Chapter 7:

Volatility Forecasting
Contents:

VII. Volatility Forecasting – ARCH/GARCH Models ................................................................................. 4

VII.1 ARCH Models .................................................................................................................................. 4
  VII.1.1 ARCH(q) Model .......................................................................................................................... 6
  VII.1.2 Testing for ARCH Effects ............................................................................................................ 10
  VII.1.3 Estimation of ARCH Models ....................................................................................................... 13

VII.2 GARCH ............................................................................................................................................. 17
  VII.2.1 GARCH(p,q) Model .................................................................................................................... 17
  VII.2.2 Unconditional Standard Deviation in GARCH ............................................................................ 19
  VII.2.2 Unconditional Standard Deviation in GARCH with Exogenous Variables .............................. 23
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>VII.3</td>
<td>Asymmetric ARCH(q) Models</td>
<td>25</td>
</tr>
<tr>
<td>VII.3.1</td>
<td>EGARCH</td>
<td>25</td>
</tr>
<tr>
<td>VII.3.2</td>
<td>TARCH</td>
<td>28</td>
</tr>
<tr>
<td>VII.4</td>
<td>Other GARCH-Models</td>
<td>30</td>
</tr>
<tr>
<td>VII.5</td>
<td>Evaluation Ratios for the Goodness of Forecast</td>
<td>32</td>
</tr>
</tbody>
</table>
VII. Volatility Forecasting – ARCH/GARCH Models

VII.1 ARCH Models

In a general form ARCH models can be described as:

\[
\ln \frac{S_t}{S_{t-1}} = r_t = \mu + \varepsilon_t
\]

\[
\varepsilon_t = \sigma_t \Omega_t
\]

\[
\Omega \sim D(0,1)
\]

\[
\sigma_t^2 | \phi_{t-1} = f(\phi_{t-1})
\]

\(r_t\) denotes the return of the financial asset, \(\mu\) stands for the mean of the returns and \(\varepsilon_t\) for the price innovation (error term). \(\Omega_t\) is a random variable, which is independently and identically distributed over time. It follows a distribution \(D\) with an expected value of 0 and a variance of 1.
For simplification we assume in the following that $\Omega_t$ follows a standard normal distribution, $\Omega_t \sim N(0,1)$. $\phi_{t-1}$ characterise the set of information that is available at time point $t-1$. The information set contains among other things the past price changes, i.e. $\phi_{t-1} = \{r_1, \ldots, r_{t-1}\}$. It also possible that the information set contains additionally the volume of trade.

Thus, the conditional variance $\sigma_t^2$ is a deterministic function of lagged price changes. This means that the variance $\sigma_t^2$ is conditional with regard to the knowledge of past price observations. This explains the third character in the abbreviation ARCH. The last one stands for heteroskedasticity – the variance is not constant over time.
VII.1.1 ARCH(q) Model

In the ARCH(q) model the conditional variance $\sigma_t^2$ is a linear function of q time lagged squared innovations (error terms):

$$\sigma_t^2 = a_0 + \sum_{i=1}^{q} a_i \epsilon_{t-i}^2$$

Due to the parameter restrictions $a_0 > 0$, $a_i \geq 0$, $t = 1, \ldots, q$ and the consideration of lagged price innovations $\epsilon_{t-i}$ in the square the conditional variance $\sigma_t^2$ is always positive. The equation above constitutes a moving average (MA) process for the variance and no AR process as the name ARCH suggests (MACH).
According to the definition of $\nu_t = \varepsilon_t^2 - \sigma_t^2$ the equation can be rewritten:

$$\varepsilon_t^2 = a_0 + \sum_{i=1}^{q} a_i \varepsilon_{t-i}^2 + \nu_t$$

where $E[\nu_t] = 0$ and $\nu_t$ are not autocorrelated. This equation represents an AR process for $\varepsilon_t^2$. An ARCH(q) process is stationary iff (if and only if) the condition

$$\sum_{i=1}^{q} a_i < 1$$

is satisfied. In this case the stationary (unconditional) variance $\overline{\sigma}^2 = \frac{a_0}{1-\sum a_i}$. 
The functionality of an ARCH(q) model can be shown with an ARCH(1) model by using the stationary variance. For the conditional variance $\sigma_t^2$ it follows:

$$\sigma_t^2 = a_0 + a_1 \varepsilon_{t-1}^2 = \frac{a_0}{1 - a_1} + a_1 \left( \varepsilon_{t-1}^2 - \frac{a_0}{1 - a_1} \right) = \bar{\sigma}^2 + a_1 (\varepsilon_{t-1}^2 - \bar{\sigma}^2)$$

The conditional variance $\sigma_t^2$ is the unconditional variance $\bar{\sigma}^2$ plus the $a_1$-weighted difference between the squared innovation $\varepsilon_{t-1}^2$ and the unconditional variance $\bar{\sigma}^2$. This means that after a large price change according to amount a high volatility is expected and respectively after a low price change a low volatility. Hence, ARCH models result in volatility clustering. The conditional variance fluctuates around the unconditional variance depending on positive or negative weighted difference. This means the conditional variance exhibits a mean reverting behaviour.
The kurtosis of the unconditional distribution of a ARCH(1) model with a conditional normal distribution for \(0 < a_1 < \sqrt{1/3}\) is:

\[
Kurt(\varepsilon_t) = \frac{E[\varepsilon_t^4]}{E[\varepsilon_t^2]^2} = 3 \frac{1 - a_1^2}{1 - 3a_1^2} > 3
\]

For \(a_1 \geq \sqrt{1/3}\) it follows that \(Kurt(\varepsilon_t) = \infty\) and for the marginal case \(a_1 = 0\) the \(Kurt(\varepsilon_t) = 3\), whereas for \(a_1 > 0\) we receive a leptokurtic unconditional distribution of returns, which are characterized by a concentration around the mean and fat tails in comparison to the normal distribution.
VII.1.2 Testing for ARCH Effects

For the identification of ARCH effects we first have to look at the descriptive statistics:

1. We have to observe a significant departure from normal distribution especially leptokurtosis by using Jarque-Bera test.

2. We have to observe volatility clustering, which we are testing by looking at the autocorrelation function. First we consider the returns and if we do not reject the null hypothesis generally we conclude that there is no autocorrelation. However, it is wrong to conclude from uncorrelatedness to independence of price changes. If there exist independent random variables then the square of the random variables are also independent.
With the ARCH Lagrange multiplier (LM) test the existence of ARCH effects can be tested. The null hypothesis of no ARCH(q) effects is equivalent to the null hypothesis the coefficients $a_1, \ldots, a_q$ in

$$\sigma_t^2 = a_0 + \sum_{i=1}^{q} a_i \varepsilon_{t-i}^2$$

are not significant different from zero:

$$H_0: a_1 = \ldots = a_q = 0$$

Engle (1984) has shown that the test statistic of the LM test is asymptotically equivalent to $T \cdot R^2$, where $T$ is the number of return observations and $R^2$ the unadjusted coefficient of determination from the auxiliary regression:

$$\varepsilon_t^2 = a_0 - a_1 \varepsilon_{t-1}^2 + \cdots + a_q \varepsilon_{t-q}^2 + u_t$$

$u_t$ denotes an error term.
The test statistic $T \cdot R^2$ follows a $\chi^2$ distribution with $q$ degrees of freedom and is equivalent to

$$T \sum_{k=1}^{q} \hat{\rho}^2(k)$$

Here $\hat{\rho}^2(k)$ denotes the estimated autocorrelation coefficient of the squared estimated price innovations $\varepsilon_t^2$ to the lag $k$. 
VII.1.3 Estimation of ARCH Models

The standard method to estimate ARCH processes is the maximum likelihood approach. Under the assumption of a conditional normal distribution of $\varepsilon_t$ we receive the conditional density of the $t^{th}$ observation as follows:

$$f_t(\varepsilon_t | \phi_{t-1}, \theta) = \frac{1}{\sqrt{2\pi}\sigma_t} \exp\left(-\frac{1}{2}\frac{\varepsilon_t^2}{\sigma_t^2}\right)$$

with $\theta = (a_0, \ldots, a_q)$ as estimated parameter vector. Thus the logarithmic conditional density $I_t$ is:

$$I_t = \ln(f_t(\varepsilon_t | \phi_{t-1}, \theta)) = -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln\sigma_t^2 - \frac{1}{2} \frac{\varepsilon_t^2}{\sigma_t^2}$$
The likelihood function for all the observations is

$$L(\theta) = \prod_{t=1}^{T} f_t(\varepsilon_t | \Phi_{t-1}, \theta),$$

with $T$ the number of observations. The logarithm of this function is called Log-Likelihood function and is calculated for conditional normal distributed $\varepsilon_t$ as

$$\ln L(\theta) = \sum_{t=1}^{T} I_t = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^{T} \left( \ln \sigma_t^2 - \frac{\varepsilon_t^2}{\sigma_t^2} \right).$$
The aim of the maximum likelihood approach is the maximization of this equation subject to the parameter vectors $\theta$. The first derivative of the log likelihood function with respect to the elements of the parameter vectors $\theta$ is

$$\frac{\partial I_t}{\partial a_i} = \frac{1}{2\sigma_t^2} \frac{\partial \sigma_t^2}{\partial a_i} \left( \frac{\varepsilon_t^2}{\sigma_t^2} - 1 \right) \forall i = 0, \ldots, q$$

with the derivatives $\frac{\partial \sigma_t^2}{\partial a_i} = \varepsilon_{t-i}^2$ for $i = 1, \ldots, q$ and $\frac{\partial \sigma_t^2}{\partial a_0} = 1$. 
The elements of the Hesse matrix are given by

\[
\frac{\partial^2 I_t}{\partial a_i \partial a_j} = - \frac{1}{2\sigma_t^2} \frac{\partial^2 \sigma_t^2}{\partial a_i \partial a_j} \sigma_t^2 + \left( \frac{\varepsilon_t^2}{\sigma_t^2} - 1 \right) \frac{\partial}{\partial a_j} \left( \frac{1}{2\sigma_t^2} \frac{\partial \sigma_t^2}{\partial a_j} \right) \quad \forall \ i, j \quad i = 0, ..., q; \ j = 0, ..., q
\]

From the inverse of the Hesse-Matrix the standard error of the estimated parameter vector can be estimated. Several empirical studies has shown that (Berndt, Hall, Hall and Haussman) BHHH-algorithm is most appropriate to maximize the Log likelihood function.
VII.2 GARCH

VII.2.1 GARCH(p,q) Model

In previous studies of ARCH models with financial data one could observe that a large number q of time lagged terms are required to represent heteroskedasticity adequately. Due to complexity of the estimation algorithm and the problem of the restriction of non-negativity for increasing number of parameters Engle (1982) proposes the following parameterisation for conditional variance:

\[ \sigma_t^2 = a_0 + a_1 \sum_{i=1}^{q} w_i \varepsilon_{t-i}^2, \]

with the weights

\[ w_i = \frac{(q + 1) - i}{\frac{1}{2} q(q + 1)}. \]
Bollerslev (1986) defines the conditional variance of a GARCH(p,q) model as:

\[ \sigma_t^2 = a_0 + \sum_{i=1}^{q} a_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} b_j \sigma_{t-j}^2 \]

Sufficient conditions for the non-negativity of \( \sigma_t^2 \) are:

\[ a_0 > 0 \]
\[ a_i \geq 0 \text{ for } i = 1, \ldots, q \]
\[ b_j \geq 0 \text{ for } j = 1, \ldots, p \]

According to the theorem 1 of Bollerslev (1986) a GARCH(p,q) process is stationary iff

\[ \sum_{i=1}^{q} a_i + \sum_{j=1}^{p} b_j < 1. \]
VII.2.2 Unconditional Standard Deviation in GARCH

The stationary (unconditional) variance $\sigma^2$ of a GARCH($p,q$) process is

$$\sigma^2 = \frac{a}{1 - \sum_{i=1}^{q} a_i - \sum_{j=1}^{p} b_j}.$$
By transforming the GARCH(p,q) model in a equivalent ARCH(∞) model we receive more insides in
the structure of the parameter restrictions of GARCH(p,q) models. By a recursive substitution of \( \bar{\sigma}^2 \)
we receive the GARCH(1,1) model:

\[
\sigma_t^2 = a_0 + a_1 \varepsilon_{t-1}^2 + b_1 \sigma_{t-1}^2 \\
= a_0 + a_1 \varepsilon_{t-1}^2 + b_1 (a_0 + a_1 \varepsilon_{t-2}^2 + b_1 \sigma_{t-2}^2) \\
... \\
= \frac{a_0}{1 - b_1} + a_1 \sum_{i=1}^{\infty} b_1^{i-1} \varepsilon_{t-i}^2
\]
Thus, a GARCH(1,1) process equals an ARCH(∞) process with geometric decreasing weights. The general GARCH(p,q) can be written as:

$$\sigma_t^2 = \frac{a_0}{1 - \sum_{j=1}^{p} b_j} + \sum_{i=1}^{\infty} \lambda_i \varepsilon_{t-i}^2$$

where $\lambda_i$ is a function of $a_i$ and $b_j$. The parameter restrictions are sufficient to guarantee a positive conditional variance. Nelson and Cao (1992) also showed that the weaker restrictions

$$\frac{a_0}{1 - \sum_{j=1}^{p} b_j} > 0$$

and $\lambda \geq 0, i = 1, ..., \infty$ also are sufficient conditions for a positive conditional variance. For a GARCH(1,2) process are the conditions $a_0 > 0$, $a_1 \geq 0$, $b_1 \geq 0$ and $a_1 b_1 + a_2 \geq 0$ sufficient for $\sigma_t^2 > 0$. The parameter $a_2$ can necessarily take negative values.
Example:

\[ \sigma_t^2 = 7.31 \cdot 10^{-6} + 0.130 \epsilon_{t-1}^2 + 0.834 \sigma_{t-1}^2 \]

The ARCH effects are clear, when you look at the high significance of the parameter estimation of \( a_1 \). Similarly, the influence of \( b_1 \) of the variance of the previous period \( \sigma_{t-1}^2 \) is quite high. The condition of stationarity is fulfilled because \( a_1 + b_1 = 0.130 + 0.834 = 0.964 < 1 \).
VII.2.2 Unconditional Standard Deviation in GARCH with Exogenous Variables

In GARCH models with exogenous variables, the asymptotic standard deviations of the residuals depend on the stationary marginal distribution of the exogenous variables. However, if the exogenous variables are deterministic functions, which could be used to account for outliers or level shifts in the variance, then the formula for the asymptotic variance can be easily modified. For example, in a GARCH(1,1) model, with the variance equation

\[ \sigma_t^2 = w + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \]

the unconditional variance is simply

\[ \frac{w}{1 - \alpha - \beta}. \]
When you now include an stationary exogenous variable $x(t)$, the variance equation becomes

$$\sigma_t^2 = w + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 + \delta x_{t-1},$$

and the unconditional variance is

$$\frac{w + \delta E[x_t]}{1 - \alpha - \beta}.$$
VII.3 Asymmetric ARCH(q) Models

VII.3.1 EGARCH

In models where ARCH processes are estimated by using the conditional variance or conditional standard deviation the problem of parameter restrictions appears. To guarantee a positive conditional variance or rather standard deviation the admissible co-domain for the estimated parameters must be restricted. To circumvent this problem Nelson (1991) introduce a transformation of the conditional variance by modeling the logarithm of the conditional variance. The EGARCH(p,q) (Exponential GARCH) model is defined as:

\[ \ln \sigma_t^2 = a_0 + \sum_{i=1}^{q} a_i \left( \frac{\varepsilon_{t-i}}{\sigma_{t-i}} - E \left[ \frac{\varepsilon_{t-i}}{\sigma_{t-i}} \right] \right) + \sum_{j=1}^{p} b_j \ln \sigma_{t-j}^2 \]
As you can see from the equation, standardized error terms $\frac{\varepsilon_{t-i}}{\sigma_{t-i}}$ are employed. In the case of a conditional normal distribution for $\varepsilon_{t-i}$ the $\frac{\varepsilon_{t-i}}{\sigma_{t-i}}$ are standard normal distributed and

$$E\left[\frac{\varepsilon_{t-i}}{\sigma_{t-i}}\right] = \sqrt{\frac{2}{\pi}} \approx 0.798$$

The parameters $a_i$ determine the influence of the value of $i$ periods past price changes and $a_i \cdot c$ determines the influence of the sign. Empirical investigations of equity data have found that already a simple EGARCH(1,1) model is adequate describing the dynamic of stock price changes.
In the EGARCH(1,1) model is the influence of price changes $\varepsilon_{t-1}$ on the logarithmic conditional variance:

$$a_1 (1 + c) \left( \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right) \text{ for } \varepsilon_{t-1} \geq 0$$

and

$$a_1 (1 - c) \left( \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right) \text{ for } \varepsilon_{t-1} < 0.$$

Typically, in empirical studies we receive for $a_1$ positive and for $c$ negative parameter values. The parameter $a_1$ represents the volatility clustering, whereas $c$ models the leverage effect. Parameter estimations of $c \leq -1$ were not previously documented. This would conflict with the economic intuition. If $c \leq -1$ then the term in the equation is smaller than zero, this means that the greater the price changes $\varepsilon_{t-1}$ the smaller the conditional variance $\sigma_t^2$ in the next period.
VII.3.2 TARCH
In the model of the Glosten, Jaganathan, and Runkle (1993) the asymmetric influence of negative interferences on the conditional variance is reproduced by a dummy variable. In the case of a negative shock, this variable takes a value of one, and in the case of a positive shock the value of zero. Thus, the conditional variance increases only by the factor of the dummy variable, if the disturbance are negative. The TGARCH(p,q) model of Glosten, Jaganathan, and Runkle (1993) is

\[ \sigma_t^2 = a_0 + \sum_{i=1}^{q} a_i \varepsilon_{t-i}^2 + \gamma \varepsilon_{t-i}^2 d_{t-1} + \sum_{j=1}^{p} b_j \sigma_{t-j} \]

where \( d_{t-i} = 1 \) if \( \varepsilon_{t-i} < 0 \), and 0 otherwise (in EViews available).
In this model, bad news ($\varepsilon_{t-i} < 0$), and good news ($\varepsilon_{t-i} > 0$), have differential effects on the conditional variance - good news has an impact of $a_i$, while bad news has an impact of $a_i$ and $\gamma$. If $\gamma > 0$ we say that the leverage effect exists. If $\gamma \neq 0$, the news impact is asymmetric.

Zakoian (1994) developed another version of the TGARCH($p$,q) model with asymmetric influences of negative disturbances. Here, instead of the conditional variance the conditional standard deviation is modelled, and the deviations are incorporated in the variance equation with positive and negative signs, in fact in the same way as the deviations actually appears. The TGARCH($p$,q) of Zakoian can be written as:

$$\sigma_t^2 = a_0 + \sum_{i=1}^{q} a_i^+ \varepsilon_{t-i}^+ + \gamma^- \varepsilon_{t-i}^- + \sum_{j=1}^{p} b_j \sigma_{t-j}^2$$

where $\varepsilon_{t-i}^+ = \max(\varepsilon_t, 0)$, $\varepsilon_{t-i}^- = \min(\varepsilon_t, 0)$, $a_0 > 0$, $a_i^+ \geq 0$, $\gamma^- \geq 0$ and $b \geq 0$. 
VII.4 Other GARCH-Models

The ARCH-in-Mean (ARCH-M) Model

Engle, Lilien and Robins (1987) give up the assumption that the conditional expected value of return $\mu$ is constant and expand the mean equation. In an ARCH-M model, the conditional mean $\mu_t$ is an explicit function of the conditional variance:

$$\mu_t = \mu + \lambda \cdot g(\sigma_t)$$

where $\lambda$ is constant. The modeling of the conditional variance does not differ from the ARCH models with a constant $\mu$. The difference between ARCH and ARCH-M model refers only to the mean equation. Typically, as functional term of $g(\sigma_t)$ linear or logarithm functions of $\sigma_t$ or $\sigma_t^2$ are used. Normally $g(\sigma_t) = \sigma_t$, $g(\sigma_t) = \sigma_t^2$ or $g(\sigma_t) = \ln(\sigma_t^2)$. The term $\lambda \cdot g(\sigma_t)$ can be interpreted as risk premium.
or rather $\lambda$ as market price of risk. However, a non-significant parameter estimation of $\lambda$ can not be interpreted as an absence of a risk premium given that there exists a nonlinear risk premium.
VII.5 Evaluation Ratios for the Goodness of Forecast

The usual ratios to check the goodness of a forecast are

1. Mean forecast error (Mean error, ME)

2. Mean quadratic forecast error (Mean squared error, MSE)

3. Mean absolute forecast error (Mean absolute error, MAE)