SCHOOL OF ECONOMICS AND MANAGEMENT
FACULTY OF SOCIAL SCIENCES
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2011-2

Nonlinear models for autoregressive conditional heteroskedasticity

Timo Teräsvirta

School of Economics and Management
Aarhus University
Bartholins Allé 10, Building 1322, DK-8000 Aarhus C
Denmark

Nonlinear models for autoregressive conditional heteroskedasticity

Timo Teräsvirta
CREATES, Aarhus University
Building 1322, DK-8000 Aarhus C, Denmark

January 5, 2011

Abstract

This paper contains a brief survey of nonlinear models of autoregressive conditional heteroskedasticity. The models in question are parametric nonlinear extensions of the original model by Engle (1982). After presenting the individual models, linearity testing and parameter estimation are discussed. Forecasting volatility with nonlinear models is considered. Finally, parametric nonlinear models based on multiplicative decomposition of the variance receive attention.

JEL Classification Codes: C22; C52

Keywords: nonlinear ARCH, nonlinear GARCH, neural network, nonlinear volatility, smooth transition GARCH, threshold GARCH

Acknowledgements: This research was supported by the Danish National Reseach Foundation. A part of the work was carried out when the author was visiting Department of Economics of the European University Institute, Florence, whose kind hospitality is gratefully acknowledged. The author wishes to thank Christian Hafner for useful comments but retains responsibility for any errors or shortcomings in this work.

Note: The paper has been prepared for Luc Bauwens, Christian Hafner and Sébastien Laurent, eds. *Wiley Handbook in Financial Engineering and Econometrics: Volatility Models and Their Applications*: New York: Wiley.

1 Introduction

One of the first occasions for at least European econometricians to hear about autoregressive conditional heteroskedasticity (ARCH) was at the 1979 Econometric Society European Meeting in Athens, where Robert Engle presented a paper with that title in a session named *Time Series II: Specification*, chaired by David Hendry. The published version of the paper (Engle, 1982) appeared three years later. The number of models for describing and forecasting autoregressive conditional heteroskedasticity has increased dramatically from those early days, as is the number of applications. Recent overviews include Bauwens, Hafner and Laurent (2011) and Teräsvirta (2009). Several chapters of Andersen, Davis, Kreiss and Mikosch (2009) are devoted to various theoretical and computational aspects of these models.

The focus of this chapter is on univariate nonlinear generalized ARCH (GARCH) models. There are examples of strongly nonlinear multivariate GARCH models, but they will not be discussed here; for recent surveys see Bauwens, Laurent and Rombouts (2006) and Silvennoinen and Teräsvirta (2009). The definition of a nonlinear GARCH model adopted in this overview is quite narrow and only covers parametric nonlinear models. The Exponential GARCH model of Nelson (1991) is excluded from the consideration, however, because the logarithm of the conditional variance is linear in parameters. The hidden Markov or Markov-switching models, such as the variance-switching model of Rydén, Teräsvirta and Åsbrink (1998) or its more complicated ARCH and GARCH variants, that can be viewed as nonlinear, are discussed in Paolella and Haas (2011) and will not be taken up here. The models to be considered include the smooth transition GARCH, the double threshold GARCH, the asymmetric power GARCH, and a particular class of time-varying GARCH models. The artificial neural network GARCH model will be mentioned as well, although the number of applications of the model appears to be very small.

A general treatment of volatility models based on multiplicative decomposition of the variance can be found in Van Bellegem (2011). This chapter contains a brief discussion of fully parametric variants of such models. A majority of them are semiparametric and are therefore not covered by this account.

The plan of the chapter is as follows. Sections 2 and 3 contain short presentations of the standard GARCH model and linear predecessors to nonlinear GARCH models. Smooth transition GARCH and related models are considered in Section 4. Testing linearity against various nonlinear GARCH models is discussed in Section 5 and parameter estimation in Section 6. Forecasting with nonlinear GARCH models is the topic of Section 7. Parametric special cases of models based on multiplicative decomposition of the variance are considered in Section 8. Section 9 contains final remarks.

2 The standard GARCH model

Assume a random variable y_t can be modelled as follows: $y_t = \mu_t + \varepsilon_t$, where $\mu_t = \mathsf{E}\{y_t|\mathcal{F}_{t-1}\}$, $\mathcal{F}_{t-1} = \{y_{t-j}, j \geq 1\}$ being the information set available at $t-1$, and ε_t is a random error with mean zero and variance σ^2. In the standard GARCH(p,q) model as defined by Bollerslev (1986) the error term is parameterised as follows:

$$\varepsilon_t = z_t h_t^{1/2} \tag{1}$$

with $z_t \sim \text{iid}(0,1)$, and the positive-valued conditional variance function is

$$h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{2}$$

where $\alpha_0 > 0$. It is seen from (1) that $\mathsf{E}\varepsilon_t\varepsilon_{t-j} = 0$ for $j \neq 0$. In what follows, we assume that the conditional mean $\mu_t \equiv 0$. Sufficient conditions for h_t being positive almost surely include $\alpha_j \geq 0$, $\beta_j \geq 0$, $j = 1, ..., q$. Identification requires that at least one $\alpha_j > 0$ for $j > 0$. These conditions are not necessary, however, unless $p = q = 1$, see Nelson and Cao (1992). The unconditional variance

$$\mathsf{E}\varepsilon_t^2 = \alpha_0(1 - \sum_{j=1}^{q} \alpha_j - \sum_{j=1}^{p} \beta_j)^{-1}$$

provided the weak stationarity condition $\sum_{j=1}^{q} \alpha_j + \sum_{j=1}^{p} \beta_j < 1$ holds.

The GARCH model, defined by (1) and (2), is linear in parameters. It has been generalised in many ways to accommodate things such as regime switches, asymmetries and the like. A number of these extensions to the standard GARCH model are described in Bauwens et al. (2011) and Teräsvirta (2009). The next section will look at ones that may be regarded as predecessors to parametric nonlinear GARCH models.

3 Predecessors to nonlinear GARCH models

Predecessors to nonlinear GARCH models, as the concept is defined here, have the property that they are linear in parameters but could be made nonlinear by assuming a certain known quantity in them to be an unknown parameter. The most frequently applied models of this kind are the GJR-GARCH model by Glosten, Jagannathan and Runkle (1993) and the Threshold GARCH (TGARCH) model by Rabemananjara and Zakoïan (1993) and Zakoïan (1994). In applications, the GJR-GARCH model is typically assumed to be a first-order GARCH model. It can be generalised to have higher-order lags although in practice this almost never seems to happen. The conditional variance of this model has the following representation:

$$h_t = \alpha_0 + \sum_{j=1}^{q} \{\alpha_j + \kappa_j I(\varepsilon_{t-j} < 0)\}\varepsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{3}$$

where $I(A)$ is an indicator variable: $I(A) = 1$ when A holds, and zero otherwise. The idea of this model is to capture the leverage effect present in stock return series. This effect manifests itself as an asymmetry: a negative shock has a greater impact on the conditional variance than the positive one with the same absolute value. This implies a positive value for κ_j. As already mentioned, in applications $p = q = 1$, so one would expect that the estimate $\widehat{\kappa}_1 > 0$.

The GJR-GARCH model can be generalised by extending the asymmetry in (3) to the other components of the model. The Volatility-Switching GARCH (VS-GARCH) model of Fornari and Mele (1997) is such an extension. The first-order VS-GARCH model is defined as follows:

$$h_t = \alpha_0 + \psi_0 \mathrm{sgn}(\varepsilon_{t-1}) + \{\alpha_1 + \psi_1 \mathrm{sgn}(\varepsilon_{t-1})\}\varepsilon_{t-1}^2 + \{\beta_1 + \psi_2 \mathrm{sgn}(\varepsilon_{t-1})\}h_{t-1} \tag{4}$$

where sgn is the sign operator: $\mathrm{sgn}(x) = 1$ for $x > 0$, $\mathrm{sgn}(x) = 0$ for $x = 0$, and $\mathrm{sgn}(x) = 0$ for $x < 0$. It is seen that by setting $\psi_0 = \psi_2 = 0$ in (4) one obtains a model that is equivalent to the first-order GJR-GARCH model.

The TGARCH model is similar to (3) with one important difference: what is being modelled is the conditional standard deviation and not the conditional variance. The model is defined by replacing h_t in (3) by its square root and each ε_{t-j}^2 by the corresponding absolute value $|\varepsilon_{t-j}|$. Some authors muddle the distinction between these two models by applying the term TGARCH model to the GJR-GARCH model.

Engle and Ng (1993) suggested the following (first-order) generalisation of the (first-order) GJR-GARCH model (3):

$$\begin{aligned} h_t &= \alpha_0 + \sum_{i=0}^{m^+} \kappa_j^+ I(\varepsilon_{t-1} > \tau_i)(\varepsilon_{t-1} - \tau_i) + \sum_{i=0}^{m^-} \kappa_j^- I(\varepsilon_{t-1} < \tau_{-i})(\varepsilon_{t-1} - \tau_{-i}) \\ &\quad + \beta_1 h_{t-1} \end{aligned} \tag{5}$$

where $\tau_i = i\sigma$, and σ is the unconditional standard deviation of ε_t. This is essentially a spline model, where τ_i, $i = m^-, m^- + 1, ..., m^+$, are the knots. There is another difference: ε_{t-1} does not appear squared in (5). As a consequence, even when the simplest case $m^+ = m^- = 0$ is considered, it seems that one cannot derive a closed form for the unconditional variance of ε_t^2, if it exists. This is due to the fact that an analytical expression for $\mathsf{E}h_t^{1/2}$ is not available. The necessary and sufficient positivity conditions for h_t in this case are $\alpha_0 > 0, \kappa_0^- < 0$, $\kappa_0^+ > 0$, and $\beta_1 \geq 0$. For the general model (5), the corresponding conditions are $\alpha_0 > 0$, $\sum_{i=0}^{n^-} \kappa_j^- < 0$, $n^- = 0, ..., m^-$, $\sum_{i=0}^{n^+} \kappa_j^+ > 0$, $n^+ = 0, ..., m^+$, and $\beta_1 \geq 0$. They are not satisfied by the estimated model of Engle and Ng (1993, Table VII) where $m^- = m^+ = 4$, due to the estimates of parameters κ_3^+ and κ_3^-. The splines make the parameterisation more flexible than that of (3), but the latter does not have the discontinuities present in the former.

Finally, it may be mentioned that there exists a threshold autoregressive stochastic volatility model in which the threshold parameter, analogously to (3), equals zero. It has been defined and applied in So, Li and Lam (2002). For overviews of stochastic volatility models, see Shephard and Andersen (2009) and Bos (2011).

4 Nonlinear ARCH and GARCH models

4.1 Engle's nonlinear GARCH model

Let $g(\varepsilon_{t-1}; \boldsymbol{\theta})$ be a generic function representing the function of ε_{t-1} in (3), (8), (5) and in the TGARCH model. For all these models, $g(\varepsilon_{t-1}; \boldsymbol{\theta}) = 0$, that is, $\varepsilon_t = 0$ does not contribute to the conditional variance at $t + 1$. This is no longer true for the nonlinear GARCH model of Engle (1990). The conditional variance of this model has the following (first-order) form:

$$h_t = \alpha_0 + \alpha_1(\varepsilon_{t-1} - \lambda)^2 + \beta_1 h_{t-1} \tag{6}$$

where $\alpha_0, \alpha_1 > 0$ and $\beta_1 \geq 0$. Thus, $g(0; \boldsymbol{\theta}) = \alpha_1\lambda^2$. When $\lambda = 0$, (6) collapses into the standard GARCH(1,1) model (2). These models share the same weak stationarity condition $\alpha_1 + \beta_1 < 1$, and for (6), $\mathsf{E}\varepsilon_t^2 = (\alpha_1 + \lambda^2)/(1 - \alpha_1 - \beta_1)$. Although (6) defines a rather simple nonlinear GARCH model, it has been less popular among users than, say, the GJR-GARCH model.

4.2 Nonlinear ARCH model

Higgins and Bera (1992) introduced a nonlinear ARCH model (NLARCH) that nests both the standard ARCH model and the logarithmic GARCH model. It is an ARCH model with Box-Cox transformed variables:

$$\frac{h_t^\delta - 1}{\delta} = \alpha_0 \frac{\omega^\delta - 1}{\delta} + \alpha_1 \frac{\varepsilon_{t-1}^{2\delta} - 1}{\delta} + ... + \alpha_q \frac{\varepsilon_{t-q}^{2\delta} - 1}{\delta} \tag{7}$$

where $0 \leq \delta \leq 1$, $\omega > 0$, $\alpha_0 > 0$, $\alpha_j \geq 0$, $j = 1, ..., q$, and $\sum_{j=0}^{q} \alpha_j = 1$. It can also be written as follows:

$$h_t = \{\alpha_0 \omega^\delta + \alpha_1 \varepsilon_{t-1}^{2\delta} + ... + \alpha_q \varepsilon_{t-q}^{2\delta}\}^{1/\delta}.$$

When $\delta \to 1$, (7) approaches Engle's ARCH(q) model. The purpose of the restriction $\sum_{j=0}^{q} \alpha_j = 1$ that also helps identify ω becomes obvious from this special case. When $\delta \to 0$, the result is the qth order logarithmic ARCH model. As the GARCH family of models have become a normal specification in applications to financial time series, the NLARCH model has been rarely used in practice.

4.3 Asymmetric Power GARCH model

Ding, Granger and Engle (1993) introduced the Asymmetric Power GARCH (APGARCH) model. The first-order APGARCH model has the following definition:

$$h_t^\delta = \alpha_0 + \alpha_1 (|\varepsilon_{t-1}| - \lambda \varepsilon_{t-1})^{2\delta} + \beta_1 h_{t-1}^\delta \tag{8}$$

where $\alpha_0 > 0$, $\alpha_1 > 0$, $\beta_1 \geq 0$, $\delta > 0$ and $|\lambda| \leq 1$, so it is nonlinear in parameters. Meitz and Saikkonen (in press) considered the special case $\delta = 1$ and called the model the Asymmetric GARCH model. Using the indicator variable they showed that for $\delta = 1$ (8) can be written as a GJR-GARCH(1,1) model with $p = q = 1$:

$$h_t = \alpha_0 + \alpha_1 (1 - \lambda)^2 \varepsilon_{t-1}^2 + 4\lambda \alpha_1 I(\varepsilon_{t-1} < 0) \varepsilon_{t-1}^2 + \beta_1 h_{t-1}. \tag{9}$$

This implies that the conditions for the existence of unconditional moments of the AGARCH model can be obtained from the corresponding conditions for the first-order GJR-GARCH model. For the general APGARCH model, the only analytic 'fourth-moment' condition available is for the fractional moment $\mathsf{E}|\varepsilon_t|^{4\delta}$, see He and Teräsvirta (1999b). Ding et al. (1993) also discussed some stylized facts of return series. Considering a number of long daily return series they found that the autocorrelations $\rho(|\varepsilon_t|^{2\delta}, |\varepsilon_{t-j}|^{2\delta})$ were maximized for $\delta = 1/2$. Fitting the APGARCH model to a long daily S&P 500 return series yielded (in our notation) $\widehat{\delta} = 0.72$. He, Malmsten and Teräsvirta (2008) fitted the model to daily return series of the 30 most actively traded stocks in the Stockholm Stock Exchange and in all cases obtained an estimate of δ that was remarkably close to the value reported by Ding, Granger and Engle. For similar results, see Brooks, Faff, McKenzie and Mitchell (2000).

4.4 Smooth transition GARCH model

As mentioned in Section 3, the GJR-GARCH model may be generalised by making it nonlinear in parameters. This is done for example substituting an unknown parameter for the zero in the argument of the indicator function of (3). The same substitution may be made for the TGARCH model. This generalisation can be further extended by replacing the indicator function by a continuous function of its argument and extending the transition to also include the intercept. In the (p, q) case this yields the following conditional variance

$$h_t = \alpha_{10} + \sum_{j=1}^{q} \alpha_{1j}\varepsilon_{t-j}^2 + (\alpha_{20} + \sum_{j=1}^{q} \alpha_{2j}\varepsilon_{t-j}^2)G(\gamma, \mathbf{c}; \varepsilon_{t-j}) + \sum_{j=1}^{p} \beta_j h_{t-j} \quad (10)$$

where the transition function

$$G_K(\gamma, \mathbf{c}; \varepsilon_{t-j}) = (1 + \exp\{-\gamma \prod_{k=1}^{K} (\varepsilon_{t-j} - c_k)\})^{-1}, \gamma > 0 \quad (11)$$

and $\mathbf{c} = (c_1, ..., c_K)'$. For $K = 1$, this model corresponds to the one Hagerud (1997) introduced. The restriction $\gamma > 0$ is an identification restriction. Furthermore, global identification requires another restriction such as $c_1 \leq ... \leq c_K$, but this restriction does not have any practical significance in the estimation of parameters. The transition function is bounded between zero and one. The parameter γ represents the slope and $\mathbf{c}$ the location of the transition(s). When $K = 1$, the 'ARCH parameters' and the intercept of the model change from α_{1j} to $\alpha_{1j} + \alpha_{2j}$ as a function of ε_{t-j}, $j = 0, 1, ..., q$. This means that the impact of each shock on the conditional variance is nonlinear. When $K = 1$, $\gamma \to \infty$ and $c_1 = 0$, (10) becomes the GJR-GARCH model (3). When $K = 2$, the transition function (11) is nonmonotonic and symmetric around $(c_1 + c_2)/2$. The model (10) could be called an additive Smooth Transition GARCH (STGARCH) model. Hagerud's choice corresponding to $K = 2$ was the exponential transition function

$$G^E(\gamma, c; \varepsilon_{t-j}) = 1 - \exp\{-\gamma\varepsilon_{t-j}^2\}. \quad (12)$$

The limiting behaviour of (12) when $\gamma \to \infty$ is different from that of (11). When $\gamma \to \infty$ in (12), $G^E(\gamma, \mathbf{c}; \varepsilon_{t-j}) = 1$ except for $\varepsilon_{t-j} = 0$, where the function equals zero. As to (11), the function equals zero for $c_1 \leq \varepsilon_{t-j} < c_2$ and is one otherwise. At least for $q = 1$, (11) may be preferred to (12), because applying (12), the STGARCH model (for all practical purposes) collapses into the standard GARCH(1,1) model when $\gamma \to \infty$. The transition function (11) is symmetric around zero if $c_1 = -c_2$. Function G^E is symmetric around zero but can easily be made asymmetric by adding a location parameter to the exponent as in Teräsvirta (1994). Lubrano (2001) considered this extension in the STGARCH context.

In the model by Gonzalez-Rivera (1998) the intercept does not switch, and the same lag of ε_t controls the transition:

$$h_t = \alpha_{10} + \sum_{j=1}^{q} \alpha_{1j}\varepsilon_{t-j}^2 + (\sum_{j=1}^{q} \alpha_{2j}\varepsilon_{t-j}^2)G(\gamma, c_1; \varepsilon_{t-d}) + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{13}$$

where $\delta > 0$. It is also assumed that $c_1 = 0$, but that is not crucial. Anderson, Nam and Vahid (1999) extended (13) by also allowing the conditional variance to switch according to the same transition function as in (13).

Smooth transition GARCH models are useful in situations where the assumption of two distinct regimes is too rough an approximation to the asymmetric behaviour of the conditional variance. Among other things, Hagerud (1997) discussed a specification strategy that allows the investigator to choose between $K = 1$ and $K = 2$ in (11). Larger values of K may also be considered, but they are likely to be less common in applications than the two simplest choices.

The standard GARCH model has the undesirable property that the estimated model often exaggerates the persistence in volatility. This means that the estimated sum of the α- and β-coefficients in (2) is close to one. Overestimated persistence results in poor volatility forecasts in the sense that following a large shock, the forecasts indicate 'too slow' a decrease of the conditional variance to more normal levels. In order to remedy this problem, Lanne and Saikkonen (2005) proposed a smooth transition GARCH model whose first-order version has the form

$$h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \delta_1 G_1(\boldsymbol{\theta}; h_{t-1}) + \beta_1 h_{t-1}. \tag{14}$$

In (14), $G_1(\boldsymbol{\theta}; h_{t-1})$ is a continuous, monotonically increasing bounded function of h_{t-1}. Since $h_{t-1} > 0$ almost surely, Lanne and Saikkonen used the cumulative distribution function of the gamma-distribution as the transition function. A major difference between (13) and (14) is that in the latter model the transition variable is a lagged conditional variance. In empirical examples given in the paper, this parameterization clearly alleviates the problem of exaggerated persistence.

Lanne and Saikkonen (2005) also considered a general family of models of the conditional variance of order $(p, 1)$:

$$h_t = g(h_{t-1}, ..., h_{t-p}) + f(\varepsilon_{t-1}) \tag{15}$$

where the functions g and f are defined in such a way that (15) contains as special cases all the GARCH$(p, 1)$ models discussed in this chapter, including the smooth transition GARCH model (13) and (6). They find conditions for geometric ergodicity of h_t defined by (15) and for the existence of moments of ε_t. These conditions are quite general in the sense that unlike the other conditions available in the literature, they cover the smooth transition GARCH models as well. For even more general results on first-order models, see Meitz and Saikkonen (2008).

4.5 The double threshold ARCH model

The TGARCH model is linear in parameters, because the threshold parameter appearing in nonlinear threshold models is assumed to equal zero. A genuine nonlinear threshold model does exist, namely the Double Threshold ARCH (DTARCH) model of Li and Li (1996). It is called a double threshold model, because both the autoregressive conditional mean and the conditional variance have a threshold-type structure. The conditional mean model is defined as follows:

$$y_t = \sum_{k=1}^{K}(\phi_{0k} + \sum_{j=1}^{p_k} \phi_{jk} y_{t-j}) I(c_{k-1}^{(\mathrm{m})} < y_{t-b} \leq c_k^{(\mathrm{m})}) + \varepsilon_t \tag{16}$$

and the conditional variance has the form

$$h_t = \sum_{\ell=1}^{L}(\alpha_{0\ell} + \sum_{j=1}^{p_\ell} \alpha_{j\ell}\varepsilon_{t-j}^2) I(c_{\ell-1}^{(\mathrm{v})} < y_{t-d} \leq c_\ell^{(\mathrm{v})}). \tag{17}$$

where $c_0^{(\mathrm{m})} = c_0^{(\mathrm{v})} = -\infty$, and $c_K^{(\mathrm{m})} = c_L^{(\mathrm{v})} = \infty$. Furthermore, b and d are delay parameters, $b, d \geq 1$. The number of regimes in (16) and (17), K and L, respectively, need not be the same, and the two delay parameters need not be equal either. Other threshold variables than lags of y_t are possible. The conditional variance model (17) differs from a typical GARCH model in the sense that the transition variable in (17) is y_{t-d} and not a function of

ε_{t-d}. This feature is somewhat analogous to the Threshold Moving Average (TMA) model, see Ling and Tong (2005) and Ling, Tong and Li (2007). The TMA model is a generalisation of the standard moving average model for the conditional mean, but the threshold variable is a lag of y_t, not ε_t.

4.6 Neural network ARCH and GARCH models

The literature on nonlinear GARCH models also comprises models based on artificial neural network (ANN) type of specifications. As an example, consider the following model by Donaldson and Kamstra (1997). It is an extension of the GJR-GARCH model, but for simplicity it is here given without the asymmetric component as an extension to the standard GARCH model. The ANN-GARCH model of the authors has the following form:

$$h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} + \sum_{j=1}^{s} \phi_j G(\mathbf{w}_{t-j}, \mathbf{\Gamma}_j)$$

where the 'hidden units' are defined as follows:

$$G(\mathbf{w}_{t-j}, \mathbf{\Gamma}_j) = (1 + \exp\{\gamma_{0j} + \sum_{i=1}^{u} (\mathbf{w}'_{t-j} \boldsymbol{\gamma}_{ji})\})^{-1} \tag{18}$$

$j = 1, ..., s$. In (18), γ_{0j} and $\mathbf{\Gamma}_j = (\boldsymbol{\gamma}_{j1} : ... : \boldsymbol{\gamma}_{ju})$ are parameters such that each $m \times 1$ vector $\boldsymbol{\gamma}_{ji} = (\gamma_{j1}, ..., \gamma_{ji}, 0, ..., 0)'$, $i = 1, ..., m$, and $\mathbf{w}_t = (w_t, w_t^2, ..., w_t^m)'$ with $w_t = \varepsilon_t/\sigma$. Note that $G(\mathbf{w}_t, \mathbf{\Gamma}_j)$ as a logistic function is bounded between zero and one. Each standardised and lagged ε_t appears in powers up to m. For a user of this model, specification of p, q, s and u is an important issue, and the authors suggest the use of BIC of Rissanen (1978) and Schwarz (1978) for this purpose. More details about this can be found in Section 6.

A simpler ANN-GARCH model can be obtained by defining the hidden unit as in Caulet and Péguin-Feissolle (2000). This results in the following ANN-GARCH model:

$$h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} + \sum_{j=1}^{s} \phi_j G(\gamma_{0j} + \boldsymbol{\varepsilon}'_t \boldsymbol{\gamma}_j) \tag{19}$$

where

$$G(\gamma_{0j} + \boldsymbol{\varepsilon}'_t \boldsymbol{\gamma}_j) = (1 + \exp\{\gamma_{0j} + \boldsymbol{\varepsilon}'_t \boldsymbol{\gamma}_j)\})^{-1} \tag{20}$$

$j = 1, ..., s$, with the $k \times 1$ parameter vector $\boldsymbol{\gamma}_j$ and $\boldsymbol{\varepsilon}_t = (\varepsilon_{t-1}, ..., \varepsilon_{t-k})'$. In fact, Caulet and Péguin-Feissolle (2000) assumed $\alpha_j = 0$, $j = 1, ..., q$, and $\beta_j = 0,\ j = 1, ..., p$, in (19), because their purpose was to set up a test of no conditional heteroskedasticity against conditional heteroskedasticity of rather general form. Nevertheless, their specification can be generalised to an ANN-GARCH model (19) with (20). Since (20) is a positive-valued function, assuming $\phi_j \geq 0,\ j = 1, ...s$, would, jointly with the restrictions $\alpha_0 > 0,\ \alpha_i, \beta_j \geq 0,\ \forall i, j$, guarantee positivity of the conditional variance. Because of the positivity of (20), one could even think of deleting the linear combination $\sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2$ from the model altogether.

4.7 Time-varying GARCH

It has been argued, see for example Mikosch and Starica (2004), that the assumption of the standard GARCH model having constant parameters may not hold in practice unless the series to be modelled are sufficiently short. The standard model may be generalised by assuming that the parameters change at specific points of time, divide the series into subseries according to the location of the break-points, and fit separate GARCH models to the subseries. The main statistical problem is then finding the number of the unknown break-points and their location. It is also possible to model the switching standard deviation regimes using the threshold GARCH model. This is done by assuming that the threshold variable is the time. A recent survey by Andreou and Ghysels (2009) expertly covers this area of research.

Another possibility is to consider the smooth transition GARCH model (13) to fit this situation. It is done by assuming the transition function in (13) to be a function of time:

$$G(\gamma, \mathbf{c}; t^*) = (1 + \exp\{-\gamma \prod_{k=1}^{K} (t^* - c_k)\})^{-1}, \gamma > 0$$

where $t^* = t/T$ is rescaled time (T is the number of observations). The resulting time-varying parameter GARCH (TV-GARCH) model has the form

$$h_t = \alpha_0(t) + \sum_{j=1}^{q} \alpha_j(t) \varepsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_j(t) h_{t-j} \quad (21)$$

where $\alpha_0(t) = \alpha_{01} + \alpha_{02} G(\gamma, c; t^*)$, $\alpha_j(t) = \alpha_{j1} + \alpha_{j2} G(\gamma, c; t^*), j = 1, ..., q$, and $\beta_j(t) = \beta_{j1} + \beta_{j2} G(\gamma, c; t^*), j = 1, ..., p$. This is quite a flexible parameterization. The TV-GARCH model is nonstationary as the unconditional variance of ε_t varies deterministically over time.

Some of the time-varying parameters in (21) may be restricted to constants *a priori*. For example, it may be assumed that only the intercept

$\alpha_0(t)$ is time-varying. This implies that while the unconditional variance is changing over time, the dynamic behaviour of volatility remains unchanged. If change is allowed in the other GARCH parameters as well, the model is capable of describing systematic changes in the amplitude of the volatility clusters. The standard weakly stationary constant-parameter GARCH model cannot accommodate such changes. For a general survey of time-varying ARCH and GARCH models, see Cizek and Spokoiny (2009).

4.8 Probabilistic properties for families of GARCH models

Some authors have defined families of GARCH models, often with the purpose of proving results on probabilistic properties of GARCH models for the whole family. The corresponding results for individual GARCH models in the family then follow as special cases. A rather general family of first-order GARCH models is defined by Duan (1997). The general process is called the augmented GARCH(1,1) process, and it is partly based on the Box-Cox transformation of the variables of the GARCH model. It accommodates several models presented in Sections 3 and 4. These include the TGARCH, the GJR-GARCH and the nonlinear GARCH model by Engle (1990). Duan (1997) derived conditions for strict stationarity for the augmented GARCH model, and so they apply to these three special cases.

Hentschel (1995) has introduced another family of first-order GARCH models that is less general than that of Duan (1997). Nevertheless, it also nests the three aforementioned models and the NARCH(1) model or its GARCH generalisation. The idea is to define a general model, estimate its parameters and consider the adequacy of nested special cases by the likelihood ratio test. It appears, however, that this specification strategy has not been often used in practice.

He and Teräsvirta (1999a) define their family of first-order GARCH processes in order to obtain general results of weak stationarity, existence of the fourth moment, and the autocorrelation function of $\{\varepsilon_t^2\}$ of the processes belonging to this family. Since in that work the integer moments of GARCH processes are the object of interest, the family is restricted to models of $h_t^{1/2}$ and h_t. The more general case h_t^δ is excluded because there do not exist analytic expressions for the integer moments of ε_t except for the two special cases $\delta = 1/2$ and $\delta = 1$. This family contains as special cases, among others, the TGARCH, the GJR-GARCH and the VS-GARCH model.

Meitz and Saikkonen (2008) consider a very general Markov chain with one observable and another unobservable process that contains a rather general family of first-order GARCH model as a special case. For example, in the standard GARCH(1,1) process (1) and (2), $\{\varepsilon_t\}$ is the observable process and $\{h_t\}$ is the unobservable one. The joint sequence $\{(\varepsilon_t, h_t)\}$ is also a Markov chain. Using the theory of Markov chains, the authors derive conditions for geometric ergodicity and β-mixing for a large number of GARCH(1,1) models. These include the Hentschel (1995) family of GARCH models but also a STGARCH model that nests both the model of Hagerud (1997) and the one considered by Lanne and Saikkonen (2005). The first-order GARCH version of the TARCH model of Li and Li (1996), $p_\ell = 1$ for $\ell = 1, ..., L$ in (17), is another special case.

5 Testing standard GARCH against nonlinear GARCH

5.1 Size and sign bias tests

After a GARCH model, for example the standard GARCH model, has been estimated, it would be wise to subject it to misspecification tests to see whether the model adequately describes the data. In this section we consider tests that are designed for alternatives that incorporate asymmetry or, more generally, missing exogenous variables or nonlinearity. The leading testing principle is the score or Lagrange multiplier principle, because then only the null model has to be estimated. As explained for example in Engle (1982b), these tests can be carried out in the so-called TR^2 form, and under the null hypothesis the test statistic has an asymptotic χ^2-distribution. When the null hypothesis is the standard GARCH model (2), the test can be carried out in stages as follows:

1. Estimate the parameters of the GARCH model (2) and compute 'the residual sum of squares' $\mathrm{SSR}_0 = \sum_{j=1}^{T}(\varepsilon_t^2/\widetilde{h}_t - 1)^2$, where $\widetilde{h}_t$ is the estimated conditional variance at t.

2. Regress $\widetilde{z}_t^2 = \varepsilon_t^2/\widetilde{h}_t$ on the gradient of the log-likelihood function and the new variables (assuming the component excluded under the null hypothesis is linear in parameters), and compute the residual sum of squares SSR_1 from this auxiliary regression.

3. Form the test statistic

$$T\frac{\mathrm{SSR}_0 - \mathrm{SSR}_1}{\mathrm{SSR}_0} \xrightarrow{d} \chi^2(m)$$

under the null hypothesis of dimension m. When the null model is (2), the gradient equals $\widetilde{\mathbf{g}}_t = \widetilde{h}_t^{-1}(\partial h_t/\partial\boldsymbol{\omega})_0$, where $\boldsymbol{\omega} = (\alpha_0, \alpha_1, ..., \alpha_q, \beta_1, ..., \beta_p)'$, and

$$(\partial h_t/\partial\boldsymbol{\omega})_0 = \widetilde{\mathbf{u}}_t + \sum_{i=1}^{p} \widetilde{\beta}_i (\partial h_{t-i}/\partial\boldsymbol{\omega})_0.$$

with $\widetilde{\mathbf{u}}_t = (1, \varepsilon_{t-1}^2, ..., \varepsilon_{t-q}^2, \widetilde{h}_{t-1}, ..., \widetilde{h}_{t-p})'$.The subscript 0 indicates that the partial derivatives are evaluated under H_0. They are available from the estimation of the null model and need not be computed separately.

The auxiliary regression is thus

$$\widetilde{z}_t^2 = a + \widetilde{\mathbf{g}}_t'\boldsymbol{\delta}_0 + \mathbf{v}_t'\boldsymbol{\delta}_1 + \eta_t \tag{22}$$

where $\widetilde{z}_t^2 = \varepsilon_t^2/\widetilde{h}_t$ and $\mathbf{v}_t$ is the $m \times 1$ vector of the variables under test, so H_0: $\boldsymbol{\delta}_1 = \mathbf{0}$. Many of the tests discussed in this section fit into this framework. Engle and Ng (1993) propose asymmetry tests such that $\mathbf{v}_t = I(\varepsilon_{t-1} < 0)$ (the sign bias test), $\mathbf{v}_t = I(\varepsilon_{t-1} < 0)\varepsilon_{t-1}$ (negative size bias test), and $\mathbf{v}_t = \{1 - I(\varepsilon_{t-1} < 0)\}\varepsilon_{t-1}$ (positive size bias test). They also suggest a joint test in which

$$\mathbf{v}_t = (I(\varepsilon_{t-1} < 0), I(\varepsilon_{t-1} < 0)\varepsilon_{t-1}, \{1 - I(\varepsilon_{t-1} < 0)\}\varepsilon_{t-1})'$$

and note that the tests can be generalised to involve more lags than the first one. These three tests are the most often used tests for asymmetry in empirical work.

5.2 Testing GARCH against smooth transition GARCH

An STGARCH model must not be fitted to a return series without first testing the standard GARCH model against it. The reason is that the STGARCH model is not identified if the data are generated from the standard GARCH model. As an example, consider the model (10). If $\alpha_{2j} = 0$, $j = 0, 1, ..., q$, the model collapses into the GARCH model. In this case, the parameters γ and $c_1, ..., c_K$ in (11) are nuisance parameters that cannot be estimated consistently. Setting $\gamma = 0$ also makes the model into a standard GARCH model, and in this case $\alpha_{2j} = 0$, $j = 0, 1, ..., q$, and $c_1, ..., c_K$ are unidentified. The problem is that the lack of identification under the null hypothesis which is the standard GARCH model, invalidates standard asymptotic inference. Consequently, the asymptotic null distribution of the customary χ^2-statistic is unknown. This is a common problem in testing linearity against many nonlinear conditional mean models such as the threshold autoregressive, the smooth transition autoregressive or the hidden Markov model. It was first considered by Davies (1977); see Teräsvirta, Tjøstheim and Granger (2010, Chapter 5) for further discussion.

A straightforward solution to the problem of testing GARCH against STGARCH would be to construct an empirical distribution of the test statistic by simulation, see Hansen (1996). Gonzalez-Rivera (1998) already mentioned this possibility in her discussion of testing for smooth transition GARCH. Since the time series in applications are often quite long, this approach, however, could be computationally demanding. Following Luukkonen, Saikkonen and Teräsvirta (1988), Hagerud (1997) and later Lundbergh and Teräsvirta (2002), suggested circumventing the identification problem by approximating the transition function (11) by a Taylor expansion around the null hypothesis $\gamma = 0$. As a simple example, let $q = 1$ in (10) and $K = 1$ in (11). A first-order Taylor expansion of the transition function becomes

$$T_1(\varepsilon_{t-1};\gamma) = 1/2 + (1/4)(\varepsilon_{t-1} - c_1)\gamma + R_1(\varepsilon_{t-1};\gamma) \tag{23}$$

where $R_1(\varepsilon_{t-1};\gamma)$ is the remainder. Substituting (23) for the transition function in (10) and reparameterising yields

$$h_t = \alpha_{10}^* + \alpha_1\varepsilon_{t-1} + \alpha_{11}^*\varepsilon_{t-j}^2 + \alpha_{111}^*\varepsilon_{t-1}^3 + R_1^*(\varepsilon_{t-1};\gamma) + \sum_{j=1}^{p}\beta_j h_{t-j}$$

where $R_1^*(\varepsilon_{t-1};\gamma) = (\alpha_{20} + \alpha_{21}\varepsilon_{t-j}^2)R_1(\varepsilon_{t-1};\gamma)$, $\alpha_1 = \gamma/4$, and $\alpha_{11}^* = \gamma\alpha_{21}/4$. Since the resulting test is a Lagrange multiplier test and $R_1(\varepsilon_{t-1};\gamma) \equiv 0$ under the null hypothesis, the remainder can be ignored in the test. The null hypothesis equals $\alpha_1 = \alpha_{111}^* = 0$. In the TR^2 version of the test, $\mathbf{v}_t = \widetilde{h}_t^{-1}(\partial h_t/\partial\boldsymbol{\omega})_0$ where

$$(\partial h_t/\partial\boldsymbol{\omega})_0 = (\varepsilon_{t-1}, \varepsilon_{t-1}^3)' + \sum_{j=1}^{p}\widetilde{\beta}_j(\partial h_{t-j}/\partial\boldsymbol{\omega})_0$$

and $\boldsymbol{\omega} = (\alpha_1, \alpha_{111}^*)'$. The asymptotically (under H_0) χ^2-distributed test statistic has two degrees of freedom. The test requires $\mathsf{E}\varepsilon_t^6 < \infty$. Generalisations to $q > 1$ are obvious. When $p = q = 1$ and $\alpha_{21} = 0$, one obtains $\mathbf{v}_t = \widetilde{h}_t^{-1}\{\varepsilon_{t-1} + \widetilde{\beta}_1(\partial h_{t-1}/\partial\alpha_1)_0\}$ and the test can be viewed as the test of the GARCH(1,1) model against the Quadratic GARCH(1,1) model of Sentana (1995).

5.3 Testing GARCH against Artificial Neural Network GARCH

Caulet and Péguin-Feissolle (2000) developed a test of the hypothesis of independent, identically distributed observations against autoregressive conditional heteroskedasticity. The model was thus (19) with $\alpha_i, \beta_j = 0$, $i, j \geq 1$, and the null hypothesis was $\phi_j = 0$, $j = 1, ..., s$. The model can be generalised for testing GARCH against (19) with (20). To solve the identification problem, the authors adopted the method introduced by Lee, White and Granger (1993). The idea was to choose a large s and draw the nuisance parameters for each hidden unit from a uniform distribution after appropriate rescaling of ε_t, $t = 1, ..., T$. It follows that

$$\mathbf{v}_t = \widetilde{h}_t^{-1}(G(\widehat{\gamma}_{01} + \boldsymbol{\varepsilon}_t'\widehat{\boldsymbol{\gamma}}_1), ..., G(\widehat{\gamma}_{0s} + \boldsymbol{\varepsilon}_t'\widehat{\boldsymbol{\gamma}}_s))' \tag{24}$$

where $\widehat{\gamma}_{0j}$ and $\widehat{\boldsymbol{\gamma}}_j$ are the parameters from the jth random draw. The power of the test is dependent on s. The dimension of the null hypothesis and $\mathbf{v}_t$ may be reduced by considering the principal components of the hidden units corresponding to the largest eigenvalues. The test of Caulet and Péguin-Feissolle (2000) is a special case of the test mentioned here. Note that in that case, $\widetilde{h}_t$ in (24) is a positive constant and can be ignored. Another possibility, discussed in Péguin-Feissolle (1999), is to develop each hidden unit into a Taylor series around the null hypothesis H_0: $\boldsymbol{\gamma}_j = \mathbf{0}$ as in Teräsvirta, Lin and Granger (1993). After merging terms and reparameterising, a third-order Taylor expansion has the form

$$\begin{aligned} T_3(\boldsymbol{\varepsilon}_t; \boldsymbol{\gamma}_1, ..., \boldsymbol{\gamma}_s) &= \phi_0^* + \sum_{j=1}^{k} \phi_j^* \varepsilon_{t-j} + \sum_{i=1}^{k} \sum_{j=i}^{k} \phi_{ij}^* \varepsilon_{t-i}\varepsilon_{t-j} \\ &\quad + \sum_{i=1}^{k} \sum_{j=i}^{k} \sum_{j=\ell}^{k} \phi_{ij\ell}^* \varepsilon_{t-i}\varepsilon_{t-j}\varepsilon_{t-\ell} + R_3(\boldsymbol{\varepsilon}_t; \boldsymbol{\Gamma}) \end{aligned}$$

where $R_3(\boldsymbol{\varepsilon}_t; \boldsymbol{\Gamma})$ is the combined remainder. Similarly to the remainder in (23), $R_3(\boldsymbol{\varepsilon}_t; \boldsymbol{\Gamma}) = 0$ under the null hypothesis H_0 and does not affect the asymptotic null distribution of the test statistic. The new null hypothesis is H_0': $\phi_j^* = 0$, $j = 1, ..., s$, $\phi_{ij}^* = 0$, $i \neq j$, and $\phi_{ij\ell}^* = 0$, $i = 1, ..., s$; $j = i, ..., s$; and $\ell = j, ..., s$. The moment condition is $\mathsf{E}\varepsilon_t^6 < \infty$. When $k = 1$, H_0': $\phi_1^* = 0$, $\phi_{111}^* = 0$, and the test is the same as the test of GARCH against STGARCH when $q = 1$ in (10). Since the power comparisons of Péguin-Feissolle (1999) only concerned the null hypothesis of no conditional heteroskedasticity, little is known about the power properties of these tests in the GARCH context. They may be regarded as rather general misspecification tests of the standard GARCH model.

6 Estimation of parameters in nonlinear GARCH models

6.1 Smooth transition GARCH

Parameters of the smooth transition GARCH models can be estimated by (quasi) maximum likelihood (QML) method. The use of QML requires, among other things, that the slope parameter γ be bounded away from both zero and infinity. General conditions for asymptotic normality of the QML estimators of nonlinear first-order GARCH models are discussed in Meitz and Saikkonen (in press). The moment conditions required for asymptotic normality include the existence of the fourth moment of ε_t. It is seen from (1) that a necessary condition for this is $\mathsf{E}z_t^4 < \infty$, but it is not sufficient. One of the examples in Meitz and Saikkonen (in press) is the STGARCH model (10) with $p = q = 1$, and it is shown how the model satisfies the general conditions given in the paper.

In practice, estimation of the parameters of the STGARCH model is quite straightforward. Numerical problems in the form of slow convergence may be expected, however, when the slope parameter γ is large, that is, much larger in magnitude than the other parameters. The reasons for this have been discussed in the context of smooth transition autoregressive model; see for example Teräsvirta, Tjøstheim and Granger (2010, Chapter 16). A major reason is that even a substantial change in the value of γ only leads to a very small change the shape of the transition function when the transition is very steep or, in other words, when γ is large.

Computing the partial derivatives is another numerical issue. Brooks, Burke and Persand (2001) already emphasised the importance of using analytical derivatives in the estimation of parameters of standard GARCH models. It is even more important to apply them when the GARCH model to be estimated is nonlinear. The reason is increased precision of the estimates. If the user chooses to use the Berndt-Hall-Hall-Hausman (BHHH) algorithm discussed in many textbooks, then only the first derivatives are needed. This algorithm has been quite popular in the estimation of GARCH models, where computing the analytic second derivatives is most often avoided despite the work of Fiorentini, Calzolari and Panattoni (1996).

The choice of initial values is important in the estimation of nonlinear models. The initial values of the conditional variance may be chosen in the same way as in linear GARCH models by setting them equal to the sample unconditional variance. The initial values $\varepsilon_0, ..., \varepsilon_{q-1}$ are set to zero as are the partial derivatives of the conditional variance. It may be harder to find appropriate initial values for the parameters of the STGARCH model. Obtaining them by applying a heuristic method such as simulated annealing or a genetic algorithm would be a possibility. That would help avoiding inferior local optima. For applications of these to standard GARCH models see for example Amilon (2003) or Adanu (2006). The latter article considers and compares several heuristic methods with each other. Finally, it can be mentioned that smooth transition GARCH models can also be estimated using Bayesian techniques; see Lubrano (2001) for discussion. Wago (2004) later showed how estimation can be carried out using the Gibbs sampler.

6.2 Neural Network GARCH

Estimation of ANN-GARCH models using analytical derivatives may be difficult due to many nonlinear parameters, unless the number of hidden units is very small. For this reason, Donaldson and Kamstra (1997) proposed to draw five sets of parameters $\mathbf{\Gamma}_j$, $j = 1, ..., s$, randomly and then estimate the remaining parameters for each of these five sets with a grid on the dimensions p, q, s, u and m, defined as $(p, q, s, u, m) \in [0, 5]$. Once this is done, a model selection criterion such as BIC is employed to choose the best model from the set of estimated models. The statistical properties of this *ad hoc* method are not known.

It seems that the parameters of the ANN-GARCH model (19) could be estimated by simulated annealing, although empirical evidence of this is scarce. Goffe, Ferrier and Rogers (1994) considered applying simulated annealing in the situation where an autoregressive single-hidden-layer neural network model was fitted to a time series generated from a chaotic model. The results showed a number of local maxima. Furthermore, in repeated experiments with different starting-values and temperatures they were not able to find the same local maximum twice. The optima found were nevertheless better than the ones obtained using derivative-based estimation methods. The estimation of GARCH models offers an extra complication since the conditional variance h_t is not observed but has to be re-estimated for each iteration. Estimation of a pure ANN-ARCH model would in this respect be a computationally easier problem. Simulations and applications would be needed to assess the usefulness of simulated annealing in the estimation of models such as (19).

7 Forecasting with nonlinear GARCH models

7.1 Smooth transition GARCH

It may be useful to begin this section by considering the first-order GJR-GARCH model, one of the predecessors of the smooth transition GARCH model. What is being forecast is the conditional variance h_t. The forecasts are conditional means, so the forecast of h_{t+1} given the information $\mathcal{F}_t$ up to t equals, see Zivot (2009),

$$h_{t+1|t} = \mathsf{E}(h_{t+1}|\mathcal{F}_t) = \alpha_0 + \alpha(z_t)h_t \qquad (25)$$

where $\alpha(z_t) = \alpha_1 z_t^2 + \kappa_1 I(z_t < 0)\} z_t^2 + \beta_1$. Accordingly, assuming that z_t has a symmetric distribution,

$$\begin{aligned} h_{t+2|t} &= \alpha_0 + \{\alpha_1 \mathsf{E} z_{t+1}^2 + \kappa_1 \mathsf{E} I(z_{t+1} < 0)\} z_{t+1}^2 + \beta_1\} h_{t+1|t} \\ &= \alpha_0 + \{\alpha_1 + (\kappa_1/2) + \beta_1\} h_{t+1|t} \\ &= \alpha_0(1 + \{\alpha_1 + (\kappa_1/2) + \beta_1\}) + \{\alpha_1 + (\kappa_1/2) + \beta_1\}\alpha(z_t)h_t. \end{aligned}$$

Generally, for $k \geq 1$,

$$h_{t+k|t} = \alpha_0 \sum_{j=0}^{k-1} \{\alpha_1 + (\kappa_1/2) + \beta_1\}^j + \{\alpha_1 + (\kappa_1/2) + \beta_1\}^{k-1}\alpha(z_t)h_t.$$

When $\alpha_1 + (\kappa_1/2) + \beta_1 < 1$, that is, when the unconditional variance of ε_t is finite,

$$h_{t+k|t} \to \alpha_0(1 - \{\alpha_1 + (\kappa_1/2) + \beta_1\})^{-1}$$

as $k \to \infty$.

Next consider the first-order smooth transition GARCH model assuming $K = 1$ and $z_t \sim \text{iid}(0, 1)$ with a continuous density $f_z(z)$. The forecast corresponding to (25) equals

$$h_{t+1|t} = \alpha_{10} + \alpha_1 z_t^2 + \beta_1 h_t + \alpha(z_t, h_t)h_t$$

where

$$\alpha(z_t, h_t) = (\alpha_{20} + \alpha_{21} z_t^2)(1 + \exp\{-\gamma(z_t h_t^{1/2} - c)\})^{-1}$$

is a nonlinear function of h_t. This makes a difference. Consider

$$h_{t+2|t} = \alpha_{10} + (\alpha_1 + \beta_1)h_{t+1|t} + \mathsf{E}\{\alpha(z_{t+1}, h_{t+1|t})|\mathcal{F}_t\}h_{t+1|t}.$$

where

$$\mathsf{E}\{\alpha(z_{t+1}, h_{t+1|t})|\mathcal{F}_t\} = \int_{-\infty}^{\infty} (\alpha_{20}+\alpha_{21}z^2)(1+\exp\{-\gamma(zh_{t+1|t}^{1/2}-c)\})^{-1} f_z(z)\mathrm{d}z. \quad (26)$$

The expectation (26) can be computed by numerical integration, but the integration becomes a multiple integral when the forecast horizon $k > 2$. A similar situation has been discussed in the context of forecasting with nonlinear conditional mean models, see for example Teräsvirta (2006), Kock and Teräsvirta (in press) or Teräsvirta, Tjøstheim and Granger (2010, Chapter 14). The suggestion has been to compute an approximation to the integral by simulation or bootstrap, for details see the aforementioned references. As a by-product, this method generates a density forecast, based on the simulated or bootstrapped values of the argument of (26).

The same procedure applies to a few other nonlinear models including the ANN-GARCH model of Caulet and Péguin-Feissolle (2000) but is not necessary for Engle's nonlinear GARCH model (6). For that model,

$$h_{t+1|t} = \alpha_0 + \alpha_1(z_t h_t^{1/2} - \lambda)^2 + \beta_1 h_t$$

is a nonlinear function of h_t. Nevertheless,

$$\begin{aligned} h_{t+2|t} &= \alpha_0 + \alpha_1 \mathsf{E}(z_{t+1}h_{t+1|t}^{1/2} - \lambda)^2 + \beta_1 h_{t+1|t} \\ &= \alpha_0 + \alpha_1\lambda^2 + (\alpha_1 + \beta_1)h_{t+1|t} \end{aligned}$$

which implies that forecasts for $k > 2$ can be obtained by a simple recursion.

7.2 Asymmetric Power GARCH

Consider first the special case of (8) with $\delta = 1$, discussed by Meitz and Saikkonen (in press). Since this model can be written as a GJR-GARCH model, forecasting several periods ahead is straightforward. The one-period-ahead forecast equals

$$h_{t+1|t} = \alpha_0 + \{\alpha_1(1-\lambda)^2 + 4\lambda\alpha_1 \mathsf{E}I(z_t < 0)z_t^2 + \beta_1\}h_t. \quad (27)$$

where, retaining the assumption that $z_t \sim \mathcal{N}(0,1)$, $\mathsf{E}I(z_t < 0)z_t^2 = (1/4)\{(\pi/2) - 1\}$. Forecasts for longer horizons are obtained by recursion. In general,

$$h_{t+h|t} = \alpha_0 \sum_{j=0}^{h-1} \alpha(z,\lambda)^j + \alpha(z,\lambda)^{h-1} h_t$$

where $\alpha(z,\lambda) = \alpha_1(1-\lambda)^2 + 4\lambda\alpha_1 \mathsf{E}I(z<0)z^2 + \beta_1$, and $z \sim \mathcal{N}(0,1)$.

When $\delta \neq 1$, the forecast $h_{t+1|t}$ already has to be computed numerically. From (8) one obtains

$$h_t = \{\alpha_0 + \alpha_1(|\varepsilon_{t-1}| - \lambda\varepsilon_{t-1})^{2\delta} + \beta_1 h_{t-1}^{\delta}\}^{1/\delta}$$

so

$$\begin{aligned} h_{t+1|t} &= \mathsf{E}[\alpha_0 + \{\alpha_1(|z_t| - \lambda z_t)^{2\delta} + \beta_1\}h_t^{\delta}]^{1/\delta} \\ &= \int_{-\infty}^{\infty}[\alpha_0 + \{\alpha_1(|z| - \lambda z)^{2\delta} + \beta_1\}h_t^{\delta}]^{1/\delta}\phi(z)dz \end{aligned}$$

where $\phi(z)$ is the density function of the standard normal random variable. The integral can be computed by simulation or by a bootstrap as in the case of the smooth transition GARCH model.

There is another special case that is of interest: $\delta = 1/2$. What is being modelled is the conditional standard deviation. In that case it is natural to also forecast the conditional standard deviation and not the conditional variance. Thus,

$$h_{t+1|t}^{1/2} = \mathsf{E}[\alpha_0 + \{\alpha_1(|z_t| - \lambda z_t) + \beta_1\}h_t^{1/2}] = \alpha_0 + \{\alpha_1(2/\pi)^{1/2} + \beta_1\}h_t^{1/2} \quad (28)$$

and the forecasts for longer horizons follow by simple recursion from (28).

8 Models based on multiplicative decomposition of the variance

So far the nonlinear extensions of the standard GARCH model considered in this review have concerned the parameterisation of the conditional variance. There is another strand of literature aiming at extending the GARCH model by decomposing the variance into two components, one which is stationary and another one that can be nonstationary. Many of the models belonging to this category are semiparametric and discussed in Van Bellegem (2011). In this section we shall look at two parametric models: one by Amado and Teräsvirta (2011) and the other one by Osiewalski (2009) and Osiewalski and Pajor (2009).

The multiplicative variance decomposition is as follows: Write

$$\varepsilon_t = z_t\sigma_t = z_t(h_t g_t)^{1/2} \quad (29)$$

where the variance $\sigma_t^2 = h_t g_t$. The first component h_t is defined in (2) or (3), and the positive-valued function

$$g_t = 1 + \sum_{l=1}^{r} \delta_l G_l(t^*; \gamma_l, \mathbf{c}_l). \quad (30)$$

with

$$G_l(t^*;\gamma_l,\mathbf{c}_l) = (1+\exp\{-\gamma_l \prod_{k=1}^{K}(t^*-c_{lk})\})^{-1},\ \gamma_l > 0,\ c_{l1} \leq c_{l2} \leq ... \leq c_{lK} \tag{31}$$

is a flexible deterministic function of time describing shifts in the unconditional variance. This makes $\{\varepsilon_t\}$ a (globally) nonstationary sequence in variance. Analogously to (11), the function (31) is a generalised logistic function. In practice, typically $K = 1$ or $K = 2$. The function h_t characterises volatility clustering as in the standard GARCH model. Rewriting (29) as follows:

$$\phi_t = \varepsilon_t/g_t^{1/2} = z_t h_t^{1/2} \tag{32}$$

it is seen that after adjusting for shifts in the unconditional variance, the first-order conditional variance process h_t has the form

$$h_t = \alpha_0 + \alpha_1\phi_{t-1}^2 + \beta_1 h_{t-1}. \tag{33}$$

The number of shifts in (30) is determined by sequential testing after fitting a GARCH or GJR-GARCH model to the time series under consideration. Maximum likelihood estimation of the parameters in (29) is carried out by maximising the log-likelihood in parts as discussed in Song, Fan and Kalbfleisch (2005); see Amado and Teräsvirta (2011) for details. Evaluation of the estimated model is carried out by misspecification tests in Lundbergh and Teräsvirta (2002) that are generalised to this situation. Examples can be found in Amado and Teräsvirta (2011). Semi- and nonparametric alternatives to (30) are considered in Van Bellegem (2011).

The hybrid volatility model of Osiewalski (2009) and Osiewalski and Pajor (2009) makes use of the decomposition (29), but g_t is stochastic and defined as a first-order autoregressive stochastic volatility process:

$$\ln g_t = \psi \ln g_{t-1} + \eta_t \tag{34}$$

where $|\psi| < 1$, and $\{\eta_t\} \sim \text{iid}\mathcal{N}(0, \sigma_\eta^2)$ and independent of $\{z_t\}$. The original version of the model is multivariate, and the conditional covariance matrix follows a BEKK-GARCH process. In this review the focus is on the univariate special case. The authors distinguish between two different specification. In the first one,

$$h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1}$$

so the conditional variance evolves independently of (34). The second version bears resemblance to the model of Amado and Teräsvirta (2011) in that h_t

is modelled as in (33) using (32). In that case, h_t is dependent on two independent sources of noise, z_t and η_t. Stationarity conditions for this model are probably not well known, and analytic expressions for unconditional moments of ε_t do not as yet seem to be available.

The treatment of these two models in Osiewalski (2009) and Osiewalski and Pajor (2009) is completely Bayesian, and the authors discuss appropriate algorithms for estimation. Multivariate applications to financial time series can be found in these two papers.

9 Final remarks

This review covers the most common nonlinear models of conditional heteroskedasticity. Parametric models in which the variance is decomposed into an unconditional and conditional component are also briefly considered. Many of these models can characterise various types of nonlinearity such as asymmetric, or symmetric but nonlinear, responses to shocks. Nevertheless, it seems that none of these models is widely used in practice. The practitioners often favour simpler models, of which the GJR-GARCH model designed for describing asymmetric response to shocks constitutes an example. Increased computational power and improved numerical methods of optimisation may be expected to change the situation in the future.

References

Adanu, K.: 2006, Optimizing the GARCH model – an application of two global and two local search methods, *Computational Economics* **28**, 277–290.

Amado, C. and Teräsvirta, T.: 2011, Modelling volatility by variance decomposition, *CREATES Research Paper 2011-1*, Aarhus University.

Amilon, H.: 2003, GARCH estimation and discrete stock prices: An application to low-priced Australian stocks, *Economics Letters* **81**, 215–222.

Andersen, T. G., Davis, R. A., Kreiss, J.-P. and Mikosch, T. (eds): 2009, *Handbook of Financial Time Series*, Springer, New York.

Anderson, H. M., Nam, K. and Vahid, F.: 1999, Asymmetric nonlinear smooth transition GARCH models, *in* P. Rothman (ed.), *Nonlinear time series analysis of economic and financial data*, Kluwer, Boston, pp. 191–207.

Andreou, E. and Ghysels, E.: 2009, Structural breaks in financial time series, *in* T. G. Andersen, R. A. Davis, J.-P. Kreiss and T. Mikosch (eds), *Handbook of Financial Time Series*, Springer, New York, pp. 839–870.

Bauwens, L., Hafner, C. and Laurent, S.: 2011, Volatility: Definitions, measurements and models, *in* L. Bauwens, C. Hafner and S. Laurent (eds), *Wiley Handbook in Financial Engineering and Econometrics: Volatility Models and Their Applications*, Wiley, New York.

Bauwens, L., Laurent, S. and Rombouts, J. V. K.: 2006, Multivariate GARCH models: A survey, *Journal of Applied Econometrics* **21**, 79–109.

Bollerslev, T.: 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* **31**, 307–327.

Bos, C.: 2011, Univariate and low dimensional stochastic volatility models, *in* L. Bauwens, C. Hafner and S. Laurent (eds), *Wiley Handbook in Financial Engineering and Econometrics: Volatility Models and Their Applications*, Wiley, New York.

Brooks, C., Burke, S. P. and Persand, G.: 2001, Benchmarks and the accuracy of GARCH model estimation, *International Journal of Forecasting* **17**, 45–56.

Brooks, R., Faff, R. W., McKenzie, M. D. and Mitchell, H.: 2000, A multi-country study of power ARCH models and national stock market returns, *Journal of International Money and Finance* **19**, 377–397.

Caulet, R. and Péguin-Feissolle, A.: 2000, Un test d'hétéroscédasticité conditionnelle inspiré de la modélisation en termes de réseaux neuronaux artificiels, *Annales d'économie et de statistique* **59**, 178–197.

Cizek, P. and Spokoiny, V.: 2009, Varying coefficient GARCH-models, *in* T. G. Andersen, R. A. Davis, J.-P. Kreiss and T. Mikosch (eds), *Handbook of Financial Time Series*, Springer, New York, pp. 169–185.

Davies, R. B.: 1977, Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika* **64**, 247–254.

Ding, Z., Granger, C. W. J. and Engle, R. F.: 1993, A long memory property of stock market returns and a new model, *Journal of Empirical Finance* **1**, 83–106.

Donaldson, R. G. and Kamstra, M.: 1997, An artificial neural network-GARCH model for international stock return volatility, *Journal of Empirical Finance* **4**, 17–46.

Duan, J.-C.: 1997, Augmented GARCH(p,q) process and its diffusion limit, *Journal of Econometrics* **79**, 97–127.

Engle, R. F.: 1982a, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* **50**, 987–1007.

Engle, R. F.: 1982b, A general approach to Lagrange multiplier model diagnostics, *Journal of Econometrics* **20**, 83–104.

Engle, R. F.: 1990, Discussion, *Review of Financial Studies* **3**, 103–106.

Engle, R. F. and Ng, V. K.: 1993, Measuring and testing the impact of news on volatility, *Journal of Finance* **48**, 1749–1777.

Fiorentini, G., Calzolari, G. and Panattoni, L.: 1996, Analytic derivatives and the computation of GARCH estimates, *Journal of Applied Econometrics* **11**, 399–417.

Fornari, F. and Mele, A.: 1997, Sign- and volatility-switching ARCH models: Theory and applications to international stock markets, *Journal of Applied Econometrics* **12**, 49–65.

Glosten, L., Jagannathan, R. and Runkle, D.: 1993, On the relation between expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* **48**, 1779–1801.

Goffe, W. L., Ferrier, G. D. and Rogers, J.: 1994, Global optimization of statistical functions with simulated annealing, *Journal of Econometrics* **60**, 65–99.

Gonzalez-Rivera, G.: 1998, Smooth transition GARCH models, *Studies in Nonlinear Dynamics and Econometrics* **3**, 161–178.

Hagerud, G.: 1997, *A New Non-Linear GARCH Model*, EFI Economic Research Institute, Stockholm.

Hansen, B. E.: 1996, Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica* **64**, 413–430.

He, C., Malmsten, H. and Teräsvirta, T.: 2008, Higher-order dependence in the general Power ARCH process and the role of power parameter, *in* Shalabh and C. Heumann (eds), *Recent Advances in Linear Models and Related Areas*, Physica-Verlag, Heidelberg, pp. 231–251.

He, C. and Teräsvirta, T.: 1999a, Properties of moments of a family of GARCH processes, *Journal of Econometrics* **92**, 173–192.

He, C. and Teräsvirta, T.: 1999b, Statistical properties of the asymmetric power ARCH process, *in* R. F. Engle and H. White (eds), *Cointegration, Causality and Forecasting. A Festschrift in Honour of Clive W.J. Granger*, Oxford University Press, Oxford, pp. 462–474.

Hentschel, L.: 1995, All in the family. nesting symmetric and asymmetric GARCH models, *Journal of Financial Economics* **39**, 71–104.

Higgins, M. L. and Bera, A. K.: 1992, A class of nonlinear ARCH models, *International Economic Review* **33**, 137–158.

Kock, A. B. and Teräsvirta, T.: in press, Forecasting with nonlinear time series models, *in* M. P. Clements and D. F. Hendry (eds), *The Oxford Handbook of Economic Forecasting*, Oxford University Press, Oxford.

Lanne, M. and Saikkonen, P.: 2005, Nonlinear GARCH models for highly persistent volatility, *Econometrics Journal* **8**, 251–276.

Lee, T.-H., White, H. and Granger, C. W. J.: 1993, Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests, *Journal of Econometrics* **56**, 269–290.

Li, C. W. and Li, W. K.: 1996, On a double threshold autoregressive heteroskedasticity time series model, *Journal of Applied Econometrics* **11**, 253–274.

Ling, S. and Tong, H.: 2005, Testing for a linear MA model against threshold MA models, *Annals of Statistics* **33**, 2529–2552.

Ling, S., Tong, H. and Li, D.: 2007, Ergodicity and invertibility of threshold moving average models, *Bernoulli* **13**, 161–168.

Lubrano, M.: 2001, Smooth transition GARCH models: A Bayesian perspective, *Recherches économiques de Louvain* **67**, 257–287.

Lundbergh, S. and Teräsvirta, T.: 2002, Evaluating GARCH models, *Journal of Econometrics* **110**, 417–435.

Luukkonen, R., Saikkonen, P. and Teräsvirta, T.: 1988, Testing linearity against smooth transition autoregressive models, *Biometrika* **75**, 491–499.

Meitz, M. and Saikkonen, P.: 2008, Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models, *Econometric Theory* **24**, 1291–1320.

Meitz, M. and Saikkonen, P.: in press, Parameter estimation in nonlinear AR-GARCH models, *Econometric Theory* .

Mikosch, T. and Starica, C.: 2004, Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects, *Review of Economics and Statistics* **86**, 378–390.

Nelson, D. B.: 1991, Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* **59**, 347–370.

Nelson, D. B. and Cao, C. Q.: 1992, Inequality constraints in the univariate GARCH model, *Journal of Business and Economic Statistics* **10**, 229–235.

Osiewalski, J.: 2009, New hybrid models of multivariate volatility (a Bayesian perspective), *Przeglad Statystycny* **56**, 15–22.

Osiewalski, J. and Pajor, A.: 2009, Bayesian analysis for hybrid MSF-SBEKK models of multivariate volatility, *Central European Journal of Economic Modelling and Econometrics* **1**, 179–202.

Paolella, M. and Haas, M.: 2011, Regime-switching models, *in* L. Bauwens, C. Hafner and S. Laurent (eds), *Wiley Handbook in Financial Engineering and Econometrics: Volatility Models and Their Applications*, Wiley, New York.

Péguin-Feissolle, A.: 1999, A comparison of the power of some tests for conditional heteroscedasticity, *Economics Letters* **63**, 5–17.

Rabemananjara, R. and Zakoïan, J. M.: 1993, Threshold ARCH models and asymmetries in volatility, *Journal of Applied Econometrics* **8**, 31–49.

Rissanen, J.: 1978, Modeling by shortest data description, *Automatica* **14**, 465–471.

Rydén, T., Teräsvirta, T. and Åsbrink, S.: 1998, Stylized facts of daily return series and the hidden Markov model, *Journal of Applied Econometrics* **13**, 217–244.

Schwarz, G.: 1978, Estimating the dimension of a model, *Annals of Statistics* **6**, 461–464.

Sentana, E.: 1995, Quadratic ARCH models, *Review of Economic Studies* **62**, 639–661.

Shephard, N. G. and Andersen, T. G.: 2009, Stochastic volatility: Origins and overview, *in* T. G. Andersen, R. A. Davis, J.-P. Kreiss and T. Mikosch (eds), *Handbook of Financial Time Series*, Springer, New York, pp. 233–254.

Silvennoinen, A. and Teräsvirta, T.: 2009, Multivariate GARCH models, *in* T. G. Andersen, R. A. Davis, J.-P. Kreiss and T. Mikosch (eds), *Handbook of Financial Time Series*, Springer, New York, pp. 201–228.

So, M. K. P., Li, W. K. and Lam, K.: 2002, A threshold stochastic volatility model, *Journal of Forecasting* **21**, 473–500.

Song, P. X.-K., Fan, Y. and Kalbfleisch, J. D.: 2005, Maximization by parts in likelihood inference, *Journal of the American Statistical Association* **100**, 1145–1157.

Teräsvirta, T.: 1994, Specification, estimation, and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association* **89**, 208–218.

Teräsvirta, T.: 2006, Forecasting economic variables with nonlinear models, *in* G. Elliott, C. W. J. Granger and A. Timmermann (eds), *Handbook of Economic Forecasting*, Vol. 1, Elsevier, Amsterdam, pp. 413–457.

Teräsvirta, T.: 2009, An introduction to univariate GARCH models, *in* T. G. Andersen, R. A. Davis, J.-P. Kreiss and T. Mikosch (eds), *Handbook of Financial Time Series*, Springer, New York, pp. 17–42.

Teräsvirta, T., Lin, C.-F. and Granger, C. W. J.: 1993, Power of the neural network linearity test, *Journal of Time Series Analysis* **14**, 309–323.

Teräsvirta, T., Tjøstheim, D. and Granger, C. W. J.: 2010, *Modelling Nonlinear Economic Time Series*, Oxford University Press, Oxford.

Van Bellegem, S.: 2011, Locally stationary volatility models, *in* L. Bauwens, C. Hafner and S. Laurent (eds), *Wiley Handbook in Financial Engineering and Econometrics: Volatility Models and Their Applications*, Wiley, New York.

Wago, H.: 2004, Bayesian estimation of smooth transition GARCH model using Gibbs sampling, *Mathematics and Computers in Simulation* **64**, 63–78.

Zakoïan, J.-M.: 1994, Threshold heteroskedastic models, *Journal of Economic Dynamics and Control* **18**, 931–955.

Zivot, E.: 2009, Practical issues in the analysis of univariate GARCH models, *in* T. G. Andersen, R. A. Davis, J.-P. Kreiss and T. Mikosch (eds), *Handbook of Financial Time Series*, Springer, New York, pp. 113–155.

Research Papers 2011

2010-65: Almut E. D. Veraart: How precise is the finite sample approximation of the asymptotic distribution of realised variation measures in the presence of jumps?

2010-66: Ole E. Barndorff-Nielsen, David G. Pollard and Neil Shephard: Integer-valued Lévy processes and low latency financial econometrics

2010-67: Shin Kanaya and Dennis Kristensen: Estimation of Stochastic Volatility Models by Nonparametric Filtering

2010-68: Dennis Kristensen and Anders Rahbek: Testing and Inference in Nonlinear Cointegrating Vector Error Correction Models

2010-69: Søren Johansen: The analysis of nonstationary time series using regression, correlation and cointegration –with an application to annual mean temperature and sea level

2010-70: Søren Johansen and Morten Ørregaard Nielsen: A necessary moment condition for the fractional functional central limit theorem

2010-71: Nektarios Aslanidis and Isabel Casas : Modelling asset correlations during the recent financial crisis: A semiparametric approach

2010-72: Søren Johansen and Katarina Juselius: An invariance property of the common trends under linear transformations of the data

2010-73: Peter Sandholt Jensen and Allan H. Würtz: Estimating the effect of a variable in a high-dimensional regression model

2010-74: Peter R. Hansen, Asger Lunde and Valeri Voev: Realized Beta GARCH: A Multivariate GARCH Model with Realized Measures of Volatility and CoVolatility

2010-75: Laurent A.F. Callot: A Bootstrap Cointegration Rank Test for Panels of VAR Models

2010-76: Peter R. Hansen, Asger Lunde and James M. Nason: The Model Confidence Set

2011-01: Cristina Amado and Timo Teräsvirta: Modelling Volatility by Variance Decomposition

2011-02: Timo Teräsvirta: Nonlinear models for autoregressive conditional heteroskedasticity