Nonparametric Estimators of GARCH Processes*

Jürgen Franke[1], Harriet Holzberger[2], Marlene Müller[3]

[1] Universität Kaiserslautern, Fachbereich Mathematik

[2] IKB Deutsche Industriebank AG

[3] Fraunhofer Institut für Techno- und Wirtschaftsmathematik (ITWM), Kaiserslautern

April 26, 2002

*This paper will appear as Chapter 17 of *Applied Quantitative Finance* edited by W. Härdle, T. Kleinow, G. Stahl, Springer-Verlag 2002. An electronic version of this book can be obtained from http://www.quantlet.de. All mentioned program code is written in XploRe available from http://www.xplore-stat.de.

1 Introduction

The generalized ARCH or GARCH model (Bollerslev, 1986) is quite popular as a basis for analyzing the risk of financial investments. Examples are the estimation of value-at-risk (VaR) or the expected shortfall from a time series of log returns. In practice, a GARCH process of order (1,1) often provides a reasonable description of the data. In the following, we restrict ourselves to that case.

We call $\{\varepsilon_t\}$ a (strong) GARCH (1,1) process if

$$\begin{aligned} \varepsilon_t &= \sigma_t Z_t \\ \sigma_t^2 &= \omega + \alpha\, \varepsilon_{t-1}^2 + \beta\, \sigma_{t-1}^2 \end{aligned} \tag{1}$$

with independent identically distributed innovations Z_t having mean 0 and variance 1. A special case is the integrated GARCH model of order (1,1) or IGARCH(1,1) model where $\alpha + \beta = 1$ and, frequently, $\omega = 0$ is assumed, i.e.

$$\sigma_t^2 = \alpha\, \varepsilon_{t-1}^2 + (1-\alpha)\, \sigma_{t-1}^2.$$

This model forms the basis for the J.P. Morgan RiskMetrics VaR analysis using exponential moving averages (Franke, Härdle and Hafner, 2001, Chapter 15). The general GARCH(1,1) process has finite variance $\sigma^2 = \omega/(1-\alpha-\beta)$ if $\alpha+\beta < 1$, and it is strictly stationary if $\mathrm{E}\{\log(\alpha\, Z_t^2 + \beta)\} < 0$. See Franke, Härdle and Hafner (2001, Chapter 12) for these and further properties of GARCH processes.

In spite of its popularity, the GARCH model has one drawback: Its symmetric dependence on past returns does not allow for including the leverage effect into the model, i.e. the frequently made observation that large negative returns of stock prices have a greater impact on volatility than large positive returns. Therefore, various parametric modifications like the exponential GARCH (EGARCH) or the threshold GARCH (TGARCH) model have been proposed to account for possible asymmetric dependence of volatility on returns. The TGARCH model, for example, introduces an additional term into the volatility equation allowing for an increased effect of negative ε_{t-1} on σ_t^2:

$$\varepsilon_t = \sigma_t Z_t\,, \quad \sigma_t^2 = \omega + \alpha\, \varepsilon_{t-1}^2 + \alpha^- \varepsilon_{t-1}^2 \cdot \mathbf{1}(\varepsilon_{t-1} < 0) + \beta\, \sigma_{t-1}^2.$$

To develop an exploratory tool which allows to study the nonlinear dependence of squared volatility σ_t^2 on past returns and volatilities we introduce a nonparametric GARCH(1,1) model

$$\begin{aligned} \varepsilon_t &= \sigma_t Z_t \\ \sigma_t^2 &= g(\varepsilon_{t-1}, \sigma_{t-1}^2) \end{aligned} \tag{2}$$

where the innovations Z_t are chosen as above. We consider a nonparametric estimator for the function g based on a particular form of local smoothing. Such an estimate may be used to decide if a particular parametric nonlinear GARCH model like the TGARCH is appropriate.

We remark that the volatility function g cannot be estimated by common kernel or local polynomial smoothers as the volatilities σ_t are not observed directly. Bühlmann and McNeil (1999) have considered an iterative algorithm. First, they fit a common parametric GARCH(1,1) model to the data from which they get sample volatilities $\widehat{\sigma}_t$ to replace the unobservable true volatilities. Then, they use a common bivariate kernel estimate to estimate g from ε_t and $\widehat{\sigma}_t^2$. Using this preliminary estimate for g they obtain new sample volatilities which are used for a further kernel estimate of g. This procedure is iterated several times until the estimate stabilizes.

Alternatively, one could try to fit a nonparametric ARCH model of high order to the data to get some first approximations $\widehat{\sigma}_t^2$ to σ_t^2 and then use a local linear estimate based on the approximate relation

$$\widehat{\sigma}_t^2 \approx g(\varepsilon_{t-1}, \widehat{\sigma}_{t-1}^2).$$

However, a complete nonparametric approach is not feasible as high-order nonparametric ARCH models based on $\sigma_t^2 = g(\varepsilon_{t-1}, \ldots, \varepsilon_{t-p})$ cannot be reliably estimated by local smoothers due to the sparseness of the data in high dimensions. Therefore, one would have to employ restrictions like additivity to the ARCH model, i.e. $\sigma_t^2 = g_1(\varepsilon_{t-1}) + \ldots + g_p(\varepsilon_{t-p})$, or even use a parametric ARCH model $\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \ldots + \alpha_p \, \varepsilon_{t-p}^2$. The alternative we consider here is a direct approach to estimating g based on deconvolution kernel estimates which does not require prior estimates $\widehat{\sigma}_t^2$.

2 Deconvolution density and regression estimates

Deconvolution kernel estimates have been described and extensively discussed in the context of estimating a probability density from independent and identically distributed data (Carroll and Hall, 1988; Stefansky and Carroll, 1990). To explain the basic idea behind this type of estimates we consider the deconvolution problem first. Let $\xi_1, \ldots, \xi_N$ be independent and identically distributed real random variables with density $p_\xi(x)$ which we want to estimate. We do not, however, observe the ξ_k directly but only with additive errors $\eta_1, \ldots, \eta_N$. Let us assume that the η_k as well are independent and identically distributed with density $p_\eta(x)$ and independent of the ξ_k. Hence, the available data are

$$X_k = \xi_k + \eta_k \ , \quad k = 1, \ldots, N.$$

To be able to identify the distribution of the ξ_k from the errors η_k at all, we have to assume that $p_\eta(x)$ is known. The density of the observations X_k is just the convolution of p_ξ with p_η:

$$p_x(x) = p_\xi(x) \star p_\eta(x) \,.$$

We can therefore try to estimate $p_x(x)$ by a common kernel estimate and extract an estimate for $p_\xi(x)$ out of it. This kind of deconvolution operation is preferably performed in the frequency domain, i.e. after applying a Fourier transform. As the subsequent inverse Fourier transform includes already a smoothing part we can start with the empirical distribution of $X_1, \ldots, X_N$ instead of a smoothed version of it. In detail, we calculate the Fourier transform or characteristic function of the empirical law of $X_1, \ldots, X_N$, i.e. the sample characteristic function

$$\widehat{\phi}_x(\omega) = \frac{1}{N} \sum_{k=1}^{N} e^{i\omega X_k} \,.$$

Let

$$\phi_\eta(\omega) = \mathrm{E}(e^{i\omega\eta_k}) = \int_{-\infty}^{\infty} e^{i\omega u} p_\eta(u)\, du$$

denote the (known) characteristic function of the η_k. Furthermore, let K be a common kernel function, i.e. a nonnegative continuous function which is symmetric around 0 and integrates up to 1: $\int K(u)\, du = 1$, and let

$$\phi_K(\omega) = \int e^{i\omega u} K(u)\, du$$

be its Fourier transform. Then, the *deconvolution kernel density estimate* of $p_\xi(x)$ is defined as

$$\widehat{p}_h(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega x} \phi_K(\omega h)\, \frac{\widehat{\phi}_x(\omega)}{\phi_\eta(\omega)}\, d\omega \,.$$

The name of this estimate is explained by the fact that it may be written equivalently as a kernel density estimate

$$\widehat{p}_h(x) = \frac{1}{Nh} \sum_{k=1}^{N} K^h\left(\frac{x - X_k}{h}\right)$$

with deconvolution kernel

$$K^h(u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega u} \frac{\phi_K(\omega)}{\phi_\eta(\omega/h)}\, d\omega$$

depending explicitly on the smoothing parameter h. Based on this kernel estimate for probability densities, Fan and Truong (1993) considered the analogous deconvolution kernel regression estimate defined as

$$\widehat{m}_h(x) = \frac{1}{Nh} \sum_{k=1}^{N} K^h\left(\frac{x - X_k}{h}\right) Y_k \;/\; \widehat{p}_h(x).$$

This Nadaraya-Watson-type estimate is consistent for the regression function $m(x)$ in an errors-in-variables regression model

$$Y_k = m(\xi_k) + W_k, \quad X_k = \xi_k + \eta_k, \quad k = 1, \dots, N,$$

where $W_1, \dots, W_N$ are independent identically distributed zero-mean random variables independent of the X_k, ξ_k, η_k which are chosen as above. The X_k, Y_k are observed, and the probability density of the η_k has to be known.

3 Nonparametric ARMA Estimates

GARCH processes are closely related to ARMA processes. If we square a GARCH (1,1) process $\{\varepsilon_t\}$ given by (1) then we get an ARMA(1,1) process

$$\varepsilon_t^2 = \omega + (\alpha + \beta)\, \varepsilon_{t-1}^2 - \beta\, \zeta_{t-1} + \zeta_t,$$

where $\zeta_t = \sigma_t^2 (Z_t^2 - 1)$ is white noise, i.e. a sequence of pairwise uncorrelated random variables, with mean 0. Therefore, we study as an intermediate step towards GARCH processes the nonparametric estimation for ARMA models which is more closely related to the errors-in-variables regression of Fan and Truong (1993). A linear ARMA(1,1) model with non-vanishing mean ω is given by

$$X_{t+1} = \omega + a\ X_t + b\ e_t + e_{t+1}$$

with zero-mean white noise e_t. We consider the nonparametric generalization of this model

$$X_{t+1} = f(X_t, e_t) + e_{t+1} \tag{3}$$

for some unknown function $f(x, u)$ which is monotone in the second argument u. Assume we have a sample $X_1, \dots, X_{N+1}$ observed from (3). If f does not depend on the second argument, (3) reduces to a nonparametric autoregression of order 1

$$X_{t+1} = f(X_t) + e_{t+1}$$

and the autoregression function $f(x)$ may be estimated by common kernel estimates or local polynomials. There exists extensive literature about that type of estimation problem, and we refer to the review paper of Härdle, Lütkepohl and Chen (1997). In the general case of (3) we again have the problem of estimating a function of (partially) non-observable variables. As f depends also on the observable time series X_t, the basic idea of constructing a nonparametric estimate of $f(x, u)$ is to combine a common kernel smoothing in the first variable x with a deconvolution kernel smoothing in the second variable u. To define the estimate we have to introduce some notation and assumptions.

We assume that the innovations e_t have a known probability density p_e with distribution function $P_e(v) = \int_{-\infty}^{v} p_e(u)\, du$ and with Fourier transform $\phi_e(\omega) \neq 0$ for all ω and

$$|\phi_e(\omega)| \geq c \cdot |\omega|^{\beta_0} \exp(-|\omega|^{\beta}/\gamma) \quad \text{for} \quad |\omega| \longrightarrow \infty$$

for some constants $c, \beta, \gamma > 0$, β_0. The nonlinear ARMA process (3) has to be stationary and strongly mixing with exponentially decaying mixing coefficients. Let $p(x)$ denote the density of the stationary marginal density of X_t.

The smoothing kernel K^x in x-direction is a common kernel function with compact support $[-1, +1]$ satisfying $0 \leq K^x(u) \leq K^x(0)$ for all u. The kernel K which is used in the deconvolution part has a Fourier transform $\phi_K(\omega)$ which is symmetric around 0, has compact support $[-1, +1]$ and satisfies some smoothness conditions (Holzberger, 2001). We have chosen a kernel with the following Fourier transform:

$$\begin{aligned}
\phi_K(u) &= 1 - u^2 && \text{for} \quad |u| \leq 0.5 \\
\phi_K(u) &= 0.75 - (|u| - 0.5) - (|u| - 0.5)^2 \\
&\quad -220\ (|u| - 0.5)^4 + 1136\ (|u| - 0.5)^5 \\
&\quad -1968\ (|u| - 0.5)^6 + 1152\ (|u| - 0.5)^7 && \text{for} \quad 0.5 \leq |u| \leq 1.
\end{aligned}$$

For convenience, we use the smoothing kernel K^x to be proportional to that function: $K^x(u) \propto \phi_K(u)$. The kernel K^x is hence an Epanechnikov kernel with modified boundaries.

Let $b = C/N^{1/5}$ be the bandwidth for smoothing in x-direction, and let $h = A/\log(N)$ be the smoothing parameter for deconvolution in u-direction where $A > \pi/2$ and $C > 0$ are some constants. Then,

$$\widehat{p}_b(x) = \frac{1}{(N+1)b} \sum_{t=1}^{N+1} K^x\left(\frac{x - X_t}{b}\right)$$

is a common Rosenblatt–Parzen density estimate for the stationary density $p(x)$.

Let $q(u)$ denote the stationary density of the random variable $f(X_t, e_t)$, and let $q(u|x)$ be its conditional density given $X_t = x$. An estimate of the latter is given by

$$\widehat{q}_{b,h}(u|x) = \frac{1}{Nhb} \sum_{t=1}^{N} K^h\left(\frac{u - X_{t+1}}{h}\right) K^x\left(\frac{x - X_t}{b}\right) \Big/ \widehat{p}_b(x) \tag{4}$$

where the deconvolution kernel K^h is

$$K^h(u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega u} \frac{\phi_K(\omega)}{\phi_e(\omega/h)}\, d\omega\,.$$

In (4) we use a deconvolution smoothing in the direction of the second argument of $f(x,u)$ using only pairs of observations (X_t, X_{t+1}) for which $|x - X_t| \le b$, i.e. $X_t \approx x$. By integration, we get the conditional distribution function of $f(X_t, e_t)$ given $X_t = x$

$$Q(v|x) = \mathrm{P}(f(x,e_t) \le v | X_t = x) = \int_{-\infty}^{v} q(u|x)\, du$$

and its estimate

$$\widehat{Q}_{b,h}(v|x) = \int_{-a_N}^{v} \widehat{q}_{b,h}(u|x) du \Big/ \int_{-a_N}^{a_N} \widehat{q}_{b,h}(u|x)\, du$$

for some $a_N \sim N^{1/6}$ for $N \to \infty$. Due to technical reasons we have to cut off the density estimate in regions where it is still unreliable for given N. The particular choice of denominator guarantees that $\widehat{Q}_{b,h}(a_N|x) = 1$ in practice, since $Q(v|x)$ is a cumulative distribution function.

To estimate the unconditional density $q(u)$ of $f(X_t, e_t) = X_{t+1} - e_{t+1}$, we use a standard deconvolution density estimate with smoothing parameter $h^* = A^*/\log(N)$

$$\widehat{q}_{h^*}(u) = \frac{1}{Nh^*} \sum_{t=1}^{N} K_{h^*}\left(\frac{u - X_t}{h^*}\right).$$

Let $p_e(u|x)$ be the conditional density of e_t given $X_t = x$, and let $P_e(v|x) = \int_{-\infty}^{v} p_e(u|x)\, du$ be the corresponding conditional distribution function. An estimate of it is given as

$$\widehat{P}_{e,h^*}(v|x) = \int_{-a_N}^{v} \widehat{q}_{h^*}(x-u)\ p_e(u) du \Big/ \int_{-a_N}^{a_N} \widehat{q}_{h^*}(x-u)\, p_e(u)\, du$$

where again we truncate at $a_N \sim N^{1/6}$.

To obtain the ARMA function f, we can now compare $Q(v|x)$ and $P_e(v|x)$. In practice this means to relate $\widehat{Q}_{b,h}(v|x)$ and $\widehat{P}_{e,h^*}(v|x)$. The nonparametric estimate for the ARMA function $f(x,v)$ depending on smoothing parameters b, h and h^* is hence given by

$$\widehat{f}_{b,h,h^*}(x,v) = \widehat{Q}_{b,h}^{-1}(\widehat{P}_{e,h^*}(v|x)\,|x)$$

if $f(x,v)$ is increasing in the second argument, and

$$\widehat{f}_{b,h,h^*}(x,v) = \widehat{Q}_{b,h}^{-1}(1 - \widehat{P}_{e,h^*}(v|x)\,|x)$$

if $f(x,v)$ is a decreasing function of v for any x. $\widehat{Q}_{b,h}^{-1}(\cdot|x)$ denotes the inverse of the function $\widehat{Q}_{b,h}(\cdot|x)$ for fixed x. Holzberger (2001) has shown that $\widehat{f}_{b,h,h^*}(x,v)$ is a consistent estimate for $f(x,v)$ under suitable assumptions and has given upper bounds on the rates of bias and variance of the estimate. We remark that the assumption of monotonicity on f is not a strong restriction. In the application to GARCH processes which we have in mind it seems to be intuitively reasonable that the volatility of today is an increasing function of the volatility of yesterday which translates into an ARMA function f which is decreasing in the second argument.

Let us illustrate the steps for estimating a nonparametric ARMA process. First we generate time series data and plot X_{t+1} versus X_t.

```
library("times")
n=1000
x=genarma(0.7,0.7,normal(n))
```

The result is shown in Figure 1. The scatterplot in the right panel of Figure 1 defines the region where we can estimate the function $f(x,v)$.

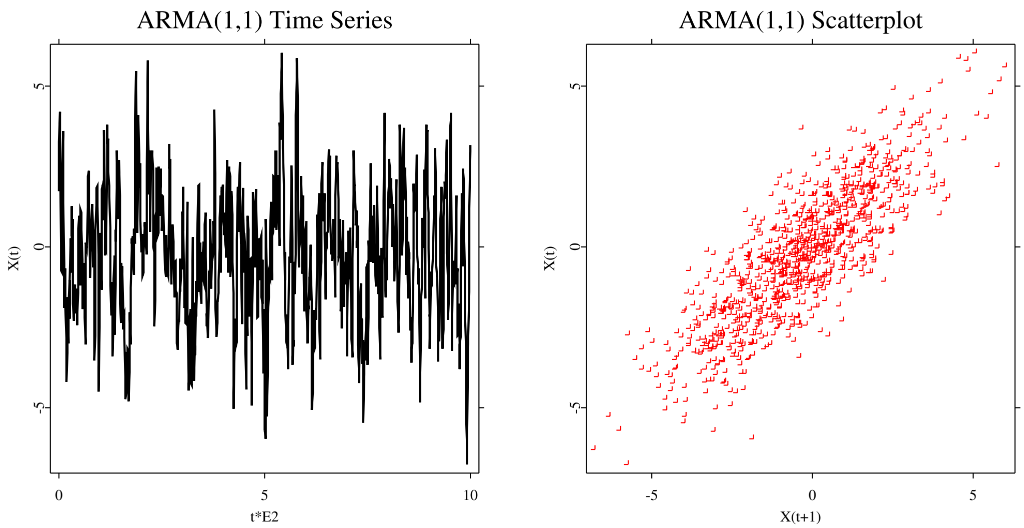

Figure 1: ARMA(1,1) process.

To compare the deconvolution density estimate with the density of $f(X_t,e_t)$ we use now our own routine (`myarma`) for generating ARMA(1,1) data from a known function (`f`):

```
proc(f)=f(x,e,c)
  f=c[1]+c[2]*x+c[3]*e
endp
```

```
proc(x,f)=myarma(n,c)
  x=matrix(n+1)-1
  f=x
  e=normal(n+1)
  t=1
  while (t<n+1)
    t=t+1
    f[t]=f(x[t-1],e[t-1],c)
    x[t]=f[t]+e[t]
  endo
  x=x[2:(n+1)]
  f=f[2:(n+1)]
endp

n=1000
{x,f}=myarma(n,0|0.7|0.7)

h=0.4
library("smoother")
dh=dcdenest(x,h)        // deconvolution estimate
fh=denest(f,3*h)        // kernel estimate
```

Figure 2 shows both density estimates. Note that the smoothing parameter (bandwidth h) is different for both estimates since different kernel functions are used.

```
f = nparmaest (x {,h {,g {,N {,R } } } } )
```

estimates a nonparametric ARMA process

The function `nparmaest` computes the function $f(x, v)$ for an ARMA process according to the algorithm described above. Let us first consider an ARMA(1,1) with $f(x, v) = 0.3 + 0.6x + 1.6v$, i.e.

$$X_t = 0.3 + 0.6X_{t-1} + 1.6e_{t-1} + e_t.$$

Hence, we use `myarma` with `c=0.3|0.6|1.6` and call the estimation routine by

```
f=nparmaest(x)
```

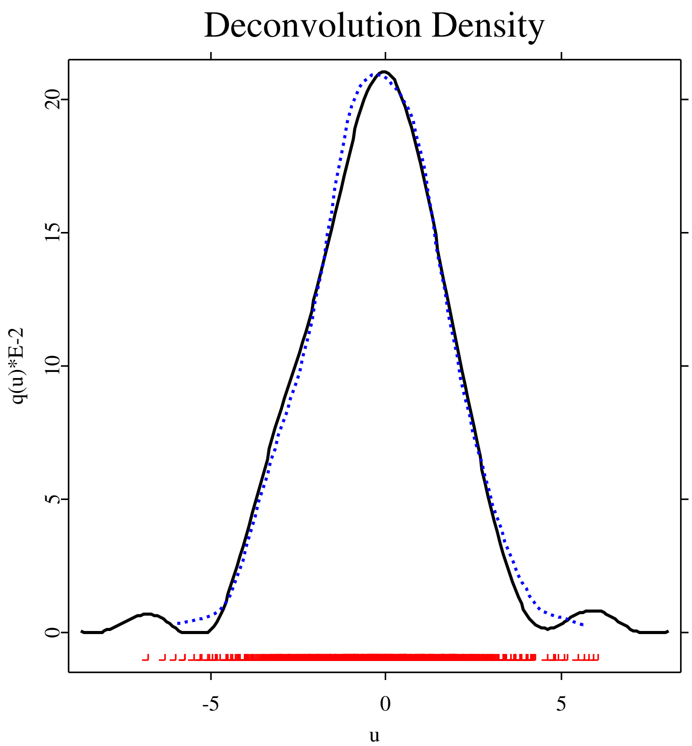

Figure 2: Deconvolution density estimate (solid) and kernel density estimate (dashed) of the known mean function of an ARMA(1,1) process.

The optional parameters `N` and `R` are set to 50 and 250, respectively. `N` contains the grid sizes used for x and v. `R` is an additional grid size for internal computations. The resulting function is therefore computed on a grid of size `N` $\times$ `N`. For comparison, we also calculate the true function on the same grid. Figure 3 shows the resulting graphs. The bandwidths `h` (corresponding to h^*) for the one-dimensional deconvolution kernel estimator $\widehat{q}$ and `g` for the two-dimensional (corresponding to h and b) are chosen according to the rates derived in Holzberger (2001).

As a second example consider an ARMA(1,1) with a truly nonlinear function $f(x,v) = -2.8 + 8F(6v)$, i.e.

$$X_t = -2.8 + 8F(6\,e_{t-1}) + e_t,$$

where F denotes the sigmoid function $F(u) = (1+e^{-u})^{-1}$ In contrast to the previous example, this function is obviously not dependent on the first argument. The code above has to be modified by using

```
proc(f)=f(x,e,c)
```

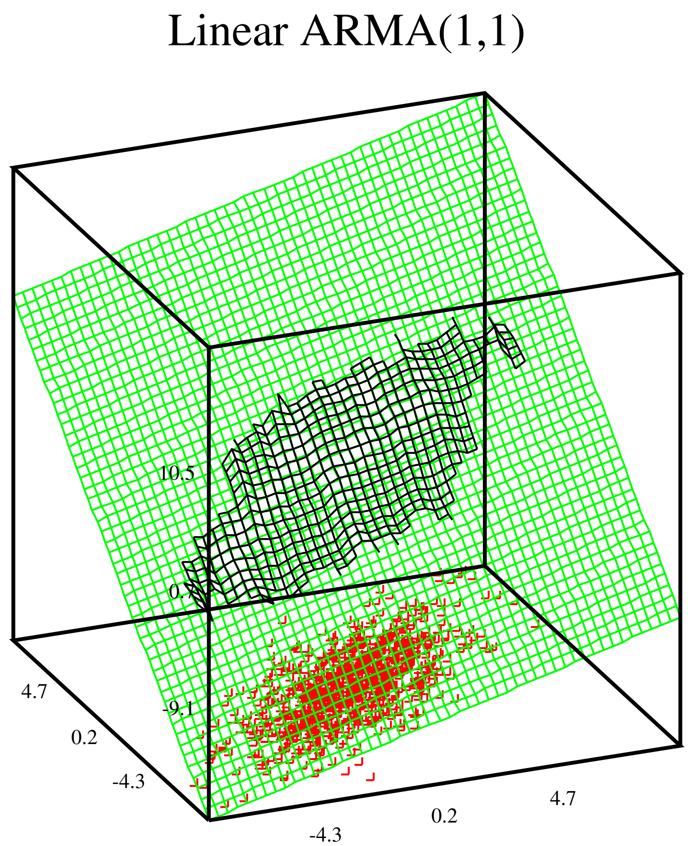

Figure 3: Nonparametric estimation of a (linear) ARMA process. True vs. estimated function and data.

```
  f=c[2]/(1+exp(-c[3]*e))+c[1]
endp
c=-2.8|8|6
```

The resulting graphs for this nonlinear function are shown in Figure 4. The estimated surface varies obviously only in the second dimension and follows the s-shaped underlying true function. However, the used sample size and the internal grid sizes of the estimation procedure do only allow for a rather imprecise reconstruction of the tails of the surface.

Nonlinear ARMA(1,1)

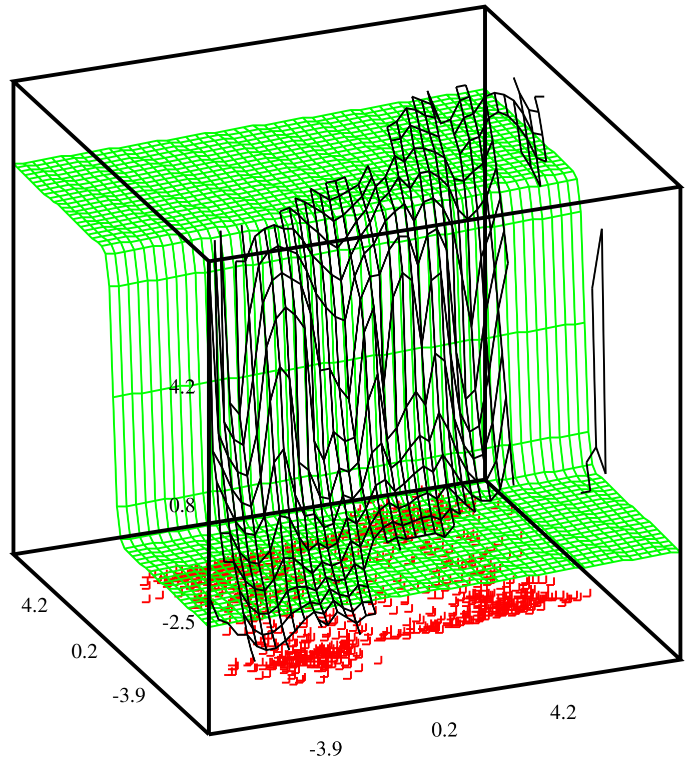

Figure 4: Nonparametric estimation of a (nonlinear) ARMA process. True vs. estimated function and data.

4 Nonparametric GARCH Estimates

In the following, we consider nonparametric GARCH(1,1) models which depend symmetrically on the last observation:

$$\begin{aligned} \varepsilon_t &= \sigma_t Z_t, \\ \sigma_t^2 &= g(\varepsilon_{t-1}^2, \sigma_{t-1}^2)\,. \end{aligned} \tag{5}$$

Here, g denotes a smooth unknown function and the innovations Z_t are chosen as in as in Section 3. This model covers the usual parametric GARCH(1,1) process (1) but does not allow for representing a leverage effect like the TGARCH(1,1) process. We show now how to transform (5) into an ARMA model. First, we define

$$X_t = \log(\varepsilon_t^2),\ e_t = \log(Z_t^2).$$

By (5), we have now

$$\begin{aligned} X_{t+1} &= \log(\varepsilon_{t+1}^2) = \log \sigma_{t+1}^2 + e_{t+1} \\ &= \log g(\varepsilon_t^2, \sigma_t^2) + e_{t+1} \\ &= \log g_1(\log(\varepsilon_t^2), \log(\sigma_t^2)) + e_{t+1} \\ &= \log g_1(X_t, X_t - e_t) + e_{t+1} \\ &= f(X_t, e_t) + e_{t+1} \end{aligned}$$

with

$$g_1(x,u) = g(e^x, e^u), \quad f(x,v) = \log g_1(x, x-v).$$

Now, we can estimate the ARMA function $f(x,v)$ from the logarithmic squared data $X_t = \log(\varepsilon_t^2)$ as in Section 4 using the nonparametric ARMA estimate $\widehat{f}_{b,h,h^*}(x,v)$ of (5). Reverting the transformations, we get

$$\widehat{g}_1(x,u) = \exp\{\widehat{f}_{b,h,h^*}(x, x-u)\},\ \widehat{g}_{b,h,h^*}(y,z) = \widehat{g}_1(\log y, \log z)$$

or, combining both equations,

$$\widehat{g}_{b,h,h^*}(y,z) = \exp\left\{\widehat{f}_{b,h,h^*}\left(\log y, \log(y/z)\right)\right\}, \quad y,z > 0,$$

as an estimate of the symmetric GARCH function $g(y,z)$.

We have to be aware, of course, that the density p_e used in the deconvolution part of estimating $f(x,v)$ is the probability density of the $e_t = \log Z_t^2$, i.e. if $p_z(z)$ denotes the density of Z_t,

$$p_e(u) = \frac{1}{2}\left\{e^{u/2} p_z(e^{u/2}) + e^{-u/2} p_z(e^{-u/2})\right\}.$$

If ε_t is a common parametric GARCH(1,1) process of form (1), then $g(y,z) = \omega + \alpha y + \beta z$, and the corresponding ARMA function is $f(x,v) = \log(\omega + \alpha e^x + \beta e^{x-v})$. This is a decreasing function in v which seems to be a reasonable assumption in the general case too corresponding to the assumption that the present volatility is an increasing function of past volatilities.

As an example, we simulate a GARCH process from

```
proc(f)=gf(x,e,c)
  f=c[1]+c[2]*x+c[3]*e
```

```
endp

proc(e,s2)=mygarch(n,c)
  e=zeros(n+1)
  f=e
  s2=e
  z=normal(n+1)
  t=1
  while (t<n+1)
    t=t+1
    s2[t]=gf(e[t-1]^2,s2[t-1]^2,c)
    e[t]=sqrt(s2[t]).*z[t]
  endo
  e=e[2:(n+1)]
  s2=s2[2:(n+1)]
endp
```

`f = npgarchest (x {,h {,g {,N {,R } } } } )` estimates a nonparametric GARCH process

The function `npgarchest` computes the functions $f(x, v)$ and $g(y, z)$ for a GARCH process using the techniques described above. Consider a GARCH(1,1) with

$$g(y, z) = 0.01 + 0.6\, y + 0.2\, z.$$

Hence, we use

```
n=1000
c=0.01|0.6|0.2
{e,s2}=mygarch(n,c)
```

and call the estimation routine by

```
g=npgarchest(e)
```

Figure 5 shows the resulting graph for the estimator of $f(x, v)$ together with the true function (decreasing in v) and the data (X_{t+1} versus X_t). As in the ARMA

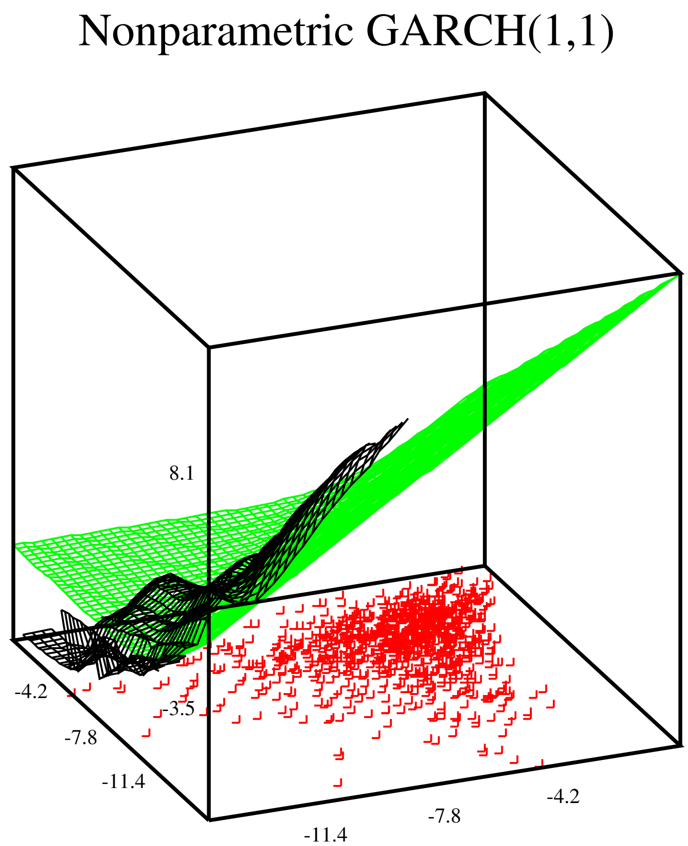

Figure 5: Nonparametric estimation of $f(x,v)$ for a (linear) GARCH process. True vs. estimated function, data $X_t = \log(\varepsilon_t^2)$.

case, the estimated function shows the underlying structure only for a part of the range of the true function.

Finally, we remark how the the general case of nonparametric GARCH models could be estimated. Consider

$$\begin{aligned} \varepsilon_t &= \sigma_t Z_t \\ \sigma_t^2 &= g(\varepsilon_{t-1}, \sigma_{t-1}^2) \end{aligned} \tag{6}$$

where σ_t^2 may depend asymmetrically on ε_{t-1}. We write

$$g(x,z) = g^+(x^2,z)\,\mathbf{1}(x \geq 0) + g^-(x^2,z)\,\mathbf{1}(x < 0).$$

As g^+, g^- depend only on the squared arguments we can estimate them as before. Again, consider $X_t = \log(\varepsilon_t^2), e_t = \log(Z_t^2)$. Let N_+ be the number of all $t \leq N$

with $\varepsilon_t \geq 0$, and $N_- = N - N_+$. Then, we set

$$\begin{aligned}
\widehat{p}_b^+(x) &= \frac{1}{N_+ b} \sum_{t=1}^{N} K^x(\frac{x - X_t}{b}) \mathbf{1}(\varepsilon_t \geq 0) \\
\widehat{q}_{b,h}^+(u|x) &= \frac{1}{N_+ hb} \sum_{t=1}^{N} K^h \left(\frac{u - X_{t+1}}{h} \right) K^x \left(\frac{x - X_t}{b} \right) \mathbf{1}(\varepsilon_t \geq 0) \ / \ \widehat{p}_b^+(x) \\
\widehat{q}_{h^*}^+(u) &= \frac{1}{N_+ h^*} \sum_{t=1}^{N} K_{h^*} \left(\frac{u - X_t}{h^*} \right) \mathbf{1}(\varepsilon_t \geq 0).
\end{aligned}$$

$\widehat{Q}_{b,h}^+(v|x)$, $\widehat{P}_{e,h^*}^+(v|x)$ are defined as in Section 3 with $\widehat{q}_{b,h}^+, \widehat{p}_b^+$ replacing $\widehat{q}_{b,h}$ and $\widehat{p}_b$, and, using both estimates of conditional distribution functions we get an ARMA function estimate $\widehat{f}_{b,h,h^*}^+(x,v)$. Reversing the transformation from GARCH to ARMA, we get as the estimate of $g^+(x^2, z)$

$$\widehat{g}_{b,h,h^*}^+(x^2, z) = \exp\left\{ \widehat{f}_{b,h,h^*}^+ \left(\log x^2, \log(x^2/z) \right) \right\}.$$

The estimate for $g^-(x^2, z)$ is analogously defined

$$\widehat{g}_{b,h,h^*}^-(x^2, z) = \exp\left\{ \widehat{f}_{b,h,h^*}^- \left(\log x^2, \log(x^2/z) \right) \right\}.$$

where, in the derivation of $\widehat{f}_{b,h,h^*}^-$, N_+ and $\mathbf{1}(\varepsilon_t \geq 0)$ are replaced by N_- and $\mathbf{1}(\varepsilon_t < 0)$.

Bibliography

Bollerslev, T.P. (1986). Generalized autoregressive conditional heteroscedasticity, *Journal of Econometrics* **31**: 307-327.

Bühlmann, P. and McNeil, A.J. (1999). *Nonparametric GARCH-models*, Manuscript, ETH Zürich, http://www.math.ethz.ch/~mcneil.

Carroll, R.J. and Hall, P. (1988). Optimal rates of convergence for deconvoluting a density, *J. Amer. Statist. Assoc.* **83**: 1184-1186.

Fan, J. and Truong, Y.K. (1993). Nonparametric regression with errors-in-variables, *Ann. Statist.* **19**: 1257-1272.

Franke, J., Härdle, W. and Hafner, Ch. (2001). *Statistik der Finanzmärkte*, Springer, ebook: http://www.quantlet.de.

Härdle, W., Lütkepohl, H. and Chen, R. (1997). A review of nonparametric time series analysis, *International Statistical Review* **65**: 49-72.

Holzberger, H. (2001). *Nonparametric Estimation of Nonlinear ARMA and GARCH-processes*, PhD Thesis, University of Kaiserslautern.

J.P. Morgan. RiskMetrics, http://www.jpmorgan.com.

Stefansky, L.A. and Carroll, R.J. (1990). Deconvoluting kernel density estimators, *Statistics* **21**: 169-184.