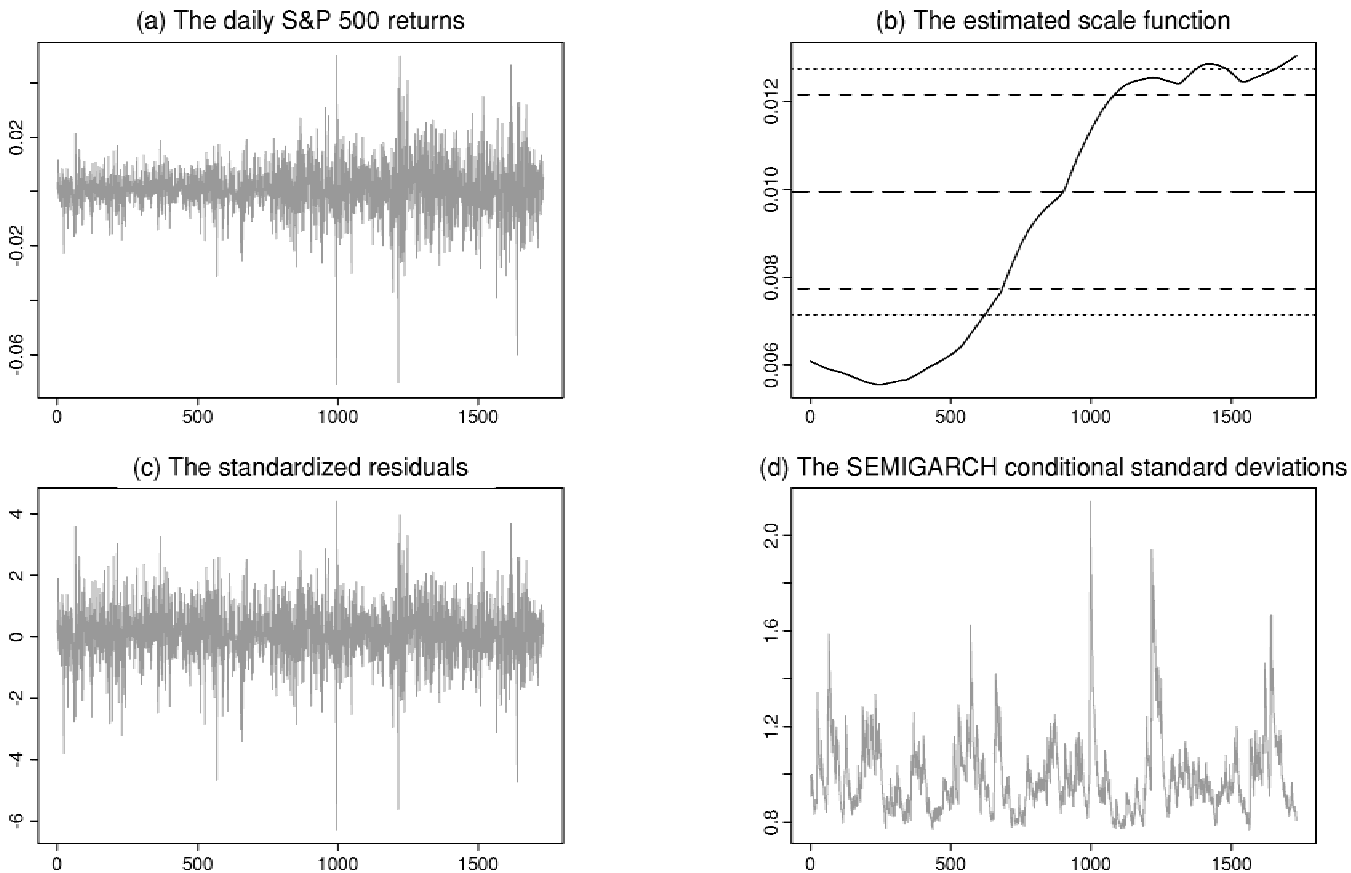
model (37) has finite moments until twelfth order. The S&P 500 and DAX 100 returns series perform quite similarly, and the conclusions on the former given previously apply to the latter.

Now, we will compare the performance of the GARCH and SEMIGARCH by predicting future volatility. The GARCH unconditional variance, $\hat{\sigma}^2$, say, is calculated following (34) or (36). For the SEMIGARCH, $\hat{\sigma}^2(t_n) = \hat{\sigma}^2(1)$ is used as the unconditional variance in the near future. The predicted (expected) conditional standard deviations $(\hat{h}^y_{n+k})^{1/2}$ following the GARCH and $\hat{\sigma}(1)(\hat{h}^{\hat{\epsilon}}_{n+k})^{1/2}$ following the SEMIGARCH, $k = 1,2,\ldots,100$, for the S&P 500 and DAX 100 returns are shown in Figure 4 together with $\hat{\sigma}$ and $\hat{\sigma}(1)$. Note that, the conditional standard deviations by both series at the right end are lower than $\hat{\sigma}(1)$. Consequently, $(\hat{h}^y_{n+k})^{1/2}$ and $\hat{\sigma}(1)(\hat{h}^{\hat{\epsilon}}_{n+k})^{1/2}$ increase for both series. The $\hat{\sigma}(1)(\hat{h}^{\hat{\epsilon}}_{n+k})^{1/2}$ look quite reasonable and converge to $\hat{\sigma}(1)$ quickly. However, $(\hat{h}^y_{n+k})^{1/2}$ in both cases seem to be underestimated, because of the inflation effect mentioned previously. Furthermore, $(\hat{h}^y_{n+k})^{1/2}$ converge very slowly to some wrongly estimated limits. The sample standard deviation for the S&P 500 returns is 0.0099. Following (34) we have $\hat{\sigma} = 0.0154$, which is clearly overestimated as a result of the instability of this model. For the DAX 100 returns, $\hat{\sigma}$ is about equal to its sample value, which is, however, clearly lower than the locally unconditional standard deviation at $t = 1$. There are two problems if the fitted parametric GARCH models from these data sets are used for predicting future volatility: (1) the unconditional variance at the current end was wrongly estimated; and (2) the predicted conditional variance converges very slowly, because these models only have finite moments of low orders. Both of these problems were overcome by applying the SEMIGARCH model.

8. DISCUSSION

The SEMIGARCH introduced in this paper provides a useful tool for estimating financial volatility in cases when the stationary assumption of a GARCH model is likely to break down, which decomposes the volatility into a smooth scale function of the location and a CH component depending on the past information. A data-driven algorithm is developed for practical implementation. Simulation and data examples show that the proposal works well in practice. There are some other recent proposals to deal with similar problems, e.g., the parametric GARCH model with change points (Mikosch and Stărică, 2004) for modeling structural breaks in the unconditional variance, which cannot be used for modeling slowly changing unconditional variance. On the other hand, structural breaks in the unconditional variance cannot be modeled by the SEMIGARCH. It is worthwhile to combine these two approaches. Another related work is Mercurio and Spokoiny (2002), where the volatility is assumed to be constant in some unknown time intervals. By this approach scale change and CH are modeled together but not separately.

https://doi.org/10.1017/S0266466604203061 Published online by Cambridge University Press



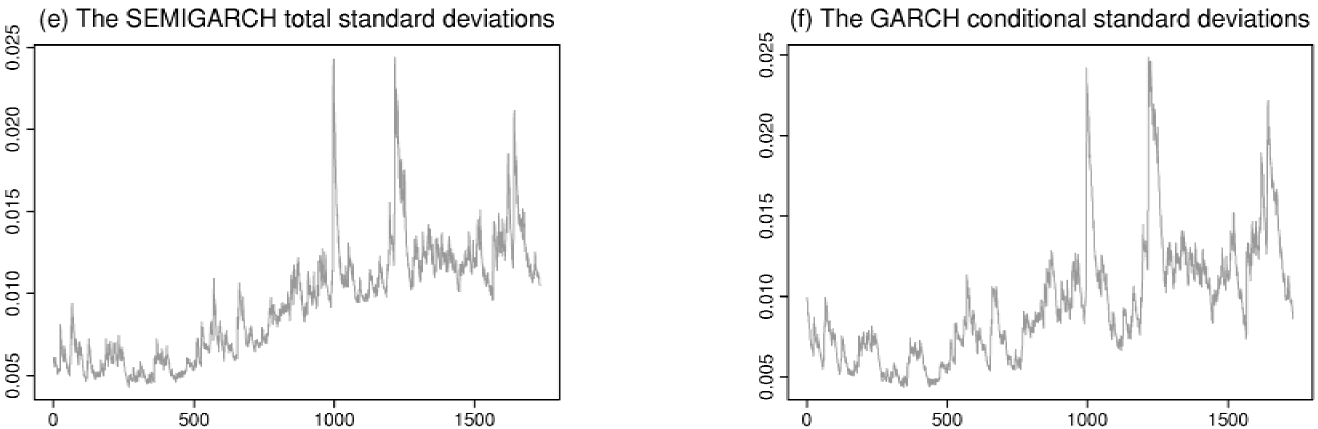


FIGURE 3. The estimation results for the S&P 500 returns.

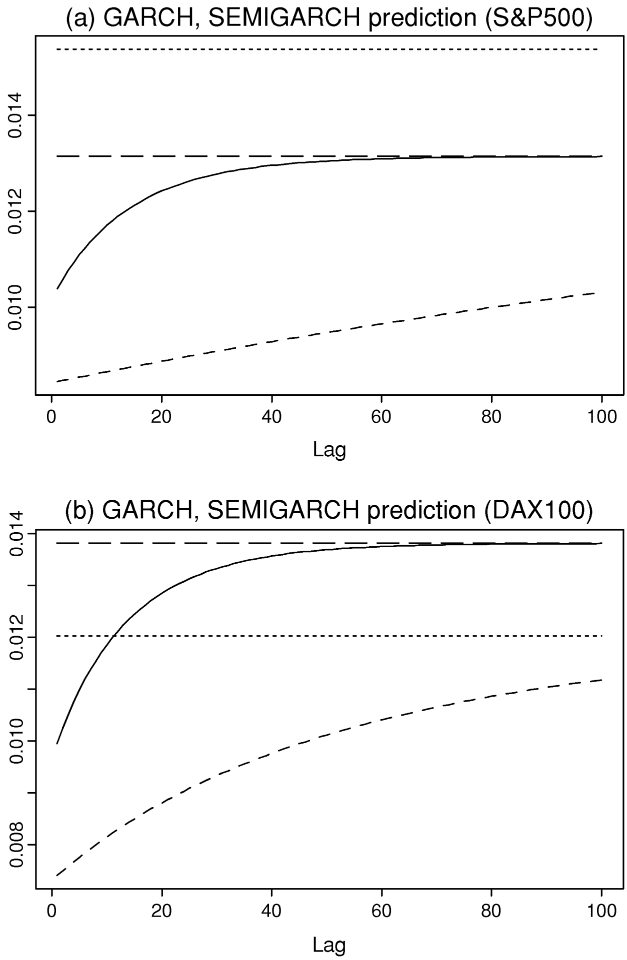


FIGURE 4. Predicted standard deviations $(\hat{h}^{y}_{n+k})^{1/2}$ (middle dashes) and $\hat{\sigma}(1)(\hat{h}^{\hat{\epsilon}}_{n+k})^{1/2}$ (solid line) together with their limits $\hat{\sigma}^2$ following the GARCH (short dashes) and $\hat{\sigma}^2(1)$ following the SEMIGARCH (long dashes) for (a) the S&P 500 and (b) the DAX 100 returns.

REFERENCES

Altman, N.S. (1990) Kernel smoothing with correlated errors. *Journal of the American Statistical Association* 85, 749–759.

Basrak, B., R.A. Davis, & T. Mikosch (2002) Regular variation of GARCH processes. *Stochastic Processes and Their Applications* 99, 95–115.

Beran, J. (1999) SEMIFAR Models: A Semiparametric Framework for Modelling Trends, Long Range Dependence, and Nonstationarity. CoFE discussion paper 99/16, University of Konstanz.
Beran, J. & Y. Feng (2001) Local polynomial estimation with a FARIMA-GARCH error process. *Bernoulli* 7, 733–750.
Beran, J. & Y. Feng (2002a) Local polynomial fitting with long-memory, short-memory and antipersistent errors. *Annals of the Institute of Statistical Mathematics* 54, 291–311.
Beran, J. & Y. Feng (2002b) Iterative plug-in algorithms for SEMIFAR models: Definition, convergence, and asymptotic properties. *Journal of Computational and Graphical Statistics* 11, 690–713.
Beran, J. & D. Ocker (2001) Volatility of stock market indices: An analysis based on SEMIFAR models. *Journal of Business and Economic Statistic* 19, 103–116.
Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.
Brockwell, P.J. & R.A. Davis (1991) *Time Series: Theory and Methods*. Springer.
Dahlhaus, R. (1997) Fitting time series models to nonstationary processes. *Annals of Statistics* 25, 1–37.
Ding, Z., C.W.J. Granger, & R.F. Engle (1993) A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83–106.
Efromovich, S. (1999) *Nonparametric Curve Estimation: Methods, Theory, and Applications.* Springer.
Engle, R.F. (1982) Autoregressive conditional heteroskedasticity with estimation of U.K. inflation. *Econometrica* 50, 987–1008.
Fan, J. & I. Gijbels (1995) Data-driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaptation. *Journal of the Royal Statistical Society, Series B* 57, 371–394.
Fan, J., J. Jiang, C. Zhang, & Z. Zhou (2002) Time-dependent diffusion models for term structure dynamics and the stock price volatility. *Statistica Sinica*, forthcoming.
Feng, Y. (2002) An iterative plug-in algorithm for nonparametric modelling of seasonal time series. CoFE discussion paper 02/04, University of Konstanz.
Feng, Y. & S. Heiler (1998) Locally weighted autoregression. In R. Galata & H. Küchenhoff (eds.), *Econometrics in Theory and Practice*, Festschrift für Hans Schneeweiß, pp. 101–117. Physica-Verlag.
Gasser, T., A. Kneip, & W. Köhler (1991) A flexible and fast method for automatic smoothing. *Journal of American Statistical Association* 86, 643–652.
Härdle, W., H. Liang, & J. Gao (2000) *Partially Linear Models*. Springer.
Härdle, W., V. Spokoiny, & G. Teyssière (2000) Adaptive Estimation for a Time Inhomogeneous Stochastic-Volatility Model. Discussion paper SFB 373, Humboldt University.
Härdle, W., A.B. Tsybakov, & L. Yang (1998) Nonparametric vector autoregression. *Journal of Statistical Planning and Inference* 68, 221–245.
Hart, J.D. (1991) Kernel regression estimation with time series errors. *Journal of the Royal Statistical Society, Series B* 53, 173–188.
He, C. & T. Teräsvirta (1999a) Fourth moment structure of the GARCH(p,q) process. *Econometric Theory* 15, 824–846.
He, C. & T. Teräsvirta (1999b) Properties of autocorrelation function of squared observations for second-order GARCH processes under two sets of parameter constraints. *Journal of Time Series Analysis* 20, 23–30.
Ibragimov, I.A. & Yu.V. Linnik (1971) *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff.
Karanasos, M. (1999) The second moment and autocovariance function of the squared errors of the GARCH model. *Journal of Econometrics* 90, 63–76.
Lee, S.-W. & B.E. Hansen (1994) Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory* 10, 29–52.
Ling, S. (1999) On probability properties of a double threshold ARMA conditional heteroskedasticity model. *Journal of Applied Probability* 36, 688–705.

Ling, S. & W.K. Li (1997) On fractional integrated autoregressive moving-average time series models with conditional heteroskedasticity. *Journal of the American Statistical Association* 92, 1184–1194.

Ling, S. & 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.

Lumsdaine, R.L. (1996) Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and GARCH(1,1) models. *Econometrica* 64, 575–596.

Mercurio, D. & V. Spokoiny (2002) Statistical Inference for Time-Inhomogeneous Volatility Models. Discussion paper SFB 373, Humboldt University.

Mikosch, T. & C. Stărică (2004) Change of structure in financial time series, long range dependence and the GARCH models. *Review of Economics and Statistics*, to appear.

Müller, H.G. (1988) *Nonparametric Analysis of Longitudinal Data*. Springer.

Ruppert, D., S.J. Sheather, & M.P. Wand (1995) An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association* 90, 1257–1270.

Ruppert, D. & M.P. Wand (1994) Multivariate locally weighted least squares regression. *Annals of Statistics* 22, 1346–1370.

Yao, Q. & B. Morgan (1999) Empirical transform estimation for indexed stochastic models. *Journal of the Royal Statistical Society, Series B* 61, 127–141.

APPENDIX: PROOFS OF RESULTS

Under model (1) and (2) $v(t)$ is integrable. This implies that $\bar{y}$ is $\sqrt{n}$-consistent. Hence, in the following discussion, $\hat{z}_i$ and $\hat{x}_i$ can be replaced by z_i and x_i respectively.

Proof of Theorem 1.

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

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

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 (11).

(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

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

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

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

https://doi.org/10.1017/S0266466604203061 Published online by Cambridge University Press

Insert this into (A.2), then we have

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

Results in (12) 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; Beran and Feng, 2002a).

(iii) Asymptotic normality. Consider the estimation problem under the model without scale change:

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

Define

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

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. Furthermore, following Theorem 4 in Beran and Feng (2001) it can be shown that the kernel estimator $\tilde{v}(t)$ is asymptotically normal if and only if the sample mean of the squared GARCH process ϵ_i^2 or equivalently the sample variance of ϵ_i is asymptotically normal. Basrak, Davis, and Mikosch (2002) show that the squared GARCH process ϵ_i^2 is strongly mixing with geometric rate. The condition $E(\epsilon_i^4) < \infty$ implies that there is a $\delta > 0$ such that $E|\epsilon_i^2|^{2+\delta} < \infty$. The conditions of Theorem 18.5.3 in Ibragimov and Linnik (1971) hold. This shows that $n^{-1}\sum \epsilon_i^2$ of a GARCH process with finite fourth moment is asymptotically normal. Theorem 1 is proved. ■

Proof of (14) and (15). Note that ξ_i has the autoregressive moving average (ARMA) representation

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

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 ξ is given by

$$f(\lambda) = \frac{\operatorname{var}(u_i)}{2\pi}\frac{|\psi(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2}$$

with

$$f(0) = \frac{\operatorname{var}(u_i)}{2\pi}\frac{(\psi(1))^2}{(\phi(1))^2}. \tag{A.8}$$

Note that $E(\epsilon_i^4) = 3E(h_i^2)$ (Bollerslev, 1986) and $\operatorname{var}(u_i) = E(u_i^2) = 2E(h_i^2)$. The last equation follows from (10). That is, $\operatorname{var}(u_i) = \frac{2}{3}E(\epsilon_i^4)$. The result in (14) is proved by inserting this formula, $\psi(1)$, and $\phi(1)$ into (A.8). The result in (15) is obtained by further inserting the explicit formula of $E(\epsilon_i^4)$ for a GARCH(1,1) model (Bollerslev, 1986) into (14). ■

The following analysis involves infinite past history of ϵ_i and $\hat{\epsilon}_i$. The presample values of ϵ_i and $\hat{\epsilon}_i$ will be assumed to be zero. The presample values of ϵ_i^2 and $h_i(\epsilon;\theta)$ (resp. $\hat{\epsilon}_i^2$ and $h_i(\hat{\epsilon};\theta)$) are chosen to be $\sum_{i=1}^n \epsilon_i^2/n$ (resp. $\sum_{i=1}^n \hat{\epsilon}_i^2/n$). For simplicity, it is also assumed that $(\hat{v}(t_i) - v(t_i))$ and $(\hat{v}(t_j) - v(t_j))$ (and hence $(\hat{\epsilon}_i^2 - \epsilon_i^2)$ and $(\hat{\epsilon}_j^2 - \epsilon_j^2)$) are of the same order of magnitude, if i and j are not far from each other. This is true if t_i and t_j are both in the interior or both in the boundary area. The preceding simplifications do not affect the asymptotic properties of $\tilde{\theta}$ and $\hat{\theta}$.

Consistency and asymptotic normality of $\tilde{\theta}$ defined in Section 3 are a part of the results of Theorem 3.2 in Ling and Li (1997). Theorems 3.1 and 3.2 therein together show that conditions of Lemma 1 are fulfilled for the log-likelihood function $L(\theta)$. In the following discussion, we will investigate the difference between $\tilde{\theta}$ and $\hat{\theta}$ caused by replacing the unobservable ϵ_i with $\hat{\epsilon}_i$. Two lemmas are introduced at first.

LEMMA A.1. *Under the assumptions of Theorem 3 we have*

$$h_i(\hat{\epsilon};\theta) - h_i(\epsilon;\theta) \doteq O_p(\hat{\epsilon}_i^2 - \epsilon_i^2) \quad \forall\theta \in \Theta. \tag{A.9}$$

Proof of Lemma A.1. For any trial value $\theta = (\alpha_0, \alpha_1, \dots, \alpha_r, \beta_1, \dots, \beta_s)' \in \Theta$, one can rewrite $h_i(\epsilon;\theta)$ as

$$h_i(\epsilon;\theta) = \alpha_0\left(1 - \sum_{j=1}^{s}\beta_j\right)^{-1} + \left(\sum_{j=1}^{r}\alpha_j B^j\right)\left(1 - \sum_{k=1}^{s}\beta_k B^k\right)^{-1}\epsilon_i^2$$

and $h_i(\hat{\epsilon};\theta)$ as

$$h_i(\hat{\epsilon};\theta) = \alpha_0\left(1 - \sum_{j=1}^{s}\beta_j\right)^{-1} + \left(\sum_{j=1}^{r}\alpha_j B^j\right)\left(1 - \sum_{k=1}^{s}\beta_k B^k\right)^{-1}\hat{\epsilon}_i^2.$$

This leads to

$$\begin{aligned} h_i(\hat{\epsilon};\theta) - h_i(\epsilon;\theta) &= \left(\sum_{j=1}^{r}\alpha_j B^j\right)\left(1 - \sum_{k=1}^{s}\beta_k B^k\right)^{-1}(\hat{\epsilon}_i^2 - \epsilon_i^2) \\ &= \left(\sum_{j=1}^{\infty} a_j B^j\right)(\hat{\epsilon}_i^2 - \epsilon_i^2) \\ &\doteq O_p(\hat{\epsilon}_i^2 - \epsilon_i^2), \end{aligned} \tag{A.10}$$

where a_j are obtained by matching the powers in B, which decay exponentially. ■

LEMMA A.2. *Under the assumptions of Theorem 3 we have, $\forall\theta \in \Theta$, the first element of*

$$\frac{\partial h_i(\hat{\epsilon};\theta)}{\partial\theta} - \frac{\partial h_i(\epsilon;\theta)}{\partial\theta}$$

is zero and the other elements of it are all of the order $O_p(\hat{\epsilon}_i^2 - \epsilon_i^2)$.

Proof of Lemma A.2. Following (21) in Bollerslev (1986) we have

$$\frac{\partial h_i(\epsilon;\theta)}{\partial\theta} = \zeta_i + \sum_{j=1}^{s}\beta_j\frac{\partial h_{i-j}(\epsilon;\theta)}{\partial\theta}, \tag{A.11}$$

where $\zeta_i = (1, \epsilon_{i-1}^2, \ldots, \epsilon_{i-r}^2, h_{i-1}(\epsilon;\theta), \ldots, h_{i-s}(\epsilon;\theta))'$. Analogously, we have

$$\frac{\partial h_i(\hat{\epsilon};\theta)}{\partial \theta} = \hat{\zeta}_i + \sum_{j=1}^{s} \beta_j \frac{\partial h_{i-j}(\hat{\epsilon};\theta)}{\partial \theta}, \tag{A.12}$$

where $\hat{\zeta}_i = (1, \hat{\epsilon}_{i-1}^2, \ldots, \hat{\epsilon}_{i-r}^2, h_{i-1}(\hat{\epsilon};\theta), \ldots, h_{i-s}(\hat{\epsilon};\theta))'$. Denote by $B\zeta_i = \zeta_{i-1}$, $B\hat{\zeta}_i = \hat{\zeta}_{i-1}$,

$$B\frac{\partial h_i(\epsilon;\theta)}{\partial \theta} = \frac{\partial h_{i-1}(\epsilon;\theta)}{\partial \theta} \quad \text{and} \quad B\frac{\partial h_i(\hat{\epsilon};\theta)}{\partial \theta} = \frac{\partial h_{i-1}(\hat{\epsilon};\theta)}{\partial \theta},$$

we have

$$\left(1 - \sum_{j=1}^{s} \beta_j B^j\right) \frac{\partial h_i(\epsilon;\theta)}{\partial \theta} = \zeta_i$$

and

$$\left(1 - \sum_{j=1}^{s} \beta_j B^j\right) \frac{\partial h_i(\hat{\epsilon};\theta)}{\partial \theta} = \hat{\zeta}_i.$$

This leads to

$$\begin{aligned} \frac{\partial h_i(\hat{\epsilon};\theta)}{\partial \theta} - \frac{\partial h_i(\epsilon;\theta)}{\partial \theta} &= \left(\sum_{j=0}^{\infty} c_j B^j\right)(\hat{\zeta}_i - \zeta_i) \\ &\doteq O_p(\hat{\zeta}_i - \zeta_i). \end{aligned} \tag{A.13}$$

Again, c_j decay exponentially. Observe that the first element of $\hat{\zeta}_i - \zeta_i$ is zero. Results of Lemma A.2 follow from (A.13) and Lemma A.1. ■

Proof of Theorem 3.

(i) Under the conditions of Theorem 3, we have $\hat{\epsilon}_i \xrightarrow{p} \epsilon_i$. Following Lemmas A.1 and A.2, $\hat{L}(\theta) \xrightarrow{p} L(\theta)\ \forall \theta \in \Theta$. Following Lemma 1 there exists a consistent approximate MLE $\hat{\theta}$ satisfying the equation $\partial \hat{L}(\theta)/\partial \theta = 0$ such that

$$(\hat{\theta} - \tilde{\theta}) = O_p(\hat{L}'(\tilde{\theta})). \tag{A.14}$$

(ii) Note that $\sqrt{n}(\tilde{\theta} - \theta_0) \xrightarrow{\mathcal{D}} N(0, \Omega_0^{-1})$ (see Ling and Li, 1997). Results given in this part hold if we can show $\sqrt{n}(\hat{\theta} - B_\theta - \tilde{\theta}) \xrightarrow{p} 0$. Because $E(\hat{\theta} - B_\theta - \tilde{\theta}) = 0$, we have to show that $\operatorname{var}(\hat{\theta} - \tilde{\theta})$, or equivalently $\operatorname{var}[\hat{L}'(\tilde{\theta})]$, is a matrix of the order $o(n^{-1})$.

Note that

$$\hat{L}'(\tilde{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{2h_i(\hat{\epsilon};\tilde{\theta})} \frac{\partial h_i(\hat{\epsilon};\theta)}{\partial \theta}\bigg|_{\theta=\tilde{\theta}} \left(\frac{\hat{\epsilon}_i^2}{h_i(\hat{\epsilon};\tilde{\theta})} - 1\right). \tag{A.15}$$

https://doi.org/10.1017/S0266466604203061 Published online by Cambridge University Press

By means of Taylor expansion and using the results of Lemmas A.1 and A.2 we have

$$\frac{1}{2h_i(\hat{\epsilon};\tilde{\theta})} \doteq \frac{1}{2h_i(\epsilon;\tilde{\theta})} + O_p(h_i(\hat{\epsilon};\tilde{\theta}) - h_i(\epsilon;\tilde{\theta}))$$

$$\doteq \frac{1}{2h_i(\epsilon;\tilde{\theta})} + O_p(\hat{\epsilon}_i^2 - \epsilon_i^2),$$

$$\left.\frac{\partial h_i(\hat{\epsilon};\theta)}{\partial\theta}\right|_{\theta=\tilde{\theta}} \doteq \left.\frac{\partial h_i(\epsilon;\theta)}{\partial\theta}\right|_{\theta=\tilde{\theta}} + O_p(\hat{\epsilon}_i^2 - \epsilon_i^2),$$

where O_p denote the order of magnitude of a random vector and

$$\frac{\hat{\epsilon}_i^2}{h_i(\hat{\epsilon};\tilde{\theta})} \doteq \frac{\epsilon_i^2}{h_i(\epsilon;\tilde{\theta})} + O_p(\hat{\epsilon}_i^2 - \epsilon_i^2).$$

Furthermore, note that

$$L'(\tilde{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2h_i(\epsilon;\tilde{\theta})} \left.\frac{\partial h_i(\epsilon;\theta)}{\partial\theta}\right|_{\theta=\tilde{\theta}} \left(\frac{\epsilon_i^2}{h_i(\epsilon;\tilde{\theta})} - 1\right) = 0.$$

Inserting these results into (A.15), we obtain

$$\begin{aligned}\hat{L}'(\tilde{\theta}) &\doteq \frac{1}{n}\left[\sum_{i=1}^{n} \frac{1}{2h_i(\epsilon;\tilde{\theta})} \left.\frac{\partial h_i(\epsilon;\theta)}{\partial\theta}\right|_{\theta=\tilde{\theta}} \left(\frac{\epsilon_i^2}{h_i(\epsilon;\tilde{\theta})} - 1\right) + O_p(\hat{\epsilon}_i^2 - \epsilon_i^2)\right] \\ &=: L'(\tilde{\theta}) + T \\ &= T, \qquad\qquad \textbf{(A.16)}\end{aligned}$$

where the random vector

$$T = O_p\left(\frac{1}{n}\sum_{i=1}^{n}(\hat{\epsilon}_i^2 - \epsilon_i^2)\right). \qquad \textbf{(A.17)}$$

Observe that $\hat{\epsilon}_i^2 \doteq \epsilon_i^2 \cdot v(t_i)/\hat{v}(t_i)$. We have that each element of T is of the order

$$\begin{aligned}O_p\left(\frac{1}{n}\sum_{i=1}^{n}(\hat{\epsilon}_i^2 - \epsilon_i^2)\right) &\doteq O_p\left(\frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2\left[\frac{v(t_i) - \hat{v}(t_i)}{\hat{v}(t_i)}\right]\right) \\ &\doteq O_p\left(\frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2[v(t_i) - \hat{v}(t_i)]\right) \\ &= O_p\left(\frac{1}{n}\sum_{i=1}^{n}o_p(\epsilon_i^2)\right). \qquad \textbf{(A.18)}\end{aligned}$$

Hence, the variance of each element of T is of the order

$$o\left(\mathrm{var}\left[\frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2\right]\right) = o(n^{-1}),$$

and so far $\hat{v}$ is consistent. This shows that all entries of $\text{var}[\hat{L}(\tilde{\theta})]$ are of the order $o(n^{-1})$.

(iii) Now, we will calculate the order of magnitude of $B_\theta = E(\tilde{\theta} - \hat{\theta}) = O[E(T)]$. Observe that $E(\epsilon_i^2) = 1$, $\text{cov}[\epsilon_i^2, \hat{v}(t_i)] = O[(nb)^{-1}]$ at any point and $E(\hat{v}(t_i) - v(t_i)) = O(b^2)$ in the interior. We have, at an interior point t_i,

$$\begin{aligned} E\{\epsilon_i^2[v(t_i) - \hat{v}(t_i)]\} &= E[v(t_i) - \hat{v}(t_i)] - \text{cov}[\epsilon_i^2, \hat{v}(t_i)] \\ &= O[b^2 + (nb)^{-1}]. \end{aligned}$$

Furthermore, note that $E(\hat{v}(t_i) - v(t_i)) = O(b)$ at the boundary and that the length of the boundary area is equal to $2b$. This shows that the expected value of each element of T is of the order $O[b^2 + (nb)^{-1}]$ and hence

$$E(T) = O[b^2 + (nb)^{-1}]. \tag{A.19}$$

Theorem 3 is proved. ■

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

$$\begin{aligned} \hat{\epsilon}_i^4 &= (z_i^2/\hat{v}(t_i))^2 \\ &\doteq \left(\frac{z_i^2}{v(t_i)} + O_p(\hat{v}(t_i) - v(t_i)) + O_p(\hat{v}(t_i) - v(t_i))^2\right)^2 \\ &\doteq \epsilon_i^4 + O_p(\hat{v}(t_i) - v(t_i)) + O_p(\hat{v}(t_i) - v(t_i))^2. \end{aligned} \tag{A.20}$$

We have

$$\begin{aligned} E[\hat{E}(\epsilon_i^4) - E(\epsilon_i^4)] &= O\left(\frac{1}{n}\sum_{i=1}^n E(\hat{v}(t_i) - v(t_i))\right) + O\left(\frac{1}{n}\sum_{i=1}^n E(\hat{v}(t_i) - v(t_i))^2\right) \\ &=: T_1 + T_2. \end{aligned} \tag{A.21}$$

Furthermore, we have $E(T_1) = O(b_\epsilon^2)$ and $T_2 \doteq \text{MISE}_{[0,1]} = O(nb_\epsilon)^{-1} + o(T_1)$, where $\text{MISE}_{[0,1]}$ denotes the MISE on [0,1]. The results given in (22) 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 ϵ_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 ϵ_i^4 exists and

$$n\,\text{var}\left(\frac{1}{n}\sum_{i=1}^n \epsilon_i^4\right) \to 2\pi c_f^\epsilon, \tag{A.22}$$

where c_f^ϵ is the value of the spectral density of ϵ_i^4 at the origin (see, e.g., Brockwell and Davis, 1991, pp. 218ff). Proposition 1 is proved. ■

A sketched proof of Proposition 2. Estimation of functionals $\int\{v^{(\nu)}(t)\}^2\,dt$, where $v^{(\nu)}$ is the νth derivative of v, was investigated by Ruppert et al. (1995) and Beran and Feng (2002b) in nonparametric regression with independent and dependent errors, respectively. 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 (2002b) together show that the orders of magnitude 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 (2002b), where k and l correspond to $\nu = 0$ and the kernel order used here and δ is the long-memory parameter, which is zero in the current context. ∎

A sketched proof of Theorem 4. Note that $\hat{b} = \hat{C}_{\mathrm{A}} n^{-1/5}$, where C_{A} is as defined in (18). Hence we have

$$(\hat{b} - b_{\mathrm{A}})/b_{\mathrm{A}} = C_{\mathrm{A}}^{-1}(\hat{C}_{\mathrm{A}} - C_{\mathrm{A}}). \tag{A.23}$$

Taylor expansion shows that

$$\hat{C}_{\mathrm{A}} - C_{\mathrm{A}} \doteq O(\hat{c}_f - c_f) + O_p(\hat{I}(v^2) - I(v^2)) + O_p(\hat{I}((v'')^2) - I((v'')^2)). \tag{A.24}$$

Observe that

$$\hat{I}((v'')^2) - I((v'')^2) \doteq O_p(n^{-2/7}). \tag{A.25}$$

The term $O_p(\hat{I}(v^2) - I(v^2)) = O_p(n^{-1/2})$ is of a much smaller order than that given in (A.25) and hence is omitted. As a result of the bias in $\hat{\theta}$ one has

$$\hat{c}_f - c_f = O(n^{-2/5}). \tag{A.26}$$

The results as given in Theorem 4 hold. ∎