Ch. 26 Autoregressive Conditional Heteroscedasticity (ARCH)

1 Introduction

Traditional econometric models assume a constant one-period forecast variance. To generalize this implausible assumption, a new stochastic processes called autoregressive conditional heteroscedasticity (ARCH) are introduced by Engle (1982). These are mean zero, serially uncorrelated processes with nonconstant variances conditional on the past, but constant unconditional variances. For such processes, the recent past gives information about the one-period forecast variance.

If a random variable u_t is drawn from the conditional density function $f(u_t|u_{t-1})$, the forecast of today's value based upon the past information, under standard assumptions, is simply $E(u_t|u_{t-1})$, which depends upon the value of the conditioning variable u_{t-1}. The variance of this one-period forecast is given by $Var(u_t|u_{t-1}) = E\{(u_t - E[u_t|u_{t-1}])^2|u_{t-1}\}$. Such an expression recognizes that the conditional forecast variance depends upon past information and may therefore be a **random variable**. For conventional econometric models, however, the conditional variance does not depend upon u_{t-1}. Engle (1982) propose a class of models where the variance does depend upon the past and argue for their usefulness in economics.

Consider initially the first-order autoregression

$$u_t = \theta u_{t-1} + \varepsilon_t,$$

where ε_t is white noise with $Var(\varepsilon_t) = \sigma^2$.

The conditional mean of u_t $(= \hat{E}(u_t|u_{t-1}))$ is θu_{t-1} while the unconditional mean $(= E(u_t))$ is zero. Clearly, the vast improvement in forecasts due to time-series models stems from the use of the conditional mean. The conditional variance of u_t $(=E\{(u_t - E[u_t|u_{t-1}])^2|u_{t-1}\} = E(\varepsilon_t^2))$ is σ^2 while the unconditional variance is $\sigma^2/(1-\theta^2)$. For real process one might expect better forecast intervals if additional information from the past were allowed to affect the forecast variance; a more general class of models seems desirable.

The standard approach of heteroscedasticity is to introduce an exogenous variable X_t which predicts the variance. With a known zero mean, the model might be

$$u_t = \varepsilon_t X_{t-1},$$

where again $Var(\varepsilon_t) = \sigma^2$. The variance of u_t is simply $\sigma^2 X_{t-1}^2$ and, therefore, the forecast interval depends upon the evolution of an exogenous variable. This standard solution to the problem seems unsatisfactory, as it requires a specification of the causes of the changing variance, rather than recognizing that both conditional means and variances may jointly evolve over time. Perhaps because of this difficulty, heteroscedasticity corrections are rarely considered in time-series data.

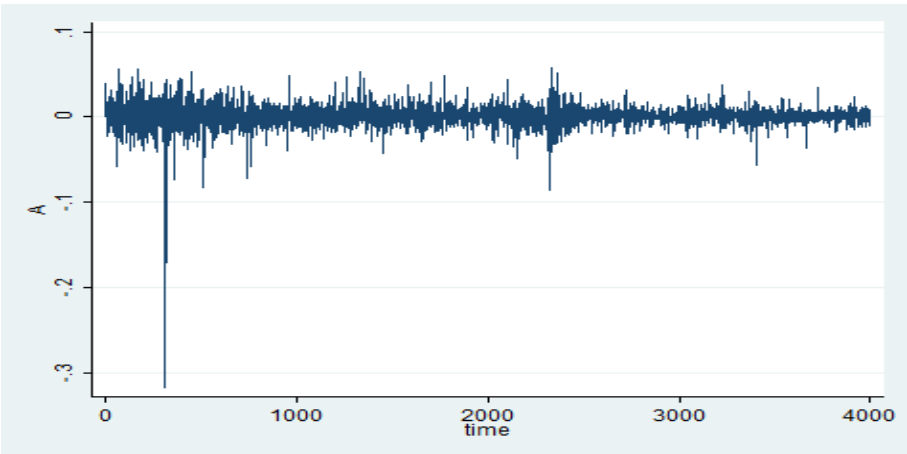

Figure (26-1). An Example of Volatility Clustering.

2 $ARCH(m)$ Model

A model which allows the conditional variance to depend on the past realization of the series is considered in the following. Suppose that

$$\begin{aligned} u_t &= \sqrt{h_t}\varepsilon_t & (1) \\ h_t &= \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2, & (2) \end{aligned}$$

with $E(\varepsilon_t) = 0$ and $Var(\varepsilon_t) = 1$, then this is a example of what will be called an **autoregressive conditional heteroscedasticity** ($ARCH(m)$) model.

2.1 Population's Properties of $ARCH$ Models

2.1.1 The Conditional Mean and Variance

Let F_{t-1} denote the information set available at time $t-1$. The conditional mean of u_t is

$$E(u_t|F_{t-1}) = \sqrt{h_t} \cdot E(\varepsilon_t|F_{t-1}) = 0. \tag{3}$$

From (3) it implies that the conditional variance of u_t is

$$\begin{aligned} \sigma_t^2 &= Var(u_t|F_{t-1}) \\ &= E\{[u_t - E(u_t|F_{t-1})]^2|F_{t-1}\} \\ &= E(u_t^2|F_{t-1}) \qquad (since\ E(u_t|F_{t-1}) = 0) \\ &= E(h_t\varepsilon_t^2|F_{t-1}) \\ &= E(\varepsilon_t^2)E(\alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2|F_{t-1}) \\ &= \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2 \\ &= h_t. \end{aligned}$$

From the structure of the model, it is seen that large past squared shocked shocks $u_{t-i}^2, i = 1, .., m$ imply a large conditional variance $\sigma_t^2 (= Var(u_t|F_{t-1}))$ for this variable u_t. Consequently, u_t tends to assume a large value. This means that, under the $ARCH$ framework, large shocks tend to be followed by another large chock. This feature is similar to the volatility clustering observed in asset returns.

2.1.2 The Conditional Density

By assuming that ε_t is a Gaussian variate, the condition density of u_t given all the information update to $t-1$ is

$$f(u_t|F_{t-1}) = \sqrt{h_t} f(\varepsilon_t|F_{t-1}) = \sqrt{h_t} \cdot N(0,1) \sim N(0, h_t).$$

2.1.3 The Unconditional Mean and Variance

The unconditional mean of u_t is $E[E(u_t|F_{t-1})] = E(0) = 0$. While u_t is conditional heteroscedastic, the unconditional variance of u_t is

$$\begin{aligned} Var(u_t) &= Var[E(u_t|F_{t-1})] + E[Var(u_t|F_{t-1})] \\ &= 0 + \alpha_0 + \alpha_1 E(u_{t-1}^2) + \alpha_2 E(u_{t-2}^2) + \cdots + \alpha_m E(u_{t-m}^2) \\ &= \alpha_0 + \alpha_1 Var(u_{t-1}) + \alpha_2 Var(u_{t-2}) + \cdots + \alpha_m Var(u_{t-m}), \end{aligned}$$

Here, we have used the fact that $E(u_t|F_{t-1}) = 0$. If the process generating u_t^2 is covariance stationary, i.e. all the roots of

$$1 - \alpha_1 z - \alpha_2 z^2 - ... - \alpha_m z^m = 0$$

lie outside the unit circle, then the unconditional variance is **not** changing over time so

$$Var(u_t) = Var(u_{t-1}) = Var(u_{t-2}) = \cdots Var(u_{t-m}) = \frac{\alpha_0}{1 - \alpha_1 - \cdots - \alpha_m}.$$

For this ratio to be finite and positive, we require that $\alpha_0 > 0$ and $\alpha_1 + \alpha_2 + \cdots + \alpha_m < 1$. Moreover, since $E(u_t|F_{t-1}) = 0$, so $E(u_t u_{t-j}) = 0$, that is, u_t is a **white noise process**.

2.1.4 Coefficients Constraint in an Conditional Gaussian $ARCH(1)$ model

For an $ARCH(1)$ model:

$$u_t = \sqrt{h_t}\varepsilon_t \tag{4}$$

$$h_t = \alpha_0 + \alpha_1 u_{t-1}^2, \tag{5}$$

with $\varepsilon_t \sim N(0,1)$. Because the variance of u_t must be positive, we need $0 \leq \alpha_1 < 1$. In some applications, we need higher order moments of u_t to exist and, hence, α_1 must also satisfy some additional constrains. For instance, to study its tail behavior, we require that the fourth moment of u_t is finite. Under the normality assumption of ε_t, we have[1]

$$E(u_t^4|F_{t-1}) = 3[E(u_t^2|F_{t-1})]^2 = 3(\alpha_0 + \alpha_1 u_{t-1}^2)^2.$$

Therefore,

$$E(u_t^4) = E[E(u_t^4|F_{t-1})] = 3E(\alpha_0 + \alpha_1 u_{t-1}^2)^2 = 3E(\alpha_0^2 + 2\alpha_0\alpha_1 u_{t-1}^2 + \alpha_1^2 u_{t-1}^4).$$

If u_t is fourth-order stationary with $m_4 = E(u_t^4)$, then we have

$$\begin{aligned} m_4 &= 3(\alpha_0^2 + 2\alpha_0\alpha_1 Var(u_t) + \alpha_1^2 m_4) \\ &= 3\alpha_0^2\left(1 + 2\frac{\alpha_1}{1-\alpha_1}\right) + 3\alpha_1^2 m_4. \end{aligned}$$

Consequently,

$$m_4 = \frac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}.$$

This result has two important implications:
(a). since the fourth moment of u_t is positive, we see that α_1 must also satisfy the condition $1 - 3\alpha_1^2 > 0$; that is $0 \leq \alpha_1^2 < 1/3$;
(b). the unconditional kurtosis of u_t is

$$\frac{E(u_t^4)}{[Var(u_t)]^2} = \frac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)} \times \frac{(1-\alpha_1)^2}{\alpha_0^2} = 3\frac{1-\alpha_1^2}{1-3\alpha_1^2} > 3.$$

Thus, the excess kurtosis of u_t is positive and the tail distribution of u_t is heavier[2] than that of a normal distribution. In other words, the shock u_t of a conditional

[1]The skewness, $S(X)$ and kurtosis, $K(X)$ of a random variable X are defined as

$$S(X) = \frac{E(X-\mu_X)^3}{\sigma_X^3}, \quad K(X) = \frac{E(X-\mu_X)^4}{\sigma_X^4},$$

where $\sigma_X^2 = E(X-\mu_X)^2$. The quantity $K(X)-3$ is called the excess kurtosis because $K(X) = 3$ for a normal distribution. Here, $f(u_t|F_{t-1}) \sim N(0, h_t)$, so the conditional density of $f(u_t|F_{t-1})$ is normal. Therefore $\frac{E(u_t^4|F_{t-1})}{[E(u_t^2|F_{t-1})]^2} = 3$.

[2]So the unconditional distribution $f(u_t)$ is not normal.

Gaussian $ARCH(1)$ model is more likely than a Gaussian white noise series to produce "outliers." This is in agreement with the empirical finding that "outliers" appear more often in asset returns than that implied by an *iid* sequence of normal random variates.

2.1.5 An Alternative Representation

It is often convenient to use an alternative representation for an $ARCH(m)$ process that imposes slightly weaker assumptions about the serial dependence of u_t. One approach is to describe the square of u_t as itself following an $AR(m)$ process:

$$u_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2 + w_t, \tag{6}$$

where w_t is a new white noise process:

$$\begin{aligned} E(w_t) &= 0 \\ E(w_t w_\tau) &= \begin{cases} \lambda^2 & for\ t = \tau \\ 0 & otherwise. \end{cases} \end{aligned}$$

Expression (6) implies that

$$E(u_t^2 | F_{t-1}) = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2. \tag{7}$$

Since u_t is random and u_t^2 cannot be negative, this can be a sensible representation only if (7) is positive and (6) is nonnegative for all realization of u_t. This can be ensured if w_t is bounded from below by $-\alpha_0$ with $\alpha_0 > 0$ and if $\alpha_i \geq 0$ for $j = 1, 2, ..., m$. In order for u_t^2 to be covariance-stationary, we further require that the roots of

$$1 - \alpha_1 z - \alpha_2 z^2 - \cdots - \alpha_m z^m = 0$$

lie outside the unit circle. If the α_i are all nonnegative, this is equivalent to the requirement that

$$\alpha_1 + \alpha_2 + \cdots + \alpha_m < 1.$$

When these conditions are satisfied, the unconditional variance of u_t is given by

$$\sigma^2 = E(u_t^2) = \frac{\alpha_0}{(1 - \alpha_1 - \alpha_2 - \cdots - \alpha_m)}.$$

2.1.6 The Intuition Behind ARCH Model

Elementary statistics tell us that $E(Y|X)$ and $Var(Y|X)$ should be a function X. But for the "classical" white noise process ε_t, we have the problematic results that

$$\begin{aligned} E(\varepsilon_t|\Omega_{t-1}) &= 0 \\ Var(\varepsilon_t|\Omega_{t-1}) &= \sigma_\varepsilon^2, \end{aligned}$$

that both are all not function of Ω_{t-1}. Consider the classical linear model (For example AR) by adding a lagged variable to ε_t, i.e. $\phi Y_{t-1} + \varepsilon_t = Y_t$ then[3]

$$\begin{aligned} E(Y_t|\Omega_{t-1}) &= E(\phi Y_{t-1} + \varepsilon_t|\Omega_{t-1}) = \phi Y_{t-1} \ \ (a\ function\ of\ \Omega_{t-1}) \\ Var(Y_t|\Omega_{t-1}) &= E(\phi Y_{t-1} + \varepsilon_t - \phi Y_{t-1}|\Omega_{t-1})^2 = \sigma_\varepsilon^2. \ \ (still\ not\ a\ function\ of\ \Omega_{t-1}) \end{aligned}$$

To conquer this problem, we can include the lag variable to the white noise nonlinearly by the following two methods.[4] The first one is the nonlinear time series model. For example $\phi Y_{t-1} + Y_{t-1}\varepsilon_t = Y_t$, then

$$\begin{aligned} E(Y_t|\Omega_{t-1}) &= E(\phi Y_{t-1} + Y_{t-1}\varepsilon_t|\Omega_{t-1}) = \phi Y_{t-1} + Y_{t-1}E(\varepsilon_t|\Omega_{t-1}) = \phi Y_{t-1} \\ Var(Y_t|\Omega_{t-1}) &= E(\phi Y_{t-1} + Y_{t-1}\varepsilon_t - \phi Y_{t-1}|\Omega_{t-1})^2 = Y_{t-1}^2\sigma_\varepsilon^2, \end{aligned}$$

both are function of Ω_{t-1} at the cost of nonlinearity of the time series model Y_t. The second one is the $ARCH$ model that define a new white noise that[5]

$$u_t = \sqrt{h_t}\varepsilon_t, \ \ h_t = \alpha_0 + \alpha_1 u_{t-1}^2,$$

[3]This results can be expected since $E(X+a) = \nu_x$ but $Var(x+a) = \sigma_x^2$.
[4]This results can be expected since $E(aX) = a\nu_x$ and $Var(ax) = a^2\sigma_x^2$.
[5]Check u_t is white noise: $E(u_t) = E[E(u_t|\Omega_{t-1})] = E(0) = 0$ and

$$\begin{aligned} Var(u_t) &= Var[E(u_t|\Omega_{t-1})] + E[Var(u_t|\Omega_{t-1})] \\ &= \alpha_0 + \alpha_1 Var(u_{t-1}). \end{aligned}$$

If the process generating u_t^2 is covariance stationary, i.e. all the roots of $1 - \alpha_1 z = 0$ lie outside the unit circle, then the unconditional variance is **not** changing over time so $Var(u_t) = Var(u_{t-1}) = \frac{\alpha_0}{1-\alpha_1}$. Moreover, since $E(u_t|\Omega_{t-1}) = 0$, so $E(u_t u_{t-j}) = 0$, that is, u_t is a **white noise process**.

such that

$$E(u_t|\Omega_{t-1}) = \sqrt{h_t} \cdot E(\varepsilon_t|\Omega_{t-1}) = 0, \quad (not\ a\ function\ of\ \Omega_{t-1})$$

$$\begin{aligned} Var(u_t|\Omega_{t-1}) &= E\{[u_t - E(u_t|F_{t-1})]^2|F_{t-1}\} \\ &= \alpha_0 + \alpha_1 u_{t-1}^2 \\ &= h_t. \ (a\ function\ of\ \Omega_{t-1}) \end{aligned}$$

The conditional mean is not a function of the past can be easily be remedy by adding a lag variable to the new white noise u_t linearly. For example $\phi Y_{t-1} + u_t = Y_t$, then

$$\begin{aligned} E(Y_t|\Omega_{t-1}) &= E(\phi Y_{t-1} + u_t|\Omega_{t-1}) = \phi Y_{t-1}, \\ Var(Y_t|\Omega_{t-1}) &= E(\phi Y_{t-1} + u_t - \phi Y_{t-1}|\Omega_{t-1})^2 = \alpha_0 + \alpha_1 u_{t-1}^2. \end{aligned}$$

Both are function of Ω_{t-1}.

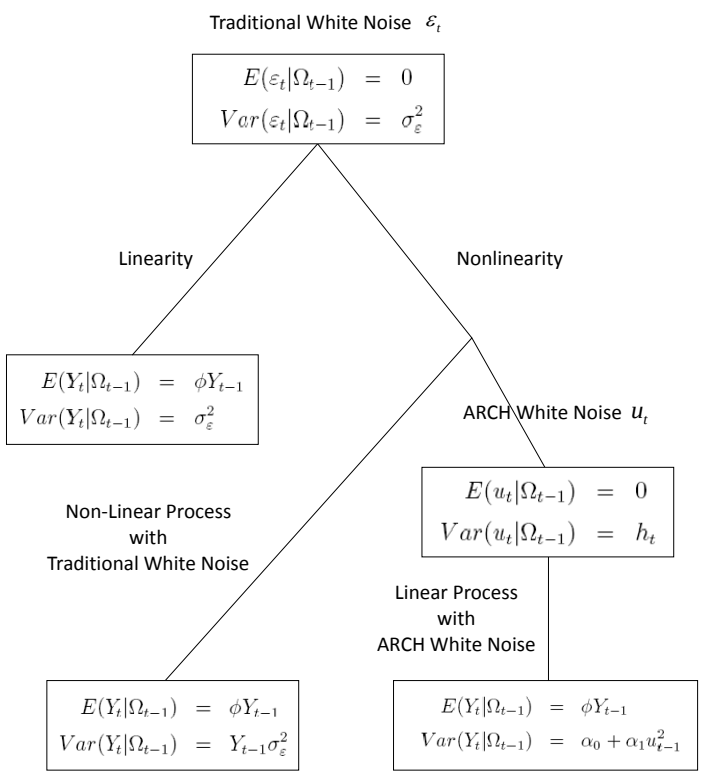

Figure 1: Summary of Intuition Behind ARCH Model

2.2 Properties of AR model with $ARCH(m)$ Disturbance

An $AR(p)$ process for an observed variable Y_t, takes the form

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + u_t, \tag{8}$$

where c, ϕ_1, ϕ_2,..., and ϕ_p are constants and u_t is an $ARCH$-free white-noise process with variance σ^2.

The process is covariance-stationary provided that the roots of

$$(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p) = 0$$

lies outside the unit circle. The optimal linear forecast of the level of Y_t for an $AR(p)$ process is

$$\begin{aligned} E(Y_t|Y_{t-1}, Y_{t-2}, ...) &= \hat{E}(Y_t|Y_{t-1}, Y_{t-2}, ...) \\ &= c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p}, \end{aligned} \tag{9}$$

where $\hat{E}(Y_t|Y_{t-1}, Y_{t-2}, ...)$ denotes the linear projection of Y_t on a constant and $(Y_{t-1}, Y_{t-2}, ...)$. While the unconditional mean of Y_t is constant:

$$E(Y_t) = c/(1 - \phi_1 - \phi_2 - \cdots - \phi_p),$$

the conditional mean of Y_t changes over time according to (9), provided that the process is covariance-stationary. However, both the unconditional and the conditional variance of Y_t is constant under this model:[6]

$$Var(Y_t) = \gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + +... + \phi_p \gamma_p + \sigma^2,$$

and

$$Var(Y_t|Y_{t-1}, Y_{t-2}, ...) = E\{[Y_t - \hat{E}(Y_t|Y_{t-1}, Y_{t-2}, ...)]^2|Y_{t-1}, Y_{t-2}, ...\} = \sigma^2.$$

If we assume that the white noise process u_t in (8) now is a $ARCH(m)$ process as that in (1) and (2), then the conditional, unconditional mean, and the unconditional variance of Y_t are the same, they are constant over time. However the conditional variance of Y_t now would be

$$\begin{aligned} Var(Y_t|Y_{t-1}, Y_{t-2}, ...) &= E\{[Y_t - \hat{E}(Y_t|Y_{t-1}, Y_{t-2}, ...)]^2|Y_{t-1}, Y_{t-2}, ...\} \\ &= E(u_t^2|Y_{t-1}, Y_{t-2}, ...) \\ &= \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2, \end{aligned}$$

which change over time.

[6]For example, $p = 2$, $\gamma_0 = \frac{(1-\phi_2)\sigma^2}{(1+\phi_2)[(1-\phi_2)^2 - \phi_1^2]}$.

2.3 Estimation

2.3.1 Pure $ARCH(m)$ Model

Under the normality assumption, the likelihood function of an $ARCH(m)$ model, given a sample of size T is

$$\begin{aligned} f(u_T, u_{T-1}, ..., u_1) &= f(u_T|\Omega_{T-1})f(u_{T-1}|\Omega_{T-2})\cdots f(u_{m+1}|\Omega_m)f(u_m, u_{m-1}, ..., u_1) \\ &= \prod_{t=m+1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} exp\left(-\frac{(u_t-0)^2}{2\sigma_t^2}\right) \times f(u_m, u_{m-1}, ..., u_1), \end{aligned}$$

where Ω_t the vector of observations obtained through date t:

$$\Omega_t = (u_t, u_{t-1}, ..., u_1),$$

and we have used the fact the conditional mean and variance of u_t, i.e. $E(u_t|\Omega_{t-1}) = 0$ and $Var(u_t|\Omega_{t-1}) = \sigma_t^2$. Since the exact form of $f(u_m, u_{m-1}, ..., u_1)$ is complicated, it is commonly dropped from the likelihood function, especially when the sample size is sufficiently large. This results in using the conditional likelihood function

$$f(u_T, u_{T-1}, ..., u_{m+1}|u_m, u_{m-1}, ..., u_1) = \prod_{t=m+1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} exp\left(-\frac{u_t^2}{2\sigma_t^2}\right), \tag{10}$$

where $\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2$ can be evaluated recursively. We refer to estimates obtained by maximizing the likelihood function as the conditional maximum likelihood estimates under normality.

Maximizing the conditional likelihood function is equivalent to maximizing its logarithm, which is easier to handle. The conditional log likelihood function is

$$\begin{aligned} & l(u_T, u_{T-1}, ..., u_{m+1}|u_m, u_{m-1}, ..., u_1) \\ = & \ln f(u_T, u_{T-1}, ..., u_{m+1}|u_m, u_{m-1}, ..., u_1) \\ = & \sum_{t=m+1}^{T} \left[-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma_t^2) - \frac{1}{2}\frac{u_t^2}{\sigma_t^2}\right]. \end{aligned} \tag{11}$$

In some applications, it is more appropriate to assume that ε_t follows a "heavy-tailed" distribution such as a standardized Student-t distribution. Let x_v be a

Student-t distribution with v degree of freedom. Then $Var(x_v) = v/(v-2)$ for $v > 2$, and we use $\varepsilon_t = x_v/\sqrt{(v/v-2)}$. The probability density of ε_t is

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

Using $u_t = \sigma_t \varepsilon_t$, we obtain the conditional likelihood function of u_ts as

$$\begin{aligned} & f(u_T, u_{T-1}, ..., u_{m+1} | u_m, u_{m-1}, ..., u_1) \\ &= \prod_{t=m+1}^{T} \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{(v-2)\pi}} \frac{1}{\sigma_t} \left(1 + \frac{u_t^2}{(v-2)\sigma_t^2}\right)^{-(v+1)/2}, \quad v > 2. \end{aligned} \tag{12}$$

We refer to the estimate that maximize the prior likelihood function as the conditional MLE under t-distribution. The degrees of freedom of the t-distribution can be specified a prior or estimated jointly with other parameters. A value between 3 and 6 is often used if it is prespecified.

2.3.2 Linear Regression With $ARCH(m)$ Disturbance

Suppose that we are interested in estimating the parameter of a regression model with $ARCH$ disturbance. Let the regression equation be

$$Y_t = \mathbf{x}_t' \boldsymbol{\beta} + u_t.$$

Here $\mathbf{x}_t$ denote a vector of predetermined explanatory variables, which could include lagged value of Y. The disturbance term u_t is assumed to satisfy (1) and (2). If $\varepsilon_t \overset{i.i.d.}{\sim} N(0,1)$, then from (11) the sample conditional log likelihood function is then

$$l(\boldsymbol{\theta}) = -\frac{T-m}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=m+1}^{T} \ln(\sigma_t^2) - \frac{1}{2}\sum_{t=m+1}^{T} \frac{(Y_t - \mathbf{x}_t'\boldsymbol{\beta})^2}{\sigma_t^2}, \tag{13}$$

where

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2 \\ &= \alpha_0 + \alpha_1 (Y_{t-1} - \mathbf{x}_{t-1}'\boldsymbol{\beta})^2 + \alpha_2 (Y_{t-2} - \mathbf{x}_{t-2}'\boldsymbol{\beta})^2 + \cdots + \alpha_m (Y_{t-m} - \mathbf{x}_{t-m}'\boldsymbol{\beta})^2 \end{aligned} \tag{14}$$

and the vector of parameters to be estimated $\boldsymbol{\theta}' = (\boldsymbol{\beta}, \alpha_0, \alpha_1, \alpha_2, \cdots, \alpha_m)'$. For a given numerical value for the parameter vector $\boldsymbol{\theta}$, the sequence of conditional

variances can be calculated from (14) and used to evaluate the log likelihood function (13). This can then be maximized numerically using the methods described in Chapter 3.[7]

The preceding formulation of the likelihood function assumed that ε_t has a Gaussian distribution. However, the unconditional distribution of many financial time series seems to have flatter tails than allowed by the Gaussian family. Some of this can be explained by the presence of $ARCH$; that is, even if ε_t in (1) has a Gaussian distribution, the unconditional distribution of u_t is non-Gaussian with heavier tails than a Gaussian distribution.

The same basic approach can be used with non-Gaussian distribution. For example, Bollerselv (1987) proposed that ε_t might be drawn from a t distribution with v degree of freedom, where v is regarded as a parameter to be estimated by maximum likelihood. The sample log likelihood conditional on the first m observations from (13) then become

$$\begin{aligned}
&\sum_{m+1}^{T} \ln f(Y_t|\mathbf{x}_t, \Omega_{t-1}; \boldsymbol{\theta}) \\
&= (T-m)\ln\left\{\frac{\Gamma[(v+1)/2]}{\pi^{1/2}\Gamma(v/2)}(v-2)^{-1/2}\right\} - (1/2)\sum_{t=m+1}^{T}\ln(\sigma_t^2) \\
&-[(v+1)/2]\sum_{t=m+1}^{T}\ln\left[1+\frac{(Y_t-\mathbf{x}_t'\boldsymbol{\beta})^2}{\sigma_t^2(v-2)}\right],
\end{aligned}$$

where

$$\begin{aligned}
\sigma_t^2 &= \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2 \\
&= \alpha_0 + \alpha_1(Y_{t-1} - \mathbf{x}_{t-1}'\boldsymbol{\beta})^2 + \alpha_2(Y_{t-2} - \mathbf{x}_{t-2}'\boldsymbol{\beta})^2 + \cdots + \alpha_m(Y_{t-m} - \mathbf{x}_{t-m}'\boldsymbol{\beta})^2
\end{aligned}$$

and the vector of parameters to be estimated $\boldsymbol{\theta}' = (v, \boldsymbol{\beta}, \alpha_0, \alpha_1, \alpha_2, \cdots, \alpha_m)'$.

[7]Imposing the stationarity condition ($\sum_{j=1}^{m}\alpha_j < 1$) and the nonnegativity conditions ($\alpha_j \geq 0, \forall j$) can be difficult in practice. Typically, either the value of m is very small or else some ad hoc structure is imposed on the sequence $\{\alpha_j\}_{j=1}^{m}$ as in Engle (1982, eq. (38)).

2.4 Testing for $ARCH(m)$ Disturbance in a Linear Regression Model

2.4.1 Engle's LM type χ^2-Test

In the linear regression model, OLS is the appropriate procedure if the disturbance are not conditional heteroscedastic. Because the $ARCH$ model requires iterative procedure, it may be desirable to test whether it is appropriate before going to the effort to estimate it. The Lagrange multiplier test procedure is ideal for this as in many cases. See for example, Breusch and Pagan (1979, 1980), Godfrey (1978) and Engle (1979).

Consider the following linear regression model

$$\begin{aligned} Y_t &= \mathbf{x}_t'\boldsymbol{\beta} + u_t, \\ u_t &= \sqrt{h_t}\varepsilon_t, \\ h_t &= \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2 = \boldsymbol{\alpha}'\mathbf{u}_t, . \end{aligned}$$

where $\boldsymbol{\alpha} = [\alpha_0, ..., \alpha_m]'$ and $\mathbf{u}_t = [1, u_{t-1}, ..., u_{t-m}]'$. Under the null hypothesis, $\alpha_1 = ... = \alpha_m = 0$. The test is based upon the score under the null and the information matrix under the null. The likelihood function, omitting the revelent constants, can be written as

$$l(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=m+1}^{T} \ln(h_t) - \frac{1}{2}\sum_{t=m+1}^{T} \frac{u_t^2}{h_t}. \tag{15}$$

The first derivative of (15) is

$$\frac{\partial l}{\partial \boldsymbol{\alpha}} = \sum_{t=m+1}^{T} \left(\frac{1}{2h_t}\frac{\partial h_t}{\partial \boldsymbol{\alpha}} \left[\frac{u_t^2}{h_t} - 1 \right] \right).$$

Under the null hypothesis that $\alpha_1 = ...\alpha_m = 0$, then the estimation of $\boldsymbol{\beta}$ is simply the OLS estimator $\hat{\boldsymbol{\beta}}$. And therefore

$$h_t = \alpha_0, \quad \frac{\partial h_t}{\partial \boldsymbol{\alpha}} = [1, e_{t-1}^2, ..., e_{t-m}^2]' = \mathbf{z}_t',$$

where $e_t = Y_t - \mathbf{x}_t'\hat{\boldsymbol{\beta}}$. So under the null hypothesis

$$\begin{aligned} \left(\frac{\partial l}{\partial \boldsymbol{\alpha}}\right)_{H_0} &= \sum_{t=m+1}^{T} \left(\frac{1}{2\alpha_0}\mathbf{z}_t' \left[\frac{e_t^2}{\alpha_0} - 1 \right] \right) \\ &= \frac{1}{2\alpha_0}\mathbf{Z}'\mathbf{g}, \end{aligned}$$

where $\mathbf{Z}' = [\mathbf{z}'_{m+1}\ \mathbf{z}'_{m+2}\ ...\ \mathbf{z}'_T]$, $\mathbf{g} = \left[(\frac{e^2_{m+1}}{\alpha_0} - 1)\ (\frac{e^2_{m+2}}{\alpha_0} - 1)\ ...\ (\frac{e^2_T}{\alpha_0} - 1)\right]'$.

The second derivative of (15) is

$$\frac{\partial^2 l}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'} = \sum_{t=m+1}^{T} \left(-\frac{1}{2h_t^2} \frac{\partial h_t}{\partial \boldsymbol{\alpha}} \frac{\partial h_t}{\partial \boldsymbol{\alpha}'} \left[\frac{u_t^2}{h_t}\right] + \left[\frac{u_t^2}{h_t} - 1\right] \frac{\partial}{\partial \boldsymbol{\alpha}'} \left[\frac{1}{2h_t} \frac{\partial h_t}{\partial \boldsymbol{\alpha}}\right] \right).$$

The condition expectation $E(u_t|F_{t-1}) = h_t$, therefore the conditional expectation of the second term of () is zero and of the last factor in the first, is just one. Hence the information matrix which is simple the negative expectation of the Hessian matrix over all observation, become

$$\mathcal{I} = \sum_{t=m+1}^{T} E\left[-\frac{1}{2h_t^2} \frac{\partial h_t}{\partial \boldsymbol{\alpha}} \frac{\partial h_t}{\partial \boldsymbol{\alpha}'} \right]$$

which is consistently estimated under H_0 by

$$\begin{aligned} \hat{\mathcal{I}}_0 &= \sum_{t=m+1}^{T} \left[-\frac{1}{2h_t^2} \frac{\partial h_t}{\partial \boldsymbol{\alpha}} \frac{\partial h_t}{\partial \boldsymbol{\alpha}'} \right]_{H_0} \\ &= \sum_{t=m+1}^{T} \left(-\frac{1}{2\alpha_0^2} \mathbf{z}'_t \mathbf{z}_t \right) \\ &= \frac{1}{2\alpha_0^2} \mathbf{Z}'\mathbf{Z}. \end{aligned}$$

The LM test statistic can be consistently estimated by

$$\begin{aligned} LM^* &= \left(\frac{1}{2\alpha_0}\mathbf{Z}'\mathbf{g}\right)' \left(\frac{1}{2\alpha_0^2}\mathbf{Z}'\mathbf{Z}\right)^{-1} \left(\frac{1}{2\alpha_0}\mathbf{Z}'\mathbf{g}\right) \\ &= \frac{1}{2}\mathbf{g}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{g} \end{aligned}$$

which under the null hypothesis is asymptotically distributed as a χ^2_m distribution.

A simplicities is to note that

$$
\begin{aligned}
plim\frac{\mathbf{g}'\mathbf{g}}{T} &= plim\frac{1}{T}\left\{\left[\left(\frac{e_{m+1}^2}{\alpha_0}-1\right)\ \left(\frac{e_{m+2}^2}{\alpha_0}-1\right)\ ...\ \left(\frac{e_T^2}{\alpha_0}-1\right)\right]'\begin{bmatrix}(\frac{e_{m+1}^2}{\alpha_0}-1)\\(\frac{e_{m+2}^2}{\alpha_0}-1)\\ \cdot\\ \cdot\\ \cdot\\(\frac{e_T^2}{\alpha_0}-1)\end{bmatrix}\right\}\\
&= plim\frac{\sum^T e_t^4/T}{\alpha_0^2}-\frac{2\sum^T e_t^2/T}{\alpha_0}+\frac{T}{T}\quad (Under\ normality)\\
&= 3-2+1\\
&= 2.
\end{aligned}
$$

Thus, an asymptotically equivalent statistic would be

$$LM = T\cdot\frac{(\mathbf{g}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{g})}{\mathbf{g}'\mathbf{g}} = TR^2$$

where R^2 is the squared multiple correlation between $\mathbf{g}$ and $\mathbf{Z}$. Since adding a constant and multiplying by a scalar will not change the R^2 of a regression, this is also the R^2 of the regression of e_t^2 on an intercept and m lagged values of e_t^2. The statistic will be asymptotically distributed as χ^2 with m degree of freedom when the null hypothesis is true. The test procedure is to run the OLS regression and save the residuals. Regress the squared residuals on a constant and m lags and test TR^2 as χ_m^2

2.4.2 Tsay's LR type F-Test

It is simple to test whether the residuals u_t from a regression model exhibit time-varying heteroskedasticity without having to estimate the $ARCH$ parameters. Engle (1982) derived the following Lagrange multiplier principle. This test is equivalent to the usual F statistic for testing $\alpha_i = 0,\ i = 1, ..., m$ in the linear regression

$$\hat{u}_t^2 = \alpha_0 + \alpha_1\hat{u}_{t-1}^2 + \alpha_2\hat{u}_{t-2}^2 + \cdots + \alpha_m\hat{u}_{t-m}^2 + e_t,\quad t = m+1, ..., T, \tag{16}$$

where e_t denote the error term, m is a prespecified positive integer, and T is the sample size. Let $SSE_0 = \sum_{t=m+1}^T(\hat{u}_t^2 - \overline{\hat{u}^2})^2$,[8] where $\overline{\hat{u}^2}$ is the sample mean of

[8]Sums of squared error under null hypothesis.

$\hat{u}^2$, and $SSE_1 = \sum_{t=m+1}^{T} \hat{e}_t^2$, where $\hat{e}_t$ is the least squares residual of the linear regression (16). Then we have

$$F = \frac{(SSE_0 - SSE_1)/m}{SSE_1/(T-m-1)},$$

which is asymptotically distributed as a χ^2 distribution with m degree of freedom under the null hypothesis.

2.5 Forecasting from an $ARCH$ Model

3 The GARCH Model

Although the $ARCH$ model is simple, it often requires many parameters to adequately describe the volatility process such as of an asset returns. Some alternative model must be sought. Bollerslev (1986) proposes a useful extension known as the generalized $ARCH$ ($GARCH$) model.

The white noise process u_t follows a $GARCH(s, m)$ if

$$u_t = \sqrt{h_t}\varepsilon_t, \quad h_t = \alpha_0 + \sum_{j=1}^{s} \beta_j h_{t-j} + \sum_{i=1}^{m} \alpha_i u_{t-i}^2 \tag{17}$$

where again ε_t is a sequence of iid random variables with mean 0 and variance 1.0, $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, and $\sum_{i=1}^{\max(m,s)}(\alpha_i + \beta_i) < 1$.

3.1 Population's Properties of $GARCH$ Models

3.1.1 $GARCH$ Can be Regarded as an $ARMA$ Idea to u_t^2

To understand properties of $GARCH$ models, it is informative to use the following representation. Let $\eta_t = u_t^2 - h_t$, so that $h_t = u_t^2 - \eta_t$. By plugging $h_{t-i} = u_{t-i}^2 - \eta_{t-i} (i = 0, ..., s)$ into (), we can rewrite the $GARCH$ models as

$$u_t^2 = \alpha_0 + \sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i) u_{t-i}^2 + \eta_t - \sum_{j=1}^{s} \beta_j \eta_{t-j}. \tag{18}$$

Here, it is understood that $\alpha_i = 0$ for $i > m$ and $\beta_j = 0$ for $j > s$. Notice that h_t is the forecast of u_t^2 based on its own lagged value and thus $\eta_t = u_t^2 - h_t$ is the error associated with this forecast. Thus, η_t is a white noise process that is fundamental for u_t^2. (18) is recognized as an $ARMA$ form for the squared series u_t^2.

3.1.2 The Conditional Mean and Variance

Let F_{t-1} denote the information set available at time $t-1$. The conditional mean of u_t is

$$E(u_t|F_{t-1}) = \sqrt{h_t} \cdot E(\varepsilon_t|F_{t-1}) = 0. \tag{19}$$

From (17) it implies that the conditional variance of u_t is

$$\begin{aligned}
\sigma_t^2 &= Var(u_t|F_{t-1}) \\
&= E\{[u_t - E(u_t|F_{t-1})]^2|F_{t-1}\} \\
&= E(u_t^2|F_{t-1}) \qquad (since\ E(u_t|F_{t-1}) = 0) \\
&= E(h_t\varepsilon_t^2|F_{t-1}) \\
&= E(\varepsilon_t^2)E(\alpha_0 + \sum_{i=1}^{m}\alpha_i u_{t-i}^2 + \sum_{j=1}^{s}\beta_j h_{t-j}|F_{t-1}) \\
&= \alpha_0 + \sum_{i=1}^{m}\alpha_i u_{t-i}^2 + \sum_{j=1}^{s}\beta_j h_{t-j} \\
&= h_t.
\end{aligned}$$

3.1.3 The Conditional Density

By assuming that ε_t is a Gaussian variate, the condition density of u_t given all the information update to $t-1$ is

$$f(u_t|F_{t-1}) = \sqrt{h_t}f(\varepsilon_t|F_{t-1}) = \sqrt{h_t}\cdot N(0,1) \sim N(0,h_t).$$

3.1.4 The Unconditional Mean and Variance

The unconditional mean of u_t is $E[E(u_t|F_{t-1})] = E(0) = 0$. While u_t is conditional heteroscedastic, from (18) the unconditional variance of u_t is

$$\begin{aligned}
E(u_t^2) &= E\left[\alpha_0 + \sum_{i=1}^{\max(m,s)}(\alpha_i+\beta_i)u_{t-i}^2 + \eta_t - \sum_{j=1}^{s}\beta_j\eta_{t-j}\right] \\
&= \alpha_0 + \sum_{i=1}^{\max(m,s)}(\alpha_i+\beta_i)E(u_{t-i}^2).
\end{aligned}$$

The unconditional variance of u_t thus is

$$\sigma^2 = E(u_t^2) = \frac{\alpha_0}{1-\sum_{i=1}^{\max(m,s)}(\alpha_i+\beta_i)}$$

provided that the denominator of the fraction is positive.

3.1.5 Coefficients Constraint in an Conditional Gaussian $GARCH(1,1)$ model

The strengths and weakness of $GARCH$ model can easily be seen by focusing on the simplest $GARCH(1,1)$ model with

$$\begin{aligned} u_t &= \sqrt{h_t}\varepsilon_t \\ h_t &= \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1}, \quad 0 \le \alpha_1, \beta_1 \le 1, \ (\alpha_1 + \beta_1) < 1. \end{aligned}$$

with $\varepsilon_t \sim N(0,1)$. First, a large u_{t-1}^2 or h_{t-1} gives rise to a large σ_t^2. This means that a large u_{t-1}^2 tends to be followed by another large u_t^2, generating, again, the well-known behavior of volatility clustering in financial time series. Second, it can be shown that if $1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 > 0$, then

$$\frac{E(u_t^4)}{[E(u_t^2)]^2} = \frac{3[1 - (\alpha_1 + \beta_1)^2]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2} > 3. \tag{20}$$

Consequently, similar to $ARCH$ models, the tail distribution of a $GARCH(1,1)$ process is heavier than that of a normal distribution. Third, the model provides a simple parametric function that can be used to describe the volatility evolution.

To see the result of (20), by definition

$$E(u_t^2|F_{t-1}) = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1} = h_t,$$

and therefore

$$E(u_t^4|F_{t-1}) = 3[E(u_t^2|F_{t-1})] \quad = \quad 3(\alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1})^2$$

Hence

$$E(u_t^4) = 3E[\alpha_0^2 + \alpha_1^2 u_{t-1}^4 + \beta_1^2 h_{t-1}^2 + 2(\alpha_0\alpha_1 u_{t-1}^2) + 2(\alpha_0\beta_1 h_{t-1}) + 2(\alpha_1\beta_1 u_{t-1}^2 h_{t-1})]. \tag{21}$$

Denote $E(u_t^4) = m_4$. Notice that

$$\begin{aligned} E(h_t) &= E[E(u_t^2|F_{t-1})] = E(u_t^2) = \sigma^2, \\ E(h_t^2) &= E[E(u_t^2|F_{t-1})^2] = \sigma^4 = \frac{m_4}{3}, \\ E(u_{t-1}^2 h_{t-1}) &= E[E(u_{t-1}^2 h_{t-1}|F_{t-1})] = E[u_{t-1}^2 E(h_{t-1}|F_{t-1})] \\ &= E\{u_{t-1}^2[E(u_{t-1}^2|F_{t-2})|F_{t-1}]\} \\ &= E\{u_{t-1}^2[E(u_{t-1}^2|F_{t-2}]\} \\ &= \sigma^4 = \frac{m_4}{3}. \end{aligned}$$

(21) become

$$E(u_t^4) = 3\alpha_0^2 + 3\alpha_1^2 m_4 + \beta_1^2 m_4 + 6\alpha_0\alpha_1\sigma^2 + 6\alpha_0\beta_1\sigma^2 + 2\alpha_1\beta_1 m_4.$$

Consequently,

$$\begin{aligned}
(1 - 3\alpha_1^2 - \beta_1^2 - 2\alpha_1\beta_1)m_4 &= 3\alpha_0^2 + 6\alpha_0\alpha_1\sigma^2 + 6\alpha_0\beta_1\sigma^2 \\
&= 3\alpha_0^2 + + \frac{6\alpha_0\alpha_1 \cdot \alpha_0}{(1-\alpha_1-\beta_1)} + \frac{6\alpha_0\beta_1 \cdot \alpha_0}{(1-\alpha_1-\beta_1)} \\
&= \frac{3\alpha_0^2(1-\alpha_1-\beta_1) + 6\alpha_0^2\alpha_1 + 6\alpha_0^2\beta_1}{(1-\alpha_1-\beta_1)} \\
&= \frac{3\alpha_0^2(1+\alpha_1+\beta_1)}{(1-\alpha_1-\beta_1)},
\end{aligned}$$

that is,

$$m_4 = \frac{3\alpha_0^2(1+\alpha_1+\beta_1)}{(1-\alpha_1-\beta_1)(1-3\alpha_1^2-\beta_1^2-2\alpha_1\beta_1)}.$$

The unconditional kurtosis of u_t is

$$\begin{aligned}
\frac{m_4}{(\sigma^2)^2} &= \frac{3\alpha_0^2(1+\alpha_1+\beta_1)}{(1-\alpha_1-\beta_1)(1-3\alpha_1^2-\beta_1^2-2\alpha_1\beta_1)} \times \frac{(1-\alpha_1-\beta_1)^2}{\alpha_0^2} \\
&= \frac{3[1-(\alpha_1+\beta_1)^2]}{1-(\alpha_1+\beta_1)^2-2\alpha_1^2} > 3.
\end{aligned}$$

3.2 Testing for $GARCH$

3.3 Estimation

3.4 Forecasting from an $GARCH$ Model

4 The integrated $GARCH$ Model

If the AR polynomial of the $GARCH$ representation in Eq. (18) has a unit root, then we have (u_t) an $IGARCH$ model. Thus, $IGARCH$ models are unit-root $GARCH$ models. Similar to $ARIMA$ models, a key feature of $IGARCH$ models is that the impact of past squared shock $\eta_{t-i} = u_{t-i}^2 - \sigma_{t-1}^2$ for $i > 0$ on u_t^2 is persistent. For example, an $IGARCH(1,1)$ model can be written as

$$u_t = \sqrt{h_t}\varepsilon_t, \quad h_t = \alpha_0 + \beta_1 h_{t-1} + (1-\beta_1)u_{t-1}^2, \tag{22}$$

where ε_t is defined as before and $1 > \beta_1 > 0$

If u_t follows an $IGARCH$ model, then the unconditional variance of u_t is infinite, so neither u_t nor u_t^2 satisfies the definition of a covariance-stationary process.

5 The Exponential $GARCH$ Model

To overcome some weaknesses of the $GARCH$ model in handling financial time series, Nelson 91991) propose the exponential $GARCH$ ($EGARCH$) model. In particular, to allow for asymmetric effect between positive and negative innovation, he consider the weighted innovation

$$g(\varepsilon_t) = \theta\varepsilon_t + \gamma[|\varepsilon_t| - E(|\varepsilon_t|)],$$

where θ and γ are real constant. Both ε_t and $|\varepsilon_t| - E(|\varepsilon_t|)$ are zero-mean iid sequence with continuous distributions. Therefore, $E[g(\varepsilon_t)] = 0$. The asymmetry of $g(\varepsilon_t)$ can be seen by rewriting it as

$$g(\varepsilon_t) = \begin{cases} (\theta+\gamma)\varepsilon_t - \gamma E(|\varepsilon_t|) & if\ \varepsilon_t \geq 0, \\ (\theta-\gamma)\varepsilon_t - \gamma E(|\varepsilon_t|) & if\ \varepsilon_t < 0. \end{cases}$$

For the standard Gaussian random variable ε_t, $E(|\varepsilon_t|) = \sqrt{2/\pi}$. For the standardized student-t distribution, we have

$$E(|\varepsilon_t)| = \frac{2\sqrt{v-2}\Gamma((v+1)/2)}{(v-1)\Gamma(v/2)\sqrt{\pi}}.$$

An $EGARCH(m,s)$model can be written as

$$u_t = \sqrt{h_t}\varepsilon_t, \quad \ln(h_t) = \alpha_0 + \frac{1+\beta_1 L + ... + \beta_s L^s}{1-\alpha_1 L - ... - \alpha_m L^m} g(\varepsilon_{t-1}),$$

where α_0 is a constant and all roots of $1+\beta(L) = 0$ and $1-\alpha(L) = 0$ are outside the unit circle. Based on this representation, some properties of the $EGARCH$ model can be obtained in a similar manner as those of the $GARCH$ model. For instance, the unconditional mean of $\ln h_t$ is α_0. However, the model differs from the $GARCH$ model in at least two ways:
(a). It uses logged conditional variance to relax the positiveness constraint of model coefficients,
(b). The use of $g(\varepsilon_t)$ enables the model to respond asymmetrically to positive and negative lagged values of u_t.

To better understand the $EGARCH$ model, let us consider the simple $EGARCH(1,0)$ model

$$u_t = \sqrt{h_t}\varepsilon_t, \quad (1-\alpha L)\ln(h_t) = (1-\alpha L)\alpha_0 + g(\varepsilon_{t-1}),$$

where ε_t are *iid* standard normal. In this case, $E(|\varepsilon_t)| = \sqrt{2/\pi}$ and the model for $\ln h_t$ become

$$(1-\alpha L)\ln(h_t) = \begin{cases} \alpha^* + (\theta+\gamma)\varepsilon_{t-1} & if\ \varepsilon_{t-1} \geq 0 \\ \alpha^* + (\theta-\gamma)\varepsilon_{t-1} & if\ \varepsilon_{t-1} < 0, \end{cases}$$

where $\alpha^* = (1-\alpha)\alpha_0 - \sqrt{2/\pi}\gamma$. This is a nonlinear function similar to that of the threshold autoregressive model (TAR) of Tong (1978,1990). It suffices to say that for this simple $EGARCH$ model the conditional variance evolves in a nonlinear manner depending on the sign of u_{t-1}. specifically, we have

$$h_t = h_{t-1}^{\alpha}\exp(\alpha^*)\begin{cases} \exp\left[(\theta+\gamma)\frac{u_{t-1}}{\sqrt{h_{t-1}^2}}\right] & if\ \ u_{t-1} \geq 0, \\ \\ \exp\left[(\theta-\gamma)\frac{u_{t-1}}{\sqrt{h_{t-1}^2}}\right] & if\ \ u_{t-1} < 0. \end{cases}$$

The coefficients $(\theta+\gamma)$ and $(\theta-\gamma)$ show the asymmetry in response to positive and negative u_{t-1}. The model is nonlinear if $\gamma \neq 0$. For higher order $EGARCH$ models, the nonlinearity become much more complicated.

6 The $GARCH - M$ Model

Finance theory suggests that an asset with a higher perceived risk would pay a higher return on average. For example, let r_t denote the expose rate of return on some asset minus the return on a safe alternative asset. Suppose that r_t is decomposed into a component anticipated by investors at date $t-1$ (denoted μ_t and a component that was unanticipated (denoted u_t):

$$r_t = \mu_t + u_t.$$

Then the theory suggests that the mean return μ_t would be related to the variance of the return h_t. In general, the $ARCH$-in mean or $ARCH_M$, regression model introduced by Engle, Lilien, and Robins (1987) is characterized by

$$\begin{aligned} Y_t &= \mathbf{x}_t'\boldsymbol{\beta} + \delta h_t + u_t \\ u_t &= \sqrt{h_t}\varepsilon_t, \quad h_t = \alpha_0 + \sum_{i=1}^{m} \alpha_i u_{t-i}^2 \end{aligned}$$

for ε_t $i.i.d.$ with zero mean and unit variance. The effect that higher perceived variability of u_t has on the level of Y_t is captured by the parameter δ.

7 Multivariate $GARCH$ Models

The preceding ideas can also be extend to an $(n \times 1)$ vector $\mathbf{u}_t$. Let $\mathbf{u}_t$ be a vector of white noise. Let $\mathbf{H}_t$ denote the $(n \times n)$ conditional variance-covariance matrix of $\mathbf{u}_t$:

$$\mathbf{H}_t = E(\mathbf{u}_t \mathbf{u}_t' | F_{t-1}).$$

Engle and Kroner (1993) proposed the following vector generalization of a $GARCH(r, m)$ specification:

$$\begin{aligned} \mathbf{H}_t &= \mathbf{K} + \mathbf{A}_1 \mathbf{H}_{t-1} \mathbf{A}_1' + \mathbf{A}_2 \mathbf{H}_{t-2} \mathbf{A}_2' + ... + \mathbf{A}_r \mathbf{H}_{t-r} \mathbf{A}_r' \\ &\quad + \mathbf{B}_1 \mathbf{u}_{t-1} \mathbf{u}_{t-1}' \mathbf{B}_1' + \mathbf{B}_2 \mathbf{u}_{t-2} \mathbf{u}_{t-2}' \mathbf{B}_2' + ... + \mathbf{B}_m \mathbf{u}_{t-1} \mathbf{u}_{t-1}' \mathbf{B}_m'. \end{aligned}$$

Here $\mathbf{K}$, $\mathbf{A}_s$, and $\mathbf{B}_s$ for $s = 1, 2, ...$ denote $(n \times n)$ matrices of parameters.

In practice, for reasonably sized n it is necessary to restrict the specification for $\mathbf{H}_t$ further to obtain a numerically traceable formulation. One useful special case restricts $\mathbf{A}_s$ and $\mathbf{B}_s$ to be diagonal matrices for $s = 1, 2, ...$ In such a model, the conditional covariance between u_{it} and u_{jt} depends only on past values of $u_{i,t-s} \cdot u_{j,t-s}$, and not on the products or squares of other disturbance.

7.1 Constant Conditional Correlations Specification

An popular approach introduced by Bollerslev (1990) assumes that the conditional correlations among the element of $\mathbf{u}_t$ are constant over time. Let $h_{ii}^{(t)}$ denote the row i, column i element of $\mathbf{H}_t$. Thus $h_{ii}^{(t)}$ represent the conditional variance of ith element of $\mathbf{u}_t$:

$$h_{ii}^{(t)} = E(u_{it}^2 | F_{t-1}).$$

This conditional variance might be modeled with a univariate $GARCH(1, 1)$ process driven by lagged innovation in variables i:

$$h_{ii}^{(t)} = \kappa_i + a_i h_{ii}^{(t-1)} + b_i u_{i,t-1}^2.$$

We might postulate n such $GARCH$ specifications $(i = 1, 2, ..., n)$, one for each element of $\mathbf{u}_t$. The conditional covariance between u_{it} and u_{jt}, or the row i, column j element of $\mathbf{H}_t$, is taken to be a constant correlation ρ_{ij} (in stead of $0\rho_{ijt}$) times the conditional standard deviation of u_{it} and u_{jt}:

$$h_{ij}^{(t)} = E(u_{it} u_{jt} | F_{t-1}) = \rho_{ij} \cdot \sqrt{h_{ii}^{(t)}} \sqrt{h_{jj}^{(t)}}.$$

Maximum likelihood estimation of this specification turns out to be quite tractable; see Bollerslev (1990) for details.