Stieltjes Institute Lectures on

ARCH Modelling and Financial Time Series

March 8, 2001

Bernard Hanzon

Department of Econometrics and Operations Research

Vrije Universiteit Amsterdam

Introduction

Yesterday: Overview of nonlinear time series models applied to finance. Today: a specific class of nonlinear time series models: ARCH/GARCH etc! Why?

(i) As you probably heard as well yesterday, linear time series models, including ARMA models and linear state space models, have a drawback when applied to financial time series: The Efficient Market Hypothesis tells us that we cannot successfully predict the price level of a financial asset, other then by using the trivial prediction that "the price of tomorrow is that of today". But linear time series models are precisely constructed to predict levels, based on observations of the levels in the past.

(ii) As you learned on Monday and Tuesday, in modern financial theory, including the theory of financial derivatives, the volatility plays a crucial role (while the drift term, which mainly determines the level, does not). Although in the Black-Scholes model the volatility is assumed to be constant, we know from the study of actual time series that it actually is *not* constant. Because derivative prices and the composition of hedge portfolios depend heavily on the volatility it is extremely important to have good models for the volatility. The ARCH models are models for the volatility.

(iii) If you pose an arbitrary stochastic nonlinear dynamic model, then in order to do prediction based on past observations, you have to calculate the conditional probability distribution function of the present and future output, based on the past observations. In general this is a hard problem in the nonlinear case. (In the linear case this can be solved using the Kalman filter that you may have heard about). In the class of ARCH models this problem is circumvented by assuming the conditional distribution functions to have a specified (and rather simple) form, which simplifies the calculations. (Whether the ARCH models present a "true" description of the phenomenon at hand is another matter. It could be that they just give a, hopefully sufficiently good, approximation of reality. Compare "Ptolemeus".)

(iv) Because ARCH/GARCH models are based on linear time series models for the *volatility*, the well-known techniques for linear time series models can still be applied, at least to some extent. This is an advantage of this model class.

(v) There is now quite a bit of software available for this model class and it is increasingly used for modelling of financial data. This afternoon we will work with the S+GARCH package which will make it easy (after some practice) to make calculations and to try several models. However, below we will give a list of potential dangers when using such software without being sufficiently careful.

(vi) One can calculate optimal hedges under the assumption of an ARCH model. However this part of the theory still appears to be somewhat less developed than the statistical part of parameter estimation and hypothesis testing.

Potential dangers when using ARCH software!

In his book on ARCH modelling and Financial Applications, Gourieroux mentions the following dangers, when using ARCH software:

- In some markets automated quote systems were initiated that allow for price announcements even if no trading takes place
- There may exist different prices on the same market, for the same asset, at the same date (market makers activities; stripping of large orders)
- Announced price is only the basis for a bargaining process that leads to the true price

- Prices take discrete values and are multiples of some basic unit like 0.05.
- Transaction costs depending on the quantities to be traded, characteristics of the investors.
- Lack of liquidity may lead to small trade volume and unclear price formation
- If the number of investors is small for a given asset on a given day, the usual economics assumptions of competitive equilibrium may be violated and one should perhaps rather find explanations in game theory etc.
- ARCH models are fitted to return series. But financial decisions do not only depend on expected returns and volatilities, but also on things like market shares, portfolio composition, and on volumes.

- ARCH model assumes a rather stable environment: it is usually fitted on data from (mostly) stable periods. In economics and finance we have little chance to experiment! The effect is that the ARCH model fails to capture irregular phenomena such as crashes, mergers, news effects, opening and closing of the markets, price evolution for an option close to maturity, etc.
- The price evolution is modelled using the common knowledge contained in lagged prices. It does not take into account the possibility of information withheld by individuals etc.

Overview of the material presented

The purpose of the lectures is to give an introduction to ARCH models and its generalizations. Although the motivation for these models may be clear, it takes a while to see that modelling the conditional mean and the conditional variance simultaneously can be done and to see how that can be done. There are a number of subtleties that have to be understood. We will start with a simple example of a nonlinear dynamical model exhibiting conditional heteroscedasticity. Then a number of models will be presented, namely ARCH(1), ARCH(q), GARCH(p,q), ARMA-GARCH, ARCH-M and a Stochastic Variance model.

Some attention will be given to the occurrence of "fat tails"(i.e. an "increased danger of outliers"), mainly by considering the kurtosis.

Then the problem of statistical inference will be treated. Here the so-called quasi-maximum likelihood method will be used. It is actually the maximum likelihood method under some additional assumptions. This is also the method used in the S+GARCH package. From the mathematical point of view the determination (with mathematical certainty) of the maximum of the likelihood function is still largely an open problem. Some remarks will be made about this. Some of the asymptotic properties of the estimator are mentioned.

In the afternoon we will work on estimation of ARCH models for actual financial series using the S+GARCH package as well as on some theoretical exercises concerning ARCH models.

Introduction to the ARCH model

Example We first start with a rather simple example of an ARCH model. To introduce it it may be interesting to know that it was shown in the beginning of the 1960's by Nisio that every strictly stationary process may be approximated as closely as desired by a polynomial in Gaussian white noise. The following is a very simple example of such a polynomial model:

$$Y_t = W_t W_{t-1}^2,$$

where W_t is Gaussian white noise with zero mean and variance σ^2.

Let us now investigate the mean, the conditional mean , the variance and the conditional variance of Y_t given the past observations $Y_{t-1}, Y_{t-2}, \ldots$. The mean is zero:

$$E(Y_t) = E(W_t)E(W_{t-1}^2) = 0.$$

The conditional mean is also zero:

$$E(Y_t|\underline{Y_{t-1}}) = E(W_t)W_{t-1}^2 = 0.$$

Here $\underline{Y_{t-1}}$ denotes the total of the observations $Y_{t-1}, Y_{t-2}, \ldots$.

Because the mean is zero, the variance is equal to the second moment:

$$E(Y_t^2) = E(W_t^2)E(W_{t-1}^4) = 3\sigma^6,$$

where $\sigma^2 = V(W_t)$, the variance of the white noise W_t. In this calculation it is used that the fourth moment of a Gaussian random variable with zero mean and variance σ^2 is equal to $3\sigma^4$. (This can be derived in a straightforward manner by differentiation in the origin of the expression for the characteristic function $\phi(\tau) = E(e^{i\tau W}) = e^{-\sigma^2\tau^2/2}$ where W is an arbitrary Gaussian random variable with zero mean and variance σ^2.)

Consider the following calculation

$$E(Y_t Y_{t-k}) = E(W_t W_{t-1}^2 W_{t-k} W_{t-k-1}^2) =$$
$$E(W_t)E(W_{t-1}^2 W_{t-k} W_{t-k-1}^2) = 0$$

It follows that the covariance between Y_t and any past value $Y_{t-k},\ k > 0,$ is zero.

Before we calculate the conditional variance, recall that the unconditional variance is time independent. The conditional variance is

$$V(Y_t|\underline{Y_{t-1}}) = V(W_t W_{t-1}^2|\underline{Y_{t-1}}) = \sigma^2 W_{t-1}^4.$$

Note that this conditional variance is depending on the lagged value of the white noise, which itself a (somewhat complicated) function of the lagged values of Y_t. This example shows that nonlinear stationary time series can have (and usually will have) conditional variance that is time dependent, and actually dependent on the past values of the output. In econometrics this is called conditional heteroscedasticity.

ARCH(1)

The acronym ARCH stands for Auto-Regressive Conditionally Heteroscedastic. The idea behind this is that in this class of models, the conditional variance satisfies an equation that closely resembles the usual equations of an autoregressive time series model. It is best to give concrete equations. Consider the following model, which is an autoregressive model of order 1 with ARCH(1) errors (i.e. the errors follow an ARCH model of order 1).

$$\begin{aligned} Y_t &= \mu + \phi Y_{t-1} + \epsilon_t, \ |\phi| < 1, \\ \epsilon_t^2 &= c + a\epsilon_{t-1}^2 + u_t \end{aligned} \qquad (1)$$

where $t \in \mathbf{Z}$,

the sequence of random variables $\epsilon = (\epsilon_t)$ is a weak white noise (i.e. a sequence of uncorrelated, homoscedastic variables with zero mean) satisfying the *martingale difference sequence condition:*

$$E(\epsilon_t | \underline{\epsilon_{t-1}}) = 0, \text{ for all} t \in \mathbf{Z}$$

Furthermore (u_t) is assumed to be martingale difference sequence i.e. $E(u_t|\underline{u_{t-1}}) = 0$ for all t.

Problems:

- The sign of ϵ_t is not determined by the second equation of (**??**). This can be solved formally by assuming $\epsilon_t = \delta_t Z_t$ where δ_t is taken to be a random variable equal to 1 or -1, each with probability $\frac{1}{2}$ and where $Z_t > 0$ is completely determined by the second equation of (**??**).

- In order to have a stationary model one needs to assume $|a| < 1$, otherwise the unconditional mean of ϵ_t^2 is not time invariant.
- The positivity of ϵ_t^2 has to be ensured! It is sufficient to assume that $a > 0$ and almost surely $c + u_t \geq 0$. Note that this rules out the possibility that (u_t) is a Gaussian white noise!

Once the model is specified correctly, one can calculate the mean, the covariances and the variance both conditional on past observations and unconditional. Here we present the formulas and leave the verification as an exercise.

First we give those for the error process (ϵ_t) :

$$E(\epsilon_t|\underline{\epsilon_{t-h}}) = 0, \ h = 1, 2, \dots$$

$$Cov((\epsilon_t, \epsilon_{t+k})|\underline{\epsilon_{t-h}}) = 0,$$

$$\text{for all } h, k \in \{1, 2, \dots\}$$

$$V(\epsilon_t|\underline{\epsilon_{t-h}}) = c\frac{1-a^h}{1-a} + a^h\epsilon_{t-h}^2, \ h = 1, 2, \dots$$

$$E(\epsilon_t) = 0$$

$$Cov(\epsilon_t, \epsilon_{t+k}) = 0, \text{ for all } k \in \{1, 2, \dots\}$$

$$V(\epsilon_t) = E(V(\epsilon_t|\underline{\epsilon_{t-h}})) = \frac{c}{1-a}$$

Remarks:

- We use the *tower law*, or law of iterated expectations, in these calculations: For example

$$E(E(\epsilon_t|\underline{\epsilon_{t-1}})|\underline{\epsilon_{t-2}}) = E(\epsilon_t|\underline{\epsilon_{t-2}})$$

  because the informational content of $\underline{\epsilon_{t-2}}$ is smaller than that of $\underline{\epsilon_{t-1}}$.

- Note that $c \geq 0$ has to hold and $c > 0$ if we want the variance to be positive (we already assumed $|a| < 1$.)
- The limit of $V(\epsilon_t|\underline{\epsilon_{t-h}}) = c\frac{1-a^h}{1-a} + a^h\epsilon_{t-h}^2$ for $h \to \infty$ is equal to the unconditional variance if the limit of $a^h\epsilon_{t-h}^2$ is zero for $h \to \infty$. This is not as easy to show as one might think. If the expectation of ϵ_{t-h}^4 exists, then it can be shown using the Chebyshev inequality combined with the Borel-Cantelli Lemma from probability theory. Note that no assumption was made up till now about the existence of the fourth moment of ϵ_t.

Secondly we give the analogous formulas for the output process (Y_t) :

$$E(Y_t|\underline{Y_{t-h}}) = \mu\frac{1-\phi^h}{1-\phi} + \phi^h Y_{t-h},$$
$$h = 1, 2, \ldots$$
$$Cov((Y_t, Y_{t+k})|\underline{Y_{t-h}}) = \frac{c\phi^k}{1-a}\,\frac{1-\phi^{2h}}{1-\phi^2}$$
$$-\frac{ca\phi^k}{1-a}\,\frac{a^h-\phi^{2h}}{a-\phi^2} + a\phi^k\epsilon_{t-h}^2\frac{a^h-\phi^{2h}}{a-\phi^2},$$
for all $h \in \{1, 2, \ldots\},\ k \in \{0, 1, 2, \ldots\}$
$$V(Y_t|\underline{Y_{t-h}}) = V(\epsilon_t|\underline{Y_{t-h}}) = c = a\epsilon_{t-1}^2,$$
$$E(Y_t) = \frac{\mu}{1-\phi}$$
$$Cov(Y_t, Y_{t+k}) = \frac{c\phi^k}{1-a}\,\frac{1}{1-\phi^2},$$
for all $k \in \{1, 2, \ldots\}$
$$V(Y_t) = \frac{c}{1-a}\,\frac{1}{1-\phi^2}$$

Note that the formula for the covariances as presented only holds if $a \neq \phi^2$. However this is merely a matter of presentation, because in those terms where the factor $a - \phi^2$ appears in the denominator, it is also a factor in the numerator and can therefore be cancelled. After this cancellation, the formula makes perfect sense, even if $a = \phi^2$.

Additional assumptions on the distribution of the error process:

The conditional mean and conditional variance of the error are $E(\epsilon_t|\epsilon_{t-1}) = 0$ and $V(\epsilon_t|\epsilon_{t-1}) = c + a\epsilon_{t-1}^2$. However up till now no further assumptions where made about the conditional distribution of ϵ_t. It is possible to assume it to be Gaussian. (This implies that u_t has a centered and scaled chi-squared distribution, where the scaling factor is $c + a\epsilon_t^2$.)

Several remarks can now be made.

- The third and fourth order moments of ϵ_t can now be calculated, under the extra assumption that $3a^2 < 1$. In fact the third moment is zero(this follows simply from the tower law). The fourth moment is given by the following formula
$$E(\epsilon_t^4) = \frac{3c^2}{(1-a)^2}\frac{1-a^2}{1-3a^2}$$
Recalling that $E(\epsilon_t^2) = \frac{c}{1-a}$ the (unconditional) kurtosis can be calculated:
$$k = \frac{E(\epsilon_t^4)}{[E(\epsilon_t^2)]^2} = 3\frac{1-a^2}{1-3a^2}$$
Clearly this is larger than 3 if a is positive and diverges to ∞ if a approaches $\frac{1}{3}\sqrt{3}$. Fat tails!(leptokurtic).

- The error process is conditionally Gaussian, but not marginally Gaussian. This follows from the fact that the kurtosis is larger than 3.

Exercise Derive the formula given above for $E(\epsilon_t^4)$. Hint: Define the vector $W_t = (\epsilon_t^2, \epsilon_t^4)$ and express its conditional expectation in terms of W_{t-1}. Then take the expectation on both sides and solve for the unconditional expectation of W_t.

GENERAL ARCH PROCESSES AND EXTENSIONS

ARCH models describe simultaneously the evolution of the *conditional mean* and the *conditional variance*. Different models have been proposed in the literature in order to obtain past dependent conditional variances. Here we give a concise presentation of some of them.

ARCH(q)models(Engle 1982)
The ARCH(1) model for the error process presented above, can be generalized by increasing the degree of the autoregressive polynomial: The ARCH(q) model is defined as

$$\epsilon_t^2 = c + \sum_{i=1}^{q} a_i \epsilon_{t-i}^2 + u_t,$$

where $E(\epsilon_t|\underline{\epsilon_{t-1}}) = 0$, and (u_t) is (also) a martingale difference sequence.

As before the sign process $\delta_t = sign(\epsilon_t)$ is assumed to be stochastically independent and δ_t is assumed to have the symmetric distribution on the set $\{-1, 1\}$.
The conditional variance of ϵ_t is

$$V(\epsilon_t|\epsilon_{t-1}) = c + \sum_{i=1}^{q} a_i \epsilon_{t-i}^2$$

and depends on the past through the q most recent values of ϵ_t^2.

GARCH(p,q) models (Bollerslev 1986)
The ARCH model is based on an autoregressive representation of the conditional variance. Just as AR models can be generalized to ARMA models (which have a nice characterization as being precisely the linear dynamical models which allow for a finite dimensional state space description), the ARCH models can also be generalized by adding an MA(moving average) part.

Stieltjes Instituut ARCH Modelling and Financial Time Series

There are two ways to present the model. The first one is:

$$\begin{aligned} E(\epsilon_t|\underline{\epsilon_{t-1}}) &= 0 \\ V(\epsilon_t|\underline{\epsilon_{t-1}}) &= h_t = \\ c + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 &+ \sum_{j=1}^{p} \beta_j h_{t-j} \end{aligned}$$

In this formulation the source of variability in the variance is in some sense hidden. It just puts conditions on the conditional mean and variance. The following formulation is more explicit in that respect. Introduce the innovation corresponding to the square of the process:

$$u_t = \epsilon_t^2 - h_t$$

Replacing h_t by $\epsilon_t^2 - u_t$ and rearranging one obtains:

$$\epsilon_t^2 = c + \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) \epsilon_{t-i}^2 + u_t - \sum_{j=1}^{p} \beta_j u_{t-j}$$

with $\alpha_i = 0$ for $i > q$ and $\beta_i = 0$ for $i > p$.
So (ϵ_t^2) process is ARMA$[\max(p, q), p]$.

Remark: It can be shown that the GARCH model is second order stationary if $c \geq 0, \alpha_i \geq 0, \beta_i \geq 0$ for all $i = 1, 2, \ldots, \max(p, q)$ and $\Sigma_{i=1}^{\max(p,q)}(\alpha_i + \beta_i) < 1$. In that case the unconditional variance of ϵ_t is equal to

$$E(\epsilon_t^2) = \frac{c}{1 - \Sigma_{i=1}^{\max(p,q)}(\alpha_i + \beta_i)}$$

ARMA-GARCH models(Weiss 1986) As in our earlier example where we had an AR model for Y_t with innovations that were modelled by an ARCH(1) model, one can also have an ARMA model for Y_t combined with a GARCH model for the innovations. Or one may have a regression model for Y_t combined with a GARCH model for the innovations.

Some examples:

- A linear regression model with GARCH errors:
$$Y_t = X_t b + \epsilon_t,$$
where (ϵ_t) satisfies a GARCH(p,q) model.
- An ARMA model with GARCH errors:
$$\Phi(L)Y_t = \Theta(L)\epsilon_t,$$
where (ϵ_t) satisfies a GARCH(p,q) model,
- A further generalization allows for more arbitrary quadratic forms in the past values of the error process on the right-hand side of the equation describing ϵ_t^2. An example of this is the following model:
$$\Phi(L)Y_t = \Theta(L)\epsilon_t,$$
$$V(\epsilon_t|\epsilon_{t-1}) = c + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2$$
$$+\gamma_0[E(Y_t|\underline{Y_{t-1}})]^2 + \sum_{i=1}^{s} \gamma_i Y_{t-i}^2$$

Stieltjes Instituut ARCH Modelling and Financial Time Series

ARCH-M models(Engle, Lilien, and Robbins 1987)

In these models the conditional variance appears as an explanatory variable in the conditional mean. For example;

$$Y_t = X_t b + \delta h_t + \epsilon_t,$$

where (ϵ_t) satisfies a GARCH model.

Stochastic Variance(SV) models

The first version of this model was pioneered by S.J. Taylor, see for example his 1980's book on Modelling Financial Time Series. It was generalized by Harvey, Ruiz, and Shepard(1991) to the following model:

$$\Phi(L)Y_t = \Theta(L)\{\exp\left(\frac{\nu_t}{2}\right)e_t\},$$
$$A(L)\nu_t = B(L)\eta_t,$$

where (e_t) and (η_t) are two independent Gaussian white noises with variance 1 and σ_η^2, respectively.

ESTIMATION

How are we going to estimate ARCH models? In order to do Maximum Likelihood Estimation one would need to specify the distribution of the process. ARCH models specify the behaviour of the conditional mean and the conditional variance. We can, but it is not required, specify the conditional distribution to be Gaussian. If we do so the Maximum Likelihood Estimator can be defined (under certain weak regularity conditions). Even if we do not make the assumption that the conditional distribution is Gaussian, the estimator based on the Gaussian assumption is still well-defined and will be called the Pseudo Maximum Likelihood estimator (PML estimator).

Consider the well-known rule that if $p(y_2, y_1)$ denotes the joint probability density function of two random variables Y_2 and Y_1, then

$$p(y_2, y_1) = p(y_2|y_1)p(y_1)$$

and similarly, in an obvious notation,

$$p(y_3, y_2, y_1) = p(y_3|y_2, y_1)p(y_2(y_1)p(y_1)$$

This holds more generally for any finite time series $Y_1, Y_2, \ldots, Y_{T-1}, Y_T$. Let θ denote the vector of unknown parameters in our model and consider

$$l_t(y; \theta) = p(y_t|\underline{y_{t-1}}; \theta),$$

where $y = (y_1, \ldots, y_T)$ denotes the vector of observations on the finite time series, then $l_t(y; \theta)$ denotes the conditional likelihood function associated with Y_t. It follows that the likelihood function for $Y_1, \ldots, Y_T$ conditional on Y_0 is given by

$$L(y; \theta) = \prod_{t=1}^{T} l_t(y; \theta)$$

The Maximum Likelihood Estimator $\hat{\theta}_T$ is defined as

$$\hat{\theta}_T = \arg\max_{\theta} \log L(y; \theta) =$$

$$\arg\max_{\theta} \sum_{t=1}^{T} \log(l_t(y; \theta))$$

Some remarks can now be made:

- If the likelihood function is well-defined, has a maximum and the maximum is attained in a unique point $\hat{\theta}_T$ then the maximum likelihood estimator is well-defined. In the case of ARCH models with Gaussian conditional distributions, the likelihood function will be well-defined. An analysis of whether the maximum is attained and unique is usually not carried out for finite values of T.

- Finding the actual maximum is in an open problem in general. For ARCH models no simple closed form solutions exist at this point. See further remarks below.

Asymptotic properties of the PLM estimator

- Under standard regularity conditions, this estimator is consistent, even if the underlying distribution is not conditionally Gaussian! In other words,

$$\lim_{T\to\infty} \hat{\theta}_T = \theta$$

where θ denotes the true parameter value.

- The PML estimator is asymptotically normal and its asymptotic covariance matrix V_{asy} is given by

$$V_{asy}[\sqrt{T}(\hat{\theta}_T - \theta)] = J^{-1} I J^{-1}$$

$$J = E_0[-\frac{\partial^2 \log l_t(Y; \theta)}{\partial\theta\partial\theta'}]$$

$$I = E_0\left[\frac{\partial \log l_t(Y; \theta)}{\partial\theta}\frac{\partial \log l_t(Y; \theta)}{\partial\theta'}\right],$$

where E_0 indicates that the expectation is taken with respect to the true distribution. Generally $I \neq J$, because l_t is not the true likelihood but a pseudo-likelihood. They do coincide if the probability density function used in the construction of the likelihood function is the true one of the process(and therefore matches with the expectation operator E_0).

General formulas for PML estimation of ARCH models

Suppose expressions are available for the conditional mean and the conditional variance in terms of the past observations of the output Y_t and past observations of other exogenous variables X_t :

$$E(Y_t|\underline{Y_{t-1}}, \underline{X_t}) = m_t(\theta)$$
$$V(Y_t|\underline{Y_{t-1}}, \underline{X_t}) = h_t(\theta)$$

Note that m_t and h_t depend on the past observations of the output and the past and present observations of the exogenous variable. ARCH models all fit into this framework, also ARCH models with exogenous variables. Now the pseudo likelihood function can be formed:

$$\log L = -\frac{1}{2}\sum_{t=1}^{T}\log(h_t(\theta)) - \frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\frac{[Y-t-m_t(\theta)]^2}{h_t(\theta)}$$

First order conditions are obtained by setting the partial derivatives of the likelihood function with respect to the parameters equal to zero. (It is implicitly assumed that the restrictions on the parameters will not be binding in the optimum)

$$0 = \frac{\partial \log L}{\partial \theta} = \frac{1}{2}\sum_{t=1}^{T}\frac{1}{h_t(\theta)}\frac{\partial h_t(\theta)}{\partial \theta} + \frac{1}{2}\sum_{t=1}^{T}\frac{[Y_t - m_t(\theta)]^2}{h_t^2(\theta)}\frac{\partial h_t(\theta)}{\partial \theta} + \sum_{t=1}^{T}\frac{Y_t - m_t(\theta)}{h_t(\theta)}\frac{\partial m_t(\theta)}{\partial \theta}$$

Remark: If $h_t(\theta)$ and $m_t(\theta)$ are rational functions of the parameters, then the first order equations lead to a system of polynomial equations in the parameters. In principle modern polynomial algebra methods could be used to solve this, but at present the algorithms for doing that can handle only relatively modest systems of polynomial equations. Therefore present-day algorithms would most likely(!) require too much memory and/or be too time-consuming.

Illustrations:

Consider the following regression model with ARCH errors:

$$Y_t = X_t b + \epsilon_t,$$
$$V(\epsilon_t|\underline{\epsilon_{t-1}} = c + a_1\epsilon_{t-1}^2 + \ldots + a_p\epsilon_{t-p}^2,$$

where $E(\epsilon_t|\underline{\epsilon_{t-1}}) = 0$.

In this model

$$m_t(\theta) = X_t b$$
$$h_t(\theta) = c + a_1(Y_{t-1} - X_{t-1}b)^2 + \ldots + (Y_{t-p} - X_{t-p}b)^2,$$

where

$$\theta' = (b', c, a_1, \ldots, a_p) = (b', c, a')$$

Next we consider a conditionally Gaussian GARCH(p,q) model:

$$Y_t|\underline{Y_{t-1}} \sim N(0, h_t)$$

where

$$h_t = c + \sum_{i=1}^{p}\alpha_i Y_{t-i}^2 + \sum_{j=1}^{q}\beta_j h_{t-j}$$

Expression in terms of parameters and observable variables is:

$$h_t = \frac{1}{1 - \Sigma_{j=1}^{q}\beta_j L^j}\left(c + \sum_{i=1}^{p}\alpha_i Y_{t-i}^2\right),$$

where L denotes the lag operator. Thus $h_t(\theta)$ depends on all past values of the Y process.

In practice it is necessary to replace $h_t(\theta)$ by its truncated approximation in which the values Y_t^2 corresponding to negative dates are set equal to zero.

Consider the stochastic variance model where the process Y_t is a white noise:

$$Y_t = \theta \exp\left(\frac{\nu_t}{2}\right) e_t$$
$$A(L)\nu_t = B(L)\eta_t$$

where $(e_t), (\eta_t)$ are two independent Gaussian processes with variance 1 and σ_η^2, respectively. The sign of Y_t has no information concerning the parameter values, therefore without loss of generality we can square the equation for Y_t.

Then taking the logarithm ($Y_t \neq 0$ with probability one) one obtains the following model:

$$\log(Y_t^2) = c + \nu_t + \left(\log(e_t^2) - E\log(e_t^2)\right),$$
$$A(L)\nu_t = B(L)\eta_t$$

where $c = \log\theta^2 + E\log(e_t^2) \approx \log(\theta^2) + 1.27$. This system is a state space model with state variable ν_t and $\log(Y_T^2)$ as the observed variable. The equality $A(L)\nu_t = B(L)\eta_t$ describes the dynamics of the state variable, while $\log(Y_t^2) = c + \nu_t + \left(\log(e_t^2 - E\log(e_t^2)\right)$ is the measurement equation. The parameters can be estimated by PML based on the normality assumption of the joint process $(\log(e_t^2) - E\log(e_t^2), \nu_t)$ using the Kalman filter. However, here it is known that $\log(e_t^2) - E\log(e_t^2)$ has a log-gamma type distribution and the assumption of normality in the PML leads to efficiency loss.

Some references

C. Gouriéroux, *ARCH Models and Financial Applications*, Springer, 1997.
R.F. Engle, *Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation,* Econometrica, vol. 50, pp.987-1008.
T. Bollerslev, *Generalized Autoregressive Conditional Heteroscedasticity,* Journal of Econometrics, vol. 31, pp. 307-327.
R.F. Engle, D. Lilien, R. Robbins, *Estimating Time Varying Rsik Premia in the Term Structure: the ARCH-M Model,* Econometrica, vol. 55, pp.391-407.
A. Harvey, E. Ruiz, N. Shepard, *Multivariate Stochastic Variance Models,* Review of Economic Studies, vol. 61, pp. 247-264.