Discussion Paper No. B-316

Axel Cron

Uniform Consistency of Modified Kernel Estimators in Parametric ARCH-Models

June 1995

Abstract:
This paper shows the uniform convergence in probability of a modified kernel estimator towards the Baire function representing the conditional variance provided the data generating process is given by a strictly stationary solution of a parametric ARCH(q)-model.

JEL Classification Numbers:C14,C22
Keywords: ARCH(q)-model,kernel estimation,nonparametric regression.

This research was supported by Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303.

Institute for Econometrics and Operations Research, University of Bonn.
Adenauerallee 24-42,D-53113 Bonn,Germany.

Contents

Chapter 1

Introduction

ARCH models have become very popular in the recent theory of financial markets because they have proven useful in explaining the empirical findings of volatility clustering and fat tailed distributions.
Since on the other hand financial time series typically consist of a large number of observations, the employment of nonparametric procedures, or more specifically of kernel estimators in order to identify the underlying data generating process is a natural approach. Not surprisingly this has been done extensively in literature. However, to our knowledge, no rigorous investigation of the asymptotic behaviour of the estimators employed in an ARCH framework is available. To fill this gap at least partially we here study the consistency of a modified kernel estimator provided the real data generating process is given by a solution of a parametric ARCH(q) model as introduced in the econometric literature by ENGLE(1982). But this can only be seen as a first step, since we have in mind an extension to nonparametric models in order to exploit the full flexibility of kernel estimators. Consequently, the main purpose of the current analysis is to reveal the particular problems of nonparametric ARCH estimation the narrow but well-known class of parametric ARCH(q) models.
The rest of the paper is organized as follows: In chapter 2 we introduce the classical ARCH(q) model and discuss some aspects of the corresponding solution theory. In chapter 3 we describe the prediction and estimation problem motivating the current analysis and propose a modified kernel estimator. The consistency of this filter in the ARCH(q) context is studied in chapter 4. The technique employed to prove the main results goes back to BIERENS(1983)

Chapter 2

The ARCH(q) Model

2.1 Basic assumptions and vector representation

Let ϵ_t denote a discrete time stochastic process of i. i. d. real-valued random variables with moments $E\epsilon_0 = 0$ and $E\epsilon_0^2 = 1$. The ARCH(q) model is then given by the stochastic difference equation

$$\begin{aligned} \eta_t &= \epsilon_t \cdot \sqrt{h_t}, \qquad t \in \mathbf{Z} \\ h_t &:= h(\eta_{t-1}, \ldots \eta_{t-q}) \\ &:= \alpha_0 + \sum_{i=1}^{q} \alpha_i \eta_{t-i}^2, \end{aligned} \tag{2.1}$$

where

$$\begin{gathered} q > 0 \\ \alpha_0 > 0 \quad \alpha_i \geq 0 \quad i = 1, \ldots, q. \end{gathered}$$

Instead of following Engle (1982) in assuming the conditional distribution of the $\eta_t's$ to be normal, we admit a more extensive class of distributions, only restricted by some mild regularity conditions.

The following set of assumptions specifies the class of models 2.1 we deal with in this paper:

(A. 1) $\sum_{i=1}^{q} \alpha_i < 1$.

(A. 2) *The distribution of the $\epsilon_t's$ is absolutely continuous with respect to Lebesgue measure. The corresponding density $f_\epsilon(u)$ is continuous and bounded on $\mathbb{R}$.*

(A. 3) $E\epsilon_0^4 < +\infty$.

(A. 4) *The density $f_\epsilon(u)$ is twice continuously differentiable on $\mathbb{R}$.*

2.2 Aspects of solution theory

Following an equivalent approach by BOUGEROL/PICARD (1992) for more general GARCH models, we start our solution analysis introducing an inflated state representation of equation 2.1 squared:

$$\eta_t^2 = \epsilon_t^2 \cdot h_t, \qquad\qquad t \in \mathbf{Z} \tag{2.2}$$

Without loss of generality we assume that $q \geq 2$ (adding some α_i equal to 0 if needed) and define, allowing for overparametrization in the regression problem we will deal with later, for some $k \geq q$, nonnegative $(k+2)$-dimensional vectors

$$Z_t := (h_{t+1}, \eta_t^2, \eta_{t-1}^2, \ldots, \eta_{t-k}^2)^T,$$

$$D := (\alpha_0, 0, \ldots, 0)^T,$$

and nonnegative $(k+2) \times (k+2)$ random matrices

$$A_t := \begin{pmatrix} \alpha_1\epsilon_t^2 & \alpha_2 & \cdots & \cdots & \cdots & \alpha_q & 0 & \cdots & 0 \\ \epsilon_t^2 & 0 & \cdots & \cdots & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 1 & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & 1 & 0 \end{pmatrix}.$$

We notice that $\{\eta_t^2\}_{t\in\mathbf{Z}}$ is a solution of 2.2 if and only if $\{Z_t\}_{t\in\mathbf{Z}}$ is a solution of

$$Z_t = A_t Z_{t-1} + D, \qquad\qquad t \in \mathbf{Z} \tag{2.3}$$

The existence of strictly stationary and integrable solutions of system 2.3 obviously depends on the limit behaviour of the product of i. i. d. random matrices

$$M_{t,\tau} := A_t A_{t-1} \ldots A_{t-\tau+1} \tag{2.4}$$

for $\tau \to \infty$. BOUGEROL/PICARD(1992) demonstrated how convergence results found by KESTEN/SPITZER(1984) can be exploited to analyze the asymptotics of $M_{t,\tau}$. We follow their procedure and present a first auxiliary result we shall extensively make use of in a later section:

Lemma 1
Suppose that (A.1) holds. Then,for some constants $m_1 > 0$, $m_2 > 0$ *and* $\forall t \in Z$

$$\begin{array}{llll} (i) & \|M_{t,\tau}\| & = & o(e^{-m_1\tau}) \quad a.\ s. \\ (ii) & E\{M_{t,\tau}\} & = & o(e^{-m_2\tau}) \\ (iii) & E\{\|M_{t,\tau}\|\} & = & o(e^{-m_2\tau}) \end{array}$$

where $\|\cdot\|$ *denotes the infinity-norm.*

Proof:
(i) The spectral radius ρ_0 of $E\{A_0\}$ is less than one under (A.1). Thus we can apply inequality (1.4) in Kesten/Spitzer (1984) to obtain $\forall t \in Z$

$$\lim_{\tau\to\infty} \frac{1}{\tau} \log \|M_{t,\tau}\| \leq \log \rho_0 < 0 \quad a.\ s.$$

which implies (i) for some $m_1 > 0$.
(ii) is obviously valid for $m_2 < -\ln \rho_0$.
(iii) The matrix $M_{t,\tau}$ is nonnegative[1]. Hence, the result follows immediatly from (ii).■

An immediate consequence of the preceding Lemma is the following result:

Lemma 2
Suppose that (A.1) holds. Then
(i) equation 2.3 has a strictly stationary and integrable solution $\{Z_t\}_{t\in Z}$,
(ii) equation 2.1 has a strictly stationary, square-integrable solution $\{\eta_t\}_{t\in Z}$.

Proof: (see Theorem 1. 3 in Bougerol/Picard (1992)).
(i) Lemma 1 (i) implies that the series

$$\sum_{s=1}^{\infty} M_{t,s} D$$

converges absolutely a. s. for any $t \in Z$. Therefore the sequence

$$Z_t = D + \sum_{s=1}^{\infty} M_{t,s} D \qquad t \in \mathbf{Z} \qquad (2.5)$$

[1] i.e. all entries are nonnegative.

is a well-defined nonnegative solution of 2.3 which can be written as

$$Z_t = F(A_t, A_{t-1}, \ldots)$$

for some measurable function F independent of t. Moreover, the process $\{A_t\}_{t\in\mathbf{Z}}$ is strictly stationary. This shows the strict stationarity of the solution 2.5. Its integrability follows from Lemma 1 (ii) and the monotone convergence theorem.
(ii) Let

$$\begin{aligned} \eta_t &:= 1_{\epsilon_t \geq 0}\sqrt{\eta_t^2} - 1_{\epsilon_t < 0}\sqrt{\eta_t^2} \qquad (2.6) \\ &= 1_{\epsilon_t \geq 0}\sqrt{Z_t^{(2)}} - 1_{\epsilon_t < 0}\sqrt{Z_t^{(2)}}, \end{aligned}$$

where $Z_t^{(2)}$ is the second [2] component of the vector Z_t that solves 2.3. Then $\{\eta_t\}_{t\in\mathbf{Z}}$ is a strictly stationary and square-integrable solution of 2.1. ∎

A further aspect of ARCH processes we are interested in is the kind of distribution of the variables that might serve as regressors in the estimation problem we shall deal with later: We ask whether the continuity and smoothness properties of the distributions of the innovations ϵ_t pass over to the column vector $x_t := (\eta_{t-1}, \ldots, \eta_{t-k})$. The affirmative answer is given in the next auxiliary result:

Lemma 3
(i) Suppose that $\{\eta_t\}_{t\in\mathbf{Z}}$ is a strictly stationary solution of 2.1 and that (A.2) holds. Then the distribution of the random vector $x_t := (\eta_{t-1}, \ldots, \eta_{t-k})$ is absolutely continuous with respect to Lebesgue measure with continuous density $f(u_1, \ldots, u_k)$.
(ii) If in addition (A.4) holds , this density is twice continuously differentiable on any compact subset $E \subset \mathbb{R}^k$.

Proof:
(i) We start showing the existence of a density: Let $\xi^l := (\eta_{-1}, \ldots, \eta_{-l})$ for any $l \geq 1$. Using algebraic induction and the usual rule for density transformation, it can be seen that for any $B \in \mathcal{B}_l^*$

$$\int_B \tilde{f}_{\xi^l}(u_1, \ldots, u_l \mid \eta_{-l-1}, \ldots, \eta_{-l-q}) du_1 \ldots du_l$$

[2] We could define $\eta_t := \epsilon_t\sqrt{Z_t^{(1)}}$ as well. For later purposes, however, representation 2.6 is advantageous.

where $\tilde{f}_{\xi^l} : I\!R^l \times I\!R^q \to I\!R$ is the nonnegative Baire function given by

$$\tilde{f}_{\xi^l}(u_1, \ldots, u_l \mid v_{-l-1}, \ldots, v_{-l-q}) := \prod_{j=1}^{l} \{h_{-j}^{-\frac{1}{2}} \times f_\epsilon(u_j h_{-j}^{-\frac{1}{2}})\},$$

$$h_{-j} := \alpha_0 + \sum_{i=1}^{q} \alpha_i \{1_{\{l-j \geq i\}} u_i^2 + 1_{\{l-j<i\}} v_{-j-i}^2\},$$

is a version of the conditional probability $P\{\xi^l \in B \mid \eta_{-l-1}, \ldots, \eta_{-l-q}\}$. Therefore, using Fubini's theorem and the Tonelli-Hobson argument, we can write for any $B \in \mathcal{B}_l^*$

$$\begin{aligned} P\{\xi^l \in B\} &= E\{P\{\xi^l \in B \mid \eta_{-l-1}, \ldots, \eta_{-l-q}\}\} \\ &= E\{\int_B \tilde{f}_{\xi^l}(u_1, \ldots, u_l \mid \eta_{-l-1}, \ldots, \eta_{-l-q}) du_1 \ldots du_l\} \\ &= \int_B E\{\tilde{f}_{\xi^l}(u_1, \ldots, u_l \mid \eta_{-l-1}, \ldots, \eta_{-l-q})\} du_1 \ldots du_l \\ &= \int_B f_{\xi^l}(u_1, \ldots, u_l) du_1 \ldots du_l, \end{aligned}$$

say. With l chosen k this shows the existence of the density . Because the function $\tilde{f}_{\xi^l}(\cdot)$ is continuous and bounded on $I\!R^l \times I\!R^q$,and l could be chosen arbitrarily in the preceding argument, we look at the representation

$$f(u_1, \ldots, u_k) = f_{\xi^k}(u_1, \ldots, u_k)$$

$$= \int_{I\!R^q} \tilde{f}_{\xi^l}(u_1, \ldots, u_k \mid u_{k+1}, \ldots, u_{k+q}) f_{\xi^q}(u_{k+1}, \ldots, u_{k+q}) du_{k+1} \ldots du_{k+q}$$

to obtain the continuity using a well known result from analysis (see e.g. APOSTOL (1977),10.38)

(ii) The second part of the Lemma follows similarly (see e.g. APOSTOL (1977), 10.39). ■

Chapter 3

Prediction and estimation in ARCH models

3.1 Prediction problems

The conditional variance $var(\eta_t \mid \mathcal{F}_{t-1})$, where $\mathcal{F}_t := \sigma(\eta_t, \eta_{t-1}, \ldots)$, of any stochastic process $\{\eta_t\}_{t\in Z}$ that fulfils 2.1 is well defined and given by

$$var(\eta_t \mid \mathcal{F}_{t-1}) = E(\eta_t^2 \mid \mathcal{F}_{t-1}) = h(\eta_{t-1}, \ldots, \eta_{t-q}) \quad a.\ s.\ , \tag{3.1}$$

since $\{\eta_t\}_{t\in Z}$ itself is a sequence of martingale differences.[1] The random variable 3.1 can be interpreted as best L_2 -approximation of η_t^2 given its predecessors if it is square-integrable or, to put it another way, if the fourth moment of η_t exists.To ensure the square-integrability of 3.1, however, rather restrictive additional assumptions which involve jointly the parameters and the distribution of the innovations are required[2]. Therefore, one has to be careful speaking of a "best" (in the mean squared error sense) prediction in connection with the conditional variance. Yet we shall accept 3.1 as a reasonable prediction, which is at least unbiased.
Instead of elaborating this point here, we turn to the main question we deal with in this paper: How can we identify the function $h : I\!R^q \to I\!R$, which is typically unknown in an economic framework, in order to be enabled to

[1] In the context of the parametric ARCH model studied here, one could equally well choose $\sigma(\eta_t^2, \eta_{t-1}^2, \ldots)$ as conditioning σ-field. However,in more general ARCH models to be analyzed in future, we possibly would give away information, since $\sigma(\eta_t^2, \eta_{t-1}^2, \ldots) \subseteq \sigma(\eta_t, \eta_{t-1}, \ldots)$.

[2] In the ARCH(1) model with normal innovations $3\alpha_1^2 < 1$ is required (see ENGLE (1982))

calculate predictions based on observed past values ? Most of the literature is devoted to (Pseudo-) Maximum-Likelihood estimators. We propose a modified kernel estimator, because there is some reason to believe that in this way we can detach ourselves from the little flexible parametric formulation of ARCH models.

3.2 A nonparametric filter

Given a data set $\{y_t, x_t\}_{t=1,\ldots,T}$, where $y_t := \eta_t^2$ and $x_t := (\eta_{t-1}, \ldots, \eta_{t-k})$, the standard kernel estimator of the regression function $m(x) := E\{y_t \mid x_t = x\}$ can be written as [3]

$$\begin{aligned} \hat{m}_T(x) &= \frac{\hat{g}_T(x)}{\hat{f}_T(x)}, \\ \hat{g}_T(x) &:= \frac{1}{T}\sum_{t=1}^{T} \gamma_T^{-k} y_t K\left(\frac{x - x_t}{\gamma_T}\right), \\ \hat{f}_T(x) &:= \frac{1}{T}\sum_{t=1}^{T} \gamma_T^{-k} K\left(\frac{x - x_t}{\gamma_T}\right), \end{aligned} \tag{3.2}$$

where the kernel $K(u)$ is a real function on $\mathbb{R}^k$ and $\{\gamma_T\}_{T\in\mathbb{N}}$ is a (possibly stochastic,i. e. data driven) decreasing sequence of positive numbers with limit zero. The class of admissible kernels is restricted by certain conditions which are listed in the introductory section of the next chapter.

We observe that whenever $k \geq q$,any restriction $g \mid_{\mathbb{R}^q \times \{u\}}, u \in \mathbb{R}^{k-q}$ of the Baire function $g(x)$ coincides with a Baire function representing the conditional variance.

When analyzing the asymptotic behaviour of the filter proposed, we are confronted with (at least) two particular problems: Firstly, we have to ensure that the underlying data generating process has an asymptotically vanishing memory. This can be achieved assuming the process to fulfil certain mixing conditions. The verification of mixing conditions is a hard task, however. Thus we follow an alternative approach introduced by Bierens(1983), that relies on the weaker concept of ν-stability.

Secondly, in view of the preceding section, we cannot assume, as is standard, that the dependent variable y_t is square integrable,or, to put it another way,

[3] The inflated representation elucidates that the denominator can be seen as an estimator for the k-dimensional marginal density of the data generating process. Its consistency is implicitly contained in the results of chapter 4.

that the data generating process 2.1 possesses fourth moments. Hence we have to use a truncation procedure. We introduce a **truncation function** $\phi : I\!R^k \to I\!R$

$$\phi(u) := \begin{cases} 1 & : \quad \|u\| \leq C \\ \frac{C^2}{\|u\|^2} & : \quad \|u\| > C \end{cases}$$

where C is a positive constant, and define the modified kernel estimator

$$\begin{aligned} \hat{m}_T(x) &= \frac{\hat{g}_T(x)}{\hat{f}_T(x)}, \qquad (3.3) \\ \hat{g}_T(x) &:= \frac{1}{T} \sum_{t=1}^{T} \gamma_T^{-k} \phi(x_t) y_t K\left(\frac{x - x_t}{\gamma_T}\right), \\ \hat{f}_T(x) &:= \frac{1}{T} \sum_{t=1}^{T} \gamma_T^{-k} K\left(\frac{x - x_t}{\gamma_T}\right). \end{aligned}$$

We notice that the truncation function together with assumption (A.1) guarantees that $\forall t \in Z$

$$\begin{aligned} \phi(x_t) y_t &= \phi(x_t) \epsilon_t^2 \{\alpha_0 + \sum_{i=1}^{q} \alpha_i \eta_{t-i}^2\} \qquad (3.4) \\ &\leq \epsilon_t^2 \{\alpha_0 + \sum_{i=1}^{q} \alpha_i C^2\} \\ &< \epsilon_t^2 \{\alpha_0 + C^2\} \\ &= \epsilon_t^2 C_0, \end{aligned}$$

say, and therefore implies $E\{\phi^2(x_t) y_t^2\} \leq C_0^2 \cdot E\epsilon_0^4 < +\infty$ under (A.3). An additional truncation of the regressors x_t is not required because this is done by the kernel. The estimator 3.3 is the usual kernel estimator applied on the data generating process with the dependent variable transformed by the truncation function.

For later purposes we notice that the truncation function is bounded by unity, continuous on $I\!R^k$, twice differentiable with continuous derivatives on $\{u \in I\!R^k : \|u\| < C\}$ and fulfils the following easily verified inequality:

$$| \phi(u) u_i^2 - \phi(v) v_i^2 | \leq | u_i^2 - v_i^2 | + \sum_{j=1}^{k} | u_j^2 - v_j^2 | \qquad (3.5)$$

for $i = 1, \ldots, k$.

Chapter 4

Consistency Results

4.1 Basic assumptions

For clarity, we start by listing all assumptions concerning the filter 3.3 needed somewhere in the current chapter:

(K. 1) *$K(u)$ is a nonnegative, bounded, continuous and symmetric real-valued function on $\mathbb{R}^k$ that integrates to one.*

(K. 2) $\|u\|^k K(u) \to 0 \quad for \quad \|u\| \to +\infty.$

(K. 3) *$K(u)$ has an absolutely integrable characteristic function $\beta(v) := \int_{\mathbb{R}^k} \exp(iv'u)K(u)du$, for which the additional integrability condition*

$$\int_{\mathbb{R}^k} \|v\| \mid \beta(v) \mid dv < +\infty$$

holds.

(K. 4)

$$\left\| \int_{\mathbb{R}^k} uu'K(u)du \right\| < +\infty.$$

(K. 5) $\|u\|^{k+2} K(u) \to 0 \quad for \quad \|u\| \to +\infty.$

To connect estimator 3.3 to the model in regard we formulate the

Working Hypothesis 1
The data set $\{(y_t, x_t)\}_{t=1,\dots,T} = \{\eta_t^2, \eta_{t-1}, \dots, \eta_{t-k}\}_{t=1,\dots,T}$ *stems from a strictly stationary solution of 2.1 with* $q \le k$ *under assumptions (A.1) to (A.3).*

Observe that in view of Lemma 2 this hypothesis is always meaningful.

4.2 The forgetfulness of ARCH(q) processes

This section has preliminary character. As indicated in section 3.2 we have to ensure that the data generating process has a vanishing memory in a sense that has to be made precise. Our definition of stochastic forgetfulness, implicitly given in the following Lemma, corresponds to the concept of exponential ν-stability in L_1.

Lemma 4
(i) Let $\mathcal{F}_{t,\tau}$ *denote the* σ*-field* $\sigma(A_t, A_{t-1}, \dots, A_{t-\tau+1})$. *Suppose that (A.1) holds. Then a constant* $m > 0$ *exists such that* $\forall t \in Z$

$$\nu(\tau) := E\|E\{(\phi(x_t)y_t, x_t) \mid \mathcal{F}_{t,\tau}\} - (\phi(x_t)y_t, x_t)\| = o(e^{-m\tau}).$$

Proof:
Step 1
Define the $\mathcal{F}_{t,\tau}$-measurable random vector $Z_{t,\tau} := \sum_{s=1}^{\tau-1} M_{t,s}D + D$ and note that

$$Z_t - Z_{t,\tau} = M_{t,\tau}Z_{t-\tau}.$$

Then we can deduce from Lemma 1 that $\forall t \in Z$

$$\begin{aligned} & E\|E\{Z_t \mid \mathcal{F}_{t,\tau}\} - Z_t\| \\ \le\ & E\|Z_t - Z_{t,\tau}\| + E\|E\{Z_t - Z_{t,\tau} \mid \mathcal{F}_{t,\tau}\}\| = o(e^{-m_2\tau}). \end{aligned} \tag{4.1}$$

Step 2
Define $\eta_{t-i,\tau} := E\{\eta_{t-i} \mid \mathcal{F}_{t,\tau}\}, i = 1, \dots, k$. It is easy to see that $\forall \epsilon > 0, \forall t \in Z, \forall \tau \in I\!N, i = 1, \dots, k$,

$$\begin{aligned} & E \mid \sqrt{\eta_{t-i,\tau}^2} - \sqrt{\eta_{t-i}^2} \mid \\ = \quad & E\{1_{\eta_{t-i,\tau}^2 \ge \epsilon \cup \eta_{t-i}^2 \ge \epsilon} \mid \sqrt{\eta_{t-i,\tau}^2} - \sqrt{\eta_{t-i}^2} \mid\} \\ + \quad & E\{1_{\eta_{t-i,\tau}^2 \le \epsilon \cap \eta_{t-i}^2 \le \epsilon} \mid \sqrt{\eta_{t-i,\tau}^2} - \sqrt{\eta_{t-i}^2} \mid\} \\ \le \quad & \epsilon^{-\frac{1}{2}} E \mid \eta_{t-i,\tau}^2 - \eta_{t-i}^2 \mid + 2\epsilon^{\frac{1}{2}} \end{aligned}$$

Now let $\epsilon = \epsilon(\tau) = e^{-\tilde{m}\tau}$ for some $\tilde{m} < m_2$. Then for any $m < \frac{1}{2}\tilde{m}$ we obtain from Step 1

$$E \mid \sqrt{\eta^2_{t-i,\tau}} - \sqrt{\eta^2_{t-i}} \mid = o(e^{-m\tau}). \tag{4.2}$$

Step 3
We note that $\eta_{t-i,\tau}, i = 1,\dots,k$, and ϵ_t are $\mathcal{F}_{t,\tau}$ measurable and recall the representation of η_t in the proof of Lemma 2 (ii). Then we can deduce from step 2 that for all $i = 1,\dots,k$,

$$\begin{aligned}
E \mid \eta_{t-i,\tau} - \eta_{t-i} \mid \quad &\leq \quad E \mid 1_{\epsilon_t \geq 0}(E\{\sqrt{\eta^2_{t-i}} \mid \mathcal{F}_{t,\tau}\} - \sqrt{\eta^2_{t-i}}) \mid \\
&\quad + \; E \mid 1_{\epsilon_t < 0}(E\{\sqrt{\eta^2_{t-i}} \mid \mathcal{F}_{t,\tau}\} - \sqrt{\eta^2_{t-i}}) \mid \\
&= \quad E \mid E\{\sqrt{\eta^2_{t-i}} \mid \mathcal{F}_{t,\tau}\} - \sqrt{\eta^2_{t-i}}) \mid \\
&\leq \quad E \mid E\{\sqrt{\eta^2_{t-i}} \mid \mathcal{F}_{t,\tau}\} - \sqrt{\eta^2_{t-i,\tau}} \mid \\
&\quad + \; E \mid \sqrt{\eta^2_{t-i,\tau}} - \sqrt{\eta^2_{t-i}}) \mid \\
&= \quad E \mid E\{\sqrt{\eta^2_{t-i}} - \sqrt{\eta^2_{t-i,\tau}} \mid \mathcal{F}_{t,\tau}\} \mid \\
&\quad + \; E \mid \sqrt{\eta^2_{t-i,\tau}} - \sqrt{\eta^2_{t-i}}) \mid \\
&\leq \quad 2E \mid \sqrt{\eta^2_{t-i,\tau}} - \sqrt{\eta^2_{t-i}}) \mid \\
&= \quad o(e^{-m\tau}).
\end{aligned}$$

Step 4
Define $x_{t,\tau} := E\{x_t \mid \mathcal{F}_{t,\tau}\}$ and $y_{t,\tau} := E\{\eta^2_t \mid \mathcal{F}_{t,\tau}\}$. With arguments identical to those in step 3 we can show that $\forall t \in Z$

$$E \mid E\{\phi(x_t)y_t \mid \mathcal{F}_{t,\tau}\} - \phi(x_t)y_t \mid \leq 2E \mid \phi(x_{t,\tau})y_{t,\tau} - \phi(x_t)y_t \mid . \tag{4.3}$$

Finally we can exploit 3.5 and the fact that ϵ_t is independent of $x_{t,\tau}$ as well as of x_t to obtain from step 1

$$\begin{aligned}
& E \mid \phi(x_{t,\tau})y_{t,\tau} - \phi(x_t)y_t \mid \\
= \; & E \mid \epsilon^2_t \sum_{i=1}^{q} \alpha_i[\phi(x_{t,\tau})\eta^2_{t-i,\tau} - \phi(x_t)\eta^2_{t-i}] \mid \\
\leq \; & (q+1)\sum_{i=1}^{k} E \mid \eta^2_{t-i,\tau} - \eta^2_{t-i} \mid \\
\leq \; & (q+1)kE\|E\{Z_t \mid \mathcal{F}_{t,\tau}\} - Z_t)\| \\
= \; & o(e^{-m_2\tau}).
\end{aligned}$$

The proof is complete if we choose the constant m as in step 2. ∎

4.3 Uniform consistency

The main result of the present analysis relies on three preparing Lemmata that describe the limit behaviour of bias and variance of the estimator 3.3. The underlying decomposition concept goes back to BIERENS(1983).

Lemma 5 (Asymptotic Unbiasedness)
Define $M(C,\epsilon) := \{u \in \mathbb{R}^k : \|u\| \leq C - \epsilon\}$ and let denote $g(x) := m(x)f(x)$. Suppose that Working Hypothesis 1 and assumptions (K.1) and (K.2) hold and that $\{\gamma_T\}_{T\in \mathbb{N}}$ is a sequence of positive numbers satisfying

$$\lim_{T\to\infty} \gamma_T = 0.$$

Then we have
(i)

$$\lim_{T\to\infty} \sup_{x\in M(C,0)} \mid E\{\hat{g_T}(x)\} - g(x) \mid = 0. \tag{4.4}$$

(ii) If in addition assumptions (A.4),(K.4) and (K.5) hold, we have for any $\epsilon \in (0, C)$

$$\sup_{x\in M(C,\epsilon)} \mid E\{\hat{g_T}(x)\} - g(x) \mid = O(\gamma_T^2). \tag{4.5}$$

Proof:
Although most of the arguments used here are standard, we present the proof in full length, since there are some small technical difficulties.
For notational convenience, we suppress the time index of the bandwidth and simply write γ.
(i) Since the data set $(y_t, x_t), t = 1, \ldots, T$ stems from a strictly stationary solution of 2.1 with density $f(u)$, we have for any $\gamma > 0$, and $\forall x \in \mathbb{R}^k$

$$\begin{aligned}
& \mid E\{\hat{g_T}(x)\} - g(x) \mid \\
= & \mid E\{E\{\hat{g}_1(x) \mid x_1\}\} - g(x) \mid \\
= & \left| E\left\{\gamma^{-k}\phi(x_1)E\{y_1 \mid x_1\}K\left(\frac{x-x_1}{\gamma}\right)\right\} - g(x)\right| \\
= & \left| \int_{\mathbb{R}^k} \gamma^{-k}\phi(z)g(z)K\left(\frac{x-z}{\gamma}\right) dz - g(x)\right|.
\end{aligned}$$

The special form of $\phi(u)$ implies

$$
\begin{aligned}
& \sup_{x \in M(C,0)} \left| \int_{\mathbb{R}^k} \gamma^{-k} \phi(z) g(z) K\left(\frac{x-z}{\gamma}\right) dz - g(x) \right| \\
= & \sup_{x \in M(C,0)} \left| \int_{\mathbb{R}^k} \gamma^{-k} \phi(z) g(z) K\left(\frac{x-z}{\gamma}\right) dz - \phi(x) g(x) \right| \\
= & \sup_{x \in M(C,0)} \left| \int_{\mathbb{R}^k} \gamma^{-k} \tilde{g}(z) K\left(\frac{x-z}{\gamma}\right) dz - \tilde{g}(x) \right|,
\end{aligned}
$$

where $\tilde{g}(u)$ is the continuous function $\phi(u)g(u)$. We can apply the classical splitting technique to obtain

$$
\begin{aligned}
& \sup_{x \in M(C,0)} \left| \int_{\mathbb{R}^k} \gamma^{-k} \tilde{g}(z) K\left(\frac{x-z}{\gamma}\right) dz - \tilde{g}(x) \right| \\
\leq & \sup_{x \in M(C,0)} \sup_{||\gamma u|| < \delta} |\ \tilde{g}(x+\gamma u) - \tilde{g}(x)\ | \\
+ & \sup_{x \in M(C,0)} \left| \int_{||\gamma u|| \geq \delta} [\tilde{g}(x+\gamma u) - \tilde{g}(x)] K(u) du \right|.
\end{aligned}
$$

Since $\tilde{g}(x)$ is continuous on $\mathbb{R}^k$, it is uniformly continuous on compact subsets. Therefore the first expression can be made arbitrarily close to zero for δ sufficiently small. Having chosen δ this way, the remaining integral can be bounded by

$$
\sup_{x \in M(C,0)} \int_{||\gamma u|| \geq \delta} K(u) \tilde{g}(x+\gamma u) du + \sup_{x \in M(C,0)} \tilde{g}(x) \int_{||\gamma u|| \geq \delta} K(u) du.
$$

The second term converges to zero for $\gamma \downarrow 0$, since $\tilde{g}(x)$ attains its maximum on $M(C,0)$ and the integral can be made arbitrarily small choosing γ sufficiently small for δ given.

The first integral is bounded by

$$
\sup_{x \in M(C,0)} \int_{||\gamma u|| \geq \delta} K(u) \tilde{g}(x+\gamma u) du
$$

$$
\begin{aligned}
&\leq \sup_{x \in M(C,0)} \sup_{\|\gamma u\| \geq \delta} (\|u\|^k K(u)) \int\limits_{\|\gamma u\| \geq \delta} (\|u\|)^{-k} \tilde{g}(x + \gamma u) du \\
&\leq \sup_{\|\gamma u\| \geq \delta} (\|u\|^k K(u)) \left\{ \delta^{-k} \gamma^k \int\limits_{I\!R^k} \gamma^{-k} \tilde{g}(z) dz \right\} \\
&= o(1) \delta^{-k} E y_1 \\
&= o(1).
\end{aligned}
$$

for $\gamma \downarrow 0$ under (K.2).

(ii)Using the same arguments as in the proof of part (i) (with the difference that ϵ is given here), we obtain

$$
\begin{aligned}
&\sup_{x \in M(C,\epsilon)} |E\{\hat{g}_T(x)\} - g(x)| \\
= &\sup_{x \in M(C,\epsilon)} \left| \int [\tilde{g}(x + \gamma u) - \tilde{g}(x)] K(u) du \right| \\
\leq &\sup_{x \in M(C,\epsilon)} \left| \int\limits_{\|\gamma u\| < \epsilon} [\tilde{g}(x + \gamma u) - \tilde{g}(x)] K(u) du \right| \\
+ &\sup_{x \in M(C,\epsilon)} \left| \int\limits_{\|\gamma u\| \geq \epsilon} \tilde{g}(x + \gamma u) K(u) du \right| \\
+ &\sup_{x \in M(C,\epsilon)} \left| \int\limits_{\|\gamma u\| \geq \epsilon} \tilde{g}(x) K(u) du \right| \\
= &I_1 + I_2 + I_3.
\end{aligned}
$$

(1) We notice that the vectors $x, x + \gamma u$ are in an open subset of $I\!R^k$ the function $\tilde{g}(z)$ is partially continuously differentiable on, as long as $x \in M(C, \epsilon)$ and $\|\gamma u\| < \epsilon$. Therefore we can make use of the arguments in Lemma 2 of Bierens(1989), together with (K.4) and the symmetry of the kernel, to see that $I_1 = O(\gamma^2)$.

(2) We have

$$
\begin{aligned}
&\frac{1}{\gamma^2} I_2 \\
\leq &\sup_{\|u\| \geq \frac{\epsilon}{\gamma}} \frac{1}{\gamma^2} \|u\|^k K(u) \left\{ \epsilon^{-k} \int \tilde{g}(z) dz \right\}
\end{aligned}
$$

$$
\begin{aligned}
&\leq \sup_{\|u\|\geq\frac{\epsilon}{\gamma}} \epsilon^{-2}\|u\|^{k+2}K(u)O(1) \\
&= o(1).
\end{aligned}
$$

under (K.5).

(3) Similarly we have $I_3 = O(\gamma^2)$ under (K.5). ■

Lemma 6

Suppose that the Working Hypothesis holds and that $\{\tau(T)\}_{T\in\mathbb{N}}$ is a sequence of natural numbers. Then, for any sequence $\{\gamma_T\}_{T\in\mathbb{N}}$ of positive numbers,

$$
\gamma_T^k E \sup_{x\in\mathbb{R}^k} \left| E\{\hat{m}_T(x) \mid \mathcal{F}_{t,\tau(T)}\} - E\{\hat{m}_T(x)\}\right| = O\left(\sqrt{\frac{\tau(T)}{T}}\right). \tag{4.6}
$$

Proof:

(cf. Bierens(1983),Lemma 1)

The measurability of the supremum on the left-hand side of 4.6 follows from a continuity argument (see Lemma 1 of Jennrich(1969)). The inversion formula for characteristic functions and Proposition 1 in the Appendix yield that $\forall\gamma > 0$ we have for the left-hand expression in 4.6

$$
\begin{aligned}
& E \sup_{x\in\mathbb{R}^k} \left| E\left\{\frac{1}{T}\sum_{t=1}^{T}\phi(x_t)y_t K\left(\frac{x-x_t}{\gamma}\right) \mid \mathcal{F}_{t,\tau(T)}\right\}\right. \\
& \qquad \left. - E\left\{\frac{1}{T}\sum_{t=1}^{T}\phi(x_t)y_t K\left(\frac{x-x_t}{\gamma}\right)\right\}\right| \\
\leq\ & \gamma^k \left(\frac{1}{2\pi}\right)^k \int_{\mathbb{R}^k} w_T(v)\left|\beta(\gamma v)\right| dv \\
\leq\ & \sup_{v\in\mathbb{R}^k} w_T(v) \left(\frac{1}{2\pi}\right)^k \int_{\mathbb{R}^k} \left|\beta(v)\right| dv,
\end{aligned}
$$

where

$$
w_T(v) := E\left|\frac{1}{T}\sum_{t=1}^{T}\left[E\left\{\phi(x_t)y_t\exp(iv'x_t) \mid \mathcal{F}_{t,\tau(T)}\right\} - E\left\{\phi(x_t)y_t\exp(iv'x_t)\right\}\right]\right|.
$$

Since for T and v given the sequences $\left\{E\left[\phi(x_t)y_t\sin(v'x_t) \mid \mathcal{F}_{t,\tau(T)}\right]\right\}_{t\in I\!N}$ and $\left\{E\left[\phi(x_t)y_t\cos(v'x_t) \mid \mathcal{F}_{t,\tau(T)}\right]\right\}_{t\in I\!N}$ are φ-mixing stochastic processes with

$$\varphi(l) = \begin{cases} 0 & : \quad l \geq \tau(T) \\ 1 & : \quad l < \tau(T), \end{cases}$$

it can be shown (Bierens(1983),Lemma 1) that under (A.1)-(A.3),

$$\sup_{v\in I\!R^k} w_T(v) = O\left(\sqrt{\frac{\tau(T)}{T}}\right).$$

Together with the integrability of the characteristic function, the proof is complete.■

The next result characterizes the rate of convergence of the approximation error that occurs when the actual random variables are replaced by conditional expectations.

Lemma 7
Suppose that the Working Hypothesis and assumptions (K.1)and (K.3) hold and that $\{\tau(T)\}_{T\in I\!N}$ is a sequence of natural numbers. Then, for any sequence $\{\gamma_T\}_{T\in I\!N}, \gamma_T > 0$, with $\lim_{T\to\infty}\gamma_T = 0$ and some constant $m > 0$,

$$E \sup_{x\in I\!R^k}\left|E\{\hat{m}_T(x) \mid \mathcal{F}_{t,\tau(T)}\} - \hat{m}_T(x)\right| = o(e^{-m\tau(T)}\gamma_T^{-(k+1)}). \qquad (4.7)$$

Proof:
We can make use of Proposition 1 in the Appendix and the inversion formula for characteristic functions to obtain

$$\begin{aligned} & E \sup_{x \in \mathbb{R}^k} \left| \gamma_T^{-k} \left[E \left\{ \phi(x_t) y_t K \left(\frac{x - x_t}{\gamma_T} \right) \mid \mathcal{F}_{t,\tau(T)} \right\} \right. \right. \\ & \qquad \left. \left. - \phi(x_t) y_t K \left(\frac{x - x_t}{\gamma_T} \right) \right] \right| \\ \leq & \left(\frac{1}{2\pi} \right)^k \int_{\mathbb{R}^k} \tilde{w}_T(v) \left| \beta(\gamma_T v) \right| dv, \end{aligned}$$

where

$$\tilde{w}_T(v) := E \left| E \left\{ \phi(x_1) y_1 \exp(iv' x_1) \mid \mathcal{F}_{1,\tau(T)} \right\} - \phi(x_1) y_1 \exp(iv' x_1) \right|.$$

We recall the notation and technique used in the proof of Lemma 4 and 3.4. Moreover we note that $\mid e^{iu} - 1 \mid \leq \mid u \mid$ for all $u \in \mathbb{R}$. We apply Lemma 4 to obtain

$$\begin{aligned} \tilde{w}_T(v) & \leq 2E \left| E\{\phi(x_1) y_1 \mid \mathcal{F}_{1,\tau(T)}\} \exp(iv' x_{1,\tau(T)}) - \phi(x_1) y_1 \exp(iv' x_1) \right| \\ & \leq 2E \left| (\exp(iv' x_{1,\tau(T)}))(E\{\phi(x_1) y_1 \mid \mathcal{F}_{1,\tau(T)}\} - \phi(x_1) y_1) \right| \\ & + 2E \left| \phi(x_1) y_1 (\exp(iv' x_{1,\tau(T)}) - \exp(iv' x_1)) \right| \\ & \leq 2E \left| E\{\phi(x_1) y_1 \mid \mathcal{F}_{1,\tau(T)}\} - \phi(x_1) y_1 \right| \\ & + 2C_0 E \| x_{1,\tau(T)} - x_1 \| \| v \| \\ & = o(e^{-m\tau(T)})(1 + \|v\|). \end{aligned}$$

This implies

$$\begin{aligned} & \int_{\mathbb{R}^k} \tilde{w}_T(v) \left| \beta(\gamma_T v) \right| dv \\ = & \; o(e^{-m\tau(T)}) \left[\gamma_T^{-k} \int_{\mathbb{R}^k} \mid \beta(v) \mid dv + \gamma_T^{-(k+1)} \int_{\mathbb{R}^k} \|v\| \mid \beta(v) \mid dv \right] \\ = & \; o(e^{-m\tau(T)} \gamma_T^{-(k+1)}), \end{aligned}$$

which completes the proof. ■

The Lemmata 5,6 and 7 enable us to formulate the following consistency result for the estimator 3.3:

Theorem 1 (Uniform Consistency)
(i) Suppose that the Working Hypothesis and (K.1),(K.2) and (K.3) hold. Then for any sequence $\{\gamma_T\}_{T\in\mathbb{N}}$ *satisfying*

$$\gamma_T = T^{-\mu}, \mu \in \left(0, \frac{1}{2k}\right)$$

and any pair

$$\left\{\delta \in \left(0, \sup_{x\in\mathbb{R}^k} f(x)\right], C > 0\right\},$$

we have

$$p\lim_{T\to\infty} \sup_{x\in\{f(u)\geq\delta\}\cap M(C,0)} \mid \hat{m}_T(x) - m(x) \mid = 0. \tag{4.8}$$

(ii) Suppose that in addition (A.4),(K.4) and (K.5) hold. Then,for any triple

$$\left\{\delta \in \left(0, \sup_{x\in\mathbb{R}^k} f(x)\right], \epsilon > 0, C > 0\right\}$$

with $C > \epsilon$ *we have*

$$p\lim \zeta_T \sup_{x\in\{f(u)\geq\delta\}\cap M(C,\epsilon)} \mid \hat{m}_T(x) - m(x) \mid = 0 \tag{4.9}$$

where $\zeta_T = o(min(\gamma_T^{-2}, \xi_T)), \xi_T = o(\gamma_T^k\sqrt{T})$.

Proof:
(i) The measurability of the supremum in 4.8 follows again from a continuity argument (Lemma 1 of Jennrich(1969)). Choosing e. g. $\tau(T) = \left[T^{\frac{1}{2}-\mu k}\right]$, we can link Lemmata 6 and 7 to obtain

$$\lim_{T\to\infty} E \sup_{x\in\mathbb{R}^k} |\hat{g}_T(x) - E\{\hat{g}_T(x)\}| = 0.$$

Combining this last implication with Lemma 5 yields

$$\lim_{T\to\infty} E \sup_{x\in M(C,0)} |\hat{g}_T(x) - g(x)| = 0$$

and, since all results remain valid if we replace y_t by $1\ \forall t \in Z$,

$$\lim_{T\to\infty} E \sup_{x\in M(C,0)} \left|\hat{f}_T(x) - f(x)\right| = 0.$$

Since we can write

$$| \hat{m}_T(x) - m(x) | = \frac{1}{f(x)} \left| \hat{m}_T(x)(f(x) - \hat{f}_T(x)) + \hat{f}_T(x)(\hat{g}_T(x) - g(x)) \right|,$$

the result follows.
(ii) follows from similar arguments. ■

Appendix A

Proposition 1 *Suppose that* x, y *are* k_1, k_2*-dimensional random vectors on a probability space* $(\Omega, \mathcal{A}, P)$ *and that* $f : \mathbb{R}^{k_2} \times \mathbb{R}^k \to \mathbb{R}$ *and* $g : \mathbb{R}^k \to \mathbb{R}$ *are Baire functions such that*

$$\sup_{y \in \mathbb{R}^{k_2}} \mid f(y,t) \mid \leq g(t), \quad \int_{\mathbb{R}^k} g(t)dt < +\infty.$$

Then

$$E\left\{ \int_{\mathbb{R}^k} f(y,t)dt \mid x \right\} = \int_{\mathbb{R}^k} E\{f(y,t) \mid x\}dt \quad F_x - a.s.,$$

where F_x *denotes the distribution of* x.

Proof:
We observe that both sides of the equation are finite because of the boundedness of $f(y,t)$.Hence, by Fubini's Theorem,$\forall C \in \sigma(x)$,

$$\begin{aligned} & \int_C [\int_{\mathbb{R}^k} E\{f(y,t) \mid x\}dt]dP \\ = \; & \int_{\mathbb{R}^k} [\int_C E\{f(y,t) \mid x\}dP]dt \\ = \; & \int_{\mathbb{R}^k} [\int_C f(y,t)dP]dt \\ = \; & \int_C [\int_{\mathbb{R}^k} f(y,t)dt]dP. \end{aligned}$$

This shows that $\int_{\mathbb{R}^k} E\{f(y,t) \mid x\}dt$ is a version of the conditional expectation $E\left\{\int_{\mathbb{R}^k} f(y,t)dt \mid x\right\}$. ■

Bibliography

[1] Apostol,T.M.(1974)
Mathematical Analysis
Addison-Wesley Publishing Company.

[2] Bierens,H.J.(1983)
Uniform Consistency of Kernel Estimators of a Regression Function Under Generalized Conditions
Journal of the American Statistical Association,**78**,699-707.

[3] Bougerol,P. and Picard,N.(1992)
Stationarity of GARCH processes and of some nonnegative time series
Journal of Econometrics,**52**,115-127.

[4] Engle,R.F.(1982)
Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation
Econometrica,**50**,987-1008.

[5] Haerdle,W.(1989)
Applied nonparametric regression
Cambridge University Press.

[6] Jennrich,R.I.(1969)
Asymptotic Properties of Nonlinear Least Squares Estimators
The Annals of Mathematical Statistics,**40**,633-643.

[7] Kesten,H. and Spitzer,F.(1984)
Convergence in Distribution of Products of Random Matrices
Zeitschrift f. Wahrscheinlichkeitstheorie und verwandte Gebiete,**67**,363-386.

[8] Singh,R.S. and Ullah,A.(1985)
Nonparametric time series estimation of joint DGP,conditional DGP and vector autoregression
Econometric Theory, **1**.