Local Likelihood for non-parametric ARCH(1) models

Paolo Vanini           Luigi Vignola

First version: May 2002  
Current version: May 2002

This research has been carried out within the NCCR FINRISK project on “Interest Rate and Volatility Risk”.

Die Nationalen Forschungsschwerpunkte (NFS) sind ein Förderinstrument des Schweizerischen Nationalfonds.  
Les Pillars de recherche nationaux (PRN) sont un instrument d’encouragement du Fonds national suisse.  
The National Centers of Competence in Research (NCCR) are a research instrument of the Swiss National Science Foundation.
Local Likelihood for non-parametric ARCH(1) models

by

Francesco Audrino
Institute of Finance
University of Southern Switzerland
Via Buffi 13
6900 Lugano
Switzerland
Local Likelihood for non-parametric ARCH(1) models

Francesco Audrino
University of Southern Switzerland

May 2002

Abstract

We propose a local likelihood estimation for the log-transformed ARCH(1) model in the financial field. Our nonparametric estimator is constructed within the likelihood framework for non-Gaussian observations: it is different from standard kernel regression smoothing, where the innovations are assumed to be normally distributed. We derive consistency and asymptotic normality for our estimators and conclude from simulation and real data analysis that the local likelihood estimator has better predictive potential than classical local regression.

Keywords: Return time series; Volatility; ARCH model; Local likelihood; Kernel regression smoothing.

1 Introduction

As a starting point, we consider the following non-parametric ARCH(1) model

\[ X_t = \sigma_t Z_t \quad (t \geq 2) \]
\[ \sigma_t^2 = f(X_{t-1}), \quad f : \mathbb{R} \rightarrow \mathbb{R}^+ \],

(1.1)

where the innovations \( Z_t \) are identically independent distributed, with distribution function \( G \), zero mean and variance one and independent from \( \{X_s; s < t\} \). We assume the time series \( \{X_t\}_{t\geq1} \) to be stationary (which is approximately true for return time series in a time-window of about 2 years; for more details see Mikosch and Starica, 1999) and that certain mixing conditions hold (see Section 3 for more details).

For our purposes, it is useful to transform the model (1.1) into regression by taking the logarithm

\[ \log(X_t^2) = \log(\sigma_t^2 Z_t^2) = \log(\sigma_t^2) + \log(Z_t^2) = \beta + \log(\sigma_t^2) + \log(Z_t^2) - \beta, \]

where \( \beta = \mathbb{E} \left[ \log(Z_t^2) \right] \). To simplify, from now on we will use the following notation:

\[ Y_t := \log(X_t^2), \]
\[ U_t := \log(Z_t^2) - \beta \] and
\[ g(X_{t-1}) := \beta + \log(f(X_{t-1})) = \beta + \log(\sigma_t^2). \]

Then, the transformed non parametric ARCH(1) model can be rewritten as

\[ Y_t = g(X_{t-1}) + U_t, \] (1.2)
where \( E[U_t] = 0 \), \( U_t \) are i.i.d. and independent from \( \{Y_s; s < t\} \).

The classical method used to find a non-parametric estimator for the function \( g(\cdot) \) in (1.2) is kernel regression (least squares) smoothing, as for example in Härdle and Vieu (1992), Härdle (1994) or Yang et al. (1999). In this paper, we propose an alternative, non-parametric estimator for \( g(\cdot) \) using local likelihood estimation in a similar way as for example in Loader (1999). For this purpose, the distribution of the innovations \( U_t \) in (1.2) must be derived from the one of the variables \( Z_t \) in (1.1). This is done as follows.

If the distribution of the innovations \( Z_t \) in (1.1) is known, the distribution of \( U_t \) can be easily calculated using a transformation formula. In the following examples, we calculate explicitly the density function of \( U_t \) in the case of standard normally and scaled \( t_\nu \)-distributed innovations \( Z_t \) in (1.1).

**Example 1.1.** Let the innovations \( Z_t \) in (1.1) be standard normally distributed, i.e. \( G = \mathcal{N}(0, 1) \). Then we have that \( Z_t^2 \sim \chi_1^2 \) with density function given by

\[
    f_{Z_t^2}(z) = \begin{cases} 
        0 & \text{if } z < 0 \\
        \frac{1}{\sqrt{2\pi}} z^{-\frac{1}{2}} e^{-\frac{z}{2}} & \text{otherwise.}
    \end{cases}
\]

Now, using a simple transformation we get

\[
    f_{U_t}(u) = f_{Z_t^2}(e^{\beta+u}) \cdot e^{\beta+u} = \frac{1}{\sqrt{2\pi}} \left( e^{\beta+u} \right)^{-\frac{1}{2}} \exp \left( - \frac{e^{\beta+u}}{2} \right) e^{\beta+u} = \\
    = c \cdot \exp \left( \frac{u}{2} - \frac{\beta e^u}{2} \right), \quad u \in \mathbb{R},
\]

where the constants \( c \) and \( \tilde{\beta} \) equal

\[
    c = \frac{1}{\sqrt{2\pi}} \tilde{\beta} \quad \text{and} \quad \tilde{\beta} = \exp(\beta) = \exp(\mathbb{E}[\log(\chi_1^2)]) > 0.
\]

**Example 1.2.** Let the innovations \( Z_t \) in (1.1) be scaled \( t_\nu \) distributed, with degrees of freedom \( \nu > 4 \), i.e. \( G = \mathcal{N}(0, 1) \). Note that the scaling factor \( \sqrt{\frac{\nu - 2}{\nu}} \) is used to satisfy the condition \( \text{Var}(Z_t) = 1 \). Now, we have that \( Z_t^2 \) has the following density

\[
    f_{Z_t^2}(z) = \begin{cases} 
        0 & \text{if } z < 0 \\
        \frac{\Gamma\left(\frac{\nu + 1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \sqrt{\pi}} \sqrt{\frac{\nu}{\nu - 2}} \frac{1}{2} \left( 1 + \frac{z}{\nu - 2} \right)^{-\frac{\nu + 1}{2}} & \text{otherwise}
    \end{cases}
\]

and analogously to Example 1.1, we get

\[
    f_{U_t}(u) = c \cdot e^{\frac{u}{2}} \left( 1 + \frac{\beta e^u}{\nu - 2} \right)^{-\frac{\nu + 1}{2}}, \quad u \in \mathbb{R},
\]

where the constant \( c = \frac{\Gamma\left(\frac{\nu + 1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \sqrt{\pi (\nu - 2)}} \tilde{\beta} \quad \text{and} \quad \tilde{\beta} \quad \text{defined as in Example 1.1.}
\]
2 The local constant log-likelihood estimation

In this section, we want to illustrate the non-parametric strategy we use to construct the estimate of the function \( g(\cdot) \) in the model (1.2). The local constant log-likelihood of the model (1.2) for a fitting point \( x \) looks like

\[
L_x(X^T, g_x) = \sum_{t=2}^n \left( w_t(x) \rho(Y_t, g_x) \right) = \\
= \sum_{t=2}^n \left( W\left(\frac{x - X_{t-1}}{h_n}\right) \left( - \log(f_{U_t}(Y_t - g_x)) \right) \right),
\]

where the kernel \( W(\cdot) : \mathbb{R} \rightarrow \mathbb{R} \) is a symmetric (i.e. \( W(u) = W(-u), \ u \in \mathbb{R} \)), nonnegative bounded function satisfying

\[
u W(u) \rightarrow 0 \text{ as } |u| \rightarrow +\infty, \ \int_{\mathbb{R}} W(u) \, du = 1.
\]

It typically assigns largest weights to observations close to \( x \) (for example Gaussian or tricube kernel) and \( h_n \) is a global bandwidth. It is also possible to choose the bandwidth \( h_n = h_n(x) \) dependent on the fitting point \( x \), but this does not belong to the scope of this work. The sequence \( \{h_n\}_{n \in \mathbb{N}} \) is such that

\[
h_n \rightarrow_{n \rightarrow \infty} 0, \ \ nh_n \rightarrow_{n \rightarrow \infty} \infty, \ \ h_n > 0, \ \forall n \geq 2 \in \mathbb{N}.
\]

For more details about the more general local polynomial log-likelihood estimation, we refer the reader to the book of Loader (1999). Minimizing (2.1) with respect to the parameter \( g_x \) leads to the local constant log-likelihood estimate \( \hat{g}_{x;n} \)

\[
\hat{g}_{x;n} = \arg\min_{g_x} \sum_{t=2}^n \left( w_t(x) \left( - \log(f_{U_t}(Y_t - g_x)) \right) \right)
\]

for every fitting point \( x \).

2.1 Local likelihood estimator with Gaussian or scaled \( t \)-distributed innovations

In this Section, we will focus on the important case of standard normally distributed innovations \( Z_t \) in (1.1), i.e. \( G = \mathcal{N}(0, 1) \), as we do in the Examples 1.1 and 2.4. The importance of this particular case is also given from the fact that the local likelihood estimator can be written in closed form. This is not the case if the innovations are scaled \( t_\nu \) distributed, for example. The general problem of the existence and uniqueness of the local polynomial log-likelihood estimate is considered for example in Loader (1999), Theorem 4.1.

**Theorem 2.1.** (Existence and uniqueness of \( \hat{g}_{x;n} \) under standard normal assumption)

Assume that the innovations \( Z_t \) in (1.1) are standard normally distributed. Then the local constant log-likelihood estimator \( \hat{g}_{x;n}^N \) given by (2.2) for the function \( g(x) \) of the transformed \( ARCH(1) \) model (1.2) exists, is unique and equals

\[
\hat{g}_{x;n}^N = \beta^N + \log \left( \frac{\sum_{t=2}^n W\left(\frac{x - X_{t-1}}{h_n}\right) X_t^2}{\sum_{s=2}^n W\left(\frac{x - X_{s-1}}{h_n}\right)} \right)
\]

for every fitting point \( x \), where \( \beta^N = \mathbb{E}[\log(Z_t^2)] \approx -1.27. \)
Proof. Differentiating, the estimator \( \hat{g}_{x:n} \) (2.2) must be a solution of the local log-likelihood equation

\[
\sum_{t=2}^{n} \left( w_t(x) \rho'(Y_t, g_x) \right) = 0,
\]

(2.4)

where \( \rho'(y, g_x) = \frac{\partial}{\partial g} \rho(y, g_x) \).

Using the results of Example 2.4 of Section 2.2 for the local constant log-likelihood equations (2.4) we get

\[
\sum_{t=2}^{n} \left( w_t(x) \rho'(Y_t, g_x) \right) = \frac{1}{2} \sum_{t=2}^{n} \left( w_t(x) \left( 1 - \beta e^{(Y_t-g_x)} \right) \right) = 0
\]

\[
\Leftrightarrow \quad \frac{1}{2} \sum_{t=2}^{n} w_t(x) = \frac{\beta e^{-g_x}}{2} \sum_{t=2}^{n} \left( w_t(x) e^{Y_t} \right).
\]

Thus, solving the last equation with respect to \( g_x \) we obtain

\[
\hat{g}_{x:n} = \log \left( \beta \right) + \log \left( \frac{\sum_{t=2}^{n} \left( W \left( \frac{x_{t-1}}{h} \right) \exp(Y_t) \right)}{\sum_{s=2}^{n} W \left( \frac{x_{s-1}}{h} \right)} \right)
\]

and (2.3) is proved. The uniqueness of this estimator follows directly from the strict convexity of \( \rho(Y_t, g_x) \) with respect to \( g_x \) (see Example 2.4 of Section 2.2).

The following Corollary 2.2 gives us the local constant log-likelihood estimator \( \hat{f}_{x:n} \) for the function \( f(x) \) in the non-parametric ARCH(1) model (1.1) with normally distributed innovations.

Corollary 2.2. Assume that the innovations \( Z_t \) in (1.1) are standard normally distributed. Then the local log-likelihood estimator \( \hat{f}_{x:n} \) for the function \( f(x) \) in the non-parametric ARCH(1) model (1.1) is given by

\[
\hat{f}_{x:n} = \log \left( \beta \right) + \log \left( \frac{\sum_{t=2}^{n} \left( W \left( \frac{x_{t-1}}{h} \right) X_t^2 \right)}{\sum_{s=2}^{n} W \left( \frac{x_{s-1}}{h} \right)} \right).
\]

(2.5)

Proof. Since \( f(x) = e^{-\beta} e^{g(x)} \) and \( e^{Y_t} = X_t^2 \), the result follows directly from (2.3).

Remark. The estimator \( \hat{f}_{x:n} \) is the same as the one that we get when we make a local regression for the quadratic ARCH(1) model

\[
X_t^2 = f(X_{t-1}) Z_t^2 + f(X_{t-1}) Z_t^2 - 1 = f(X_{t-1}) + \eta_t,
\]

where \( \eta_t \) is a martingale difference with \( \mathbb{E} [\eta_t] = 0 \), which follows from the definition of \( Z_t \) in (1.1). In other words, the estimator (2.5) is the popular kernel estimator in the quadratic ARCH model. What is new here is the insight that it is a (transformed) local likelihood estimator and, as we will argue in Section 4, it is efficient under Gaussian innovations. The same result can be found also for more than one predictor, i.e. \( p > 1 \), and standard normally distributed innovations \( Z_t \). For more details, we refer the reader to Section 6.

As we have already said at the beginning of this Section, in the case of scaled \( t_\nu \)-distributed innovations the local likelihood estimator can not be written in a closed form. In spite of this, the following theorem proves the existence and the uniqueness of the local likelihood estimator under scaled \( t_\nu \)-distributed innovations.
**Theorem 2.3.** (Existence and uniqueness of \( \hat{g}_{x,n} \) under scaled \( t_\nu \)-distributed innovations)

Assume that the innovations \( Z_t \) in (1.1) are scaled \( t_\nu \)-distributed, \( \nu > 4 \). Then the local constant log-likelihood estimator \( \hat{g}_{x,n} \) given as the solution of the local likelihood equation (2.4) exists and is unique.

**Proof.** The result follows directly from Theorem 4.1 in Loader (1999). Note that in Example 2.5 of Section 2.2 we prove the convexity of \( \rho(y, g_x) \) in the case of scaled \( t_\nu \)-distributed innovations.

\[ \Box \]

### 2.2 The convexity of \( \rho(y, g_x) = -\log \left( f_{U_t}(y - g_x) \right) \)

We want to find here a sufficient condition to determine whether the function \( \rho(y, \cdot) \) is convex for all \( y \). Naturally, if the function \( \rho(y, \cdot) \) is twice differentiable in the second argument (which is true in many cases), it is well-known that a sufficient condition for the convexity is that the second partial derivative of \( \rho(y, \cdot) \) with respect to the second argument must be positive, i.e.

\[
\frac{\partial^2}{\partial y^2} \rho(y, g_x) > 0.
\]

Therefore, we only have to calculate an explicit formula for the second partial derivative of \( \rho(y, g_x) = -\log \left( f_{U_t}(y - g_x) \right) \) in our particular case. The first partial derivative of \( \rho(y, g_x) \) with respect to \( g_x \) looks like

\[
\frac{\partial}{\partial g_x} \rho(y, g_x) = -\frac{1}{f_{U_t}(y - g_x)} \cdot f'_{U_t}(y - g_x),
\]

where \( f'_{U_t}(y - g_x) = \frac{\partial}{\partial g_x} f_{U_t}(y - g_x) \). Differentiating a second time with respect to \( g_x \) we obtain the following condition for the convexity

\[
\frac{\partial^2}{\partial g_x^2} \rho(y, g_x) = f''_{U_t}(y - g_x)^2 - \frac{f'_{U_t}(y - g_x)}{f_{U_t}(y - g_x)} > 0 \quad \forall y,
\]

(2.6)

where \( f''_{U_t}(y - g_x) = \frac{\partial^2}{\partial g_x^2} f_{U_t}(y - g_x) \).

**Example 2.4.** We verify whether the condition (2.6) holds, and consequently whether the function \( \rho(y, \cdot) \) is convex, in the case of standard normally distributed innovations \( Z_t \) in (1.1). In the Example 1.1 we have already calculated the density function \( f_{U_t}(\cdot) \) in the case \( G = N(0, 1) \)

\[
f_{U_t}(y - g_x) = c \cdot \exp \left( \frac{y - g_x}{2} - \frac{\beta e(y - g_x)}{2} \right).
\]

Thus, the first and second partial derivative of \( f_{U_t}(\cdot) \) with respect to \( g_x \) are given by:

\[
f'_{U_t}(y - g_x) = c \cdot \exp \left( \frac{y - g_x}{2} - \frac{\beta e(y - g_x)}{2} \right) \cdot \left( -\frac{1}{2} + \frac{\beta e(y - g_x)}{2} \right) = f_{U_t}(y - g_x) \cdot \frac{1}{2} \left( \beta e(y - g_x) - 1 \right) \quad \text{and}
\]

\[
f''_{U_t}(y - g_x) = f'_{U_t}(y - g_x) \cdot \frac{1}{2} \left( \beta e(y - g_x) - 1 \right) - f_{U_t}(y - g_x) \cdot \frac{1}{2} \beta e(y - g_x) = f_{U_t}(y - g_x) \cdot \left( \frac{1}{4} \left( \beta e(y - g_x) - 1 \right)^2 - \frac{1}{2} \beta e(y - g_x) \right).
\]
Using these results in (2.6), we obtain that
\[
\frac{\partial^2}{\partial g_x^2} \rho(y, g_x) = \frac{1}{4} (\beta \varepsilon(y-g_x) - 1)^2 - \frac{1}{4} (\beta \varepsilon(y-g_x) - 1)^2 - \frac{1}{2} \beta \varepsilon(y-g_x) = \frac{1}{2} \beta \varepsilon(y-g_x) > 0 \forall y
\]
and hence in the case of standard normally distributed innovations $Z_t$ the function $\rho(y, \cdot)$ is convex with respect to $g_x$.

An identical result can (in some cases easily) be found directly from the calculation of the partial derivatives of $\rho$ with respect to $g_x$.

Example 2.5. We calculate explicitly $\rho$ in the case of scaled $t_\nu$-distributed innovations $Z_t$.

From the result of Example 1.2, we have
\[
\rho(y, g_x) = -\log \left( f_U(y-g_x) \right) = -\log \left( c e^{\frac{y-g_x}{\nu}} \left( 1 + \frac{\beta \varepsilon(y-g_x)}{\nu-2} \right)^{-\frac{\nu+1}{2}} \right)
\]
\[
= \frac{\nu + 1}{2} \log \left( 1 + \frac{\beta \varepsilon(y-g_x)}{\nu-2} \right) - \frac{y-g_x}{2} - \log(c).
\]
The first and second partial derivative of $\rho$ with respect to $g_x$ equal
\[
\rho'(y, g_x) = \frac{1}{2} - \frac{\nu + 1}{2} \frac{1}{1 + \frac{\nu-2}{\beta \varepsilon(y-g_x)}} := \psi(y - g_x)
\]
and
\[
\rho''(y, g_x) = \frac{\nu + 1}{2} \left( \frac{1}{1 + \frac{\nu-2}{\beta \varepsilon(y-g_x)}} \right)^2 \frac{\nu - 2}{\beta \varepsilon(y-g_x)},
\]
respectively. Since clearly the condition (2.6) is satisfied, the convexity of $\rho$ also holds in this case.

3 Consistency of the local likelihood estimator

In this Section, we want to prove the consistency of the normal local constant log-likelihood estimator $\hat{g}_{x,n}^N$ given by (2.3) in the “right” case of standard normally distributed innovations $Z_t$ and in the misspecified case of innovations $Z_t$ with distribution function $G \neq N$, having zero mean and variance one (pseudo-likelihood estimation). Moreover, we also want to show the consistency of the local constant log-likelihood estimator $\hat{g}_{x,n}^\nu$ given as the unique solution of (2.4) in the case of scaled $t_\nu$-distributed innovations.

In order to study the consistency of our estimators, we have to consider certain dependence or mixing conditions for stochastic processes. The process $\{X_t\}_{t \geq 1}$ satisfies the following strong mixing or $\alpha$-mixing condition (3.1), firstly introduced by Rosenblatt (1956):

there exists a sequence $\alpha(n)$ of positive numbers such that $\lim_{n \to \infty} \alpha(n) = 0$ and for any $A \in \mathcal{M}_{1,j}$, $B \in \mathcal{M}_{j+n,+\infty}$ we have
\[
| \mathbb{P}(A \cap B) - \mathbb{P}(A) \mathbb{P}(B) | \leq \alpha(n),
\]
where we denote by $\mathcal{M}_{a,b}$ the $\sigma$-algebra generated by the random variables $\{X_t : a \leq t \leq b\}$, $1 \leq a \leq \infty$;
or for the more often studied $\phi$-mixing or uniform mixing condition (3.2) (Billingsley, 1968) there exist coefficients $\phi(n)$ such that $\lim_{n\to\infty} \phi(n) = 0$ and the inequality

$$| \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) | \leq \phi(n)\mathbb{P}(A)$$

holds for any $A \in \mathcal{M}_{1,j}$, $B \in \mathcal{M}_{j+n,+\infty}$.

A process which satisfies the condition (3.1) or (3.2) is called to be $\alpha$-mixing or $\phi$-mixing respectively. Note that the $\phi$-mixing condition is considerably stronger than the $\alpha$-mixing condition.

Note also that if the process $\{X_t\}_{t\geq 1}$ is $\alpha$- or $\phi$-mixing, then also the process $\{X_{t-1}, Y_t\}_{t\geq 2}$ is mixing with coefficients bounded by the ones from $\{X_t\}_{t\geq 1}$ but with separation lag being smaller by one.

We assume that the mixing coefficients $\alpha(n)$ (or $\phi(n)$) satisfy the following condition (A1)

$$\sum_{n=2}^{\infty} \alpha(n) \frac{2^{n+\tau}}{2^{n+\tau}} < \infty$$

and $\mathbb{E}[|Y_t|^{2+\tau} \mid X_{t-1} = x] < \infty$ for some $\tau > 0$ \textit{(A1)}

or

$$\sum_{n=2}^{\infty} \phi(n) \frac{1+\tau}{1+\tau} < \infty$$

and $\mathbb{E}[|Y_t|^{2+\tau} \mid X_{t-1} = x] < \infty$ for some $\tau > 0$. \textit{(A1)}

This condition ensures that versions of the law of large numbers and of the central limit theorem for the dependent variables $\{X_{t-1}, Y_t\}_{t\geq 2}$ exist. Note that in the particular case where the innovations $Z_t$ in (1.1) are normally or scaled $t_\nu$-distributed, (A1) holds (for more details, see for example Doukhan, 1994).

### 3.1 Consistency of $\hat{g}^N_{x,n}$ and $\hat{f}^N_{x,n}$

The consistency in the “right” case of normally distributed innovations $Z_t$ is proved in the following Theorem 3.1.

**Theorem 3.1.** (Consistency of $\hat{g}^N_{x,n}$)

Suppose that the process $\{X_{t-1}, Y_t\}_{t\geq 2}$ is from the model (1.1)-(1.2) with standard normally distributed innovations $Z_t$. Then, for every fitting point $x$ the estimator $\hat{g}^N_{x,n}$ for the function $g(x)$ of the transformed ARCH(1) model (1.2) given by (2.3) is consistent, i.e.

$$\hat{g}^N_{x,n} \xrightarrow{P} g(x) = \beta^N + \log(f(x)) \quad \text{as } h_n \to 0, \ n h_n \to \infty.$$

In the proof of Theorem 3.1 we need the results of the following Proposition 3.2.

**Proposition 3.2.** Denote by $\gamma(g(x), \overline{g}) := \mathbb{E}_g [\rho(Y_t, g(x)) \mid X_{t-1} = x]$ and with $\mathcal{C}$ the compact set in $\mathbb{R}$ (i.e. a closed interval) of all possible constant functions $g(x)$. Let we assume that the following two conditions are satisfied:

\begin{enumerate}[i)]  
\item $\gamma(g(x), g(x)) < \gamma(g(x), \overline{g}(x))$ for all $\overline{g}(x) \neq g(x) \in \mathcal{C}$;  
\item for all $g, \overline{g}$: $\mathbb{E}_g [\inf_{\tau \in U(\overline{g})} \rho(Y_t, \tau) \mid X_{t-1} = x] \longrightarrow \gamma(g(x), \overline{g}(x))$ for $U(\overline{g}) \downarrow \{\overline{g}\}$, where $U(\overline{g})$ is a neighborhood of $\overline{g}$.
\end{enumerate}

Then:

$$\hat{g}_{x,n} \xrightarrow{P} g(x),$$

with $\hat{g}_{x,n}$ defined as in (2.2).
Proof. The proof can be found for example in Künsch (1997), Satz 6.6.

The second assumption of Proposition 3.2 is satisfied if the function \( \rho(y, g_x) \) is continuous \( \forall y \) and if we have that \( \inf_{\tau \in U(\overline{g}_x)} \rho(Y_t, \tau) \) is bounded from below. In this case, using the monotone convergence, we can exchange expectation and the limiting operation and assumption ii) holds.

Proof of Theorem 3.1. For every fitting point \( x \), denote by \( \hat{C}_n(g_x) \) the kernel estimator

\[
\hat{C}_n(g_x) := \frac{1}{(n-1)h_n} \sum_{t=2}^{n} \left( \frac{w_t(x)\rho(Y_t, g_x)}{\sum_{s=2}^{n} w_s(x)} \right).
\]

Thanks to the regularity conditions assumed for the process \( \{X_t, Y_t\}_{t \geq 2} \) and to (A1), using the WLLN for dependent observations with mixing structure (see for example Doukhan, 1994) we have that

\[
\forall g_x \in \mathcal{C} : \quad \hat{C}_n(g_x) \xrightarrow{P} \gamma(g(x), g_x) \quad \text{as} \quad nh_n \to \infty,
\]

where \( \gamma(\cdot, \cdot) \) is defined as in Proposition 3.2. Now, we have to show that

\[
\hat{g}^N_{x:n} = \arg\min_{g_x} \hat{C}_n(g_x) \xrightarrow{P} \arg\min_{g_x} \gamma(g(x), g_x) \xrightarrow{\text{law}} g(x)
\]

holds for every fitting point \( x \). This follows directly from the result of Proposition 3.2. Thus, what we need is to prove that the assumptions of Proposition 3.2 hold. Firstly, we prove that assumption i) is satisfied, which corresponds to show that \( \arg\min_{g_x} \gamma(g(x), g_x) = g(x) \). From the model (1.2), we know that

\[
g(x) = \beta^N + \log(f(x)) = \mathbb{E}_g[Y_t \mid X_{t-1} = x].
\]

Now, from the definition of \( \gamma \) it follows that

\[
\arg\min_{g_x} \gamma(g(x), g_x) = \arg\min_{g_x} \mathbb{E}_g[\rho(Y_t, g_x) \mid X_{t-1} = x]
\]

and differentiating we obtain that the minimum is a solution of the following equation

\[
\mathbb{E}_g[\rho'(Y_t, g_x) \mid X_{t-1} = x] = 0.
\]

In the case of standard normally distributed innovations \( Z_t \), we have already calculated in Examples 1.1 and 2.4 that \( \rho'(Y_t, g_x) = \frac{1}{2} \frac{-\beta^N}{2} e^{(Y_t-g_x)} \). Introducing the explicit form of \( \rho' \) in the equation above, we obtain that

\[
\arg\min_{g_x} \gamma(g(x), g_x) = \log(\hat{\beta}^N) + \log(\mathbb{E}_g[e^{Y_t} \mid X_{t-1} = x]).
\]

Using the definition of \( \hat{\beta} \), we rewrite \( \log(\hat{\beta}^N) \) as \( \beta^N \) and considering that from the model (1.1) follows \( f(x) = \mathbb{E}[X_t^2 \mid X_{t-1} = x] = \mathbb{E}_g[e^{Y_t} \mid X_{t-1} = x] \) we have proved assumption i) of Proposition 3.2.

To end the proof, we just have to show the continuity of the function \( \rho(y, \cdot) \) \( \forall y \) and to verify if the assumption

\[
\inf_{\tau \in U(\overline{g}_x)} \rho(Y_t, \tau)
\]

is satisfied for a neighborhood \( U(\overline{g}_x) \subset \mathcal{C} \) of the form \( U(\overline{g}_x) = (a_1, a_2) \), \( a_1 < a_2 \), \( a_1 \neq -\infty \). This follows directly from the explicit form of \( \rho(Y_t, g_x) \) as

\[
\rho(Y_t, g_x) = \frac{\hat{\beta}^N}{2} e^{(Y_t-g_x)} - \frac{Y_t - g_x}{2} - \log(c),
\]

c as in Example 1.1. 

\[\Box\]
Remark. It is also possible using stronger assumptions and robust theory (note that the estimator \( \hat{g}_{g,N} \) must satisfy the local log-likelihood equation (2.4)) to prove the almost sure convergence of the normal local constant log-likelihood estimator (2.3) to \( g(x) \), the solution of

\[
E[\rho'(Y_t, g(x)) \mid X_{t-1} = x] = \int_{\mathbb{R}} \rho'(Y_t, g(x))dF(Y_t \mid X_{t-1} = x) = 0.
\]

See for example Boente and Fraiman (1989), Theorem 2.1.

The following Corollary 3.3 gives us the consistency of the local constant log-likelihood estimator \( \hat{f}_{x,n}^N \) given by (2.5) for the function \( f(x) \) in the non-parametric ARCH(1) model (1.1) with standard normally distributed innovations \( Z_t \).

**Corollary 3.3.** (Consistency of \( \hat{f}_{x,n}^N \))

Under the same assumptions of Theorem 3.1, the estimator \( \hat{f}_{x,n}^N \) given by (2.5) is consistent for every fitting point \( x \).

**Proof.** Thanks to the result of continuous mapping theorem (\( \exp \) is a continuous function) and since \( f(x) = e^{-\beta^N}e^{\delta} \), the consistency of \( \hat{f}_{x,n}^N \) follows directly from Theorem 3.1. \( \square \)

### 3.2 Consistency of \( \hat{g}_{g,N} \) and \( \hat{f}_{x,n}^N \) under pseudo-likelihood estimation

We want to investigate here the consistency of the estimators \( \hat{g}_{g,N} \) and \( \hat{f}_{x,n}^N \) in the misspecified case of innovations \( Z_t \) as in (1.1) with distribution function \( G \neq N \).

**Theorem 3.4.** (Consistency under pseudo-likelihood estimation)

Suppose that the process \( \{X_{t-1}, Y_t\}_{t \geq 2} \) is from the model (1.1)-(1.2) and let the innovations \( Z_t \) have a distribution function \( G \neq N \). Denote by

\[
\delta := E[\log(Z_t^2)] - \beta^N,
\]

where \( \beta^N \) is defined as in Theorem 2.1. Then for every fitting point \( x \):  

i) \( \hat{g}_{g,N} \) given by (2.3) is a consistent estimator for \( g(x) - \delta \), where \( g(x) = E[Y_t \mid X_{t-1} = x] \) in the model (1.2);

ii) \( \hat{f}_{x,n}^N \) is a consistent estimator for \( f(x) \) in the model (1.1).

A very important consequence of the results of Theorem 3.4 is that even in the misspecified case of non-Gaussian innovations \( Z_t \) the normal estimator \( \hat{f}_{x,n}^N \) is consistent. On the other side, we see that the estimator \( \hat{g}_{g,N} \) is not consistent for the less interesting function \( g(x) \).

**Proof of Theorem 3.4.** We rewrite \( g(X_{t-1}) \) as \( \delta + g^N(X_{t-1}) \) and \( U_t \) as \( U_t^N - \delta \) in the model (1.2)

\[
Y_t = g(X_{t-1}) + U_t = \delta + g^N(X_{t-1}) + U_t^N - \delta,
\]

where \( g^N(X_{t-1}) \) and \( U_t^N \) are the variables in the misspecified case of normally distributed innovations \( Z_t \). Note that in the “right” model is \( E[U_t] = E[U_t^N - \delta] = 0 \), but \( \delta = E[U_t^N] = E[\log(Z_t^2) - \beta^N] \neq 0 \).

Assuming standard normally distributed innovations \( Z_t \), we proceed with a pseudo-likelihood estimation for the function \( g(x) = \delta + g^N(x) = \delta + \beta^N + \log(f(x)) \). As before, the optimal normal local constant log-likelihood estimator is given by (2.3). The proof of i) is analogous to the one of Theorem 3.1.
From our construction we have that \( g(x) = \delta + \beta N + \log(f(x)) \). Solving the equation with respect to \( f(\cdot) \) we get that
\[
f(x) = e^{-\beta N} e^{(g(x) - \delta)}
\]
and from i) and Corollary 2.2 follows ii).

3.3 Consistency of \( \hat{g}_{x;n}^{t_{\nu}} \)

Since in this case we do not know the explicit form of the estimator \( \hat{g}_{x;n}^{t_{\nu}} \), we have to show the consistency in a different way as the one used for the normal constant log-likelihood estimator. As a starting point, we know that \( \hat{g}_{x;n}^{t_{\nu}} \) is the unique solution of the constant local log-likelihood equation (2.4) with \( \rho'(y,g_x) = \psi(y-g_x) : \mathbb{R} \to \mathbb{R} \) already calculated in Example 2.5 and given by
\[
\psi(y-g_x) = 2 - \frac{\nu + 1}{2} \frac{1}{1 + \frac{\nu - 2}{\beta e^{(y-g_x)}}}.
\]

The most important condition for consistency is that the function \( g(x) \) must satisfy
\[
\mathbb{E}_g[\psi(Y_t - g(x)) | X_{t-1} = x] = 0
\]
for every fitting point \( x \). The consistency of the estimator \( \hat{g}_{x;n}^{t_{\nu}} \) follows from the general result of Proposition 3.5.

**Proposition 3.5.** Suppose that the process \( \{X_{t-1}, Y_t\}_{t \geq 2} \) is stationary, \( \alpha \)- or \( \phi \)-mixing and that assumption (A1) holds. Let \( \psi : \mathbb{R} \to \mathbb{R} \) be continuous and \( \mathbb{E}_g[|\psi(Y_t - g_x)| | X_{t-1} = x] < \infty \) for every fitting point \( x \). If:
\[
\begin{align*}
\mathbb{E}_g[\psi(Y_t - g_x) | X_{t-1} = x] &> 0 \text{ for } g_x > g(x) \text{ and } \\
\mathbb{E}_g[\psi(Y_t - g_x) | X_{t-1} = x] &< 0 \text{ for } g_x < g(x)
\end{align*}
\]
\tag{3.3}

then there exists a sequence of estimators \( \{\hat{g}_{x;n}\}_n \) which converges almost surely to \( g(x) \) and satisfies
\[
\sum_{t=2}^{n} (w_t(x) \psi(Y_t - \hat{g}_{x;n})) = 0.
\]

**Proof.** The proof can be found in Künsch (1997), Satz 6.7. Note that the mixing assumption ensures that also for the dependent variables \( \psi(Y_t - g_x) \), a general version of the SLLN holds; for more details, see Doukhan (1994)

Now, we are in position to prove the consistency of the estimator \( \hat{g}_{x;n}^{t_{\nu}} \) in the case of scaled \( t_{\nu} \)-distributed innovations \( Z_t \).

**Theorem 3.6.** (Consistency of \( \hat{g}_{x;n}^{t_{\nu}} \))

Suppose that the process \( \{X_{t-1}, Y_t\}_{t \geq 2} \) is from the model (1.1)-(1.2) and let the innovations \( Z_t \) in (1.1) be scaled \( t_{\nu} \)-distributed with degrees of freedom parameter \( \nu > 4 \). Then, the local constant log-likelihood estimator \( \hat{g}_{x;n}^{t_{\nu}} \) given as the unique solution of the likelihood equation (2.4) is consistent for \( g(x) \) satisfying
\[
\mathbb{E}_g[\psi(Y_t - g(x)) | X_{t-1} = x] = 0
\]
for every fitting point \( x \).
We have already shown that for scaled \( t_{\nu} \)-distributed innovations the function \( \psi \) takes the values

\[
\psi(u) = \frac{1}{2} - \frac{\nu + 1}{2} \frac{1}{1 + \frac{\nu - 2}{\beta x^u}}
\]

and is clearly continuous for all \( u \in \mathbb{R} \). It also directly follows that \( \psi(u) \) is bounded for all \( u \in \mathbb{R} \), since:

\[
\text{for } u \to +\infty : \psi(u) \longrightarrow \frac{1}{2} - \frac{\nu + 1}{2} \quad \text{and for } u \to -\infty : \psi(u) \longrightarrow \frac{1}{2}.
\]

Therefore, we have that \( |\psi(u)| \leq C(\nu) := -\left( \frac{1}{2} - \frac{\nu + 1}{2} \right) \), for \( 4 < \nu < \infty \) and \( x \in \mathbb{R} \) and consequently, thanks to the monotonicity of the integral, we get that \( \mathbb{E}_g[|\psi(Y_t - g_x)| | X_{t-1} = x] < \infty \) for every fitting point \( x \).

We end the proof by showing that the condition (3.3) of Proposition 3.5 is satisfied. The result follows then directly from Proposition 3.5. Define

\[
\lambda(g_x) := \mathbb{E}_g[\psi(Y_t - g_x) | X_{t-1} = x] = \int_{\mathbb{R}} \psi(y - g_x) dF_{Y_t|X=x}(y).
\]

Now, if \( g_x > g(x) \) it follows that \( y - g_x < y - g(x) \). Since \( \psi(u) \) is strictly monotone, we get that \( \psi(y - g_x) > \psi(y - g(x)) \). According to the monotonicity of the integral, we finally have that \( \lambda(g_x) > \lambda(g(x)) = 0 \) by definition. On the other hand, if \( g_x < g(x) \) we get analogously that \( \lambda(g_x) < \lambda(g(x)) = 0 \) for all \( x \in \mathbb{R} \) and (3.3) holds.

## 4 Asymptotic normality of the local likelihood estimator

In this Section, we investigate the asymptotic representation of the normal and scaled \( t_{\nu} \) local constant log-likelihood estimator \( \hat{g}_{x:n}^N \) given by (2.3) and \( \hat{g}_{x:n}^{t_{\nu}} \), respectively. For this purpose, we will need the following general assumptions:

**H1.** \( \mathbb{E} \left[ \rho''(Y_t, g(x)) | X_{t-1} = x \right] \neq 0 \).

**H2.** The process \( \{X_{t-1}, Y_t\}_{t \geq 2} \) is a stationary \( \alpha \)-mixing process, with mixing coefficients \( \alpha(n) \) defined analogously to (3.1), such that

\[
\sum_{i=2}^{\infty} \alpha(i) \tau ^{-i} < \infty \quad \text{and} \quad \mathbb{E} \left[ |\psi(Y_t - g_x)|^{2+\tau} | X_{t-1} = x \right] < \infty \quad \text{for some } \tau > 0.
\]

(If the process is \( \phi \)-mixing, an analogous condition exists for the coefficients \( \phi(n) \).)

**H3.** The stationary density \( d \) of the vector \( X_t \) is continuous and positive at \( x \).

**H4.** For all \( s \geq 2 \), the density \( d_s(u, v) \) of \( (X_t, X_{t+s}) \) is bounded uniformly in \( s \).

**H5.** The kernel \( W(\cdot) \) is defined as in Section 2.

**H6.** There exists a constant \( 0 \leq s < \infty \) such that \( h_n n^{-\frac{1}{2}} \rightarrow s \) as \( n \rightarrow \infty \).

**H7.** We assume that the function \( g(x) \) in the model (1.2) verifies a Lipschitz condition of order 1, i.e.

\[
|g(u) - g(x)| \leq C |u - x| \quad \text{for some } C > 0,
\]
and that the following limit exists
\[ \lim_{\epsilon \to 0} \frac{g(x + \epsilon u) - g(x)}{\epsilon} = g'(x, u). \]

**Remark.** Note that assumption H7 on the function \( g(x) = \beta + \log f(x) \) is satisfied if we assume that
\[ \sup_x \left| \frac{f'(x)}{f(x)} \right| \leq C < \infty \]
implied for example by \( \sup_x |f'(x)| < \infty \) and \( \inf_x f(x) > 0 \).

This condition holds for example for the parametric ARCH(1) model, where \( f(x) = \alpha_0 + \alpha_1 x^2, \alpha_0 > 0, \alpha_1 \geq 0 \).

**H8.** Consider the function \( \psi : \mathbb{R} \to \mathbb{R} \) defined by
\[ \psi(Y_t - g_x) = \frac{\partial}{\partial g_x}(Y_t, g_x). \]
Assume that \( \psi(\cdot) \) is differentiable with finite second moment and satisfying
\[ \left| \frac{\partial \psi(Y_t - t_1)}{\partial t_1} - \frac{\partial \psi(Y_t - t_2)}{\partial t_2} \right| \leq H(Y_t) |t_1 - t_2|, \]
for all \( t_1, t_2 \in \mathcal{C} \), a compact set \( \subset \mathbb{R} \). Moreover, analogously to H2 we assume that
\[ \mathbb{E} \left[ \left| \frac{\partial \psi(Y_t - s)}{\partial s} \right|_{s=g(x)}^2 \right] < \infty \text{ for some } \tau > 0. \]

In the proof of the following Theorem 4.2, we will need the results of Lemma 4.1. For simplicity from now on, we will denote by \( n' = n - 1 \).

**Lemma 4.1.** Let \( \{X_t, Z_t\}_{t \geq 1} \) be a stationary random process verifying H2. Denote by \( F(u \mid X_1 = x) \) the conditional distribution of \( Z_1 \) given that \( X_1 = x \), by \( \mu_1(x) = \mathbb{E}[Z_1 \mid X_1 = x] \) and by \( \sigma^2_1(x) = \mathbb{E}\left[ (Z_1 - \Phi(x))^2 \mid X_1 = x \right] \). Let us suppose that
i) \( \mu_1 \) is Lipschitz and \( \lim_{\epsilon \to 0} \frac{\mu_1(x + \epsilon u) - \mu_1(x)}{\epsilon} = \mu'_1(x, u) \) exists;
ii) \( \sigma^2 \) is continuous in a neighborhood of \( x \).

For a fitting point \( x \), denote by \( \hat{\mu}^{K}_{1,n} = \sum_{i=2}^{n} (k_t(x)Z_t) \) the kernel estimator, i.e.
\[ k_t(x) = \frac{K\left(\frac{X_t-x}{h_n}\right)}{\sum_{s=2}^{n} K\left(\frac{X_s-x}{h_n}\right)}, \]
where \( K(\cdot) \) is a kernel function as in Section 2. Then H3 to H6 imply that \( \sqrt{n}h_n(\hat{\mu}^{K}_{1,n} - \mu_1(x)) \)
is asymptotically normally distributed with mean \( \mu_2 = s^3 \int_{\mathbb{R}} \mu'_1(x, u)K(u)du \) and variance \( \sigma^2_2 = \sigma^2_1(x)K_0 \), with \( K_0 = \int_{\mathbb{R}} K^2(u)du / d(x) \).

**Proof.** The proof can be found in Boente and Fraiman (1990), Lemma 2. \( \square \)

First of all, we give the asymptotic representation of the general local constant log-likelihood estimator \( \hat{g}_{x,n} \) given by (2.2), i.e. without knowing the distribution function \( G \) of the innovations \( Z_t \).
**Theorem 4.2.** (Asymptotic normality of $\hat{g}_{x,n}$)
Assume that the conditions H1 to H8 are satisfied and that $\hat{g}_{x,n} \xrightarrow{P} g(x)$ as $n'h_n \to \infty$. Then for every fitting point $x$
\[
\sqrt{n'h_n} (\hat{g}_{x,n} - g(x)) \xrightarrow{d} N(\mu, \sigma^2) \quad \text{as} \quad n \to \infty,
\]
where the asymptotic mean $\mu$ is given by
\[
\mu = s^2 \int_{\mathbb{R}} g'(x, u) W(u) du
\]
and the asymptotic variance $\sigma^2$ by
\[
\sigma^2 = V(\psi, G) \frac{\int_{\mathbb{R}} W^2(u) du}{d(x)},
\]
where the factor $V(\psi, G)$ equals
\[
V(\psi, G) = \frac{\int_{\mathbb{R}} \psi^2(u) dF_{U_t}(u)}{\left( \int_{\mathbb{R}} \psi'(u) dF_{U_t}(u) \right)^2}.
\]

Note that if the function $g$ is smooth, we have that $g'(x, u) = g'(x) \cdot u$ and the asymptotic mean $\mu = 0$ (since the kernel $W(\cdot)$ is assumed to be symmetric). On the other hand, the asymptotic variance $\sigma^2$ strictly depends from the factor $V(\psi, G)$ and, as we can see from the explicit calculation for the examples with $G = \mathcal{N}(0, 1)$ and $G = \sqrt{\frac{\nu-2}{\nu}} t_\nu$, this can lead to big differences if we confront the asymptotic variance of the local likelihood estimator with the one of other non-parametric estimators like for example classical local regression.

**Proof of Theorem 4.2.** As a starting point, we remind the reader of the fact that the local constant log-likelihood estimator $\hat{g}_{x,n}$ given by (2.2) must be a solution of the log-likelihood equation (2.4). Since the function $\psi(Y_t - g_x) = \rho'(Y_t, g_x)$ is assumed to be differentiable for all $Y_t$ (see H8) and since $\hat{g}_{x,n}$ should be an estimator for $g(x)$, the idea is to consider the Taylor expansion of $\hat{g}_{x,n}$ around $g(x)$. This leads to the following result:
\[
0 = \frac{1}{n'h_n} \sum_{t=2}^{n} (w_t(x) \psi(Y_t - \hat{g}_{x,n})) = \frac{1}{n'h_n} \sum_{t=2}^{n} \left( w_t(x) \psi(Y_t - g(x)) \right) + \\
+ (\hat{g}_{x,n} - g(x)) \frac{1}{n'h_n} \sum_{t=2}^{n} \left( w_t(x) \frac{\partial \psi}{\partial \xi} (Y_t - \xi_n) \right),
\]
where $w_t(x) = W(\frac{x - X_{t-1}}{h_n})$ and we have that $|\xi_n - g(x)| \leq |\hat{g}_{x,n} - g(x)|$. We denote by
\[
M_n := \frac{1}{n'h_n} \sum_{t=2}^{n} \left( w_t(x) \frac{\partial \psi}{\partial \xi} (Y_t - \xi_n) \right)
\]
and by
\[
M((g(x)) = \mathbb{E}_g \left[ \frac{\partial \psi}{\partial t} (Y_t - t) \right]_{t=g(x)} | X_{t-1} = x] .
\]
Since the estimator $\hat{g}_{x,n}$ is assumed to be consistent, it follows from assumption H8 and the LLN that

$$
|M_n - M(g(x))| \leq \left| \frac{1}{n'h_n} \sum_{t=2}^{n} \left( w_t(x)\psi'(Y_t - g(x)) \right) - M(g(x)) \right| +
+ |\hat{g}_{x,n} - g(x)| \cdot \left| \frac{1}{n'h_n} \sum_{t=2}^{n} \left( w_t(x)H(Y_t) \right) \right|,
$$

where the right side converges to zero in probability. Thus, we have proved that

$$M_n \xrightarrow{P} M(g(x)) \text{ as } n'h_n \to \infty, \ h_n \to 0.$$

Therefore, assuming that $M^{-1}$ exists (see H1), we can approximate the difference $\hat{g}_{x,n} - g(x)$ by an arithmetic mean as follows:

$$\hat{g}_{x,n} - g(x) = -(M_n)^{-1} \frac{1}{n'h_n} \sum_{t=2}^{n} \left( w_t(x)\psi(Y_t - g(x)) \right)
= - (M(g))^{-1} \frac{1}{n'h_n} \sum_{t=2}^{n} \left( w_t(x)\psi(Y_t - g(x)) \right) + R_n,$$

where $R_n$ is a rest term such that $\sqrt{n'h_n}R_n \xrightarrow{P} 0$ as $n'h_n \to \infty$.

Rewriting the last equation without using the introduced notation, we get that

$$\hat{g}_{x,n} - g(x) = \frac{1}{n'h_n} \sum_{t=2}^{n} \left( w_t(x)\psi(Y_t - g(x)) \right) + \mu_1(x) + R_n,$$

Now we can use the result of Lemma 4.2 for the term

$$\frac{\sqrt{n'h_n}}{n'h_n} \sum_{t=2}^{n} \left( w_t(x)\psi(Y_t - g(x)) \right)$$

and the dependent variables $Z_t = \psi(Y_t - g(x))$. Note that in our case

$$\mu_1(x) = \mathbb{E}[\psi(Y_t - g(x) \mid X_{t-1} = x] = 0$$

and

$$\sigma_1^2(x) = \int_{\mathbb{R}} \psi'(y - g(x))^2 dF_{U_t}(y - g(x)).$$

The result follows since by Slutsky’s Lemma the error term $R_n$ can be neglected, and since from the definition of $\mu'_1(x,u)$ we have that

$$\mu'_1(x,u) = g'(x,u) \mathbb{E}[\psi'(Y_t - g(x)) \mid X_{t-1} = x].$$
4.1 The asymptotic variance of $\hat{\theta}_{x,n}^N$

Using the results of Theorem 4.2, we calculate the asymptotic variance of the normal local constant log-likelihood estimator $\hat{\theta}_{x,n}^N$ given by (2.3). This is done in the following theorem.

Theorem 4.3. (Asymptotic variance of $\hat{\theta}_{x,n}^N$)

Suppose that the process $\{X_{t-1}, Y_t\}_{t \geq 2}$ is from the model (1.1)-(1.2) and let the innovations $Z_t$ be standard normally distributed. Under the assumptions H4 to H7, the asymptotic variance of the normal local constant log-likelihood estimator $\hat{\theta}_{x,n}^N$ given by (2.3) of the function $g(x)$ equals

$$\sigma^2 = 2 \cdot \int_{\mathbb{R}} W^2(u)du \frac{d(x)}{dx} (4.2)$$

for every fitting point $x$, where the kernel $W(\cdot)$ is defined as in Section 2.

Proof. The result follows directly from Theorem 4.2. In the case of standard normally distributed innovations $Z_t$ we have that assumptions H2-H3 hold (see Doukhan, 1994). Moreover, we have already shown in Examples 1.1 and 2.4 that the function $\psi(Y_t - g(x)) = \rho'(Y_t, g(x))$ equals

$$\psi(Y_t - g(x)) = \frac{1}{2} - \bar{\beta}_N e^{(Y_t - g(x))}.$$

Knowing the explicit form of $\psi$, it is easy to verify that assumption H8 is satisfied (the exponential function is Lipschitz-continuous in a compact set $C$ and for example for $\tau = 1$, the second condition holds). Now, we want to show that H1 holds, too:

$$\mathbb{E} \left[ \frac{\bar{\beta}_N}{2} e^{(Y_t - g(x))} \mid X_{t-1} = x \right]_{u = y - g(x)} = \frac{\bar{\beta}_N}{2} \int_{\mathbb{R}} e^u \exp \left( \frac{u}{2} - \frac{\bar{\beta}_N e^u}{2} \right) du =$$

$$= \frac{\bar{\beta}_N}{2} \int_{\mathbb{R}} e^u \exp \left( - \frac{\bar{\beta}_N e^u}{2} \right) du =$$

$$= \frac{\bar{\beta}_N}{2} \int_{0}^{+\infty} \sqrt{z} \exp \left( - \frac{\bar{\beta}_N z}{2} \right) dz > 0.$$

Thus, we can use the result of Theorem 4.2 to calculate the asymptotic variance of $\hat{\theta}_{x,n}^N$, taking into account also that this estimator is consistent for $g(x)$ (see Theorem 3.1). For simplicity, we denote by $k := \frac{\bar{\beta}_N}{2}$ in the following calculations. We have already shown that

$$\int_{\mathbb{R}} \psi'(y - g(x))dF_{U_i}(y - g(x)) = k \int_{0}^{+\infty} e^{-\frac{kz}{c}} \sqrt{z} dz =$$

$$= k \left[ \frac{-c}{k} e^{-\frac{kz}{c}} \right]_{0}^{+\infty} + \frac{c}{2} \int_{0}^{+\infty} \frac{e^{-\frac{kz}{c}}}{\sqrt{z}} dz =$$

$$= c \int_{0}^{+\infty} e^{-\frac{kz}{c}} dv = c \int_{0}^{+\infty} e^{-\frac{k^2}{2} dv = \frac{c}{\sqrt{2\pi}} \sqrt{\frac{2\pi}{2}} \bar{\beta}_N.$$

Using the definition of $c$ made in Example 1.1, we finally get that

$$\int_{\mathbb{R}} \psi'(y - g(x))dF_{U_i}(y - g(x)) = \frac{1}{2}.$$
Analogously we obtain that

\[
\int_{\mathbb{R}} \psi^2(y - g(x)) dF_{U_t}(y - g(x)) =
\]

\[
= c \int_{\mathbb{R}} \left( \frac{1}{2} - \frac{3 \nu}{2} e^{Y_t - g_x} \right)^2 \exp \left( \frac{u}{2} - \frac{3 \nu}{2} e^u \right) du = \frac{1}{2}
\]

and the result is proved. \(\square\)

4.2 The asymptotic variance of \(\hat{g}_{x,n}^{t_\nu}\)

We are interested here in the explicit calculation from (4.1) of the asymptotic variance of \(\hat{g}_{x,n}^{t_\nu}\) in the case of scaled \(t_\nu\), \(\nu > 4\), distributed innovations \(Z_t\).

**Theorem 4.4.** (Asymptotic variance of \(\hat{g}_{x,n}^{t_\nu}\))

Suppose that the process \(\{X_{t-1}, Y_t\}_{t \geq 2}\) is from the model (1.1)-(1.2) and let the innovations \(Z_t\) in (1.1) be scaled \(t_\nu\), \(\nu > 4\), distributed. Under the assumptions H4 to H7, the asymptotic variance of the local constant log-likelihood estimator \(\hat{g}_{x,n}^{t_\nu}\), given as the unique solution of (2.4) equals

\[
\sigma^2 = \frac{2(\nu + 3)}{\nu} \int_{\mathbb{R}} W^2(u) du \frac{d(x)}{d(x)} \quad (4.3)
\]

for every fitting point \(x\), where the kernel \(W(\cdot)\) is defined as in Section 2.

**Remark.** Note that for \(\nu \to \infty\), the factor \(V(\psi, G) = \frac{2(\nu + 3)}{\nu}\) in (4.3) converges to 2, which is consistent with the result found in (4.2) for a standard normal distribution.

**Proof of Theorem 4.4.** The result follows directly from Theorem 4.2. In the case of scaled \(t_\nu\)-distributed innovations \(Z_t\) we have that assumptions H2-H3 hold (see Doukhan, 1994). Moreover, we have already shown in Examples 1.2-2.5 that the function \(\psi(Y_t - g_x) = \rho'(Y_t, g_x)\) equals

\[
\psi(Y_t - g_x) = \frac{1}{2} - \frac{\nu + 1}{2} \frac{1 + \frac{1}{\beta t_\nu e^{Y_t - g_x}}}{1 + \frac{\nu - 2}{\beta t_\nu e^{Y_t - g_x}}}.
\]

Like in the normal case, knowing the explicit form of \(\psi\), it is easy to verify that H8 is satisfied. Now, we want to show that H1 holds, too. This is done by proving that

\[
E \left[ \frac{\nu + 1}{2} \left( \frac{1}{1 + \frac{\nu - 2}{\beta t_\nu e^{(y - g(x))}}} \right)^2 \cdot \frac{\nu - 2}{\beta t_\nu e^{(y - g(x))}} \right| X_{t-1} = x \right] =
\]

\[
= \frac{\nu + 1}{2} \int_{\mathbb{R}} \left( \frac{1}{1 + \frac{\nu - 2}{\beta t_\nu e^u}} \right)^2 \frac{\nu - 2}{\beta t_\nu e^u} dF_{U_t}(u) > 0.
\]

Therefore we can use the result of Theorem 4.2 to calculate the asymptotic variance of \(\hat{g}_{x,n}^{t_\nu}\), considering also that this estimator is consistent for \(g(x)\) (see Theorem 3.6). For simplicity, we
denote by \( k := \frac{\nu + 1}{2} \frac{\Gamma\left(\frac{\nu + 1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\nu}} \) in the following calculations. We have already shown that

\[
\int_R \psi'(y - g(x)) dF_{U_t}(y - g(x)) = \frac{\nu + 1}{2} \int_R \left(\frac{1}{1 + \frac{\nu - 2}{\beta^2 e^u}}\right)^2 \frac{\nu - 2}{\beta^2 e^u} \frac{\nu - 2}{\beta^2 e^u} dF_{U_t}(u)
\]

\[
= c \frac{\nu + 1}{2} \int_R \left(\frac{1}{1 + \frac{\nu - 2}{\beta^2 e^u}}\right)^2 \left(1 + \frac{\beta^2}{\nu - 2} e^u\right)^{-(\frac{\nu + 1}{2})} e^u du =
\]

\[
= k \sqrt{\pi} \frac{\Gamma\left(\frac{\nu + 1}{2}\right)}{2} \frac{1}{\Gamma\left(\frac{\nu + 3}{2}\right)} = \frac{\nu + 1}{4} \frac{\Gamma\left(\frac{\nu + 1}{2}\right)\Gamma\left(\frac{\nu + 1}{2} + 1\right)}{\Gamma\left(\frac{\nu + 3}{2}\right)\Gamma\left(\frac{\nu + 5}{2}\right)} = \frac{\nu}{2(\nu + 3)},
\]

because of the definition of \( \Gamma(x + 1) = x\Gamma(x), \forall x > 0 \). Analogously we obtain that

\[
\int_R \psi^2(u) dF_{U_t}(u) =
\]

\[
= c \int_R \left(\frac{\nu + 1}{2} - \frac{\nu - 2}{2} \frac{1}{1 + \frac{\nu - 2}{\beta^2 e^u}}\right)^2 \left(1 + \frac{\beta^2}{\nu - 2} e^u\right)^{-(\frac{\nu + 1}{2})} e^u du =
\]

\[
= 1 - \frac{\nu + 1}{4} \frac{\Gamma\left(\frac{\nu + 1}{2}\right)}{\Gamma\left(\frac{\nu + 3}{2}\right)} + \frac{3(\nu + 1)^2}{16} \frac{\Gamma\left(\frac{\nu + 1}{2}\right)}{\Gamma\left(\frac{\nu + 5}{2}\right)} = \frac{\nu}{2(\nu + 3)}
\]

and the result is proved. \( \square \)

### 4.3 Asymptotic variance of the local regression estimator

We now want to compare now the asymptotic variances given by (4.2) and (4.3) in the normal and in the scaled \( t_\nu, \nu > 4 \), case respectively with the one obtained from a local regression estimation. In the local regression, we have that the estimator must be a solution of the constant local log-likelihood equation (2.4) with \( \rho(Y_t, g_x) = (Y_t - g_x)^2 \). Note that this happens exactly in the misspecified case when the variables \( U_t \) are standard normally distributed. Therefore, in this case we get that

\[
\int_R \psi'(y - g(x)) dF_{U_t}(y - g(x)) = 2 \int_R dF_{U_t}(y - g(x)) = 2
\]

and

\[
\int_R \psi^2(y - g(x)) dF_{U_t}(y - g(x)) = 4 \int_R u^2 dF_{U_t}(u) = 4 E[U_t^2] = 4 \sigma_{U_t}^2,
\]

since \( E[U_t] = 0 \), and from the result of Theorem 4.2 follows that the asymptotic variance equals

\[
\sigma_{reg}^2 = \frac{\int_R K^2(u) du}{d(x)}. \tag{4.4}
\]

An approximation of the term \( V(\psi, G) = \sigma_{U_t}^2 \) in (4.4) can be easily calculated by simulating. Since the estimator \( \hat{g}_N \) given by (2.3) is consistent for \( \beta^N + \log(f(x)) \) also under pseudo-likelihood estimation (see the results of Theorem 3.4), we simulate the values for the factor \( V(\psi, G) \) in (4.1) for \( G \neq N(0, 1) \) and \( \psi(u) = \frac{1}{2} - \frac{\beta^N}{2} e^u \).  

17
The behaviour of the factor $V(\psi, G)$ of the asymptotic variance of the local constant log-likelihood estimators $\hat{g}_N^n$ and $\hat{g}_{t,\nu}^n$, and of the local regression estimator in the case of scaled $t_\nu$ distributed innovations $Z_t$ with degrees of freedom $\nu > 4$ is illustrated in Figure 4.1.

As we expect, the asymptotic variance of the estimator from a local regression is bigger than the one of the normal and $t_\nu$ local log-likelihood estimator in the range $\nu \in [6, +\infty)$. The reason is that in the case of innovations $Z_t$ with distribution $G$ approximately standard normal, as for example the scaled $t_\nu$-distribution with $\nu$ big enough, the variables $U_t$ have a distribution function very different from a normal one as assumed by a local regression estimation. We also see that the “right” estimator (i.e. the estimator $\hat{g}_{t,\nu}^n$ with the true parameter $\nu$) is always minimal. A main disadvantage occurring when we work with the normal local log-likelihood estimator in the misspecified case of scaled $t_\nu$-distributed innovations $Z_t$ is that the asymptotic variance increases more rapidly than for example when we work with the local regression estimator. A consequence of this fact is that the local regression estimator has a smaller factor $V$ for degrees of freedom $\nu < 5.5$. In other words, the local regression estimator and the $t_\nu$ local log-likelihood estimator are more stable than the normal one under pseudo-likelihood estimation. The big improvement in the asymptotic variance of the local log-likelihood estimators over the local regression happens when the innovations are standard normally distributed (2 vs 4.94).

5 Some numerical results

We consider here the performance of the local constant log-likelihood estimators for simulated and real data. We compare the results with the ones from a local regression as we have done in
the last section. The problem of the choice of an optimal bandwidth is not considered here (see for example Loader, 1999, or Hart and Vieu, 1990), but we always report with the use of the bandwidth that minimize our performance measure.

5.1 Simulations

The model that we use for simulating data is as in (1.1) with function \( f(\cdot) \) given by

\[
  f(x) = 0.2 + 0.8x^2 \cdot \exp(1 - 0.5|x|). 
\] (5.1)

The distribution of the innovations \( Z_t \) in (1.1) is either standard normal or scaled \( t_\nu, \sqrt{\frac{\nu}{\nu - 2}}Z_t \sim t_\nu, \nu > 4 \), so that \( Z_t \) has variance one.

For quantifying the goodness of fit, we consider the following measure:

\[
\text{OS-L}_2 = \frac{1}{n - 1} \sum_{t=2}^{n} \left| \sigma^2_t - \hat{f}_{Q_{t-1:n}} \right|^2, \quad Q^n_1 \text{ a new test set (out-sample loss)}, 
\] (5.2)

where \( \hat{f}_{Q_{t-1:n}} \) uses the model estimated from the data \( X^n_1 \) but evaluates it on new test data \( Q^n_1 \) that is another independent realization of the data. The out-sample OS-L\(_2\) statistic (5.2) is a measure for predictive performance. The results for five independent realizations of \( n = 500 \) days from the model (1.1) with \( f(\cdot) \) given by (5.1) with various degrees of freedom parameters \( \nu > 4 \) for the normal local log-likelihood estimator \( \hat{f}_{x:n}^N \) given by (2.5), the \( t_\nu \) local log-likelihood estimator \( \hat{f}_{x:n}^t = \exp(\hat{g}_{x:n}^t - \hat{\beta}) \) coming out from (2.2) and the local regression estimator are reported in Table 5.1.

<table>
<thead>
<tr>
<th>degrees of freedom ( \nu )</th>
<th>OS-L(_2) measure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>local reg.</td>
</tr>
<tr>
<td>( \infty )</td>
<td>20.262</td>
</tr>
<tr>
<td>10</td>
<td>8.359</td>
</tr>
<tr>
<td>8</td>
<td>5.283</td>
</tr>
<tr>
<td>6</td>
<td>1.325</td>
</tr>
<tr>
<td>5</td>
<td>6.639</td>
</tr>
</tbody>
</table>

Table 5.1: The OS-L\(_2\) measure for five independent realizations of \( n = 500 \) days from the model (1.1) with \( f(\cdot) \) given by (5.1) with various degrees of freedom parameters \( \nu > 4 \). The relative gains over the classical local regression estimation are given between parenthesis.

Note that the differences in the first row (i.e. for \( G = N \)) between the normal and the \( t_\nu \) local estimator are caused by the estimation of \( \hat{\beta} \) (with an analogous method as in Audrino and Bühlmann, 2001) needed by the construction of \( \hat{f}_{x:n}^t \).

As we expect, the local log-likelihood estimators consistently outperform the classical local regression estimation for all degrees of freedom parameters \( \nu \) considered: the relative gain in the out-sample OS-L\(_2\) measure is always bigger than 60%.
5.2 Three real data examples

We consider three financial instruments with 1000 daily negative log-returns \( X_t = \log \left( \frac{P_t}{P_{t-1}} \right) \) (in percentages): from the German DAX index between January 18, 1994 and November 17, 1997; from the US DJIA index between December 11, 1995 and November 24, 1999; and from the BMW stock price between September 23, 1992 and July 23, 1996. We consider the normal local log-likelihood estimator \( \hat{f}_{x\alpha_n} \), again in comparison with a local regression estimation, and with the parametric ARCH(1) model, i.e. the function \( f(\cdot) \) in (1.1) is given by

\[
f(x) = \alpha_0 + \alpha_1 x^2,
\]

where \( \alpha_0, \alpha_1 \) are real positive parameters.

Since the OS-L\(_2\) measure introduced by (5.2) can not be calculated for real data, we measure goodness of fit with the following OS-PL\(_2\) (out-of-sample prediction loss) statistic:

\[
\text{OS-PL}_2 = \frac{1}{n-1} \sum_{t=2}^{n} \left| \hat{f}_{Q_{t-1};n} - Q_t^2 \right|^2, \quad Q_1^n \text{ a new test set},
\]

where \( Q_1^n = X_{2n+1}^n \) with \( n = 500 \). Note that the OS-PL\(_2\) criterion (and also others) for real data allows only to discriminate between volatility forecasts with performance different in large orders of magnitude. Otherwise, small differences could be obscured by low signal to noise ratio; see for example Audrino and Bühlmann (2001) for more details. The results are summarized in the following Table 5.2.

<table>
<thead>
<tr>
<th>financial instrument</th>
<th>OS-PL(_2) measure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>parametric ARCH(1)</td>
</tr>
<tr>
<td>BMW share price</td>
<td>3.80</td>
</tr>
<tr>
<td>DAX index</td>
<td>17.79</td>
</tr>
<tr>
<td>DJIA index</td>
<td>8.44</td>
</tr>
</tbody>
</table>

Table 5.2: The OS-PL\(_2\) measure for three financial instruments: the German DAX index, the US DJIA index and the BMW share price. The relative gains over the parametric ARCH(1) model are given between parenthesis.

On the contrary to the results of Table 5.1, the differences between the models of Table 5.2 are very small. For this reason, we can not reach any definitive conclusion, even if the normal local log-likelihood estimator yields better forecast results than a classical local regression, also for real data.

6 Multiplicative ARCH models with \( p > 1 \) predictor variables

Analogously to Section 1, we consider as a starting point the non-parametricARCH(p) model

\[
X_t = \sigma_t Z_t \quad (t \geq p + 1)
\]

\[
\sigma_t^2 = f(X_{t-1}, \ldots, X_{t-p}), \quad f : \mathbb{R}^p \rightarrow \mathbb{R}^+,
\]

(6.1)
where the innovations $Z_t$ are independent identically distributed, with distribution function $G$, zero mean, variance one and independent from $\{X_s; s \leq t\}$. We make the same assumptions on the process $\{X_t\}_{t \geq p+1}$ as in Section 1.

To estimate the function $f$, we proceed here as follows. We assume that the function $f$ is of the form

$$f(X_{t-1}, \ldots, X_{t-p}) = f_1(X_{t-1})f_2(X_{t-2}) \cdots f_p(X_{t-p}),$$

(6.2)
i.e. the product of $p$ functions depending on one predictor variable only. Analogous to Section 1, it is useful to transform the model (6.1)-(6.2) logarithmically. We get

$$Y_t = \beta + \sum_{i=1}^{p} g_i(X_{t-i}) + U_t,$$

(6.3)
where $Y_t$, $\beta$ and $U_t$ are defined as in (1.2), and $g_i(X_{t-i}) = \log \left(f_i(X_{t-i})\right)$ for $i = 1, \ldots, p$.

Note that (6.3) belongs to the class of the generalized additive models (GAM), see Hastie and Tibshirani (1990). For this reason we can estimate the different functions $g_i$ (and consequently the functions $f_i$ and the squared volatility $f$ of (6.1)) using a backfitting algorithm.

The goal is to minimize

$$\sum_{t=p+1}^{n} \rho(Y_t, \{g_i(X_{t-i})\}^p_t) = \sum_{t=p+1}^{n} \left(- \log \left(f_{t:t}(Y_t - \beta - \sum_{i=1}^{p} g_i(X_{t-i}))\right)\right).$$

Proceeding with a constant local log-likelihood estimation we get that for a given fitting point $x = (x_1, \ldots, x_p) \in \mathcal{C} \subset \mathbb{R}^p$

$$\hat{g}_{i:n}(x) = \arg\min_{t=p+1}^{n} \sum_{t=p+1}^{n} W \left(\frac{x_i - X_{t-i}}{h_n}\right) \cdot R_{t,i}, \ i = 1, \ldots, p,$$

(6.4)
where the residuals $R_{t,i}$ are given by $R_{t,i} = \rho(Y_t - \sum_{j \neq i} \hat{g}_{j:n}(X_{t-j}) - g_i)$.

Now, we want to show that the Remark at the end of Corollary 2.2 is still true also for $p > 1$. This is done in the next section.

### 6.1 The normal case

We assume here that the innovations $Z_t$ are standard normally distributed. From the result of Theorem 2.1, we have that the local constant log-likelihood estimator $\hat{g}_{i:n}(x)$ for the function $g_i$, $i = 1, \ldots, p$, and every fitting point $x = (x_1, \ldots, x_p)$ is given by

$$\hat{g}_{i:n}(x_i) = \log \left(\frac{\sum_{t=p+1}^{n} \left(W(\frac{x_i - X_{t-i}}{h_n}) \exp \left(Y_t - \sum_{j \neq i} \hat{g}_j(X_{t-j})\right)\right)}{\sum_{s=p+1}^{n} W(\frac{x_i - X_{s-i}}{h_n})}\right).$$

With a back transformation, we find that the estimator $\hat{f}_{i:n}(x_i)$ for the original function $f_i$ equals

$$\hat{f}_{i:n}(x_i) = \exp \left(\hat{g}_{i:n}(x_i)\right) = \frac{\sum_{t=p+1}^{n} W(\frac{x_i - X_{t-i}}{h_n}) \exp \left(Y_t - \sum_{j \neq i} \hat{g}_j(X_{t-j})\right)}{\sum_{s=p+1}^{n} W(\frac{x_i - X_{s-i}}{h_n})} \cdot \frac{X_{i}^2}{\prod_{j \neq i} f_j(X_{t-j})}, \ i = 1, \ldots, p.$$
Note that, analogous to the case where only one predictor variable is involved, multiplicative models of the form (6.1)-(6.2) with Gaussian innovations $Z_t$ can be fitted with closed form expressions.

Analogously to the remark at the end of Corollary 2.2, exactly the same result can be obtained rewriting the model (6.1) as

$$X_t^2 = \sigma_t^2 + \eta_t = f_1(X_{t-1})f_2(X_{t-2}) \cdots f_p(X_{t-p}) + \eta_t,$$

where $\eta_t = \sigma_t^2(Z_t^2 - 1)$, and estimating the functions $f_i$ with a local regression using a backfitting algorithm for the current residuals

$$\frac{X_t^2}{\prod_{j \neq i} f_j(X_{t-j})}, \ i = 1, \ldots, p.$$

### 6.2 Numerical results

We test the GAM model with normal local log-likelihood estimation of the last section on the same real financial instruments of Section 5.2, in comparison with a GAM model again given by (6.3) but with local regression estimation in (6.4) (i.e. $\rho(\cdot) = (\cdot)^2$). For quantifying the goodness of fit, we consider the same OS-PL$_2$ statistic introduced by (5.3). The results for different values of $p$ are summarized in Table 6.1.

The improvements of the GAM model with normal local log-likelihood over the GAM model with local regression are bigger than the ones of Table 5.2. On the other hand, we see that using more than one predictor in the estimation does not yield a relevant improvement in the OS-PL$_2$ statistic.

### 7 Concluding remarks

We have proposed a non-parametric local likelihood estimator for volatility in the ARCH(1) model (1.1) which leads to more accurate predictions than classical kernel regression smoothing. We have also shown how our estimation procedure can be generalized for multiplicative ARCH models with $p > 1$ predictor variables (6.1)-(6.2).

As supporting asymptotics, we have presented consistency and asymptotic normality results for our local likelihood estimator in the general situation where the distribution function of the innovations $Z_t$ may be mis-specified.

We analyze the results of our local likelihood estimator on simulated and real return time series and we confront the performance of the volatility forecasts with the ones from a classical local regression with respect to the out-of-sample loss (simulations) and out-of-sample prediction loss (real data sets). More specifically, we have the following:

- Predicting the volatility using classical local regression for the variables $Y_t = \log(X_t^2)$ versus $X_{t-1}, \ldots, X_{t-p}$ in the transformed ARCH model (1.2) and back-transforming is not a good idea.
- A better strategy is to locally regress $X_t^2$ versus $X_{t-1}, \ldots, X_{t-p}$ directly in the quadratic ARCH model. We found that this is equal to making a local likelihood estimation under Gaussian innovations in the transformed ARCH model (1.2).
- The volatility local likelihood estimator under Gaussian innovations has closed form solution.
<table>
<thead>
<tr>
<th>financial instrument</th>
<th>order</th>
<th>OS-PL₂ measure</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>GAM with local</td>
<td>GAM with normal local likelihood</td>
<td></td>
</tr>
<tr>
<td>BMW share price</td>
<td>$p = 2$</td>
<td>3.97</td>
<td>3.83 (-3.5%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p = 3$</td>
<td>3.96</td>
<td>3.94 (-0.5%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p = 4$</td>
<td>4.21</td>
<td>3.94 (-6.4%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p = 5$</td>
<td>4.74</td>
<td>4.11 (-13.3%)</td>
<td></td>
</tr>
<tr>
<td>DAX index</td>
<td>$p = 2$</td>
<td>20.62</td>
<td>18.52 (-10.2%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p = 3$</td>
<td>20.35</td>
<td>18.34 (-9.9%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p = 4$</td>
<td>20.47</td>
<td>18.31 (-10.6%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p = 5$</td>
<td>20.41</td>
<td>18.30 (-10.3%)</td>
<td></td>
</tr>
<tr>
<td>DJIA index</td>
<td>$p = 2$</td>
<td>16.70</td>
<td>8.22 (-50.8%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p = 3$</td>
<td>20.81</td>
<td>8.72 (-58.1%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p = 4$</td>
<td>17.62</td>
<td>8.95 (-49.2%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p = 5$</td>
<td>17.33</td>
<td>8.70 (-49.8%)</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.1: The OS-PL₂ measure for the same three financial instruments of Table 5.2. The relative gains over the GAM model with local regression are given in parenthesis.

- For real data, we found that the local likelihood estimator gives best (or at least equally good) volatility forecasts than the local regression one.
- For simulated data, the local likelihood estimator consistently outperforms the classical local regression one; the relative gain with respect to the out-of-sample loss (OS-L₂) statistic is always bigger than 60%.

References


