Efficient Estimation in Semiparametric GARCH Models

FEIKE C. DROST AND CHRIS A.J. KLAASSEN *
Tilburg University and University of Amsterdam

Abstract

It is well-known that financial data sets exhibit conditional heteroskedasticity. GARCH type models are often used to model this phenomenon. Since the distribution of the rescaled innovations is generally far from a normal distribution, a semiparametric approach is advisable. Several publications observed that adaptive estimation of the Euclidean parameters is not possible in the usual parametrization when the distribution of the rescaled innovations is the unknown nuisance parameter. However, there exists a reparametrization such that the efficient score functions in the parametric model of the autoregression parameters are orthogonal to the tangent space generated by the nuisance parameter, thus suggesting that adaptive estimation of the autoregression parameters is possible. Indeed, we construct adaptive and hence efficient estimators in a general GARCH in mean type context including integrated GARCH models.

Our analysis is based on a general LAN Theorem for time-series models, published elsewhere. In contrast to recent literature about ARCH models we do not need any moment condition.

Keywords: LAN in time-series, semiparametrics, adaptivity, (integrated) GARCH (in mean).

1 Introduction.

It is a well established empirical fact in financial economics that time-series like exchange rates and stock prices exhibit conditional heteroskedasticity. Big shocks are clustered together. The original paper of Engle (1982) proposes the ARCH model to incorporate conditional heteroskedasticity in econometric modeling of financial data sets. Bollerslev (1986) introduces the GARCH model as a generalization of ARCH. This facilitates a parsimonious parametrization which is particularly useful when shocks are important for a longer period (the idea corresponds

*Department of Econometrics, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands and Department of Mathematics, University of Amsterdam, Plantage Muidergracht 24, 1018 TV Amsterdam, The Netherlands. The first author is research fellow of the Royal Netherlands Academy of Arts and Sciences (K.N.A.W.). Part of this research was done by the second author at the Euler International Mathematical Institute, St. Petersburg. The authors gratefully acknowledge the helpful comments of Bas Werker, the editor Helmut Lütkepohl, and three anonymous referees.

to the generalization of AR to ARMA models). Several variations and extensions have been proposed in the literature. Nelson (1991) proposes the exponential GARCH model to capture the fact that the stock market is smoother in upward directions than in the opposite case (because of the leverage effect). Gourieroux and Monfort (1992) suggest a nonparametric approach. They do not restrict attention to conditional variances that depend only upon past squared observations, but they try to estimate the functional form of the conditional heteroskedastic variance from the data. Another important extension is the GARCH-M type model [cf. Engle, Lilien, and Robins (1987)]. According to the Capital Asset Pricing Model one expects higher returns due to risk premia if the asset is more risky. To model this phenomenon the conditional variance is also included in the mean equation. Lots of applications have shown the strength of the GARCH type of modeling. In this paper we do not refer to original application papers but we want to draw attention to the monograph of Diebold (1988) and the survey paper of Bollerslev, Chou, and Kroner (1992).

Despite the success of the GARCH history there are several topics that require attention. In this paper we consider the distributional assumptions on the rescaled innovations. The original formulations of GARCH type models assume that these residuals are standard normal. Diebold (1988), however, shows that this assumption is often violated in empirical examples. Typically, the innovations have fat-tailed distributions and they are also non-symmetric in several applications. Drost and Werker (1995) provide an explanation for high kurtosis if the observations arise from a GARCH data generating process in continuous time [see also Drost and Nijman (1993) and Nelson (1990a)]. Diebold (1988) suggests that the errors will be 'more normal' if the process is more and more aggregated. Despite the observed non-normality of the error structure Weiss (1986) and Lee and Hansen (1994) have shown that Quasi Maximum Likelihood Estimation (QMLE), based upon the false assumption of normality, yields $\sqrt{n}$-consistent estimators, see also Lumsdaine (1989). However, the efficiency loss may be considerable. Therefore, several authors try to avoid efficiency loss, allowing the error structure to belong to some flexible parametric family of distributions. Student t-distributions are very popular, cf., e.g., Baillie and Bollerslev (1989). As a drawback of the introduction of such parametric models of the innovation distribution we mention that the results of Weiss (1986) and Lee and Hansen (1994) do not carry over to general error distributions. While QMLE based upon the normal distribution yields $\sqrt{n}$-consistent estimators, QMLE based upon other distributions (e.g., the student distributions) generally even fails to be consistent if the true distribution is different.

In the approaches mentioned above, the stochastic error structure is still described by some finite dimensional statistical model. To avoid the introduction of a wrong parametric family of innovation distributions leading to inconsistent estimators and to be more flexible, a semiparametric approach is to be preferred. We want to estimate the conditional heteroskedastic character of the GARCH process but we do not want to restrict the class of error distributions too much. Apart from some regularity conditions we will assume the distribution of the innovations to be completely unknown. In passing we will consider the case of symmetrically distributed innovations. At first sight, these types of estimation problems seem to be much harder than the corresponding parametric ones and one would expect that optimal semiparametric estimators are less precise asymptotically than optimal parametric estimators. For lots of interesting econometric models this presumption turns out to be too pessimistic. Adaptive estimation is often possible. Adaptive estimation is just a special instance of semiparametric efficient

estimation. Just as in parametric models, in semiparametric models an efficient estimator is an asymptotically normal estimator with minimal variance. If this minimal variance is the same as when the error distribution is known, one calls the efficient estimator adaptive since it adapts, so to say, to the underlying error distribution. Typically, an estimator based on a (wrongly) specified error distribution is not efficient in the semiparametric sense. For i.i.d. observations a lot of adaptive and semiparametric results are available [cf., e.g., Bickel, Klaassen, Ritov, and Wellner (1993) [BKRW(1993) from now on] and the survey papers of Robinson (1988) and Newey (1990)]. Rigorous results are sparse in a time-series context. ARMA models are considered in detail by Kreiss (1987a,b). Some results for GARCH are obtained in Engle and González-Rivera (1991) and Steigerwald (1992) [see also Pötscher (1995) and Steigerwald (1995)]. Linton (1993) discusses the semiparametric properties of ARCH models in more detail. However, these papers impose rather high moment conditions. The parameter estimates obtained in empirical work generally fail these moment conditions and, therefore, the scope for application seems to be limited.

In Drost, Klaassen, and Werker (1994b) [henceforth DKW(1994b)] a general LAN Theorem for time-series models is presented together with conditions guaranteeing the existence of efficient estimators. We will apply these results to GARCH type models, including, e.g., I-GARCH and GARCH-M, thus avoiding severe moment conditions. We do not need the existence of moments neither of the rescaled innovations in the GARCH model (admitting for example Cauchy errors) nor of the observations (as is clear from the inclusion of integrated GARCH models). Since we only assume the existence of a stationary solution of the GARCH equations, our approach captures the models commonly used. A general LAN Theorem for time-series models is also contained in Theorem 13, Section 4, of Jeganathan (1988). Based hereon is his Theorem 17, Section 4, which yields adaptive estimators for ARMA-type location models. However, this result is not directly applicable to GARCH scale models and, moreover, it heavily leans on symmetry of the innovations.

To keep notation simple we restrict attention to the popular and most commonly used GARCH(1,1) type models. This preserves the essential difficulty of GARCH (with respect to ARCH) since both the AR and the MA part are present in the conditional variance equation. All past observations show up (at an exponentially decaying rate). The statement of our theorem with respect to GARCH(1,1) is easily generalized to the general case of GARCH(p,q).

The first semiparametric results in a GARCH context were only partially successful. Engle and González-Rivera (1991) state "Monte Carlo evidence suggest that this semiparametric method [*i.e. the discrete maximum penalized likelihood estimation technique of Tapia and Thompson* (1978)] can improve the efficiency of the parameter estimates up to 50% over QMLE, but it does not seem to capture the total potential gain in efficiency. In this sense we say that the estimator is not adaptive in the class of densities with mean 0 and variance 1; that is, the estimator is not fully efficient, and it does not achieve the Cramér-Rao lower bound. The information matrix is not block-diagonal between the parameters of interest (the ones in the mean and in the variance equation) and the nuisance parameters (the knots of the density). If we choose the parametric form of the model with a conditional parametric density defined by a shape parameter, this one being part of the parameters to estimate, we can show easily that the expectation of the cross-partial derivatives of the log-likelihood function respects the parameter of interest and the shape parameter is different from 0. In other words, the estimation of the shape parameter affects the efficiency of the estimates of the parameters of interest"

(pp. 355–356). These statements imply that the finite dimensional parameter describing the GARCH model (with the standardized error distribution as nuisance parameter) is not adaptively estimable. This is not surprising since the classical GARCH formulation contains a scale parameter, and in most models the variance is not adaptively estimable. Therefore the scale parameter is often included into the (infinite dimensional) nuisance parameter. For the GARCH model this procedure does not work: the scores w.r.t. the remaining autoregression parameters are still not orthogonal to the tangent space. Hence, complete adaptive estimation of the conditional heteroskedastic character is not possible in GARCH models. This explains the efficiency loss observed by Engle and González-Rivera (1991). However, calculation of the scores w.r.t. the parameters of the GARCH model shows that there are several orthogonality relations between the score space and the tangent space generated by the unknown shape. Linton (1993) and Drost, Klaassen, and Werker (1994a) [henceforth DKW(1994a)] obtain along different lines a reparametrization of the ARCH and GARCH model respectively such that the autoregression parameters are adaptively estimable and the location-scale parameters generate the most difficult one-dimensional subproblems. So, knowledge of the shape of the error distribution does not help to construct better estimators of the conditional heteroskedastic character of the GARCH process. This resembles the regression model with unknown location $\mu \in R$, regression parameter $\beta \in R^k$ and completely unknown error distribution, where the regression parameter β is adaptively estimable if the location parameter is included into the nuisance [see Bickel (1982)].

The paper is organized along the following lines. In Section 2 we state the LAN Theorem for a large set of GARCH(1,1) type models, including all stationary classical GARCH models such as, e.g., I-GARCH and GARCH-M. This LAN property is derived for the parametric model with the shape of the innovations known, and it implies the Convolution Theorem of Hájek (1970) which we will state next. This Convolution Theorem yields a bound on the asymptotic performance of estimators in the parametric model and is valid, a fortiori, for the semiparametric model as well. Section 3 is devoted to the construction of an estimator of the autoregression parameters on the assumption that the shape of the innovations is unknown, i.e. within the semiparametric model. This estimator happens to attain the bound from the parametric Convolution Theorem and therefore is asymptotically efficient in the parametric model and hence in the semiparametric model, since it does not use knowledge about the shape of the innovations. Such an estimator, which attains the parametric bound in a semiparametric model, is called adaptive. The proofs of these results are based on DKW(1994b) and most of them are given in the Appendix.

A small simulation study is presented in Section 4. It turns out that the suggested optimal estimator performs as expected: the estimator performs better than QMLE and the difference with MLE (if the error distribution is known) becomes negligible when the sample size is growing large. The empirical illustration in this section shows that the efficiency loss by using QMLE may be considerable. Some conclusions are drawn in Section 5.

2 LAN and Convolution Theorem.

We consider a generalization of the reparametrized GARCH(p, q) model as given in Linton (1993), with $p = 0$, and motivated by adaptation arguments in DKW(1994a). For notational

simplicity, we take $p = q = 1$. In this manner the essential difficulty of an infinite number of lags is retained. To obtain the corresponding results for the general case (with $p, q \in \mathbb{N}$ fixed) a careful replacement of coefficients by vectors suffices.

Let $\mu \in \mathbb{R}$, $\sigma > 0$, $\alpha > 0$, and $\beta > 0$ be parameters and let $\{\varepsilon_t : t \in \mathbb{Z}\}$ be an i.i.d. sequence of innovation errors with location zero, scale one, and density g. Put $\xi_t = \mu + \sigma\varepsilon_t$ and note that ξ_t is a random variable with location μ, scale σ, and density $\sigma^{-1}g(\{\cdot - \mu\}/\sigma)$. We introduce the following convention: random variables, like ε and ξ, denote a typical element of the corresponding sequences $\{\varepsilon_t : t \in \mathbb{Z}\}$ and $\{\xi_t : t \in \mathbb{Z}\}$.

Consider the model with observations

$$y_t = h_t^{1/2}\xi_t = \mu h_t^{1/2} + \sigma h_t^{1/2}\varepsilon_t, \tag{2.1}$$

where the unobservable heteroskedasticity factors h_t depend on the past via

$$h_t = 1 + \beta h_{t-1} + \alpha y_{t-1}^2 = 1 + h_{t-1}(\beta + \alpha\xi_{t-1}^2). \tag{2.2}$$

Observe that the Euclidean parameter $\theta = (\alpha, \beta, \mu, \sigma)'$ is identifiable. Throughout we assume that equation (2.2) admits a stationary solution $\{h_t : t \in \mathbb{Z}\}$. A necessary and sufficient condition is given by [Nelson (1990b), Theorem 2]

ASSUMPTION A

$$E\ln\{\beta + \alpha\xi^2\} < 0. \tag{2.3}$$

Our semiparametric analysis treats the density g as an infinite dimensional nuisance parameter and includes all strictly stationary GARCH models of type (2.1)–(2.2). These equations contain, e.g., the classical Engle (1982)-Bollerslev (1986) GARCH model with a different parametrization and with finite second moments ($\beta + \alpha\sigma^2 < 1$ and $\mu = 0$), and the I-GARCH model of Engle and Bollerslev (1986) ($\beta + \alpha\sigma^2 = 1$ and $\mu = 0$). Furthermore, our model resembles the GARCH-M model of Engle, Lilien, and Robins (1987). In the mean equation (2.1) we have included the conditional standard deviation of y_t while Engle, Lilien, and Robins (1987) include a kind of conditional variance. More precisely stated, their model is given by $z_t = \delta h_t + y_t$ and $\mu = 0$, i.e., $z_t = \delta h_t + \sigma h_t^{1/2}\varepsilon_t$. Inserting $\mu = 0$ in (2.1) or $\mu = \delta = 0$ in the GARCH-M model yields the classical GARCH model. Generally, risk aversion is stronger pronounced in the original GARCH-M model than in our formulation.

Suppose that we observe $y_1, \ldots, y_n$, and some starting value h_{01} initializing (2.2). It is not needed that h_{01} arises from the stationary solution of (2.2). We are considering estimation of θ, based on $h_{01}, y_1, \ldots, y_n$, in the presence of the infinite dimensional nuisance parameter g. However, in this section we will fix the nuisance parameter g and in the resulting parametric model we will derive a bound on the asymptotic performance of regular estimators of θ, a so-called Convolution Theorem. To that end we choose local submodels and we will study estimation of θ locally asymptotically. The above mentioned Convolution Theorem holds once the log-likelihood ratios of the observed random variables are locally asymptotically normal (LAN).

Observe that the model with the autoregression parameters α and β fixed too, corresponds to the location-scale model for i.i.d. random variables since the information provided by the observations $h_{01}, y_1, \ldots, y_n$ is equal to the information contained in the i.i.d. random variables

$\xi_1, \ldots, \xi_n$. Consequently, the location-scale model is a parametric submodel of our time-series model and it makes sense to assume that this submodel is regular, i.e. [see Hájek and Šidák (1967)]

ASSUMPTION B The distribution of ε possesses an absolutely continuous Lebesgue density g with derivative g' and finite Fisher information for location

$$I_l(g) = \int \{g'/g\}^2 g(\varepsilon) d\varepsilon \tag{2.4}$$

and for scale

$$I_s(g) = \int \{1 + \varepsilon g'/g(\varepsilon)\}^2 g(\varepsilon) d\varepsilon. \tag{2.5}$$

Moreover, the random variable ε has location zero and scale one.

To be able to derive an asymptotic lower bound we have to rely on semiparametric methods as presented in, e.g., BKRW(1993) and DKW(1994a,b). So we fix θ at $\theta_0 = (\alpha_0, \beta_0, \mu_0, \sigma_0)'$ and choose local parametrizations $\theta_n = (\alpha_n, \beta_n, \mu_n, \sigma_n)'$ and $\tilde{\theta}_n = (\tilde{\alpha}_n, \tilde{\beta}_n, \tilde{\mu}_n, \tilde{\sigma}_n)'$ such that $|\theta_n - \theta_0| = O(n^{-1/2})$, $|\tilde{\theta}_n - \theta_0| = O(n^{-1/2})$, and even

$$\lambda_n = \sqrt{n}(\tilde{\theta}_n - \theta_n) \to \lambda, \text{ as } n \to \infty. \tag{2.6}$$

In the remainder expectations, convergences, etc. are implicitly taken under θ_n and g (unless otherwise indicated).

To obtain a uniform LAN Theorem we consider the log-likelihood ratio Λ_n of $h_{01}, y_1, \ldots, y_n$ for $\tilde{\theta}_n$ with respect to θ_n under θ_n (and g fixed). Observe that the residuals and the conditional variances up to time t can be recursively calculated from θ and the observations $h_{01}, y_1, \ldots, y_t$: with $h_1(\theta) = h_{01}$, obtain for $t = 1, 2, \ldots$

$$\begin{aligned} \xi_t(\theta) &= y_t / h_t^{1/2}(\theta), & (2.7)\\ \varepsilon_t(\theta) &= \{\xi_t(\theta) - \mu\}/\sigma, & (2.8)\\ h_{t+1}(\theta) &= 1 + \beta h_t(\theta) + \alpha y_t^2. & (2.9) \end{aligned}$$

Conditionally on h_{01} the density of $y_1, \ldots, y_n$ under θ_n is

$$\prod_{t=1}^{n} \sigma_n^{-1} h_{nt}^{-1/2} g(\sigma_n^{-1}\{h_{nt}^{-1/2} y_t - \mu_n\}) = \prod_{t=1}^{n} \sigma_n^{-1} h_{nt}^{-1/2} g(\{\xi_{nt} - \mu_n\}/\sigma_n) = \prod_{t=1}^{n} \sigma_n^{-1} h_{nt}^{-1/2} g(\varepsilon_{nt}),$$

where $h_{nt} = h_t(\theta_n)$, $\xi_{nt} = \xi_t(\theta_n)$, and $\varepsilon_{nt} = \varepsilon_t(\theta_n)$.

To enhance the interpretation of this formula and to stress the link between the present time-series model and the i.i.d. location-scale model we introduce the notation $\tilde{h}_{nt} = h_t(\tilde{\theta}_n)$,

$$l\{\mu, \sigma\}(x) = \log g(\{x - \mu\}/\sigma) - \log \sigma,$$

$$\begin{pmatrix} M_{nt} \\ S_{nt} \end{pmatrix} = n^{1/2} \sigma_n^{-1} h_{nt}^{-1/2} \begin{pmatrix} \tilde{\mu}_n \tilde{h}_{nt}^{1/2} - \mu_n h_{nt}^{1/2} \\ \tilde{\sigma}_n \tilde{h}_{nt}^{1/2} - \sigma_n h_{nt}^{1/2} \end{pmatrix}, \tag{2.10}$$

and $\tilde{\varepsilon}_{nt} = \varepsilon_t(\tilde{\theta}_n)$. With Λ_n^s the log-likelihood ratio for h_{01}, the log-likelihood ratio Λ_n may be written as

$$\begin{aligned}\Lambda_n &= \log\left\{\prod_{t=1}^n \tilde{\sigma}_n^{-1}\tilde{h}_{nt}^{-1/2} g(\tilde{\varepsilon}_{nt}) / \prod_{t=1}^n \sigma_n^{-1} h_{nt}^{-1/2} g(\varepsilon_{nt})\right\} + \Lambda_n^s \\ &= \sum_{t=1}^n \Big\{ l\{(\mu_n,\sigma_n) + \sigma_n n^{-1/2}(M_{nt},S_{nt})\}(\xi_{nt}) - l\{\mu_n,\sigma_n\}(\xi_{nt})\Big\} + \Lambda_n^s \\ &= \sum_{t=1}^n \Big\{ l\{(0,1) + n^{-1/2}(M_{nt},S_{nt})\}(\varepsilon_{nt}) - l\{0,1\}(\varepsilon_{nt})\Big\} + \Lambda_n^s. \qquad (2.11)\end{aligned}$$

This expression resembles the log-likelihood ratio statistic for the i.i.d. location-scale model but here the deviations M_{nt} and S_{nt} are random. In the i.i.d. case the LAN Theorem is obtained with deterministic sequences. We will apply the results of DKW(1994b) which allow for such random sequences.

To get rid of the starting condition in the log-likelihood ratio statistic we will use the following regularity condition [compare assumption (A.3) of Kreiss (1987a) and Assumption A of DKW(1994b)].

ASSUMPTION C The density $\bar{g}_\theta$ of the initial value h_{01} satisfies, under θ_n,

$$\Lambda_n^s = \log\{\bar{g}_{\tilde{\theta}_n}/\bar{g}_{\theta_n}(h_{01})\} \xrightarrow{P} 0, \text{ as } n \to \infty. \qquad (2.12)$$

To make an appropriate expansion of Λ_n it will be handy to introduce the notation $\dot{l}_{nt}$ for the four-dimensional conditional score at time t. To be more precise, denote the two-dimensional vector derivative of the conditional variance by

$$H_t(\theta) = \frac{\partial}{\partial(\alpha,\beta)} h_t(\theta) = \beta H_{t-1}(\theta) + \begin{pmatrix} y_{t-1}^2 \\ h_{t-1}(\theta) \end{pmatrix}, \qquad (2.13)$$

with $H_1(\theta) = 0_2$. Define the (4×2)-derivative matrix $W_t(\theta)$ [motivated by differentiation of (M_{nt}, S_{nt}) with respect to $\tilde{\theta}_n$ at θ_n] by

$$W_t(\theta) = \sigma^{-1}\begin{pmatrix} \frac{1}{2}h_t^{-1}(\theta) H_t(\theta)(\mu,\sigma) \\ I_2 \end{pmatrix}, \qquad (2.14)$$

denote the location-scale score by (with $l' = g'/g$)

$$\psi_t(\theta) = -\begin{pmatrix} l'(\varepsilon_t(\theta)) \\ 1 + \varepsilon_t(\theta) l'(\varepsilon_t(\theta)) \end{pmatrix}, \qquad (2.15)$$

and put

$$\dot{l}_t(\theta) = W_t(\theta)\psi_t(\theta).$$

Then, the conditional score at time t may be denoted by $\dot{l}_{nt} = \dot{l}_t(\theta_n)$. Observe that $\dot{l}$ is just the heuristic score. An expansion of (2.11) shows that the log-likelihood ratio Λ_n may be alternatively written as

$$\Lambda_n = \lambda' n^{-1/2} \sum_{t=1}^n \dot{l}_{nt} - \frac{1}{2} n^{-1} \sum_{t=1}^n \{\lambda' \dot{l}_{nt}\}^2 + R_n. \qquad (2.16)$$

The LAN result for the parametric version of model (2.1)–(2.2) is stated in the following theorem. The proof is deferred to Appendix A.

Theorem 2.1 (LAN) *Suppose that Assumptions* A–C *are satisfied. Then the local log-likelihood ratio statistic* Λ_n*, as defined by* (2.11) *and* (2.16)*, is asymptotically normal. More precisely, under* θ_n*,*

$$R_n \xrightarrow{P} 0,\ \Lambda_n \xrightarrow{D} N\left(-\frac{1}{2}\lambda' I(\theta_0)\lambda, \lambda' I(\theta_0)\lambda\right),\ \text{as } n \to \infty, \tag{2.17}$$

where $I(\theta_0)$ *is the probability limit of the averaged score products* $\dot{l}_{nt}\dot{l}'_{nt}$*.*

We are now in a position to apply the Convolution Theorem of Hájek (1970); cf. Theorem 2.3.1, p. 24, of BKRW(1993).

Theorem 2.2 (Convolution Theorem) *Under the assumptions of the LAN Theorem* 2.1 *let* $\{T_n : n \in \mathbb{N}\}$ *be a regular sequence of estimators of* $q(\theta)$*, where* $q : \mathbb{R}^4 \to \mathbb{R}^k$ *is differentiable with total differential matrix* $\overset{\circ}{q}$*. As usual, regularity at* $\theta = \theta_0$ *means that there exists a random* k*-vector* Z *such that for all sequences* $\{\theta_n : n \in \mathbb{N}\}$*, with* $n^{1/2}(\theta_n - \theta_0) = O(1)$*,*

$$n^{1/2}\{T_n - q(\theta_n)\} \xrightarrow{D} Z,\ \text{as } n \to \infty, \tag{2.18}$$

where the convergence is under θ_n*. Let* $\tilde{l} = \overset{\circ}{q}(\theta_0) I(\theta_0)^{-1} \dot{l}(\theta_0)$ *be the efficient influence function, then, under* θ_0*,*

$$\begin{pmatrix} n^{1/2}\{T_n - q(\theta_0) - n^{-1}\sum_{t=1}^n \tilde{l}_t\} \\ n^{-1/2}\sum_{t=1}^n \tilde{l}_t \end{pmatrix} \xrightarrow{D} \begin{pmatrix} \Delta_0 \\ Z_0 \end{pmatrix},\ \text{as } n \to \infty, \tag{2.19}$$

where Δ_0 *and* Z_0 *are independent and* Z_0 *is* $N(0, \overset{\circ}{q}(\theta_0) I(\theta_0)^{-1}\ \overset{\circ}{q}(\theta_0)')$*. Moreover,* $\{T_n : n \in \mathbb{N}\}$ *is efficient if* $\{T_n : n \in \mathbb{N}\}$ *is asymptotically linear in the efficient influence function, i.e. if* $\Delta_0 = 0$ (a.s.)*.*

As a conclusion from the Convolution Theorem we obtain that a regular estimator $\hat{\theta}_n$ of θ satisfies, under θ_0,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} \Delta_0 + Z_0,$$

i.e. the limit distribution of $\hat{\theta}_n$ is the convolution of the random vector Δ_0 and a Gaussian random vector with mean zero and variance the inverse of the information matrix $I(\theta_0)$. Since Δ_0 adds noise to the Gaussian vector Z_0, it is clear that $\Delta_0 = 0$ would be preferred. This motivates the usual terminology (as lower bound, etc.) because $\Delta_0 = 0$ is attainable in lots of situations.

In the remainder of this paragraph we simplify exposition by supposing that the scores given above are stationary such that we may restrict attention to just one specific element, compare DKW(1994a). In this way it is easier to comprehend the specific adaptiveness features in the GARCH model. These results are derived along the lines of Sections 2.4 and 3.4 of BKRW(1993). This expository simplification will be suppressed again in the next section when deriving a (semiparametric) efficient estimator. This optimal estimator satisfies the properties obtained in I–IV below.

In a stationary setting the Fisher information matrix defined in the LAN Theorem 2.1 simplifies to

$$I(\theta_0) = E\dot{l}\dot{l}' = EW\psi\psi'W' = EW I_{ls}(g)W',$$

where $I_{ls}(g)$ is the information matrix in the location-scale model,

$$I_{ls}(g) = E\psi\psi' = \begin{pmatrix} E(l')^2 & E\varepsilon(l')^2 \\ E\varepsilon(l')^2 & E(1+\varepsilon l')^2 \end{pmatrix}.$$

If the location parameter μ is known to be zero, as in the classical GARCH case, this formula simplifies even further to

$$I(\theta_0) = I_s(g) E W_s W_s', \tag{2.20}$$

where W_s is the 3-dimensional subvector of W concerning the relevant derivatives with respect to the scale parameter σ and where $I_s(g) = E(1+\varepsilon l')^2$ is the information for scale in the i.i.d. scale model.

I. If g is known and if we want to estimate the autoregression parameter $\nu = (\alpha, \beta)'$ in the presence of the nuisance parameter $\eta = (\mu, \sigma)'$ then we see that the efficient influence function, as defined in the Convolution Theorem 2.2, equals

$$\tilde{l} = (I_2, 0_{2\times 2})[E\dot{l}\dot{l}']^{-1}\dot{l} = (I_2, 0_{2\times 2}) I(\theta_0)^{-1}\dot{l}.$$

As in Proposition 2.4.1.A and formula (2.4.3), pp. 28,30, of BKRW(1993) we may write

$$\tilde{l} = [E l_1^* l_1^{*\prime}]^{-1} l_1^*, \tag{2.21}$$

where the so-called efficient score function l_1^* of ν is obtained by the componentwise projection of $\dot{l}_1$, the first two elements of $\dot{l}$, onto the orthocomplement of $[\dot{l}_2]$, the linear span of the last two components of $\dot{l}$. Here the inner product is the covariance and the orthocomplement is taken in the linear space spanned by all components of $\dot{l}$. It is easy to verify that

$$l_1^* = \frac{1}{2}\sigma^{-1}\{(H/h) - E(H/h)\}(\mu, \sigma)\psi \tag{2.22}$$

and that l_1^* is orthogonal to $\dot{l}_2$ indeed, since $H/h = H_t/h_t$ depends on the past only and is independent of the present innovation ε_t.

II. If g is unknown and if we want to estimate ν in the presence of the nuisance parameters η and g then we obtain the same efficient influence function. To see this note that the components of l_1^* as given in (2.22) are orthogonal to every element of $L_2^0(\varepsilon)$ by the independence of present (ξ and ε) and past (h and H). By (3.4.2) and Corollary 3.4.1.A, pp. 70,72, of BKRW(1993) we obtain

$$I(P_0 \mid \nu, Q) \geq E l_1^* l_1^{*\prime} \tag{2.23}$$

for all regular parametric submodels Q of our semiparametric model $\mathcal{P}$, i.e. the information at P_0 in estimation of ν within the parametric submodel $\mathcal{Q}$ equals at least the information at P_0 in estimation of ν within the parametric model, studied in I, with g known. In other words, as far as estimation of ν is concerned, no parametric model $\mathcal{Q}$ is asymptotically more difficult to first order (contains less information) than the model from I. Consequently the semiparametric model

$\mathcal{P}$ itself is asymptotically to first order as difficult as the parametric model with g known, i.e. the information matrix with respect to ν evaluated at P_0 for the semiparametric model $\mathcal{P}$ equals the lower bound in the parametric model with g known (case I),

$$I(P_0 \mid \nu, \mathcal{P}) = El_1^* l_1^{*\prime}.$$

Once more, the efficient influence function is given by (2.21). Apparently, introduction of the nuisance parameter g in the presence of the Euclidean nuisance parameter η does not change the efficient influence function for ν. Hence, estimation of ν is asymptotically as hard not knowing g as knowing g. One usually calls this *adaptivity*. Observe, however, that the presence of the nuisance parameter η is important to derive this result. If η is known adaptive estimation of ν is not possible! The same conclusion applies if η is included into the "big" infinite dimensional nuisance parameter g. So, the nuisance parameter η is treated in another way than the nuisance parameter g. Since location-scale parameters are almost always present in econometric models a different treatment is not unreasonable and the usage of the protected notion "adaptivity" is legalized. However, with the comments above in mind, a more appropriate way of saying this is to call the parameter ν η-adaptive, explicitly referring to the remaining nuisance parameters present in the model. [Of course, a similar remark applies to, e.g., the non-symmetric regression model as discussed in Bickel (1982), where the regression parameter β is not fully adaptively estimable. In fact β is μ-adaptive.]

III. Estimation of the remaining parameter η is completely analogous to the location-scale problem for i.i.d. variables. Obtain the well-known lower bound for η in the semiparametric location-scale model. It suffices to construct a sequence of estimators $\{\hat{\eta}_n, n \in \mathbb{N}\}$ for η attaining this bound. Let $\hat{\theta}_n$ be some initial $\sqrt{n}$-consistent estimator of θ, calculate $\hat{h}_{nt} = h_t(\hat{\theta}_n)$ by plugging in $\hat{\theta}_n$ into (2.9) and obtain the residuals $\hat{\xi}_{nt} = \xi_t(\hat{\theta}_n) = y_t/\hat{h}_{nt}^{1/2}$, similarly. If one proceeds as if the $\hat{\xi}_{nt}$ are i.i.d. observations from some location-scale model, one obtains a semiparametric efficient estimator for η in our model (as is easily verified from the Convolution Theorem 2.2 by choosing an appropriate function q). To be more explicit, we assume that g has finite second moment and we define the location and scale parameters by standardizing g via the equations $E_g\varepsilon = 0$, $E_g\varepsilon^2 = 1$. Then the square root of the sample variance is optimal for σ both in the symmetric and non-symmetric case. The sample mean is optimal for μ if no symmetry is assumed and under the assumption of symmetry one has to use an efficient estimator for the symmetric location-problem [cf. Example 7.8.1, p. 400, of BKRW(1993)]. If one wants to avoid moment conditions on ε one may define the location-scale parameter in another way, see the discussion of the M-estimator in Section 3.

IV. Finally, when estimating the whole Euclidean parameter θ, the efficient score is simply obtained from II and III. Following the arguments leading to (2.23) in II, this score function yields a lower bound indeed. Optimality of this bound follows from III by choosing the most difficult direction from the location-scale problem.

Obvious substitutions in Theorems 2.1 and 2.2 show that the conclusions above are also valid for the classical GARCH model with $\mu = E_g\xi = 0$. An optimal estimator of σ in the non-symmetric case is given then by the square root of [cf. Example 3.2.3, pp. 53–55, of BKRW(1993)]

$$n^{-1}\sum_{t=1}^{n}\hat{\xi}_{nt}^2 - n^{-1}\frac{\sum_{t=1}^{n}\hat{\xi}_{nt}^3}{\sum_{t=1}^{n}\hat{\xi}_{nt}^2}\sum_{t=1}^{n}\hat{\xi}_{nt}.$$

In the symmetric case the limiting behavior of this estimator and the square root of the sample variance are the same.

3 Adaptive Estimators.

In classical parametric models the Maximum Likelihood Estimator is asymptotically efficient, typically. In semiparametric models such an estimation principle yielding efficient estimators does not exist. However, there exist methods to upgrade $\sqrt{n}$-consistent estimators to efficient ones by a Newton-Raphson technique, provided it is possible to estimate the relevant score or influence functions sufficiently accurately. In Klaassen (1987) such a method based on "sample splitting" is described for i.i.d. models. Schick (1986) uses both "sample splitting" and Le Cam's "discretization", again in i.i.d. models. See, e.g., Section 7.8 of BKRW(1993) for details. Schick's (1986) method has been adapted to time-series models in Theorem 3.1 of DKW(1994b). We assume the existence of such a preliminary, $\sqrt{n}$-consistent estimator.

ASSUMPTION D There exists a $\sqrt{n}$-consistent estimator $\hat{\theta}_n$ of θ_n (under θ_n and g).

For our GARCH model a natural candidate for such an initial estimator is the MLE based on the assumption of normality of the innovations ε_t. One often calls this estimator the Quasi MLE. Probably, this QMLE is $\sqrt{n}$-consistent under every density g with $E_g\varepsilon^4 < \infty$; this has been shown by Weiss (1986) for ARCH models and under restrictions by Lee and Hansen (1994) for GARCH models, which are slightly different from ours, see also Lumsdaine (1989). The additional moment condition on ε is needed there since a quadratic term appears in the score function of the scale parameter. To avoid moment conditions altogether, one could use, e.g., another preliminary M-estimator, instead. Let $\chi : \mathbb{R} \to \mathbb{R}^2$ be a sufficiently smooth bounded function with monotonicity properties. As an example we mention $\chi = (\chi_1, \chi_2)'$ with

$$\chi_1(x) = \frac{2}{1+\exp\{-x\}} - 1,\ x \in \mathbb{R},$$

the location score function for the logistic distribution and

$$\chi_2(x) = \int_0^x 2y\frac{\exp\{-y\}}{(1+\exp\{-y\})^2}dy - 1,\ x \in \mathbb{R}.$$

The M-estimator will solve the equations [cf. (2.7)–(2.9) and (2.13)–(2.14)]

$$\sum_{t=1}^{n} W_t(\theta)\chi(\varepsilon_t(\theta)) = 0. \tag{3.1}$$

Use of this M-estimator implies that one standardizes g at location 0 and scale 1 by the equation $E_g\chi(\varepsilon) = 0$; in the normal case with QMLE this yields μ as expectation and σ as standard deviation.

To prove that estimation via (3.1) shows validity of Assumption D we have to prove existence of this M-estimator and its $\sqrt{n}$-consistency. It should be possible to show existence along the lines of Scholz (1971) by studying the 4 by 4 pseudo information matrix $EW\chi\chi'W'$; see also Huber (1981), pages 138-139. Here we will not attempt to do this, since the situation is much more complicated than the location-scale problem studied in the literature. At the cost of some generality we suppose here that $\sqrt{n}$-consistent estimators $\hat{\alpha}_n$ and $\hat{\beta}_n$ are given. The $\sqrt{n}$-consistency of $\hat{\alpha}_n$ and $\hat{\beta}_n$ together with the contiguity obtained from the LAN Theorem 2.1 implies that we may treat the parameters α and β as given. So, we are in fact in the i.i.d. location-scale model and the M-estimators for μ and σ solving the latter two equations in (3.1) are $\sqrt{n}$-consistent, see Huber (1981) and Bickel (1982). We conjecture that the proof of the more general M-estimator solving (3.1) can be given along similar lines.

Here we will focus on efficient and hence adaptive estimation of the autoregression parameters α and β (cf. Subsections I–IV of Section 2); alternatively, in view of (2.14), note that the score $\dot{l}_{nt}$ satisfies the form discussed in Example 3.1 of DKW(1994b). In the Appendix we verify the conditions of Theorem 3.1 in DKW(1994b), this yields the following theorem.

Theorem 3.1 *Under Assumptions* A–D *adaptive estimators of α and β do exist.*

To describe our adaptive estimator more accurately, let $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n, \hat{\mu}_n, \hat{\sigma}_n)'$ be a $\sqrt{n}$-consistent estimator of θ and compute $W_t(\hat{\theta}_n)$ via (2.13) and (2.14). Let $\hat{\varepsilon}_{n1}, \ldots, \hat{\varepsilon}_{nn}$ be the residuals computed from $h_1, y_1, \ldots, y_n$ and $\hat{\theta}_n$ using (2.8). Via a kernel estimate based on $\hat{\varepsilon}_{n1}, \ldots, \hat{\varepsilon}_{nn}$ with the logistic kernel, say $k(\cdot)$, and bandwidth b_n we estimate $g(\cdot)$ by

$$\hat{g}_n(\cdot) = \frac{1}{n}\sum_{t=1}^{n} \frac{1}{b_n} k\left(\frac{\cdot - \hat{\varepsilon}_{nt}}{b_n}\right)$$

and subsequently $\psi(\cdot)$ by $\hat{\psi}_n(\cdot)$; here $b_n \to 0$ and $nb_n^4 \to \infty$. Now our estimator may be written as

$$\begin{aligned}(\hat{\alpha}_n, \hat{\beta}_n)' \;&+\; (I_2, 0_{2\times 2})\Big(\frac{1}{n}\sum_{t=1}^{n} W_t(\hat{\theta}_n)\hat{\psi}_n(\hat{\varepsilon}_{nt})\hat{\psi}_n(\hat{\varepsilon}_{nt})'W_t(\hat{\theta}_n)'\Big)^{-1} \\ &\times\; \frac{1}{n}\sum_{t=1}^{n}\left\{W_t(\hat{\theta}_n) - \frac{1}{n}\sum_{s=1}^{n} W_s(\hat{\theta}_n)\right\}\hat{\psi}_n(\hat{\varepsilon}_{nt}). \qquad (3.2)\end{aligned}$$

With $\hat{\theta}_n$ the QMLE this is the estimator used in the simulations of Section 4. To prove that such estimators are adaptive we need the following two technical modifications.

- **Discretization.** $\hat{\theta}_n$ is discretized by changing its value in $(0,\infty)\times(0,\infty)\times\mathbb{R}\times(0,\infty)$ into (one of) the nearest point(s) in the grid $\frac{c}{\sqrt{n}}(\mathbb{N}\times\mathbb{N}\times\mathbb{Z}\times\mathbb{N})$. This technical trick enables one to consider $\hat{\theta}_n$ to be non-random, and therefore independent of $\hat{\varepsilon}_{nt}$, y_t, and h_1.

- **Sample Splitting.** The set of residuals $\hat{\varepsilon}_{n1}, \ldots, \hat{\varepsilon}_{nn}$ is split into two samples, which may be viewed as independent now. For $\hat{\varepsilon}_{nt}$ in the first sample, the second sample is used to

estimate $\psi(\cdot)$ by $\hat{\psi}_{n2}(\cdot)$ and $\hat{\psi}_n(\hat{\varepsilon}_{nt})$ in (3.2) is replaced by $\hat{\psi}_{n2}(\hat{\varepsilon}_{nt})$. Similarly for $\hat{\varepsilon}_{nt}$ in the second sample, the first sample is used to estimate $\psi(\cdot)$. In this way, again some independence is introduced artificially to make the proof work.

This approach has been adopted in DKW(1994b). It should be emphasized that both tricks are merely introduced as a technical device to make proofs work. Other approaches have also been studied in the literature. Klaassen (1987) has shown that discretization may be avoided at the cost of an extra sample splitting. Schick (1986) and Koul and Schick (1995) show that sample splitting may be avoided at the cost of some extra conditions.

4 Simulations and an Empirical Example.

To enhance the interpretation and validity of the theoretical results of the previous sections we present a small simulation experiment. Furthermore, a case study concerning some exchange rate series is given.

We simulated several GARCH(1,1) series of length $n = 1000$, parameters $(\alpha, \beta, \sigma) = (.3, .6, 1)$, $(.1, .8, 1)$, and $(.05, .9, 1)$ [the parameter μ is set to zero and is not estimated to allow for a better comparison with previous simulation studies], and eight different innovation distributions: normal, a balanced mixture of two standard normals with means 2 and -2, respectively, double exponential, student distributions with $\nu = 5$, 7, and 9 degrees of freedom, and (skew) chi-squared distributions with $\nu = 6$ and 12 degrees of freedom. These densities are rescaled such that they have the required zero mean and unit variance.

It is the purpose of the simulations to evaluate the moderate sample properties of the autoregression parameters α and β which are adaptively estimable, in principle. For each series we estimated these parameters with MLE, QMLE, and a one-step semiparametric procedure. For the latter estimation method we made two estimates: one under general assumptions on the innovation distribution and one under the extra assumption of symmetry. The theoretical results imply that there should be no difference between these two semiparametric methods if the true underlying density is symmetric indeed but small sample properties may differ. In the semiparametric part we used standardized logistic kernels with a bandwidth of $h = .5$. Reasonable changes of the bandwidth, say $.25 \leq h \leq .75$, or another kernel like the normal one do not alter the conclusions below.

In the first part of the simulation experiment we compared the ML estimator with the semiparametric ones (with the MLE as initial starting value). Asymptotically both semiparametric estimators should behave as well as the MLE but one may expect that the small sample properties of the semiparametric estimators are worse due to the inherent problems of choosing the bandwidth. These results are not reported here but they are comparable to those given in Table 4.1, from which MLE can be compared with the semiparametric procedure with the less efficient QMLE starting value.

Of course ML estimation is not feasible in practice since the underlying distribution is not known. Therefore, we used the QMLE as starting point. Since μ vanishes for the situation chosen here and $\alpha\sigma^2 + \beta < 1$, Theorems 2 and 3 of Lee and Hansen (1994) are applicable and the QMLE is $\sqrt{n}$-consistent. This estimator has been improved by the one-step Newton method. For convenience we also report the behavior of the unfeasible MLE in Table 4.1. The

		α	β			α	β			α	β		
		.300	.600			.100	.800			.050	.900		
		$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}_\alpha$	$\hat{\sigma}_\beta$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}_\alpha$	$\hat{\sigma}_\beta$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}_\alpha$	$\hat{\sigma}_\beta$
N	ML=QML	.298	.593	.071	.056	.099	.786	.035	.073	.047	.891	.022	.051
	1-step	.298	.593	.072	.057	.098	.786	.036	.074	.047	.892	.022	.050
	1-step(sym)	.298	.593	.072	.056	.099	.786	.036	.073	.047	.891	.022	.051
DE	ML	.299	.592	.080	.070	.099	.782	.038	.083	.048	.885	.023	.061
	QML	.303	.588	.089	.079	.100	.776	.043	.094	.048	.880	.026	.073
	1-step	.294	.593	.085	.074	.097	.784	.040	.087	.046	.886	.024	.067
	1-step(sym)	.295	.592	.083	.073	.097	.783	.039	.086	.046	.885	.024	.065
NM	ML	.295	.595	.058	.041	.098	.790	.029	.054	.047	.898	.018	.030
	QML	.295	.595	.059	.042	.097	.790	.030	.054	.046	.897	.018	.032
	1-step	.295	.595	.060	.043	.098	.793	.030	.056	.047	.901	.018	.032
	1-step(sym)	.295	.595	.059	.042	.099	.793	.030	.056	.047	.901	.018	.032
t_5	ML	.295	.592	.076	.067	.100	.787	.036	.071	.048	.888	.021	.054
	QML	.296	.586	.098	.086	.101	.777	.047	.101	.048	.879	.027	.083
	1-step	.284	.594	.080	.071	.094	.791	.037	.081	.044	.890	.022	.064
	1-step(sym)	.285	.594	.079	.070	.095	.791	.037	.081	.045	.889	.022	.063
t_7	ML	.296	.595	.075	.060	.100	.782	.037	.079	.047	.885	.021	.063
	QML	.298	.592	.086	.070	.101	.776	.042	.094	.047	.882	.024	.076
	1-step	.291	.597	.078	.064	.096	.784	.038	.082	.045	.886	.022	.068
	1-step(sym)	.292	.597	.077	.063	.097	.783	.038	.082	.045	.886	.022	.068
t_9	ML	.298	.592	.076	.060	.098	.783	.037	.077	.047	.887	.022	.058
	QML	.300	.591	.083	.066	.099	.781	.040	.085	.048	.886	.024	.064
	1-step	.295	.593	.079	.062	.096	.785	.038	.080	.046	.889	.022	.057
	1-step(sym)	.295	.593	.077	.062	.096	.784	.038	.079	.046	.889	.022	.057
χ^2_6	ML	.297	.596	.042	.034	.099	.796	.020	.036	.050	.899	.012	.022
	QML	.299	.589	.091	.073	.101	.780	.042	.096	.048	.884	.024	.072
	1-step	.283	.603	.062	.051	.092	.801	.030	.061	.045	.898	.017	.047
χ^2_{12}	ML	.298	.596	.057	.045	.099	.794	.029	.048	.048	.893	.016	.036
	QML	.299	.592	.084	.064	.100	.782	.041	.079	.047	.881	.023	.071
	1-step	.289	.598	.065	.051	.095	.796	.032	.061	.045	.891	.018	.049

Table 4.1: Comparison of MLE, QMLE, and two semiparametric one-step estimators in the GARCH(1,1) model with eight different standardized innovation distributions. Number of observations $n = 1000$, true parameters $(\alpha, \beta) = (.3, .6)$, $(.1, .8)$, and $(.05, .9)$, respectively. The sample means of 2500 independent replications and their sample standard deviations are given.

mean values of the estimates in 2500 replications are given together with their sample standard deviations.

To calculate the efficiency of the QMLE, observe that the asymptotic variance of the QMLE is equal to the well-known variance formula $A^{-1}BA^{-1}$, where A is the expectation under $(\alpha, \beta, \sigma, g)$ of the second derivative of the pseudo log-likelihood (with a wrongly specified normal density) and B the expectation of the squared first derivative. With W_s as defined just below (2.20), straightforward calculations show

$$\begin{aligned} A &= 2EW_sW_s', \\ B &= (\kappa - 1)EW_sW_s', \end{aligned}$$

where $\kappa = \int \varepsilon^4 g(\varepsilon)d\varepsilon$. Except for the normal distribution, the matrices A^{-1} and B^{-1} are generally not equal. Since the asymptotic variance of the QMLE is equal to the lower bound up to a constant, the asymptotic efficiency of each component of the QMLE is given by

$$\frac{4}{(\kappa - 1)I_s(g)} = \frac{4}{\int(\varepsilon^2 - 1)^2 g(\varepsilon)d\varepsilon \int(1 + \varepsilon l'(\varepsilon))^2 g(\varepsilon)d\varepsilon} \leq 1.$$

The latter inequality follows from Cauchy-Schwarz applied to the following identity

$$-2 = E(\varepsilon^2 - 1)(1 + \varepsilon l'(\varepsilon)) = \int (\varepsilon^2 - 1)(1 + \varepsilon l'(\varepsilon))g(\varepsilon)d\varepsilon.$$

Since the lower bound for α and β does not change in the semiparametric setting, this expression also entails the loss in the semiparametric model and shows the (potential) gain of the semiparametric estimator (3.2).

Except for the mixture distribution we can exactly calculate the efficiency of QMLE with respect to MLE. For the standardized double exponential the relative efficiency is $\frac{4}{5}$, for standardized student distributions with ν degrees of freedom it is $1 - \frac{12}{\nu(\nu-1)}$, and for standardized chi-squared distributions with ν degrees of freedom it is $\frac{\nu-4}{\nu+6}$. For these heavy-tailed distributions the efficiency losses of QMLE with respect to MLE show up in Table 4.1 and we see that the semiparametric methods regain most of the loss caused by the inefficient QMLE method. For light-tailed alternatives, as in the mixture case, the situation is less clear cut. There the efficiency is approximately .94 and the performance of the estimators is not much different. For the normal distribution MLE and QMLE are of course equivalent. The use of the additional symmetry information hardly improves the estimated standard deviation of the semiparametric estimator (maximal .002), just as expected from our general theory. In empirical data sets one often observes outlier type innovation distributions with high kurtoses. Therefore, it seems worthwhile to apply the semiparametric estimation programs in these situations.

We conclude this section with a simple empirical example based on daily data. We applied our estimation methods to fifteen logarithmic differenced exchange rate series for the period January 1, 1980 to April 1, 1994 ($n = 3719$): Austrian Schilling (AS), Australian Dollar (AD), Belgium Franc (BF), British Pound (BP), Canadian Dollar (CD), Dutch Guilder (DG), Danish Kroner (DK), French Franc (FF), German Mark (GM), Italian Lire (IL), Japanese Yen (JY), Norwegian Kroner (NK), Swiss Franc (SF), Swedish Kroner (SK), and Spanish Peseta (SP), all with respect to US Dollar. These data are taken from Datastream. To facilitate the interpretation of the autoregression parameters we have standardized the series such that the

		estimates based on							
		original data				bootstrap samples			
		$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}_\alpha$	$\hat{\sigma}_\beta$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}_\alpha$	$\hat{\sigma}_\beta$
AD	QMLE	.129	.843	.075	.034	.116	.828	.044	.063
	1-step	.253	.867			.112	.844	.027	.024
AS	QMLE	.075	.891	.013	.016	.073	.888	.018	.018
	1-step	.113	.897			.072	.890	.014	.013
BF	QMLE	.068	.902	.011	.023	.068	.899	.019	.018
	1-step	.093	.906			.065	.903	.013	.014
BP	QMLE	.052	.932	.008	.013	.051	.931	.015	.012
	1-step	.055	.932			.050	.931	.012	.010
CD	QMLE	.138	.798	.042	.062	.139	.793	.032	.031
	1-step	.169	.809			.133	.797	.021	.021
DG	QMLE	.078	.888	.013	.016	.077	.886	.018	.018
	1-step	.107	.916			.076	.887	.015	.014
DK	QMLE	.067	.902	.011	.016	.065	.898	.015	.016
	1-step	.095	.920			.064	.901	.012	.014
FF	QMLE	.088	.873	.016	.017	.088	.869	.023	.022
	1-step	.119	.913			.085	.872	.017	.017
GM	QMLE	.073	.894	.012	.013	.073	.891	.017	.018
	1-step	.095	.925			.072	.893	.014	.015
IL	QMLE	.093	.869	.016	.031	.092	.864	.022	.020
	1-step	.109	.896			.090	.867	.019	.017
JY	QMLE	.059	.891	.017	.025	.060	.888	.016	.026
	1-step	.078	.912			.057	.891	.012	.021
NK	QMLE	.080	.907	.009	.014	.078	.906	.023	.015
	1-step	.092	.916			.075	.908	.017	.010
SF	QMLE	.059	.904	.012	.013	.058	.901	.014	.019
	1-step	.064	.922			.057	.903	.012	.015
SK	QMLE	.221	.754	.035	.119	.210	.751	.049	.038
	1-step	.185	.839			.209	.756	.032	.020
SP	QMLE	.106	.871	.014	.032	.104	.868	.027	.020
	1-step	.171	.912			.100	.870	.021	.014

Table 4.2: Comparison of QMLE and a semiparametric one-step estimator for several logarithmic differenced daily exchange rate series. Observation period January 1, 1980 to April 1, 1994 ($n = 3719$). The first part of the table gives the estimates based on the original data set. Estimated standard deviations are deleted for the semiparametric estimators. The sample means and sample standard deviations of 500 bootstrap replications are given in the second half of the table.

QMLE of σ is one. In all series both the QMLE method and the semiparametric procedure estimate the persistence $\alpha\sigma^2 + \beta$ less than one (for the semiparametric estimates this cannot be inferred from Table 4.2 since the semiparametric estimate of σ is not constrained to be one). The estimates based on the original data sets are given in the first four columns of Table 4.2. Of course we used the variance formula $A^{-1}BA^{-1}$ for the direct estimate of the standard deviation of the QMLE. As described above, the parameter estimates produced by the semiparametric procedure are not very sensitive to the choice of the bandwidth. However, it turns out that the direct variance estimates change dramatically (even for small changes of the bandwidth). Therefore, these estimates are not reliable and they have been deleted from the table.

For the simulation study above the situation was quite different since we estimated the variance of the semiparametric one-step estimators from independent parameter estimates in the replications. Here we have only one data set. Independent replications are not available. This leads to the following paradox. On the one hand one may have the imprecise QML estimate with quite large estimated standard deviations. So it may be possible that the hypotheses of integrated GARCH or no conditional heteroskedasticity cannot be rejected. On the other side one has the improved semiparametric estimate which allows for more powerful tests. But since the estimated standard deviations are unreliable one can get any answer one wants by changing the bandwidth. To avoid this paradox, we propose to use the bootstrap. I.e. simulate replications of the original data set with the estimated parameter and the estimated innovation distribution as inputs and proceed as in the case of simulations described above. Then we have several parameter estimates available from which we calculate the straightforward sample estimate of the variance. In this manner we only rely upon the parameter estimates and not on direct estimates of the variance. Hence, the variability of the variance due to different bandwidth choices is greatly reduced. Some simulation experiments show that this procedure works quite well. We apply the bootstrap procedure to our data sets and we report the sample means and sample standard deviations in the final four columns of Table 4.2. Observe that the estimated standard deviations of the semiparametric estimators of the heteroskedastic parameters are between four tenth (AD) and nine tenth (IL) of the estimated ones for the QMLE method. This implies the efficiency of the QMLE method lies approximately in the interval $(.15, .80)$ in these special examples. The efficiency gain is also supported by the plots in Figure 4.1 of the nonparametric density estimates and the corresponding score estimates which are far away from the normal density and score. Although these figures suggest some skewness of the exchange rate densities, they are close to the densities of student t_ν-distributions with ν between 4.1 and 5.4. If the true underlying density would be symmetric, we expect from the simulation study that the symmetric nonparametric procedure performs slightly better in moderate samples. However, in the exchange rate applications the latter procedure yields somewhat larger standard deviations (.003 for AS and less than .002 for the others, these values are not reported here). This indicates that the true densities are not fully symmetric and hence the symmetric semiparametric approach may lead to wrong conclusions. Since the possible moderate sample loss is very small it seems to be safer to use the ordinary non-symmetric improvement.

Finally, we note that the simulation results of Table 4.1 show that all estimators, even the unfeasible MLE, tend to underestimate the heteroskedasticity parameters. This negative bias explains why in Table 4.2 on the average the bootstrap estimates are less in value than the original estimates.

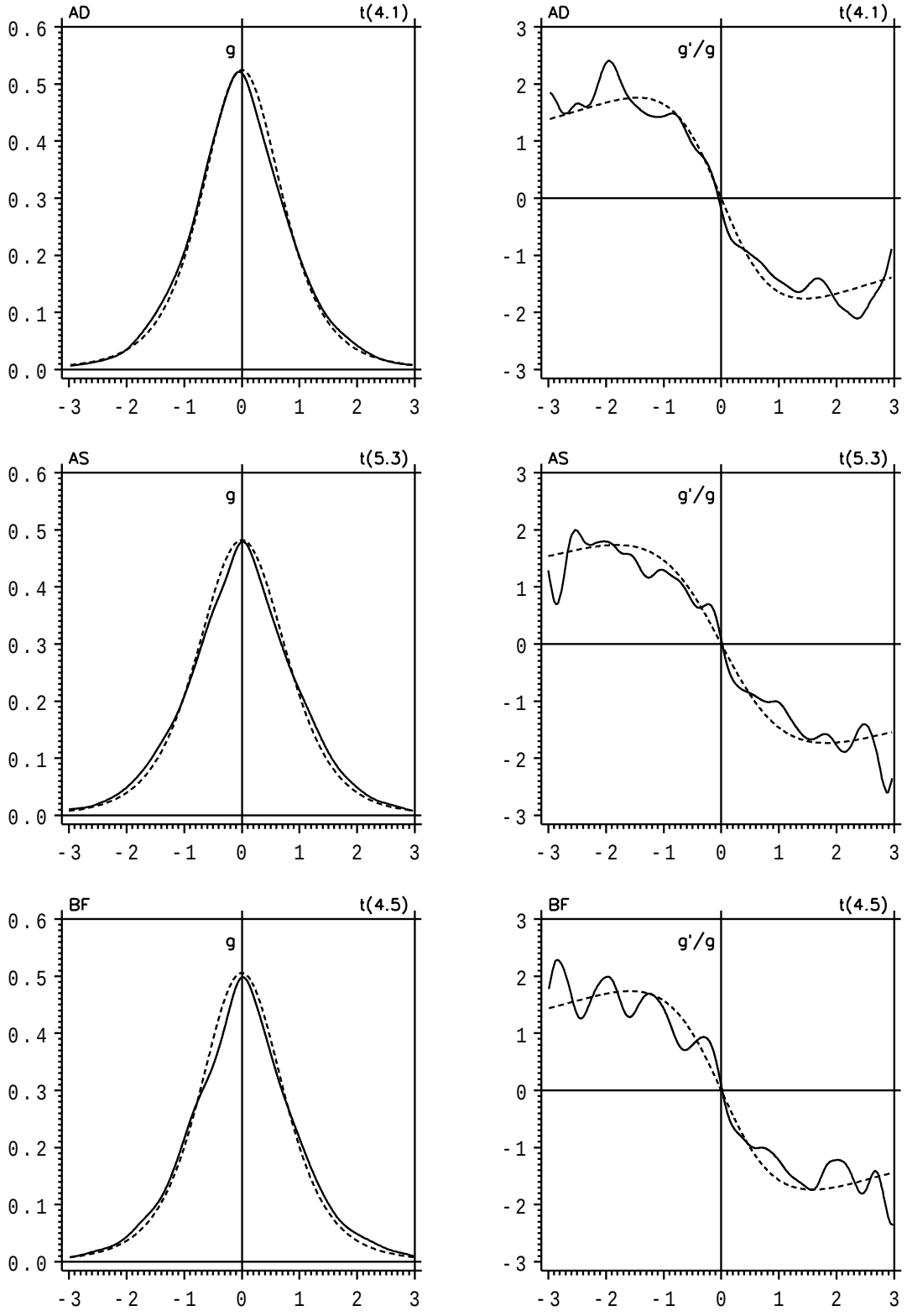

Figure 4.1: Comparison of estimated densities and scores with t_ν-densities and scores for several logarithmic differenced daily exchange rate series. Observation period January 1, 1980 to April 1, 1994 ($n = 3719$).

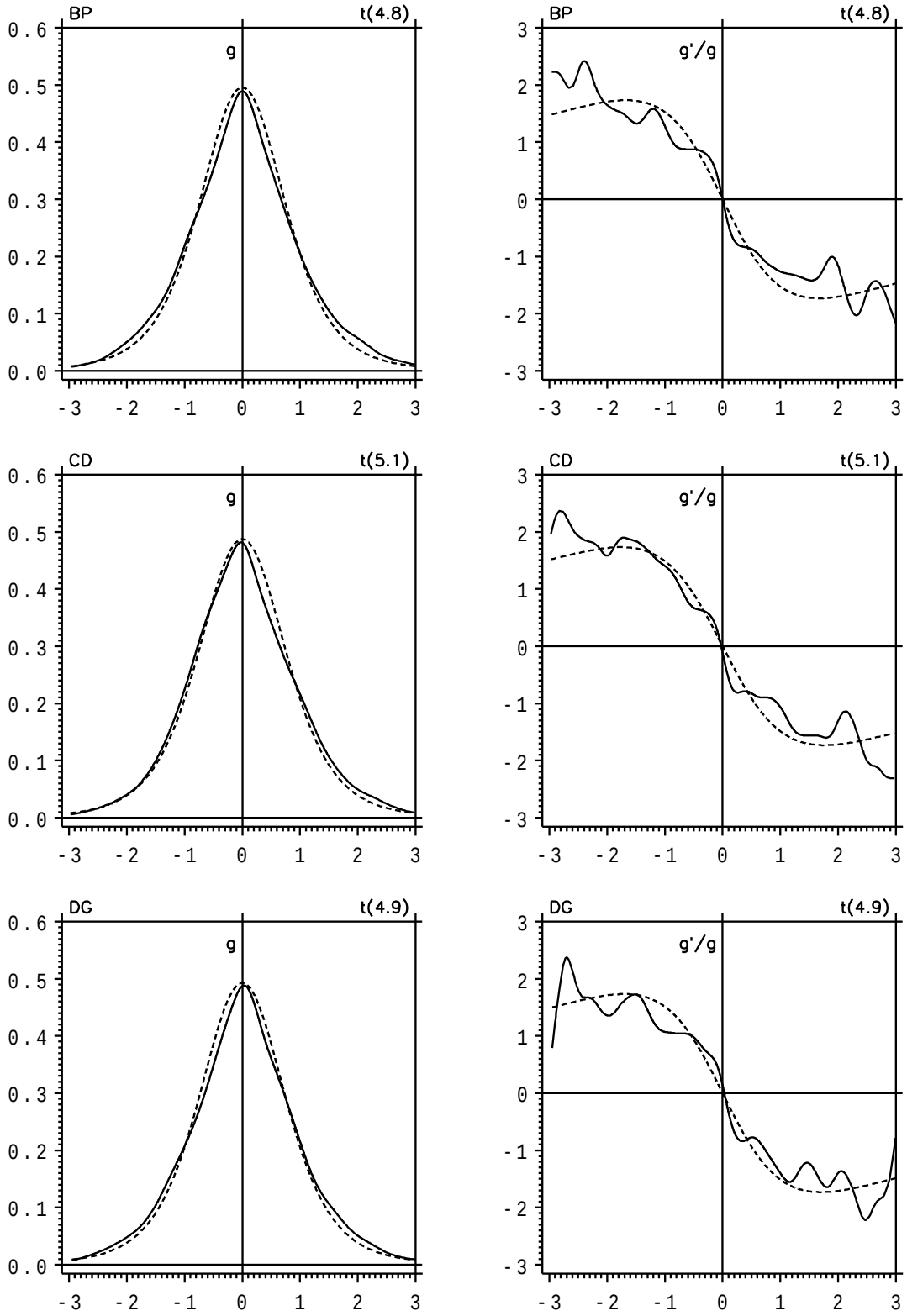

Figure 4.1: (CONTINUED)

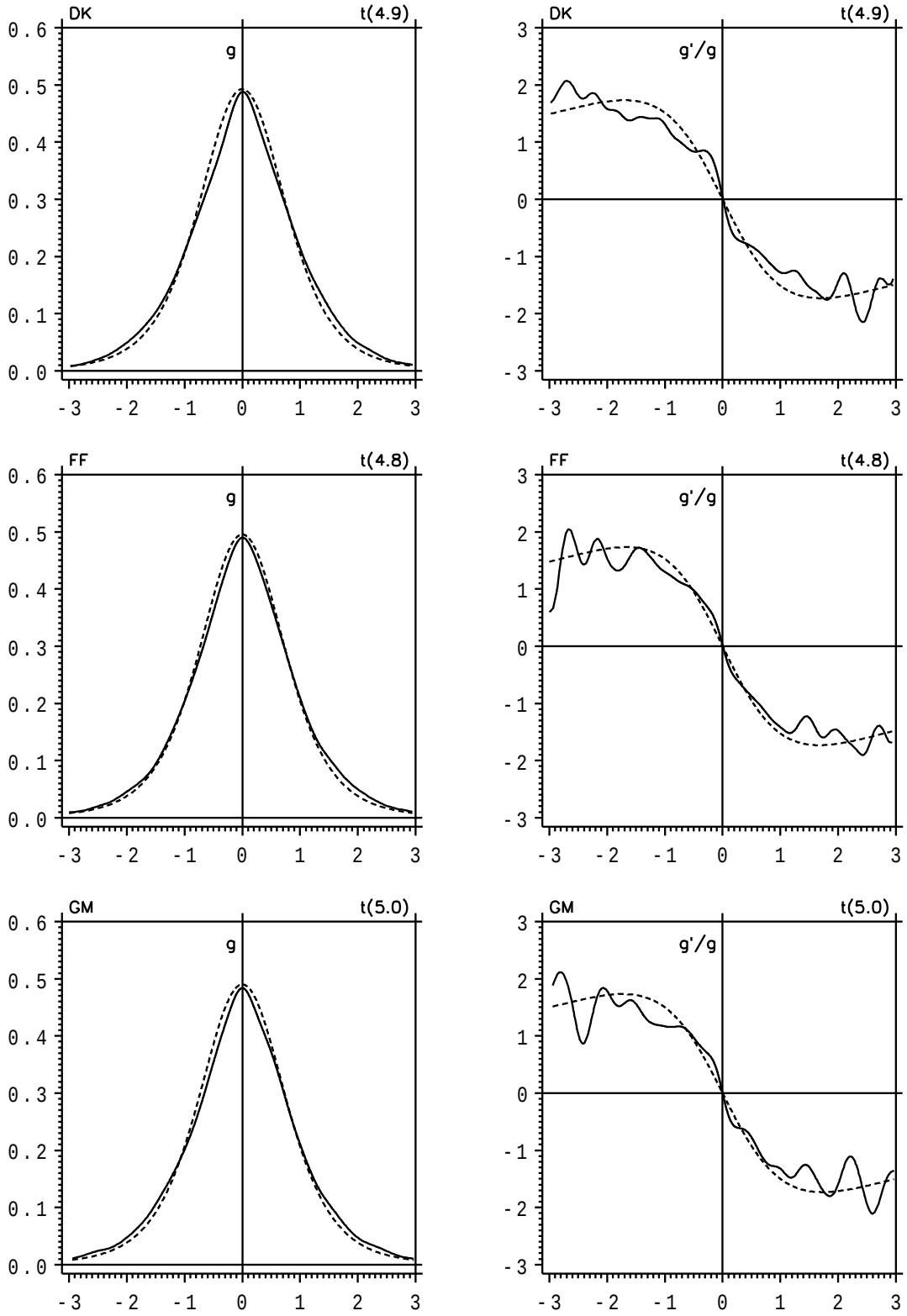

Figure 4.1: (CONTINUED)

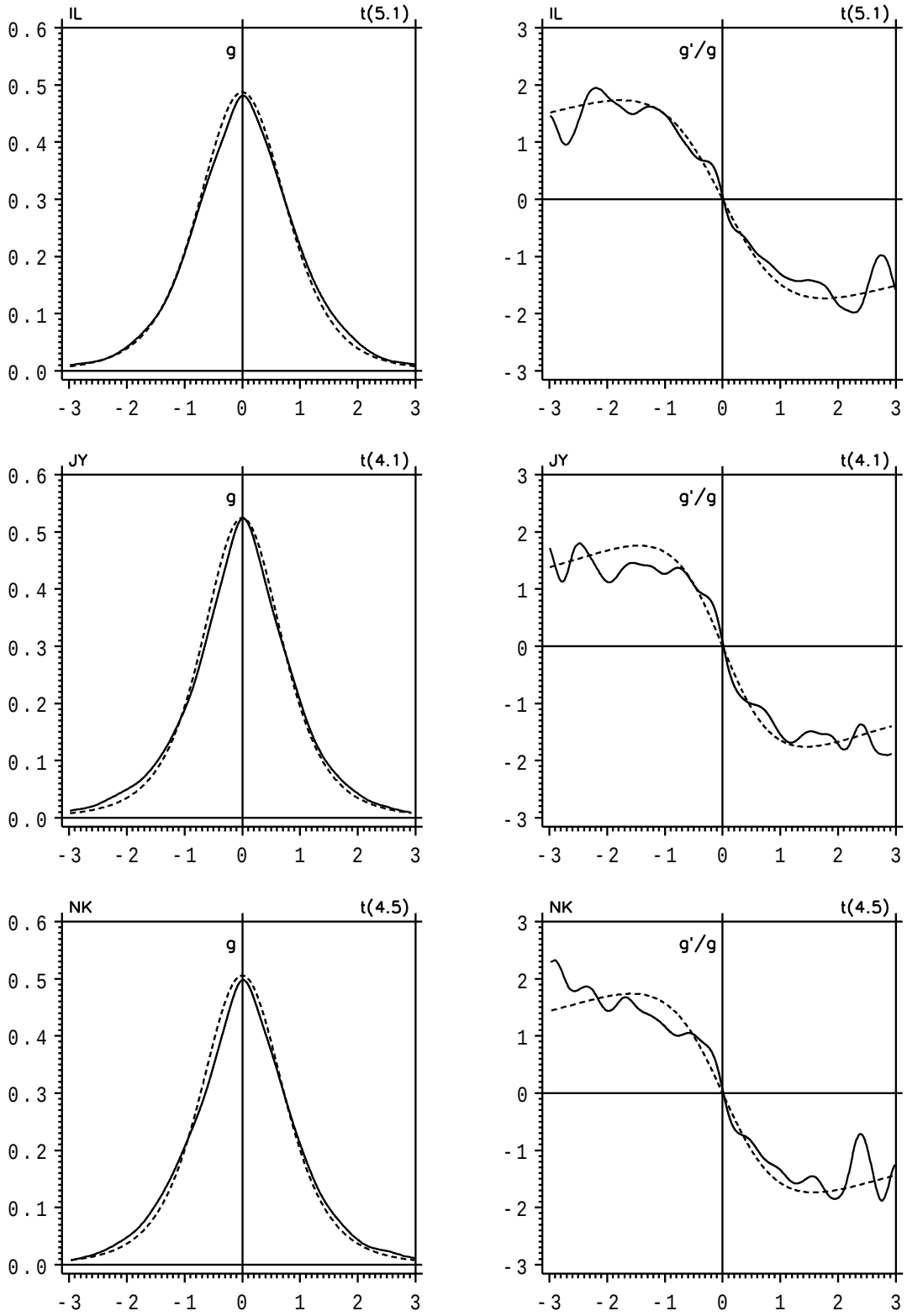

Figure 4.1: (CONTINUED)

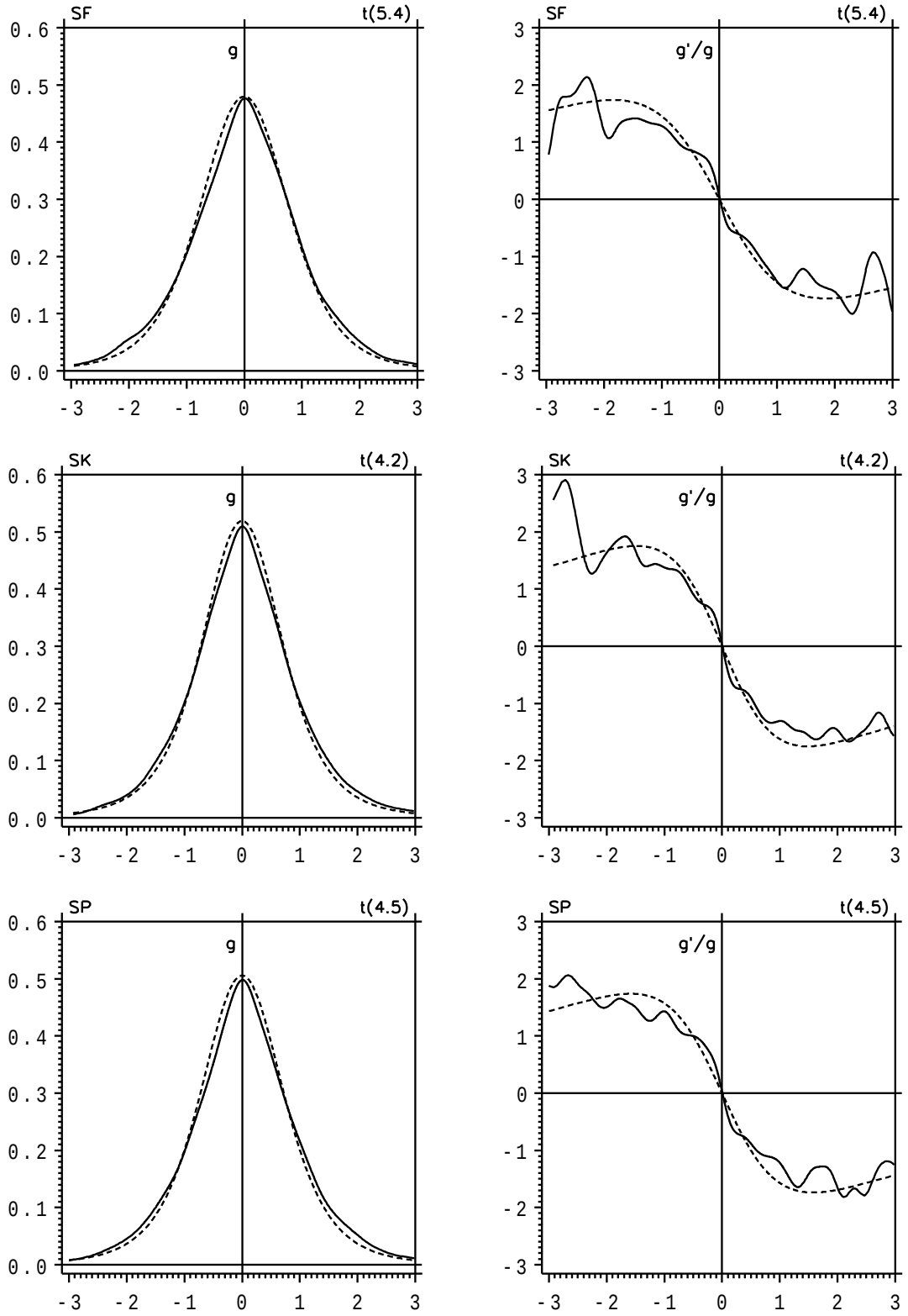

Figure 4.1: (CONTINUED)

5 Conclusions.

In this paper we studied the semiparametric properties of (integrated) GARCH-M type models. In this model, adaptive estimation is not possible. This fact is completely caused by a location-scale parameter. After a suitable reparametrization of the model we showed that the estimation problem of the parameters characterizing the conditional heteroskedastic character of the process is equally difficult in cases where the innovation distribution is known or unknown, respectively. In that sense we may call these parameters still adaptively estimable. This property is derived in a general GARCH context avoiding moment conditions and including integrated GARCH models. The simulations showed that this property is not only interesting from a theoretical point of view. In moderate sample sizes with $n = 1000$ observations, usually available in financial time-series, the semiparametric procedures work reasonably well. Most of the loss caused by the QMLE method (instead of the infeasible MLE method) is regained by the one-step estimator in case of the interesting group of heavy-tailed alternatives. Moreover, the empirical example showed that the efficiency loss caused by the QMLE method may be considerable.

It is clear from the exposition in this paper that the adaptivity results carry over to complicated models with time dependent mean and variance structures, e.g., ARMA with GARCH errors. The basic conditions given in DKW(1994b) do not seem to put serious restrictions on the models. However, a complete verification of the technical details may be much more demanding.

A Appendix.

PROOF OF THE LAN THEOREM 2.1: Since the general GARCH model (2.1)–(2.2) is a location-scale model in which the location-scale parameter only depends on the past, our model fits into the general time-series framework of DKW(1994b), especially Section 4. Therefore, it suffices to verify the conditions (2.3′), (A.1), and (2.4) of DKW(1994b). In passing we also prove (3.3′) of DKW(1994b) which we will need in the proof of Theorem 3.1. I.e., with the notation introduced in (2.10), (2.14), and (2.15), and $I_{ls}(g)$ the expectation under θ of the product $\psi(\theta)\psi(\theta)'$, we have to show, under θ_0,

$$n^{-1}\sum_{t=1}^{n} W_t(\theta_0) I_{ls}(g) W_t(\theta_0)' \xrightarrow{P} I(\theta_0) > 0,\ n^{-1}\sum_{t=1}^{n} |W_t(\theta_0)|^2 1_{\{n^{-1/2}|W_t(\theta_0)|>\delta\}} \xrightarrow{P} 0, \quad \text{(A.1)}$$

$$n^{-1}\sum_{t=1}^{n} W_t(\theta_0) \xrightarrow{P} W(\theta_0), \quad \text{(A.2)}$$

$$n^{-1}\sum_{t=1}^{n} |W_t(\theta_n) - W_t(\theta_0)|^2 \xrightarrow{P} 0, \quad \text{(A.3)}$$

and, under θ_n,

$$\sum_{t=1}^{n} |n^{-1/2}(M_{nt}, S_{nt})' - W_t(\theta_n)'(\tilde{\theta}_n - \theta_n)|^2 \xrightarrow{P} 0, \quad \text{(A.4)}$$

for some positive definite matrix $I(\theta_0)$ and some random matrix $W(\theta_0)$. Together with their Lemma A.1, these four relations yield the desired conclusions. We prepare the proof by deriving some helpful results.

Although $W_t(\theta_0)$ is not stationary under θ_0, the following proposition shows that these variables can be approximated by a stationary sequence.

Proposition A.1 *Let $h_t(\theta)$, $H_t(\theta)$, and $W_t(\theta)$ be given by* (2.9), (2.13), *and* (2.14), *respectively, and let $h_{st}(\theta)$, $H_{st}(\theta)$, and $W_{st}(\theta)$ be their corresponding stationary solutions under θ, i.e.*

$$h_{st}(\theta) = \sum_{j=0}^{\infty} \prod_{k=1}^{j} \{\beta + \alpha \xi_{t-k}^2\},\ H_{st}(\theta) = \sum_{i=0}^{\infty} \beta^i h_{s,t-1-i}(\theta) \begin{pmatrix} \xi_{t-1-i}^2 \\ 1 \end{pmatrix},$$

$$W_{st}(\theta) = \sigma^{-1} \begin{pmatrix} \frac{1}{2} h_{st}^{-1}(\theta) H_{st}(\theta)(\mu, \sigma) \\ I_2 \end{pmatrix}.$$

Then, under θ_0,

$$n^{-1} \sum_{t=1}^{n} |W_t(\theta_0) - W_{st}(\theta_0)|^2 \to 0 \text{ (a.s.)}, \text{ as } n \to \infty. \tag{A.5}$$

PROOF: We adopt the convention that empty sums are equal to zero while empty products are equal to one. Iterating $h_t(\theta)$ yields

$$\begin{aligned} h_t(\theta) &= 1 + \beta h_{t-1}(\theta) + \alpha y_{t-1}^2 = 1 + h_{t-1}(\theta)\{\beta + \alpha \xi_{t-1}^2(\theta)\} \\ &= \sum_{j=0}^{i-1} \prod_{k=1}^{j} \{\beta + \alpha \xi_{t-k}^2(\theta)\} + h_{t-i}(\theta) \prod_{k=1}^{i} \{\beta + \alpha \xi_{t-k}^2(\theta)\},\ 0 \le i \le t-1, \end{aligned} \tag{A.6}$$

and hence

$$h_{t-i}(\theta)/h_t(\theta) \le \prod_{k=1}^{i} \{\beta + \alpha \xi_{t-k}^2(\theta)\}^{-1},\ 0 \le i \le t-1. \tag{A.7}$$

Under θ, the calculated variables $\xi_t(\theta)$ simply are the true innovations ξ_t in (A.6) and (A.7). For the stationary random variables $h_{st}(\theta)$ we obtain similar relations,

$$\begin{aligned} h_{st}(\theta) &= \sum_{j=0}^{i-1} \prod_{k=1}^{j} \{\beta + \alpha \xi_{t-k}^2\} + h_{s,t-i}(\theta) \prod_{k=1}^{i} \{\beta + \alpha \xi_{t-k}^2\},\ 0 \le i, \\ h_{s,t-i}(\theta)/h_{st}(\theta) &\le \prod_{k=1}^{i} \{\beta + \alpha \xi_{t-k}^2\}^{-1},\ 0 \le i, \end{aligned}$$

and hence, under θ, we obtain

$$\begin{aligned} &|h_{st}(\theta) h_{t-i}(\theta) - h_t(\theta) h_{s,t-i}(\theta)| = |h_{s,t-i}(\theta) - h_{t-i}(\theta)| \sum_{j=0}^{i-1} \prod_{k=1}^{j} \{\beta + \alpha \xi_{t-k}^2\} \\ &\le\ h_t(\theta) |h_{s1}(\theta) - h_1(\theta)| \prod_{k=1}^{t-1-i} \{\beta + \alpha \xi_{t-i-k}^2\} \\ &=\ h_t(\theta) |h_{s1}(\theta) - h_1(\theta)| \prod_{k=1}^{i} \{\beta + \alpha \xi_{t-k}^2\}^{-1} \prod_{k=1}^{t-1} \{\beta + \alpha \xi_k^2\},\ 0 \le i \le t-1. \end{aligned}$$

With C some generic constant only depending on θ we obtain, under θ,

$$\begin{aligned}
|W_t(\theta) - W_{st}(\theta)| &\le C|H_t(\theta)/h_t(\theta) - H_{st}(\theta)/h_{st}(\theta)| \\
&\le C\sum_{i=0}^{t-2}\beta^i|\frac{h_{t-1-i}(\theta)}{h_t(\theta)} - \frac{h_{s,t-1-i}(\theta)}{h_{st}(\theta)}|\left|\begin{pmatrix}\xi^2_{t-1-i}\\ 1\end{pmatrix}\right| + C\sum_{i=t-1}^{\infty}\beta^i\frac{h_{s,t-1-i}(\theta)}{h_{st}(\theta)}\left|\begin{pmatrix}\xi^2_{t-1-i}\\ 1\end{pmatrix}\right| \\
&\le C|h_{s1}(\theta) - h_1(\theta)|\prod_{k=1}^{t-1}\{\beta+\alpha\xi_k^2\}\sum_{i=0}^{t-2}\prod_{k=1}^{i}\frac{\beta}{\beta+\alpha\xi^2_{t-k}} + C\sum_{i=t-1}^{\infty}\prod_{k=1}^{i}\frac{\beta}{\beta+\alpha\xi^2_{t-k}} \\
&\le C|h_{s1}(\theta) - h_1(\theta)|(t-1)\prod_{k=1}^{t-1}\{\beta+\alpha\xi_k^2\} + C\prod_{k=1}^{t-1}\frac{\beta}{\beta+\alpha\xi_k^2}\sum_{i=-1}^{\infty}\prod_{k=0}^{i}\frac{\beta}{\beta+\alpha\xi^2_{-k}}.
\end{aligned}$$

By (2.3) the right-hand side tends to zero (a.s.), as $t\to\infty$. Cesàro's Theorem completes the proof of the proposition. □

Intuitively it is clear that slight perturbations of the parameters yield solutions of equations (2.1) and (2.2) that are close. The following proposition makes this more precise. Just as expected from heuristic formal calculations, the leading term of $\tilde{h}_{nt}/h_{nt} - 1$ is a linear combination of the components of H_{nt}/h_{nt} which appears in the score $\dot{l}_{nt}$.

Proposition A.2 *Let $h_t(\theta)$ and $H_t(\theta)$ be given by* (2.9) *and* (2.13)*, respectively, and define*

$$Q_t(\theta) = H_t(\theta)/h_t(\theta) = \sum_{i=0}^{t-2}\beta^i\begin{pmatrix}y^2_{t-1-i}\\ h_{t-1-i}(\theta)\end{pmatrix}/h_t(\theta) = \sum_{i=0}^{t-2}\beta^i\frac{h_{t-1-i}(\theta)}{h_t(\theta)}\begin{pmatrix}\xi^2_{t-1-i}(\theta)\\ 1\end{pmatrix},$$

$$R_t(\theta,\tilde{\theta}) = h_t(\tilde{\theta})/h_t(\theta) - 1 - (\tilde{\alpha}-\alpha,\tilde{\beta}-\beta)Q_t(\theta).$$

Let θ_n and $\tilde{\theta}_n$ satisfy the conditions just above (2.6)*. Put $Q_{nt} = Q_t(\theta_n)$ and $R_{nt} = R_t(\theta_n,\tilde{\theta}_n)$. Then, under θ_n,*

$$n^{-1}\sum_{t=1}^{n}|Q_{nt}|^2 = O_P(1),\ n^{-1}\sum_{t=1}^{n}|Q_{nt}|^2 1_{\{n^{-1/2}|Q_{nt}|>\delta\}}\to 0,\ \text{(a.s.)},\ \text{as } n\to\infty, \tag{A.8}$$

$$\sum_{t=1}^{n}R^2_{nt}\to 0,\ \text{(a.s.)},\ \text{as } n\to\infty. \tag{A.9}$$

PROOF: By equation (A.7) we obtain

$$Q_t(\theta)\le\beta^{-1}\sum_{i=0}^{t-2}\prod_{k=1}^{i+1}\frac{\beta}{\beta+\alpha\xi^2_{t-k}(\theta)}\begin{pmatrix}\xi^2_{t-1-i}(\theta)\\ 1\end{pmatrix}.$$

For n sufficiently large, this latter relation shows that, under θ_n, $|Q_{nt}|$ may be bounded by the product of a constant only depending on θ_0 and the stationary sequence

$$S_t = \sum_{i=0}^{\infty}\prod_{k=1}^{i}\frac{\beta_0}{\beta_0+\frac{1}{2}\alpha_0\xi^2_{t-k}}.$$

Note that all moments of S_t exist. The relations concerning Q_{nt} are easily obtained.

To prove the result concerning the remainder term R_{nt} note that an explicit relationship for the difference of $h_t(\tilde{\theta})$ and $h_t(\theta)$ is given by [compare (2.3) of Kreiss (1987a)]

$$h_t(\tilde{\theta}) - h_t(\theta) = \sum_{i=0}^{t-2} \tilde{\beta}^i h_{t-1-i}(\theta)(\tilde{\alpha} - \alpha, \tilde{\beta} - \beta) \begin{pmatrix} \xi_{t-1-i}^2(\theta) \\ 1 \end{pmatrix}.$$

Hence, the remainder term $R_t(\theta, \tilde{\theta})$ is given by

$$R_t(\theta, \tilde{\theta}) = \sum_{i=0}^{t-2} (\tilde{\beta}^i - \beta^i) \frac{h_{t-1-i}(\theta)}{h_t(\theta)} (\tilde{\alpha} - \alpha, \tilde{\beta} - \beta) \begin{pmatrix} \xi_{t-1-i}^2(\theta) \\ 1 \end{pmatrix}.$$

Choose $c > 1$ such that $Ec\beta_0/(\beta_0 + \frac{1}{2}\alpha_0\xi_1^2) < 1$. By the mean value theorem, there exists a $\tilde{\beta}_{ni}$ in between $\tilde{\beta}_n$ and β_n such that, for n sufficiently large,

$$|\tilde{\beta}_n^i - \beta_n^i| = |\tilde{\beta}_n - \beta_n| i \tilde{\beta}_{ni}^{i-1} \leq |\tilde{\beta}_n - \beta_n| i c^{i-1} \beta_n^{i-1}, \ i \geq 0.$$

Just as for Q_{nt}, we may bound R_{nt} by the product of a constant times n^{-1} and the stationary sequence

$$S_t^* = \sum_{i=0}^{\infty} i \prod_{k=1}^{i} \frac{c\beta_0}{\beta_0 + \frac{1}{2}\alpha_0 \xi_{t-k}^2}.$$

The proof of the proposition can be easily completed. □

Now we are ready to prove (A.1)–(A.4). Define $I(\theta_0) = E_{\theta_0} W_{s1}(\theta_0) I_{ls}(g) W_{s1}(\theta_0)'$ and $W(\theta_0) = E_{\theta_0} W_{s1}(\theta_0)$ (the existence of these quantities can be obtained along the lines of the proof of Proposition A.2 since $|W_{st}(\theta_0)|$ is bounded by the product of S_t and a constant depending on θ_0, only). Obviously the relations (A.1) and (A.2) hold true if $W_t(\theta_0)$ is replaced by the stationary ergodic sequence $W_{st}(\theta_0)$. Consequently, Proposition A.1 implies the validity of these relations for $W_t(\theta_0)$ itself.

To prove (A.4) we will use Proposition A.2. Writing $\lambda_n = (\lambda_{1n}', \lambda_{2n}')'$ with λ_{1n} (λ_{2n}) the first (latter) two components of λ_n, and defining

$$\chi(x) = \left\{-1 + 2(\sqrt{1+x} - 1)/x\right\} 1_{\{x \geq -1\}},$$

we see

$$\begin{aligned} &\sum_{t=1}^{n} |n^{-1/2}(M_{nt}, S_{nt})' - W_t(\theta_n)'(\tilde{\theta}_n - \theta_n)|^2 = \\ &\quad = \sigma_n^{-2} n^{-1} \sum_{t=1}^{n} \Big| (\tilde{\mu}_n, \tilde{\sigma}_n)' \left\{ \frac{1}{2}(\lambda_{1n}' Q_{nt} + \sqrt{n} R_{nt}) \chi(n^{-1/2} \lambda_{1n}' Q_{nt} + R_{nt}) + \frac{1}{2}\sqrt{n} R_{nt} \right\} \\ &\qquad + n^{-1/2} \lambda_{2n} \frac{1}{2} \lambda_{1n}' Q_{nt} \Big|^2. \end{aligned}$$

Together with Proposition A.2, Lemma 2.1 of DKW(1994b) [with $Y_{nt} = \lambda_{1n}' Q_{nt}$, $X_{nt} = \lambda_{1n}' Q_{nt} + \sqrt{n} R_{nt}$, and the function $\phi = \chi^2$ as above] yields (A.4).

Finally, we have to prove (A.3). Note that

$$|W_t(\theta_n) - W_t(\theta_0)|^2 \leq C|Q_t(\theta_n) - Q_t(\theta_0)|^2 + C|Q_t(\theta_0)|^2 |\theta_n - \theta_0|^2$$

and obtain contiguity of P_{θ_n} and P_{θ_0} from (A.1) and (A.4), and Theorem 2.1 of DKW(1994b). Then the required result is easily obtained from

$$\begin{aligned} Q_t(\tilde{\theta}) - Q_t(\theta) &= \sum_{i=0}^{t-2} (\tilde{\beta}^i - \beta^i) \frac{h_{t-1-i}(\tilde{\theta})}{h_t(\tilde{\theta})} \begin{pmatrix} \xi^2_{t-1-i}(\tilde{\theta}) \\ 1 \end{pmatrix} \\ &+ Q_t(\theta) \left\{ (\theta_1 - \tilde{\theta}_1)' Q_t(\tilde{\theta}) + R_t(\tilde{\theta}, \theta) \right\} \\ &- \begin{pmatrix} 0 \\ \sum_{i=0}^{t-2} \beta^i \frac{h_{t-1-i}(\tilde{\theta})}{h_t(\tilde{\theta})} \left\{ (\theta_1 - \tilde{\theta}_1)' Q_{t-1-i}(\tilde{\theta}) + R_{t-1-i}(\tilde{\theta}, \theta) \right\} \end{pmatrix} \end{aligned}$$

along the lines of the proofs of the propositions above. This completes the proofs of the theorems in Section 2. □

PROOF OF THEOREM 3.1: It suffices to verify the conditions of DKW(1994b). These reduce to (A.1)–(A.4) above, which are verified there, and the existence of an estimator $\hat{\psi}_n(\cdot)$, based on $\varepsilon_1, \ldots, \varepsilon_n$, of $\psi(\cdot) = -(l'(\cdot), 1 + \cdot l'(\cdot))'$, from (2.15), satisfying the consistency condition

$$\int |\hat{\psi}_n(x) - \psi(x)|^2 g(x) dx \xrightarrow{P} 0, \text{ under } g.$$

Indeed, such an estimator exists in view of Proposition 7.8.1, p. 400, of BKRW(1993) with $k = 0$ and $k = 1$; see also Lemma 4.1 of Bickel (1982). The estimator $\hat{\psi}_n(\cdot)$ in Section 3 is based on these constructions. □

References

BAILLIE, R.T., and BOLLERSLEV, T. (1989), The message in daily exchange rates: a conditional-variance tale, *Journal of Business and Economic Statistics*, **7**, 297–305.

BICKEL, P.J. (1982), On adaptive estimation, *Annals of Statistics*, **10**, 647–671.

BICKEL, P.J., KLAASSEN, C.A.J., RITOV, Y., and WELLNER, J.A. (1993), *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopkins University Press, Baltimore.

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

BOLLERSLEV, T., CHOU, R.Y., and KRONER, K.F. (1992), ARCH modeling in finance: a review of the theory and empirical evidence, *Journal of Econometrics*, **52**, 5–59.

DIEBOLD, F.X. (1988), *Empirical Modeling of Exchange Rate Dynamics*, Springer Verlag, New York.

DROST, F.C., KLAASSEN, C.A.J., and WERKER, B.J.M. (1994a), Adaptiveness in time-series models, in MANDL, P. and HUŠKOVÁ, M., *Asymptotic Statistics*, 203–212, Physica Verlag, New York.

DROST, F.C., KLAASSEN, C.A.J., and WERKER, B.J.M. (1994b), Adaptive estimation in time-series models, CentER Discussion Paper 9488, Tilburg University.

DROST, F.C., and NIJMAN, T.E. (1993), Temporal aggregation of GARCH processes, *Econometrica*, **61**, 909–927.

DROST, F.C., and WERKER, B.J.M. (1995), Closing the GARCH gap: continuous time GARCH modeling, *Journal of Econometrics*, forthcoming.

ENGLE, R.F. (1982), Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, **50**, 987–1007.

ENGLE, R.F., and BOLLERSLEV, T. (1986), Modelling the persistence of conditional variances, *Econometric Reviews*, **5**, 1–50, 81–87.

ENGLE, R.F., and GONZÁLEZ-RIVERA, G. (1991), Semiparametric ARCH models, *Journal of Business and Economic Statistics*, **9**, 345–359.

ENGLE, R.F., LILIEN, D.M., and ROBINS, R.P. (1987), Estimating time varying risk premia in the term structure: the ARCH-M model, *Econometrica*, **55**, 391–407.

GOURIEROUX, C., and MONFORT, A. (1992), Qualitative threshold ARCH models, *Journal of Econometrics*, **52**, 159–199.

HÁJEK, J. (1970), A characterization of limiting distributions of regular estimates, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **14**, 323–330.

HÁJEK, J. and ŠIDÁK, Z. (1967), *Theory of Rank Tests*, Academia, Prague.

HUBER, P.J. (1981), *Robust Statistics*, Wiley, New York.

JEGANATHAN, P. (1988), Some aspects of asymptotic theory with applications to time series models, Technical Report 166, Department of Statistics, University of Michigan.

KLAASSEN, C.A.J. (1987), Consistent estimation of the influence function of locally asymptotically linear estimators, *Annals of Statistics*, **15**, 1548–1562.

KOUL, H.L. and SCHICK, A. (1995), Efficient estimation in nonlinear time series models, Technical Report, Department of Statistics, University of Michigan.

KREISS, J.-P. (1987a), On adaptive estimation in stationary ARMA processes, *Annals of Statistics*, **15**, 112–133.

KREISS, J.-P. (1987b), On adaptive estimation in autoregressive models when there are nuisance functions, *Statistics and Decisions*, **5**, 112-133.

LEE, S.-W. and HANSEN, B.E. (1994), Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator, *Econometric Theory*, **10**, 29–52.

LINTON, O. (1993), Adaptive estimation in ARCH models, *Econometric Theory*, **9**, 539–569.

LUMSDAINE, R.L. (1989), Asymptotic properties of the maximum likelihood estimator in GARCH(1,1) and IGARCH(1,1) models, Manuscript, Department of Economics, Harvard University.

NELSON, D.B. (1990a), ARCH models as diffusion approximations, *Journal of Econometrics*, **45**, 7–38.

NELSON, D.B. (1990b), Stationarity and persistence in the GARCH(1,1) model, *Econometric Theory*, **6**, 318–334.

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

NEWEY, W.K. (1990), Semiparametric efficiency bounds, *Journal of Applied Econometrics*, **5**, 99–135.

PÖTSCHER, B.M. (1995), Comment on 'Adaptive estimation in time series regres-

sion models' by D.G. Steigerwald, *Journal of Econometrics*, **66**, 123–129.

ROBINSON, P.M. (1988), Semiparametric econometrics: a survey, *Journal of Applied Econometrics*, **3**, 35–51.

SCHICK, A. (1986), On asymptotically efficient estimation in semiparametric models, *Annals of Statistics*, **14**, 1139–1151.

SCHOLZ, F.-W. (1971), Comparison of Optimal Location Estimators, PhD-thesis, University of California, Berkeley.

STEIGERWALD, D.G. (1992), Adaptive estimation in time-series regression models, *Journal of Econometrics*, **54**, 251–275.

STEIGERWALD, D.G. (1995), Reply to B.M. Pötscher's comment on 'Adaptive estimation in time series regression models', *Journal of Econometrics*, **66**, 131–132.

TAPIA, R.A. and THOMPSON, J.R. (1978), *Nonparametric probability density estimation*, Johns Hopkins University Press, Baltimore.

WEISS, A.A. (1986), Asymptotic theory for ARCH models: estimation and testing, *Econometric Theory*, **2**, 107–131.