Lecture 14: Conditional Heteroscedasticity

Bus 41910, Time Series Analysis, Mr. R. Tsay

The introduction of conditional heteroscedastic autoregressive (ARCH) models by Engle (1982) popularizes conditional variance modeling in the literature. The idea of allowing the conditional variance to depend on the conditional mean of a random variable has been around for some time, even in linear regression analysis. However, the ARCH model represents a neat and parsimonious way to describe the conditional heteroscedasticity. It leads to the development of various conditional heteroscedastic models.

The linear time series models considered so far in this course, such as the ARIMA models, assume that both the conditional and unconditional variances of a process are time invariant. In reality, we often see clusters of high variabilities. For instance, Mandelbrot (1963) states ".... large changes tend to be followed by large changes – of either sign – and small changes by small changes ...". The ARCH model is a simple parametric model to describe such clustering phenomena.

For a time series Z_t, the linear models discussed so far focus on modeling $E(Z_t|\psi_{t-1})$, where ψ_{t-1} denotes the information available at $t-1$, inclusive. For the ARCH-type models, we focus on $\mathrm{Var}(Z_t|\psi_{t-1})$. For a linear model with innovation e_t, we have $\mathrm{Var}(Z_t|\psi_{t-1}) = \mathrm{Var}(e_t|\psi_{t-1})$. Thus, all the models discussed below deal with $e_t = \sigma_t a_t$, where $\{a_t\}$ is a sequence of independent and identically distributed random variables with mean zero and variance 1 and $\sigma_t^2 = \mathrm{Var}(e_t|\psi_{t-1})$. In other words, we are studying the time evolution of the conditional variance σ_t^2. See Chapters 3 and 4 of Tsay (2005) for further information and real examples.

A. ARCH Model

An innovational process $\{e_t\}$ is a Gaussian ARCH(r) process if it can be written as

- $e_t = \sigma_t a_t$, where a_t's are *iid* N(0,1) random variates;
- $\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \cdots + \alpha_r e_{t-r}^2$, where $\alpha_0 > 0$ and $\alpha_i \geq 0$ for $i > 0$.

Remark: The non-negativeness condition $\alpha_0 > 0$, $\alpha_i \geq 0$ can be relaxed. It is a condition to ensure that the conditional variance is postive. In fact, a general form for the ARCH(r) process can be written as

$$\begin{aligned} \sigma_t^2 &= \text{Quadratic function of } e_{t-1}, \cdots, e_{t-r} \\ &= \alpha_0 + \boldsymbol{e}'_{r,t-1} \boldsymbol{G} \boldsymbol{e}_{r,t-1} \end{aligned}$$

where $\boldsymbol{e}_{r,t-1} = (e_{t-1}, \cdots, e_{t-r})'$ and $\boldsymbol{G}$ is a $r \times r$ non-negative definite matrix. Engle's ARCH model requires that $\boldsymbol{G}$ is diagonal so that it is a parsimonious approach to have a quadratic function.

A special case: Consider the case of $r = 1$. Then,

$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2.$$

It is easily seen that

$$E(e_t) = 0, \quad \text{Var}(e_t) = \frac{\alpha_0}{1-\alpha_1}$$

provided that $\alpha_1 < 1$. Further, it can be shown that

$$E(e_t^4) = \frac{3\alpha_0^2}{(1-\alpha_1)^2}\frac{1-\alpha_1^2}{1-3\alpha_1^2} = 3[\text{Var}(e_t)]^2 \times \frac{1-\alpha_1^2}{1-3\alpha_1^2},$$

provided that $3\alpha_1^2 < 1$. This equation shows that for $\alpha_1 \neq 0$ the fourth moment of e_t is greater than that of a normal random variable, implying that the e_t process is heavy-tailed and, hence, it is capable of producing clusters of outliers.

Next, the likelihood function of $e_1, \cdots, e_n$ is

$$\begin{aligned} f(e_1, \cdots, e_n|\alpha_0, \alpha_1) &= f(e_n|e_{n-1})f(e_{n-1}|e_{n-2})...f(e_2|e_1)f(e_1) \\ &= \prod_{t=2}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp[-\frac{1}{2\sigma_t^2}e_t^2] \times \frac{\sqrt{1-\alpha_1}}{\sqrt{2\pi\alpha_0}} \exp[-\frac{1-\alpha_1}{2\alpha_0}e_1^2], \end{aligned}$$

where we assume that the marginal distribution of e_0 is Gaussian. The log-likelihood function is

$$\ell(\alpha_0, \alpha_1) \propto -\frac{1}{2}\sum_{t=2}^{n} \ln \sigma_t^2 - \frac{1}{2}\sum_{t=2}^{n} \frac{e_t^2}{\sigma_t^2} - \frac{1}{2}\ln\frac{\alpha_0}{1-\alpha_1} - \frac{(1-\alpha_1)e_1^2}{2\alpha_0}$$

where $\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2$. This log-likelihood function can be maximized by non-linear optimization routines. In practice, it may be informative to parameterize α_0 and α_1 as β_0^2 and β_1^2 so that the non-negativeness condition is satisfied. It is common that the marginal distribution of e_1 is omitted from the likelihood function, resulting in using conditional maximum likelihood estimates.

In summary, this simple model is informative. It says that the process e_t is weakly stationary if $\alpha_1 < 1$ and is heavy-tailed if $3\alpha_1^2 < 1$. The log-likelihood function of the model is no more complicated than those of ARMA models. For the asymptotic distribution of MLE of ARCH models, see Weiss (1986, JBES).

Remark: The innovation a_t of an ARCH model may assume other distributions such as Student-t, skewed Student-t, generalized error distribution. In particular, these latter distributions have heavy tails, which are commonly seen in financial time series.

Remark: Recall that a random variable X has a Student-t distribution with v degrees of freedom if the probability density function of X is

$$f(x) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{v\pi}\Gamma(\frac{v}{2})}\left(1+\frac{x^2}{v}\right)^{-(v+1)/2}, \quad x \in (-\infty, \infty).$$

For such a distribution, we have $E(X) = 0$ and $\text{Var}(X) = v/(v-2)$ for $v > 2$. This is a symmetric distribution. But it can be generalized to *Skewed* t-distribution as follows:

A random variable X has a skewed Student-t distribution with degrees of freedom v if its probability density function is

$$g(x) = \begin{cases} \frac{2}{\gamma+\frac{1}{\gamma}} f(\gamma x) & \text{if } x < 0 \\ \frac{2}{\gamma+\frac{1}{\gamma}} f(x/\gamma) & \text{if } x \geq 0 \end{cases}$$

where $f(x)$ is the pdf of a Student-t distribution with v degrees of freedom.See Fernandez and Steel (1998,JASA, p. 359-371.) There is a R package, called **skewt**, that deals with this skewed Student-t distribution. To gain some insights, see the demonstration below:

```
> library(fBasics)  <== To compute summary statistics.
> library(skewt)
> x=rskt(1000,5,0.5)   <== Gamma = 0.5
> y=rskt(1000,5,2)     <== Gamma = 2.0

> basicStats(x)
                        x
nobs          1000.000000
NAs              0.000000
Minimum        -12.675984
Maximum          2.504241
1. Quartile     -2.308930
3. Quartile     -0.214185
Mean            -1.468837
Median          -1.122484
Sum          -1468.836694
SE Mean          0.055915
LCL Mean        -1.578561
UCL Mean        -1.359113
Variance         3.126468
Stdev            1.768182
Skewness        -1.483462
Kurtosis         4.074447
> basicStats(y)
                       y
nobs         1000.000000
NAs             0.000000
Minimum        -2.036304
Maximum        10.233795
1. Quartile     0.223086
3. Quartile     2.370258
Mean            1.477367
Median          1.060711
```

```
Sum          1477.367416
SE Mean         0.056367
LCL Mean        1.366756
UCL Mean        1.587979
Variance        3.177263
Stdev           1.782488
Skewness        1.317663
Kurtosis        2.396833
> hist(x)  <== Not shown.
> hist(y)
```

For a further generalization of skewed t-distribution, see (1) Jones and Faddy (2003, JRSSB, Vol. 65, p. 159-174.) and (2) Theodossiou (1998, *Management Science*, Vol. 44, p. 1650-1661.).

Remark: The ARCH model can be treated as a random coefficient model because

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{r} \alpha_i e_{t-i}^2 \\ &= \alpha_0 + \sum_{i=1}^{n} \alpha_i a_{t-i}^2 \sigma_{t-i}^2 \end{aligned}$$

where a_{t-i}'s are *iid* N(0,1) so that a_{t-i}^2 are *iid* χ_1^2. We shall consider this link later.

Exercise: Generalize the above discussion to the pure ARCH(r) model.

B. GARCH Models

Bollerslev (1986) generalizes the pure ARCH model to the GARCH model. An innovational process $\{e_t\}$ is a GARCH(r, s) model if

- $e_t = \sigma_t a_t$, where a_t's are *iid* N(0,1).
- $\sigma_t^2 = \alpha_0 + \sum_{i=1}^r \alpha_i e_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2$, where $\alpha_0 > 0$, $\alpha_i \geq 0$ and $\beta_j \geq 0$ for $i, j > 0$.

This generalization is similar in spirit to that of MA models to ARMA models. Intuitively, consider the pure ARCH(∞) model

$$\sigma_t^2 = \beta_0 + \beta_1 e_{t-1}^2 + \beta_2 e_{t-2}^2 + \cdots.$$

If $\beta_i = \beta_1^i$ for $i \geq 1$, then we have

$$\begin{aligned} \sigma_t^2 &= \beta_0 + [\beta_1 + \beta_1^2 B + \beta_1^3 B^2 + \cdots] e_{t-1}^2 \\ &= \beta_0 + \beta_1 [1 + \beta_1 B + \beta_1^2 B^2 + \cdots] e_{t-1}^2 \\ &= \beta_0 + \frac{\beta_1}{1 - \beta_1 B} e_{t-1}^2, \end{aligned}$$

which implies

$$(1-\beta_1 B)\sigma_t^2 = (1-\beta_1 B)\beta_0 + \beta_1 e_{t-1}^2$$

that is

$$\sigma_t^2 = \alpha_0 + \beta_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

where $\alpha_0 = \beta_0(1-\beta_1)$. The above equation is in the form of GARCH(1,1) model, even though the coefficients of e_{t-1}^2 and σ_{t-1}^2 are identical.

This also highlights certain features of GARCH models. First of all, rewrite the above equation as

$$\sigma_t^2 = \alpha_0 + \beta_1 a_{t-1}^2 \sigma_{t-1}^2 + \beta_1 \sigma_{t-1}^2 = \alpha_0 + \beta_1(1 + a_{t-1}^2)\sigma_{t-1}^2.$$

This equation says that the GARCH model in effect is a *Random Coefficient ARCH* model with χ_1^2 random variates, because a_t's are *iid* N(0,1). In general, let $p = \max\{r, s\}$ and $\alpha_i = 0$ for $i > r$ and $\beta_j = 0$ for $j > s$, then we have

$$\begin{aligned}
\sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i a_{t-i}^2 \sigma_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \\
&= \alpha_0 + \sum_{i=1}^{p} [\beta_i + \alpha_i a_{t-i}^2] \sigma_{t-i}^2
\end{aligned}$$

which is a random coefficient AR-type of equation.

Alternatively, let $\eta_t = e_t^2 - \sigma_t^2$. It is easily seen that $E(\eta_t|\psi_{t-1}) = E(e_t^2 - \sigma_t^2|\psi_{t-1}) = E(e_t^2|\psi_{t-1}) - \sigma_t^2 = \sigma_t^2 - \sigma_t^2 = 0$. Consequently, we have

$$\begin{aligned}
E(\eta_t) &= E_{t-1}[E(\eta_t|\psi_{t-1})] = 0 \\
E(\eta_t \eta_{t-j}) &= E_{t-1}[\eta_{t-j} E(\eta_t|\psi_{t-1})] = 0, \quad j > 0.
\end{aligned}$$

Thus, $\{\eta_t\}$ is a sequence of uncorrelated random variables with mean zero. The unconditional variance of η_t, however, is not constant over time.

Let $m = \max\{r, s\}$ and define $\alpha_j = 0$ for $j > r$ and $\beta_j = 0$ for $j > s$. Using $\sigma_{t-j}^2 = e_{t-j}^2 - \eta_{t-j}$, the GARCH$(r, s)$ model becomes

$$e_t^2 = \alpha_0 + \sum_{j=1}^{m} (\alpha_j + \beta_j) e_{t-j}^2 + \eta_t - \sum_{j=1}^{s} \beta_j \eta_{t-j},$$

which is in the form of an ARMA(m, s) model for e_t^2 with $\{\eta_t\}$ as the innovations. Thus, GARCH models can be regarded as an ARMA model for the sequared innovation series $\{e_t^2\}$. Some properties of GARCH models can then be obtained in a similar manner as those of ARMA models. For instance, consider the GARCH(1,1) model. The process e_t will have a unconditional finite variance if $1 > \alpha_1 + \beta_1 > 0$ and $\alpha_0 > 0$.

A special case: The conditional variance of a GARCH(1,1) model is

$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$

Take expectation of the equation. If $\alpha_1 + \beta_1 < 1$, then

$$\text{Var}(e_t) = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}.$$

If $3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$, then

$$\begin{aligned} E(e_t^4) &= 3\frac{\alpha_0^2}{1 - \alpha_1 - \beta_1} \times \frac{1 + \alpha_1 + \beta_1}{1 - \beta_1^2 - 2\alpha_1\beta_1 - 3\alpha_1^2} \\ &= 3[\text{Var}(e_t)]^2 \times \frac{1 - (\alpha_1 + \beta_1)^2}{1 - (\beta_1 + \alpha_1)^2 - 2\alpha_1^2} \end{aligned}$$

which again shows that the fourth moment of a GARCH(1,1) model is greater than that of a normal random variable.
For estimation, it is also informative to write down the log-likelihood function for a GARCH(1,1) model.

Testing for ARCH and GARCH Model

The conventional way to test for conditional heteroscedasticity is to use Lagrange multiplier tests. Asymptotically, the Lagrange multiplier test is equivalent to

$$\xi = nR^2$$

where n is the sample size and R^2 is the coefficient of determination of the multiple linear regression

$$e_t^2 = \gamma_0 + \gamma_1 e_{t-1}^2 + \cdots + \gamma_r e_{t-r}^2 + \epsilon_t.$$

This test statistic is asymptotically χ_r^2. Alternatively, one can simply use the usual Ljung-Box Q-statistic of the $\{e_t^2\}$ process.

For GARCH model, a natural hypothesis is to assume that the null model is ARCH(r). Then, to test for a GARCH(r, s) model is asymptoically equivalent to testing for an ARCH($r + s$) model. Consequently, we can use multiple regressions

$$e_t^2 = \gamma_0 + \gamma_1 e_{t-1}^2 + \cdots + \gamma_r e_{t-r}^2 + \epsilon_t$$

and

$$e_t^2 = \gamma_0 + \gamma_1 e_{t-1}^2 + \cdots + \gamma_r e_{t-r}^2 + \gamma_{r+1} e_{t-r-1}^2 + \cdots + \gamma_{r+1} e_{t-r-s}^2 + \eta_t$$

and test for $\gamma_{r+1} = \cdots = \gamma_{r+s} = 0$.

Note: In practice, GARCH(1,1) or GARCH(2,1) or GARCH(1,2) is often used to analyze daily or monthly financial return series. Thus, one can test the need for ARCH effects and proceed to build an GARCH(1,1) model. If the fitted model is inadequate, e.g. the squared standardized residuals show significant serial correlations, then a higher-order GARCH model is employed.

Model checking: Let $\hat{e}_t$ be the residual of the mean equation, e.g.

$$\hat{e}_t = Z_t - \sum_{i=1}^{p} \hat{\phi}_i Z_{t-i} + \sum_{j=1}^{q} \hat{\theta}_j \hat{e}_{t-j},$$

of an ARMA(p, q) model. Let $\hat{\sigma}_t^2 = \hat{\alpha}_0 + \sum_{i=1}^{r} \hat{\alpha}_i e_{t-i}^2 + \sum_{j=1}^{s} \hat{\beta}_j \hat{\sigma}_{t-j}^2$ be the estimated conditional variance. Define the standardized residual as

$$\hat{a}_t = \frac{\hat{e}_t}{\hat{\sigma}_t}.$$

The ACF of $\{\hat{a}_t\}$ can be used to check the adequacy of the mean equation of Z_t, i.e. the ARMA model, and the ACF of the squared residuals $\{\hat{e}_t^2\}$ can be used to check the adequacy of the volatility equation, i.e. the conditional heteroscedastic model.

C. CHARMA Model

From a random-coefficient-model view point, conditional heteroscedasticity of an innovational process $\{e_t\}$ can be described by the model

$$e_t = \omega_{1t} e_{t-1} + \omega_{2t} e_{t-2} + \cdots + \omega_{rt} e_{t-r} + \epsilon_t$$

where $\{\epsilon_t\}$ is a sequence of *iid* random variates with mean zero and variance α_0, $\boldsymbol{\omega}_t = (\omega_{1t}, \cdots, \omega_{rt})'$ is a sequence of *iid* random vectors with mean zero and covariance matrix $\boldsymbol{G}$, and $\{\boldsymbol{\omega}_t\}$ and $\{\epsilon_t\}$ are independent. It is then easily seen that the conditional mean and variance of e_t are

$$\begin{aligned} E(e_t|\psi_{t-1}) &= 0 \\ \sigma_t^2 &= \alpha_0 + \boldsymbol{e}_{r,t-1}' \boldsymbol{G} \boldsymbol{e}_{r,t-1} \end{aligned}$$

where $\sigma_t^2 = E(e_t^2|\psi_{t-1})$, ψ_{t-1} denotes the information set available at time $t-1$, and $\boldsymbol{e}_{r,t-1} = (e_{t-1}, \cdots, e_{t-r})'$. If $\boldsymbol{G}$ is diagonal, then

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{r} g_{ii} e_{t-i}^2$$

which is the ARCH model. As mentioned earlier, this equation connects a pure ARCH model with a pure random-coefficient model.

Obviously, if we introduce another random vector $\boldsymbol{\delta}_t = (\delta_{1t}, \cdots, \delta_{st})'$, which is independent of $\boldsymbol{\omega}_t$ and e_t and $E(\boldsymbol{\delta}_t) = \boldsymbol{O}$ and $\text{Cov}(\boldsymbol{\delta}_t) = \boldsymbol{\Delta}$, and modify the above model as

$$e_t = \boldsymbol{\omega}_t' \boldsymbol{e}_{r,t-1} + \boldsymbol{\delta}_t' \boldsymbol{\sigma}_{s,t-1} + \epsilon_t$$

where $\boldsymbol{\sigma}_{s,t-1} = (\sigma_{t-1}, \cdots, \sigma_{t-s})'$, then

$$\begin{aligned} E(e_t|\psi_{t-1}) &= \;= 0 \\ \sigma_t^2 &= \alpha_0 + \boldsymbol{e}_{r,t-1}' \boldsymbol{G} \boldsymbol{e}_{r,t-1} + \boldsymbol{\sigma}_{s,t-1} \boldsymbol{\Delta} \boldsymbol{\sigma}_{s,t-1}. \end{aligned}$$

If both $\boldsymbol{G}$ and $\boldsymbol{\Delta}$ are diagonal, then

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{r} g_{ii} e_{t-i}^2 + \sum_{j=1}^{s} \Delta_{jj} \sigma_{t-1}^2,$$

which is a GARCH model. For random coefficient models, the non-negativeness condition holds automatically as the coefficients g_{ii} and Δ_{jj} are variances of random coefficients. Using such an idea of random coefficient, Tsay (1987) considers a class of conditional heteroscedastic autoregressive moving-average (CHARMA) models. An advantage of looking conditional heteroscedasticity via random coefficient models is that certain properties of GARCH or ARCH models can easily be obtained by using results of the random coefficient models. A decent reference for random coefficient autoregressive (RCA) models is Nicholls and Quinn (1982, book).

So far, we have only discussed the evolution of conditional variance of a time series. In practice, we can also consider the evolution of conditional mean. For instance, for an ARMA time series Z_t with ARCH-type conditional heteroscedasticity , we have

$$\begin{aligned} E(Z_t|\psi_{t-1}) &= \phi_1 Z_{t-1} + \cdots + \phi_p Z_{t-p} + c - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q} \\ V(Z_t|\psi_{t-1}) &= \alpha_0 + \alpha_1 e_{t-1}^2 + \cdots + \alpha_r e_{t-r}^2. \end{aligned}$$

This is an ARMA-ARCH model with the linear ARMA model governing the evolution of the conditional mean and an ARCH(r) model governing the evolution of the conditional variance. It generalizes easily to ARMA-GARCH models.

IGARCH Model:
A particular GARCH(1,1) model that has received some attention (and has some empirical evidence) is the *integrated* GARCH(1,1) model

- $e_t = \sigma_t a_t$ where a_t's are *iid* N(0,1),
- $\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2$ with $\alpha_1 + \beta_1 = 1$.

By rewriting the conditional variance equation as

$$\sigma_t^2 = \alpha_0 + (\alpha_1 a_{t-1}^2 + \beta_1)\sigma_{t-1}^2 \tag{1}$$

and considering the fact that $E(\alpha_1 a_{t-1}^2 + \beta_1) = \alpha_1 + \beta_1 = 1$, we have an equation similar to that of a "random walk" model. For this reason, the model is referred to as the integrated GARCH (IGARCH(1,1)) model.
There are, however, some differences between a "random walk" and the IGARCH(1,1) model. First, the conditional variance equation in (1) for the IGARCH model is in effect a random coefficient equation. Secondly, the equation does not have an "innovational term". Consequently, there are certain differences between a random walk and an IGARCH(1,1) model.

Consider first the concept of persistence. Recall that in introducing the random walk model, we mentioned that any shock a_{t-i} to a random walk has a long lasting effect, i.e. the shock is persistent. For the IGARCH(1,1), the situation could be very different. Assuming that the IGARCH(1,1) model starts at time $t = 1$ with initial values a_0 and σ_0. Then, by repeatedly substitution, we have

$$\sigma_t^2 = \alpha_0[1 + \sum_{i=1}^{t-1}\prod_{j=i}^{t-1}(\alpha_1 a_{t-j}^2 + \beta_1)] + \prod_{i=0}^{t-1}(\alpha_1 a_i^2 + \beta_1)\sigma_0^2.$$

If $\alpha_0 = 0$, then

$$\sigma_t^2 = \prod_{i=0}^{t-1}(\alpha_1 a_i^2 + \beta_1)\sigma_0^2,$$

or equivalently,

$$\ln(\sigma_t^2) = \sum_{i=0}^{t-1}\ln(\alpha_1 a_i^2 + \beta_1) + \ln(\sigma_0^2).$$

Since a_i's are *iid* random variables, $\ln(\alpha_1 a_i^2 + \beta_1)$'s are *iid* random variables, too. [In our case, we assume that a_i's are *iid* N(0,1). However, the normaility assumption is not critical in our discussion here.] Therefore, provided that $E[\ln(\alpha_1 a_i^2 + \beta_1)]$ is well defined, we obtain, by the Law of Large Numbers, that

- if $E[\ln(\alpha_1 a_i^2 + \beta_1)] > 0$, then $\ln(\sigma_t^2)$ approaches infinity as $t \to \infty$, implying that σ_t^2 diverges;
- if $E[\ln(\alpha_1 a_i^2 + \beta_1)] < 0$, then $\ln(\sigma_t^2)$ approaches minus infinity so that σ_t^2 goes to zero almost surely;
- if $E[\ln(\alpha_1 a_i^2 + \beta_1)] = 0$, then the process $\ln(\sigma_t^2)$ is a random walk without drift so that σ_t^2 is a random variable wondering around $[0, \infty)$. Mathematically, we have $\limsup \sigma_t^2 = \infty$ and $\liminf \sigma_t^2 = 0$.

Next, for $\alpha_0 > 0$, the results are more involved. The conditional variance equations gives

$$E(\sigma_{t+k}^2|\sigma_t^2) = \alpha_0 k + \sigma_t^2.$$

Thus, σ_t^2 is a martingale with a drift. If $E\ln(\alpha_1 a_i^2 + \beta_1) \geq 0$, we have $\sigma_t^2 \to \infty$ as $t \to \infty$. For $E\ln(\alpha_1 a_i^2 + \beta_1) < 0$, the result is relatively complicated, See Nelosn (1990, ET).

D. The Exponential GARCH (EGARCH) Model:

One of the main weaknesses of the GARCH models is that the conditional variance responds symmetrically to "positive" and "negative" past innovations. On the other hand, in finance literature, there is evidence that current return is negatively correlated with the future return volatility. To remedy such weaknesses, Nelson (1991) proposed the EGARCH models. For simplicity, we consider the EGARCH(1,1) model first. The model is

- $e_t = \sigma_t a_t$ with a_t's *iid* N(0,1);
- $\ln(\sigma_t^2) = \alpha + \frac{1-\omega B}{1-\delta B} g(a_{t-1})$ where the function $g(.)$ is defined by

  $$g(a_{t-1}) = \theta a_{t-1} + \gamma[|a_{t-1}| - E|a_{t-1}|],$$

  and $\alpha, \omega, \delta, \theta$ and γ are parameters.

Since log-transformation is taken, there is no "positiveness" requirement on the parameters. The function $g(.)$ is a weighted linear combination of a_{t-1} and the mean-corrected $|a_{t-1}|$. Treating $g(.)$ as a function of its argument a_{t-1}, we see that for $a_{t-1} \geq 0$, $g(.)$ is a linear function of a_{t-1} with slope $(\theta + \gamma)$, but for $a_{t-1} < 0$, it is a linear function of a_{t-1} with slope $\theta - \gamma$. Consequently, the conditional variance responds asymmetrically to "positive" and "negative" past innovations.
Rewriting the conditional variance as

$$(1 - \delta B) \ln(\sigma_t^2) = (1 - \delta B)\alpha + (1 - \omega B) g(a_{t-1}),$$

we see that for the EGARCH model the logarithm of the conditional variance has a simple linear ARMA structure with $g(.)$ playing the role of an innovation. Consequently, as long as the zeros of $(1 - \delta B)$ and $(1 - \omega B)$ are outside the unit circle, the process of conditional variances is stationary and invertible which in turn implies that $\sigma_t^2 = \exp[\ln(\sigma_t^2)]$ is well defined. Thus, the stationarity condition of EGARCH model is the same as that of linear ARMA processes. For further properties of EGARCH models, see Nelson (1991).

Exercise: Generalize the EGARCH(1,1) model to the EGARCH(p, q) model.

Next, we briefly examine the parameterization of EGARCH models. For simplicity, consider the simplest EGARCH(0,0) model, for which the conditional variance equation is

$$\ln(\sigma_t^2) = \alpha + g(a_{t-1}) = \alpha + \theta a_{t-1} + \gamma[|a_{t-1}| - E|a_{t-1}|].$$

Under the normality assumption, we have

$$\begin{aligned} \ln(\sigma_t^2) &= \alpha + \theta a_{t-1} + \gamma |a_{t-1}| - \gamma \sqrt{\frac{2}{\pi}} \\ &= (\alpha + \gamma \sqrt{\frac{2}{\pi}}) + \theta a_{t-1} + \gamma |a_{t-1}| \\ &= (\alpha + \gamma \sqrt{\frac{2}{\pi}}) + \begin{cases} (\theta + \gamma) a_{t-1} & \text{if } a_{t-1} \geq 0 \\ (\theta - \gamma) a_{t-1} & \text{if } a_{t-1} < 0, \end{cases} \end{aligned}$$

which is equivalent to the equation

$$\ln(\sigma_t^2) = \alpha_0 + \begin{cases} \alpha_1 a_{t-1} & \text{if } a_{t-1} \geq 0 \\ \alpha_2 a_{t-1} & \text{if } a_{t-1} < 0. \end{cases}$$

This links the EGARCH model to the idea of threshold models, a class of nonlinear time series models. In either case, three parameters are used. The latter parameterization, however, seems to be more direct. Of course, there is a one-to-one linear transformation between the parameters.

Exercise: Consider an EGARCH-type of model with conditional variance

$$\ln(\sigma_t^2) = \alpha_0 + \alpha_1 e_{t-1} + \alpha_2 e_{t-1} I(e_{t-1} < 0),$$

where $I(.)$ is the usual indicator variable. What is the difference, if any, between this model and the EGARCH(0,0) model?

Consider next the EGARCH(0,1) model with

$$\ln(\sigma_t^2) = \alpha + g(a_{t-1}) - \omega g(a_{t-2}).$$

Under the normality assumption, this conditional variance is equivalent to

$$\ln(\sigma_t^2) = \alpha_0 + \theta a_{t-1} + \gamma |a_{t-1}| - \omega(\theta a_{t-2} + \gamma |a_{t-2}|).$$

This model can alternatively be written as

$$\ln(\sigma_t^2) = \alpha_0 + \gamma_1 a_{t-1} I(a_{t-1} \geq 0) + \gamma_2 a_{t-1} I(a_{t-1} < 0) - \omega\gamma_1 a_{t-2} I(a_{t-2} \geq 0) - \omega\gamma_2 I(a_{t-2} < 0).$$

The above equation shows that (a) the conditional variance responds asymmetrically to “positive” and “negative” past innovations; (b) the effect of the innovation a_{t-2} is discounted by a factor ω; and (c) more importantly, the responses to “positive” and “negative” innovations are independent of the lags $t-i$. The plausibility of the last point should be carefully checked in practice. Obviously, a more general setup is

$$\ln(\sigma_t^2) = \alpha_0 + \gamma_1 a_{t-1} I(a_{t-1} \geq 0) + \gamma_2 a_{t-1} I(a_{t-1} < 0) + \gamma_3 a_{t-2} I(a_{t-2} \geq 0) + \gamma_4 a_{t-2} I(a_{t-2} < 0).$$

This equation, however, uses one more parameter in order to obtain greater flexibility.
In summary, the EGARCH models employ a threshold structure to capture the asymmetric responses of past innovations in the evolution of conditional variance.

E. Forecasts of Conditional variances.

We consider two simple models to illustrate the multi-step ahead forecasts of conditional variance for GARCH and EGARCH models. The first model is the GARCH(1,1) model and the second model is an EGARCH(1,0) process.

Example 1: An GARCH(1,1) Model: In the first example, we present the formula used to calculate the multi-step ahead forecasts of the conditional variance for the GARCH(1,1) model. For such a model, the variance equation is

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$

Denote the forecast origin by n and the forecast horizon by ℓ. Let I_n be the information set available at time n. For $\ell = 1$, the 1-step ahead forecast of the conditional variance is simply

$$\begin{aligned} E(\sigma_{n+1}^2 | I_n) &= E(\alpha_0 + \alpha_1 \epsilon_n^2 + \beta_1 \sigma_n^2 | I_n) \\ &= \alpha_0 + \alpha_1 \epsilon_n^2 + \beta_1 \sigma_n^2 \end{aligned}$$

For $\ell = 2$, by using the assumption that z_t's are *i.i.d.* N(0,1), we have

$$\begin{aligned} E(\sigma_{n+2}^2 | I_n) &= E(\alpha_0 + \alpha_1 \epsilon_{n+1}^2 + \beta_1 \sigma_{n+1}^2 | I_n) \\ &= \alpha_0[1 + (\alpha_1 + \beta_1)] + \alpha_1(\alpha_1 + \beta_1)\epsilon_n^2 + \beta_1(\alpha_1 + \beta_1)\sigma_n^2 \\ &= \alpha_0 + (\alpha_1 + \beta_1) \cdot E(\sigma_{n+1}^2 | I_t). \end{aligned}$$

By the same argument, it is easily seen that for $\ell = j$, the j-step ahead forecast of the conditional variance of the GARCH(1,1) model is

$$\begin{aligned} E(\sigma_{n+j}^2 | I_n) &= \alpha_0 \sum_{k=0}^{j-1} (\alpha_1 + \beta_1)^k + \alpha_1(\alpha_1 + \beta_1)^{j-1}\epsilon_n^2 + \beta_1(\alpha_1 + \beta_1)^{j-1}\sigma_n^2 \\ &= \alpha_0 + (\alpha_1 + \beta_1) \cdot E(\sigma_{n+j-1}^2 | I_n). \end{aligned}$$

Therefore, the forecasts of the conditional variances of an GARCH(1,1) model can be computed recursively.

Example 2: An EGARCH(1,0) model: We derive next the formula for obtaining multi-step ahead forecasts of the conditional variance for the EGARCH model. For simplicity, we assume that the process is stationary so that it starts with $t = -\infty$. The conditional variance of the estimated EGARCH model is

$$\begin{aligned} \ln(\sigma_t^2) &= \alpha + \frac{1}{1 - \Delta B} g(z_{t-1}) \\ &= \alpha + \sum_{k=1}^{\infty} \Delta^{k-1} g(z_{t-k}) \end{aligned}$$

where $g(z_t) = \theta z_t + \gamma[|z_t| - E|z_t|]$. Therefore,

$$\sigma_t^2 = \exp[\alpha + \sum_{k=1}^{\infty} \Delta^{k-1} g(z_{t-k})].$$

Denote the forecast origin of the conditional variance by n and the forecast horizon by ℓ. For $\ell = 1$, the 1-step ahead forecast of the conditional variance is

$$E(\sigma_{n+1}^2 | I_n) = \exp[\alpha + \sum_{k=1}^{\infty} \Delta^{k-1} g(z_{n+1-k})],$$

where, again, I_n denotes the information set available at time index n. For $\ell = 2$, the 2-step ahead forecast of the conditional variance is

$$\begin{aligned} E(\sigma^2_{n+2}|I_n) &= E[\exp(\alpha + \sum_{k=1}^{\infty} \Delta^{k-1} g(z_{n+2-k})|I_t)] \\ &= \exp(\alpha)\exp[\sum_{k=2}^{\infty} \Delta^{k-1} g(z_{n+2-k})] E[\exp(g(z_{n+1}))|I_n]. \end{aligned}$$

In general, using the assumption that $\{z_t\}$ are *i.i.d.* N(0,1), we have, for $j \geq 2$,

$$\begin{aligned} E(\sigma^2_{n+j}|I_n) &= E[\exp(\alpha + \sum_{k=1}^{\infty} \Delta^{k-1} g(z_{n+j-k}))|I_n] \\ &= \exp(\alpha)\exp[\sum_{k=j}^{\infty} \Delta^{k-1} g(z_{n+j-k})] E[\exp(\sum_{k=1}^{j-1} \Delta^{k-1} g(z_{n+j-k}))|I_n] \\ &= \exp[\alpha + \sum_{k=j}^{\infty} \Delta^{k-1} g(z_{n+j-k})] \\ & \prod_{k=1}^{j-1} \exp(-\Delta^{k-1}\gamma\sqrt{\frac{2}{\pi}})[\exp(\frac{1}{2}(\gamma-\theta)^2\Delta^{2(k-1)})\Phi((\gamma-\theta)\Delta^{k-1}) + \\ & \exp(\frac{1}{2}(\gamma+\theta)^2\Delta^{2(k-1)})\Phi((\gamma+\theta)\Delta^{k-1})] \end{aligned}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal random variable.

Estimation of GARCH models in R

To estimate volaitlity models in R, one needs the following packages: **fSeries** and **fGarch**. They are available from R CRAN.
The command is **garchFit**. Suppose the return series is *rt*. Then, to fit an ARMA(p, q) model for the mean equation and a GARCH(r, s) model for the volaitlity, one use the command:

```
m1=garchFit(formula=~arma(p,q)+garch(r,s),rt,trace=F)
```

where "m1" is the name of the model.
One can also specify the conditional distribution by adding the statement cond.dist=c("std") for Student-t distribution.
To check the fitted model, use the command

```
plot(m1)
```

where, again,"m1" is the name of the model.
Below is a demonstration created on December 1, 2008. The data are daily log return of FedEx from 1997 to 2006. Log returns in percentages are used.

```
> library(fSeries)
> library(fGarch)

> da=read.table("d-fdx9706.txt")
> rt=log(da[,2]+1)*100     <== compute log returns in percentages.
> plot(rt,type='l')
> acf(rt)
> Box.test(rt,lag=10,type='Ljung')
        Box-Ljung test
data:  rt
X-squared = 16.8427, df = 10, p-value = 0.07792

> m1=garchFit(formula=~arma(0,0)+garch(1,1),rt,trace=F)
> m1
Title:
 GARCH Modelling
Call:
 garchFit(formula = ~arma(0, 0) + garch(1, 1), data = rt, trace = F)

Mean and Variance Equation:
 data ~ arma(0, 0) + garch(1, 1)
 [data = rt]

Conditional Distribution:
 norm

Coefficient(s):
       mu      omega     alpha1      beta1
0.0750712  0.0044111  0.0134347  0.9854139

Std. Errors:
 based on Hessian

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.075071    0.036113    2.079   0.0376 *
omega   0.004411    0.002008    2.197   0.0280 *
alpha1  0.013435    0.001772    7.580 3.46e-14 ***
beta1   0.985414    0.001724  571.624  < 2e-16 ***
---
Log Likelihood:
 -5273.391    normalized:  -2.095942
```

```
> plot(m1)
Make a plot selection (or 0 to exit):

 1:   Time Series
 2:   Conditional SD
 3:   Series with 2 Conditional SD Superimposed
 4:   ACF of Observations
 5:   ACF of Squared Observations
 6:   Cross Correlation
 7:   Residuals
 8:   Conditional SDs
 9:   Standardized Residuals
10:   ACF of Standardized Residuals
11:   ACF of Squared Standardized Residuals
12:   Cross Correlation between r^2 and r
13:   QQ-Plot of Standardized Residuals

Selection: 3  <== Plot is not shown in this handout.

> m1=garchFit(formula=~arma(0,0)+garch(1,1),rt,trace=F,cond.dist=c("std"))
> m1
Title:
 GARCH Modelling
Call:
 garchFit(formula = ~arma(0, 0) + garch(1, 1), data = rt, cond.dist = c("std"),
    trace = F)

Mean and Variance Equation:
 data ~ arma(0, 0) + garch(1, 1)
 [data = rt]

Conditional Distribution:
 std

Coefficient(s):
       mu      omega     alpha1      beta1      shape
0.0377494  0.0047011  0.0196153  0.9791496  5.5732740

Std. Errors:
 based on Hessian
```

```
Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.037749    0.032687    1.155    0.248
omega   0.004701    0.003640    1.291    0.197
alpha1  0.019615    0.003833    5.118 3.09e-07 ***
beta1   0.979150    0.003855  253.997  < 2e-16 ***
shape   5.573274    0.569025    9.794  < 2e-16 ***
---
Log Likelihood:
 -5161.691    normalized:  -2.051547

> plot(m1)
Make a plot selection (or 0 to exit):

 1:   Time Series
 2:   Conditional SD
 3:   Series with 2 Conditional SD Superimposed
 4:   ACF of Observations
 5:   ACF of Squared Observations
 6:   Cross Correlation
 7:   Residuals
 8:   Conditional SDs
 9:   Standardized Residuals
10:   ACF of Standardized Residuals
11:   ACF of Squared Standardized Residuals
12:   Cross Correlation between r^2 and r
13:   QQ-Plot of Standardized Residuals

Selection: 13

> predict(m1,5)
  meanForecast meanError standardDeviation
1   0.03774941  2.107909          1.309133
2   0.03774941  2.107909          1.310120
3   0.03774941  2.107909          1.311105
4   0.03774941  2.107909          1.312087
5   0.03774941  2.107909          1.313068
```