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

\[
\begin{align*}
\varepsilon_t &= \sigma_t Z_t \\
\sigma_t^2 &= \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2
\end{align*}
\] (17.1)

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 \( \omega + \beta = 1 \) 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 \( \text{E}\{\log(\alpha Z_t^2 + \beta)\} < 0 \). 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 $\varepsilon_{t-1}$ on $\sigma^2_t$:

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

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

$$\varepsilon_t = \sigma_t Z_t$$

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

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 $\sigma_t$ 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 $\hat{\sigma}_t$ to replace the unobservable true volatilities. Then, they use a common bivariate kernel estimate to estimate $g$ from $\varepsilon_t$ and $\hat{\sigma}_t^2$. 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 $\hat{\sigma}^2_t$ to $\sigma^2_t$ and then use a local linear estimate based on the approximate relation

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

However, a complete nonparametric approach is not feasible as high-order nonparametric ARCH models based on $\sigma^2_t = g(\varepsilon_{t-1}, \ldots, \varepsilon_{t-p})$ 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. $\sigma^2_t = g_1(\varepsilon_{t-1}) + \ldots + g_p(\varepsilon_{t-p})$, or even use a parametric ARCH model $\sigma^2_t = \omega + \alpha_1 \varepsilon^2_{t-1} + \ldots + \alpha_p \varepsilon^2_{t-p}$. The alternative we consider here is a direct approach to estimating $g$ based on deconvolution kernel estimates which does not require prior estimates $\hat{\sigma}^2_t$. 