TOPIC 5
MODELLING ASSET RETURN VOLATILITY

1. Introduction

We introduce the ARCH\GARCH class of models which were developed to account for the persistence in squared returns which, as we have seen, is a typical feature of asset return data. ARCH and GARCH refer, respectively, to an Autoregressive Conditional Heteroscedastic and a Generalized Autoregressive Conditional Heteroscedastic model. These models are heteroscedatic and autoregressive because they model time-varying volatility in returns and do that by taking account of past volatility.

2. ARCH Processes

Consider a very simple model for returns $\{y_t\}$, which in practice is not an unreasonable characterization for mean returns, namely,

$$y_t = c + \varepsilon_t \tag{1}$$

where $\varepsilon_t \sim WN(0,\sigma^2)$ and *c* is the intercept. This assumption means that the unconditional mean and variance of the innovation ε_t are constant and finite (equal to zero and σ^2, respectively) and that the innovations are serially uncorrelated at all leads and lags. In topic 3, we made the further assumption that the innovations were independently and identically distributed so that the innovations were characterized as a strict white noise process, that is, $\varepsilon_t \sim iid\,WN(0,\sigma^2)$. The assumption of independence means that the conditional probability density function of $\varepsilon_t \mid \Omega_{t-1}$ where $\Omega_{t-1} = \{\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$ is the same as the unconditional probability density function of ε_t. This has the implication that the conditional variance of the innovation, namely, $\mathrm{var}(\varepsilon_{t+j} \mid \Omega_t)$ depends only on *j* and not on the conditioning information set Ω_t. Thus, for example, when *j*=1

$$\mathrm{var}(\varepsilon_t \mid \Omega_{t-1}) = \mathrm{var}(\varepsilon_{t+1} \mid \Omega_t)$$

The arrival of "news" at t, namely ε_t, does not affect the value of conditional variance of the innovation one period ahead, namely, at *t*+1. In other words, the conditional variance does not adapt to the arrival of new conditioning information. In financial markets, however, return volatility responds to the arrival of "news" and the present specification cannot capture this feature of financial data.

To do so, we will continue to assume that the innovations are white noise but that they are *not* strictly white noise by dropping the assumption of independence. In particular, we will assume a particular dependence structure in the conditional variance of the innovation.

The Autoregressive Conditional Heteroscedastic Model of order one (ARCH(1)) for the innovation ε_t is

$$
\begin{aligned}
&\varepsilon_t \mid \Omega_{t-1} \sim N(0, \sigma_t^2) \\
&\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \qquad \alpha_0 > 0, \quad 0 \le \alpha_1 < 1
\end{aligned}
\tag{2}
$$

where $\sigma_t^2 = \mathrm{var}(\varepsilon_t \mid \Omega_{t-1})$, the conditional variance of ε_t based on the information set $\Omega_{t-1} = \{\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$. The ε_t's are serially uncorrelated but they are not serially independent because the current conditional variance of ε_t depends on the one period lagged value of ε_t^2. The conditional distribution is assumed normal so that equations (1) and (2) can be estimated jointly by maximum likelihood estimation.

It is important to stress that ε_t is a weak white noise process and is covariance stationary. The unconditional mean and the unconditional variance of ε are constant and finite and are, respectively,

$$
\begin{aligned}
&E(\varepsilon_t) = 0 \\
&\mathrm{var}(\varepsilon_t) = E(\varepsilon_t - E(\varepsilon_t))^2 = \frac{\alpha_0}{1 - \alpha_1}
\end{aligned}
$$

The restrictions $\alpha_0 > 0$, $0 \le \alpha_1 < 1$, ensure that the unconditional variance is positive and finite and, also, that the conditional variance is always positive. The important point is that the unconditional mean and variance are finite and constant as required for covariance stationarity. Also, as stated earlier the ε_t's are serially uncorrelated. The conditional mean and variance are

$$
\begin{aligned}
E(\varepsilon_t \mid \Omega_{t-1}) &= 0 \\
\mathrm{var}(\varepsilon_t \mid \Omega_{t-1}) &= E[(\varepsilon_t - E(\varepsilon_t \mid \Omega_{t-1}))^2 \mid \Omega_{t-1}] \\
&= E(\varepsilon_t^2 \mid \Omega_{t-1}) \\
&= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \\
&= \sigma_t^2
\end{aligned}
$$

The important point here is that the conditional variance is time-varying and thus heteroscedastic. Note that

$$
\mathrm{var}(\varepsilon_{t+1} \mid \Omega_t) = \alpha_0 + \alpha_1 \varepsilon_t^2
$$

and it follows that $\mathrm{var}(\varepsilon_t \mid \Omega_{t-1}) \ne \mathrm{var}(\varepsilon_{t+1} \mid \Omega_t)$. The arrival of news, namely, ε_t, will affect the conditional variance for next period (at time *t*+1). This is the key feature of ARCH type models: the conditional variance adapts to the conditioning information set. Clearly, if $\alpha_1 = 0$, there are no conditional variance dynamics and we are back to case of independence where $\varepsilon_t \sim iid\,WN(0, \alpha_0)$ from which it follows, for example, that $\mathrm{var}(\varepsilon_t \mid \Omega_{t-1}) = \mathrm{var}(\varepsilon_{t+1} \mid \Omega_t)$.

The ARCH model is capable of capturing the feature of volatility clustering that is observed in financial data where large changes in returns tend to be followed by large changes, and small changes in returns by small changes, of either sign. A large observed ε_{t-1}^2 produces a large conditional variance at time *t*, that is, a large σ_t^2, thereby increasing the likelihood of a large ε_t^2 at time *t*. The ARCH process models conditional variance dynamics in an autoregressive fashion.

Clearly, in the case of equation (1), the unconditional and conditional mean of y_t is just c. The conditional variance of y_t is time-varying and equal to:

$$
\begin{aligned}
\operatorname{var}(y_t \mid \Omega_{t-1}) &= E[(y_t - E(y_t \mid \Omega_{t-1}))^2 \mid \Omega_{t-1}] \\
&= E(c + \varepsilon_t - c)^2 \mid \Omega_{t-1}) \\
&= E(\varepsilon_t^2 \mid \Omega_{t-1}) \\
&= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \\
&= \sigma_t^2
\end{aligned}
$$

Equation (1) is referred to as the mean equation and equation (2) is the variance equation. In practice, both equations are estimated jointly using maximum likelihood techniques and an estimate is obtained for each parameter, namely, for c, α_0 and α_1. There is no need for the mean equation to take the particularly simple form of equation (1). The mean equation could easy be:

$$y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \varepsilon_t \quad \text{with} \quad E(\varepsilon_t \mid X_{1,t}, X_{2,t}) = 0$$

or

$$y_t = c + \varepsilon_t + \theta \varepsilon_{t-1}$$

or

$$y_t = c + \rho y_{t-1} + \varepsilon_t$$

where, in each case, $\varepsilon_t \sim WN(0, \sigma^2)$. Similarly, there is no need to restrict the variance equation to an ARCH specification with one autoregressive lag. The ARCH(q) specification for the variance equation is:

$$
\begin{aligned}
&\varepsilon_t \mid \Omega_{t-1} \sim N(0, \sigma_t^2) \\
&\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \ldots + \alpha_q \varepsilon_{t-q}^2 \\
&\alpha_0 > 0,\ \ \alpha_i \geq 0 \ \text{ for all } i = 1 \ to \ q,\ \ \textstyle\sum_{i=1}^{q} \alpha_i < 1
\end{aligned}
$$

The stated restrictions are sufficient to ensure that the conditional and unconditional variances are positive and finite and that ε_t is covariance stationary. Note that for the ARCH(q) process, the unconditional mean, variance and covariance are,

$$
\begin{aligned}
&E(\varepsilon_t) = 0 \\
&\operatorname{var}(\varepsilon_t) = \frac{\alpha_0}{1 - \sum_{i=1}^{q} \alpha_i} \\
&\operatorname{cov}(\varepsilon_t, \varepsilon_{t-j}) = 0 \quad for \quad j > 0.
\end{aligned}
$$

while the conditional mean and variance are,

$$
\begin{aligned}
E(\varepsilon_t \mid \Omega_{t-1}) &= 0 \\
\operatorname{var}(\varepsilon_t \mid \Omega_{t-1}) &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \ldots + \alpha_q \varepsilon_{t-q}^2 \\
&= \sigma_t^2
\end{aligned}
$$

Consider now the y_t process. It is important to realize that no matter the specific form of the mean equation for y_t, it will always be the case that

$$
\begin{aligned}
\operatorname{var}(y_t \mid \Omega_{t-1}) &= \operatorname{var}(\varepsilon_t \mid \Omega_{t-1}) \\
&= \sigma_t^2
\end{aligned}
$$

For example, suppose y_t follows the AR(1) process given above. Then the conditional mean of y_t is

$$
\begin{aligned}
E(y_t \mid \Omega_{t-1}) &= E[(c + \rho y_{t-1} + \varepsilon_t) \mid \Omega_{t-1}] \\
&= c + \rho y_{t-1} + E(\varepsilon_t \mid \Omega_{t-1}) \\
&= c + \rho y_{t-1}
\end{aligned}
$$

which is the forecast value of y_t on the basis of time *t*-1 information.
Now

$$
\begin{aligned}
\operatorname{var}(y_t \mid \Omega_{t-1}) &= E[(y_t - E(y_t \mid \Omega_{t-1}))^2 \mid \Omega_{t-1}] \\
&= E[(c + \rho y_{t-1} + \varepsilon_t - (c + \rho y_{t-1}))^2 \mid \Omega_{t-1}] \\
&= E[\varepsilon_t^2 \mid \Omega_{t-1}] \\
&= \sigma_t^2
\end{aligned}
$$

The two-standard error confidence interval for the one-step ahead forecast of y_t, based on information available at time *t*-1 (i.e. Ω_{t-1}) for the AR(1) model is:

$$c + \rho y_{t-1} \pm 2\sqrt{\sigma_t^2}$$

In the case of an ARCH(1) process it is

$$c + \rho y_{t-1} \pm 2\sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2}$$

On the basis of time *t* information, the two-standard error confidence interval for the one-step ahead forecast of y_{t+1}, based on information available at time *t* (i.e. Ω_t) is:

$$c + \rho y_t \pm 2\sqrt{\sigma_{t+1}^2}$$

In the case of an ARCH(1) process it is

$$c + \rho y_t \pm 2\sqrt{\alpha_0 + \alpha_1 \varepsilon_t^2}$$

Notice that not only is the forecasted value of *y* updated, as it now depends on y_t, but also the standard error of the forecast is updated as it depends on ε_t^2. Thus, the arrival of new information at time *t* influences not only the forecast but the confidence interval associated with the forecast.

2. GARCH Processes

The generalized ARCH or GARCH(1,1) model is:

$$y_t = c + \varepsilon_t \tag{3}$$

$$\varepsilon_t \mid \Omega_{t-1} \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{4}$$

$$\alpha_0 > 0,\ \ \alpha_1 \geq 0,\ \ \beta_1 \geq 0,\ \ \alpha_1 + \beta_1 < 1$$

Here the conditional variance at time *t* (σ_t^2) depends not only on last period's squared innovation (ε_{t-1}^2) but also on the conditional variance last period (σ_{t-1}^2). The parameter restrictions ensure that the unconditional variance and the conditional variance are positive and finite and that y_t is covariance stationary. Equation (3) is the mean equation. In this case, it is very simple but it can assume any form, for example, a regression equation or an MA or AR process. Equation (4) is the variance equation. The unconditional mean and the unconditional variance of ε are constant and finite and are, respectively,

$$E(\varepsilon_t) = 0$$

$$\text{var}(\varepsilon_t) = E(\varepsilon_t - E(\varepsilon_t))^2 = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$$

Note that the unconditional covariance is $\text{cov}(\varepsilon_t, \varepsilon_{t-j}) = 0$ for j>0. The conditional mean and variance are

$$E(\varepsilon_t \mid \Omega_{t-1}) = 0$$

$$\begin{aligned}\text{var}(\varepsilon_t \mid \Omega_{t-1}) &= E[(\varepsilon_t - E(\varepsilon_t \mid \Omega_{t-1}))^2 \mid \Omega_{t-1}] \\ &= E(\varepsilon_t^2 \mid \Omega_{t-1}) \\ &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\ &= \sigma_t^2\end{aligned}$$

The important point is that the unconditional variance is constant, as must be the case under covariance stationarity, whereas the conditional variance is time-varying.

The two-standard error confidence interval for the one-step ahead forecast of y_t, based on information available at time t-1 (i.e. Ω_{t-1}) for the AR(1) model is:

$$c + \rho y_{t-1} \pm 2\sqrt{\sigma_t^2}$$

In the case of a GARCH(1,1) process it is

$$c + \rho y_{t-1} \pm 2\sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2}$$

On the basis of time t information, the two-standard error confidence interval for the one-step ahead forecast of y_t, based on information available at time t (i.e. Ω_t) is:

$$c + \rho y_t \pm 2\sqrt{\sigma_{t+1}^2}$$

In the case of an GARCH(1,1) process it is

$$c + \rho y_t \pm 2\sqrt{\alpha_0 + \alpha_1 \varepsilon_t^2 + \beta_1 \sigma_t^2}$$

Notice that not only is the forecasted value of y updated, as it now depends on y_t, but also the standard error of the forecast is updated as it depends on ε_t^2 and σ_t^2. Thus, the arrival of new information at time t influences not only the forecast but the confidence interval associated with the forecast.

In practice, it was found that to adequately model volatility, the ARCH(q) specification required a very long number of lags, that is, q was found to be very large. Not only do a large number of lags use up degrees of freedom but with so many parameters to estimate it is more probable that one or more of the estimated parameters may be negative, violating the sign restrictions on the parameters. This led to the development of the GARCH model. We will now show that the GARCH(1,1) model is equivalent to a restricted ARCH model of infinite order. Now

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

Substitute for σ_{t-1}^2 to get

$$\begin{aligned}\sigma_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 (\alpha_0 + \alpha_1 \varepsilon_{t-2}^2 + \beta_1 \sigma_{t-2}^2) \\ &= \alpha_0 (1 + \beta_1) + \alpha_1 (\varepsilon_{t-1}^2 + \beta_1 \varepsilon_{t-2}^2) + \beta_1^2 \sigma_{t-2}^2\end{aligned}$$

Continue in this way to get,

$$\sigma_t^2 = \alpha_0 (1 + \beta_1 + \ldots + \beta_1^{j-1}) + \alpha_1 (\varepsilon_{t-1}^2 + \beta_1 \varepsilon_{t-2}^2 + \ldots + \beta_1^{j-1} \varepsilon_{t-j}^2) + \beta_1^j \sigma_{t-j}^2$$

As $j \to \infty$,

$$\sigma_t^2 = \frac{\alpha_0}{1 - \beta_1} + \alpha_1 (\varepsilon_{t-1}^2 + \beta_1 \varepsilon_{t-2}^2 + \ldots + \beta_1^{j-1} \varepsilon_{t-j}^2 + \ldots)$$

This follows because $\beta_1 < 1$, given the sign restrictions of the GARCH(1,1) model. Thus, the GARCH(1,1) model is equivalent to an ARCH(∞) where the coefficients on the lagged squared innovations decline geometrically. The GARCH(1,1) is a parsimonious model since there are only three parameters to estimate: $\alpha_0, \alpha_1, \beta_1$.

In many finance applications, particularly in Value-at-Risk calculations and the pricing of options, forecasts of the conditional variance are required. We will now derive an expression for the *h*-step ahead forecast of the conditional variance for the GARCH(1,1) specification. By definition

$$\sigma_t^2 = E(\varepsilon_t^2 \mid \Omega_{t-1})$$

Updating by *h* periods, we obtain

$$\sigma_{t+h}^2 = E(\varepsilon_{t+h}^2 \mid \Omega_{t+h-1})$$

The optimal forecast of σ_{t+h}^2 based on information available at time *t* is $E(\sigma_{t+h}^2 \mid \Omega_t)$. By construction,

$$\begin{aligned}E[\sigma_{t+h}^2 \mid \Omega_t] &= E[E(\varepsilon_{t+h}^2 \mid \Omega_{t+h-1}) \mid \Omega_t] \\ &= E(\varepsilon_{t+h}^2 \mid \Omega_t) \qquad (5)\end{aligned}$$

where the last equality follows from the law of iterative expectations. The GARCH(1,1) model is

$$\sigma_{t+h}^2 = \alpha_0 + \alpha_1 \varepsilon_{t+h-1}^2 + \beta_1 \sigma_{t+h-1}^2$$

Take the conditional expectation on the basis of the information set Ω_t to get

$$\begin{aligned} E(\sigma_{t+h}^2 \mid \Omega_t) &= \alpha_0 + \alpha_1 E(\varepsilon_{t+h-1}^2 \mid \Omega_t) + \beta_1 E(\sigma_{t+h-1}^2 \mid \Omega_t) \\ &= \alpha_0 + (\alpha_1 + \beta_1) E(\sigma_{t+h-1}^2 \mid \Omega_t) \qquad (6) \end{aligned}$$

where the last equality follows from equation (5). By recursive substitution, we can write equation (6) as

$$\begin{aligned} E(\sigma_{t+h}^2 \mid \Omega_t) &= \alpha_0[1 + (\alpha_1 + \beta_1) + (\alpha_1 + \beta_1)^2 + \ldots + (\alpha_1 + \beta_1)^{h-2}] + (\alpha_1 + \beta_1)^{h-1} E(\sigma_{t+1}^2 \mid \Omega_t) \\ &= \alpha_0[1 + (\alpha_1 + \beta_1) + (\alpha_1 + \beta_1)^2 + \ldots + (\alpha_1 + \beta_1)^{h-2}] + (\alpha_1 + \beta_1)^{h-1} \sigma_{t+1}^2 \end{aligned}$$

where the last step follows because σ_{t+1}^2 is known at time *t*. This formula can also be used to calculate a two standard error confidence band around the optimal *h*-step ahead forecast of y. Fortunately, EViews automatically calculates forecasts of the *h*-step ahead conditional variance using this formula for the case of a GARCH(1,1) process. Provided $\alpha_1 + \beta_1 < 1$, the forecast of the conditional variance will converge to the unconditional variance of ε_t as $h \to \infty$. That is

$$\lim_{h \to \infty} E(\sigma_{t+h}^2 \mid \Omega_t) = \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)} = \mathrm{Var}(\varepsilon_t)$$

Further if ε_t follows a GARCH(1,1) process, then it can be shown that ε_t^2 has an ARMA(1,1) representation, namely,

$$\varepsilon_t^2 = \alpha_0 + (\alpha_1 + \beta_1)\varepsilon_{t-1}^2 - \beta_1 v_{t-1} + v_t$$

where $v_t = \varepsilon_t^2 - \sigma_t^2$ is the difference between the squared innovation and the conditional variance at time *t*. (You will be asked to show this result in a tutorial exercise). The term ε_t^2 is a noisy proxy for the conditional variance σ_t^2: ε_t^2 is an unbiased predictor of σ_t^2 but it is more volatile. Also, recall that real world financial asset returns are typically unconditionally symmetric but leptokurtic (that is, more peaked in the centre and with fatter tails than a normal distribution). It turns out that the implied unconditional distribution of the conditionally normal GARCH process is also symmetric and leptokurtic.

Finally, the GARCH(1,1) can be extended to a GARCH(p,q) process given by

$$\varepsilon_t \mid \Omega_{t-1} \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2$$

$$\alpha_0 > 0, \alpha_i \geq 0, \beta_i \geq 0, \sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i < 1$$

As before, the stated conditions ensure that ε_t is covariance stationary and that the conditional variance is positive. The unconditional mean and the unconditional variance of ε are constant and finite and are, respectively,

$$E(\varepsilon_t) = 0$$

$$\text{var}(\varepsilon_t) = E(\varepsilon_t - E(\varepsilon_t))^2 = \frac{\alpha_0}{1 - \sum_{i=1}^{q} \alpha_i - \sum_{i=1}^{p} \beta_i}$$

The conditional mean and variance are

$$E(\varepsilon_t \mid \Omega_{t-1}) = 0$$

$$\begin{aligned}\text{var}(\varepsilon_t \mid \Omega_{t-1}) &= E[(\varepsilon_t - E(\varepsilon_t \mid \Omega_{t-1}))^2 \mid \Omega_{t-1}] \\ &= E(\varepsilon_t^2 \mid \Omega_{t-1}) \\ &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \\ &= \sigma_t^2\end{aligned}$$

Also, as before, no matter the specific form of the mean equation for y_t, it will always be the case that

$$\begin{aligned}\text{var}(y_t \mid \Omega_{t-1}) &= \text{var}(\varepsilon_t \mid \Omega_{t-1}) \\ &= \sigma_t^2\end{aligned}$$

3. Maximum Likelihood Estimation of the GARCH(1,1) Process

In lecture note 2, we discussed maximum likelihood estimation of the simple linear regression model

$$Y_t = \beta_0 + \beta_1 X_t + u_t \text{ where } u_t \sim iid\, N(0, \sigma^2).$$

The log-likelihood function for this model is

$$l(\beta_0, \beta_1, \sigma^2 \mid Y_t, X_t \;\; t = 1, \ldots, T) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(Y_t - \beta_0 - \beta_1 X_t)^2$$

Now lets assume the conditional mean equation is

$$Y_t = \delta_0 + \delta_1 X_t + \varepsilon_t$$

and the conditional variance equation is

$$\varepsilon_t \mid \Omega_{t-1} \sim N(0, \sigma_t^2)$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$
$$\alpha_0 > 0,\ \ \alpha_1 \geq 0,\ \ \beta_1 \geq 0,\ \ \alpha_1 + \beta_1 < 1$$

The log-likelihood function for this model is

$$l(\delta_0, \delta_1, \alpha_0, \alpha_1, \beta_1)$$
$$= -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\ln(\alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2) - \frac{1}{2}\sum_{t=1}^{T}\frac{(Y_t - \delta_0 - \delta_1 X_t)^2}{(\alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2)}$$

Once we substitute $\varepsilon_{t-1}^2 = (Y_{t-1} - \delta_0 - \delta_1 X_{t-1})^2$ and σ_{t-1}^2 in the log-likelihood function, it is then possible to maximize the log-likelihood with respect to each of the parameters ($\delta_0, \delta_1, \alpha_0, \alpha_1, \beta_1$). The values of ε_0^2 and σ_0^2 have to be set and they are usually set equal to the unconditional variance of the $\varepsilon's$ from the estimated mean equation, which is

$\hat{\sigma}_{ols}^2 = \frac{1}{T}\sum_{t=1}^{T} e_t^2$ where the e's are the OLS residuals. EViews actually uses a more sophisticated approach. Because the first-order conditions are nonlinear, they are solved using an iterative search procedure. In order to implement such a procedure, starting values for the parameters need to be specified. The OLS estimates of δ_0 and δ_1 from the mean equation are used as the starting values for these parameters, respectively. In the absence of GARCH effects, α_0 can be interpreted as the unconditional variance of the $\varepsilon's$. Thus, an appropriate starting value of α_0 is $\hat{\sigma}_{ols}^2$. For α_1 and β_1, arbitrarily select a small number (say, 0.05) for both as a starting value. The search procedure will iterate from these starting values and continue until a convergence criterion is satisfied. Hopefully, the resulting estimates will correspond to a global and not just a local maximum of the log-likelihood function.

4. Tests for ARCH\GARCH Effects

To see whether there are ARCH\GARCH effects evident in the data, first estimate the mean equation by OLS and save the residuals. Denote the OLS residuals from the mean equation as e_t. Consider the autocorrelations of the squared OLS residuals, that is, of e_t^2. If the autocorrelations for the squared residuals are large and exceed the Bartlet bands, that is an indication of the presence of ARCH\GARCH effects in the data. Alternatively, one may use an LM test for ARCH\GARCH. Estimate the auxiliary regression

$$e_t^2 = \gamma_0 + \gamma_1 e_{t-1}^2 + \ldots + \gamma_q e_{t-q}^2 + v_t$$

where the *e's* are the OLS residuals from the mean equation and v_t is an error term. Obtain the R^2 from this regression. The LM test statistic is

$$TR^2 \sim \chi_q^2$$

where T is number of OLS residuals. The null and alternative hypotheses are

$$H_0 : \gamma_i = 0 \text{ for all } i = 1, 2, \ldots, q.$$
$$H_1 : \gamma_i \neq 0 \text{ for at least one } i = 1, 2, \ldots, q.$$

Large values of TR^2 lead to a rejection of the null hypothesis of no ARCH\GARCH effects. Alternatively, an F-test for $\gamma_i = 0$ for all $i = 1, 2, \ldots, q$ could also be used.

5. Tests of Model Adequacy

It can be shown that if $\varepsilon_t \mid \Omega_{t-1} \sim N(0, \sigma_t^2)$, then

$$\frac{\varepsilon_t}{\sigma_t} = v_t \sim iidN(0,1)$$

This provides a means to check on the adequacy of the estimated ARCH\GARCH model. Standardize the residuals e_t by the conditional standard deviation from the fitted ARCH\GARCH model $\hat{\sigma}_t$ to get $(e_t / \hat{\sigma}_t)$. If the ARCH\GARCH model is adequate, then the standardized residuals should be uncorrelated and homoscedastic. Standard tests can be applied to the standardized residuals to see whether that is the case.

Finally, the standardized residuals should appear normally distributed. When the assumption of conditional normality does not hold, the ARCH\GARCH parameter estimates will still be consistent provided the mean and variance equations are correctly specified. However, the standard errors are no longer correct. In this case, we treat the log-likelihood function as being approximately correct but calculate robust standard errors. This approach is known as Quasi-Maximum Likelihood and the resulting standard errors are known as Bollerslev-Wooldridge standard errors. The Jarque-Bera test can be applied to the standardized residuals to test for normality.

6. An Application to the NYSE Composite Index

Figure 1 shows a graph of the percentage daily logarithmic change in the NYSE Composite Index (which is the percentage daily log return, denoted sr_t) over the period

January 3, 1995 to August 30, 2002, a total of 1,931 observations. Figure 2 shows a graph of the squared return (that is, sr_t^2). Volatility clustering is apparent in this graph whereby large changes tend to be followed by large changes and small changes by small changes.

Figure 1: Daily Log Return in the NYSE Composite Index
January 3, 1995 to August 30, 2002

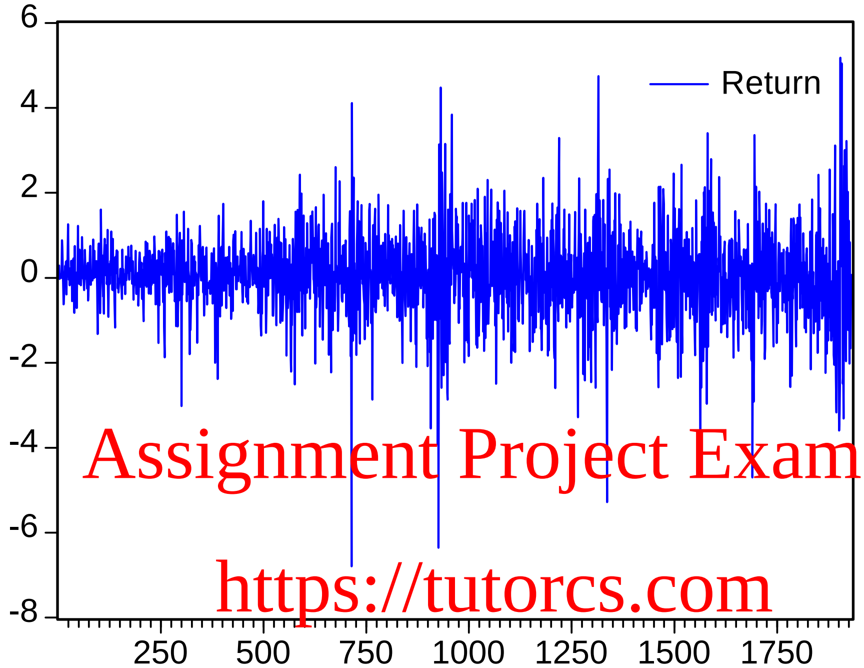
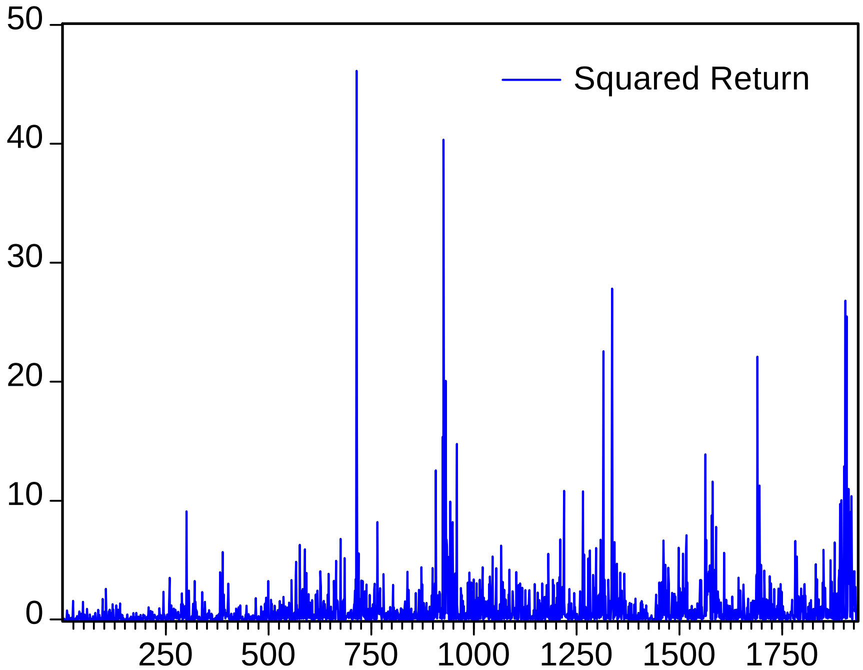


Figure 2: NYSE Composite Log Squared Returns



On the basis of the autocorrelation and partial autocorrelation function for returns (that is, for sr_t), it was decided to fit an MA(1) model for returns. The results are reported in Table 1 where it can be seen that the coefficient on the MA(1) term is statistically significant. An LM test for ARCH\GARCH effects is applied to the estimated residuals from the MA(1) model and the results are reported in Table 2.

Table 1: Estimation of MA(1) Model for NYSE Composite Stock Returns

Dependent Variable: SR

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.035311	0.024568	1.437292	0.1508
MA(1)	0.075177	0.02271	3.310311	0.0009

R-squared	0.005155	Mean dependent var	0.0353
Adjusted R-squared	0.004639	S.D. dependent var	1.006207
S.E. of regression	1.003871	Akaike info criterion	2.84664
Sum squared resid	1942.955	Schwarz criterion	2.852407
Log likelihood	-2745.007	F-statistic	9.989889
Durbin-Watson stat	2.006224	Prob(F-statistic)	0.001598

Table 2: LM Test for ARCH Effects in Residuals from MA(1) Model for Stock Returns

ARCH Test:

F-statistic	28.5059	Probability	0.0000
Obs*R-squared	181.4663	Probability	0.0000

Dependent Variable: RESID^2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.462217	0.068106	6.786703	0.0000
RESID^2(-1)	0.142459	0.022803	6.247313	0.0000
RESID^2(-2)	0.075486	0.023027	3.278114	0.0011
RESID^2(-3)	0.11154	0.022954	4.859341	0.0000
RESID^2(-4)	0.010104	0.023093	0.43754	0.6618
RESID^2(-5)	0.110557	0.022955	4.81633	0.0000
RESID^2(-6)	0.026927	0.023041	1.168689	0.2427
RESID^2(-7)	0.066393	0.022816	2.909882	0.0037

R-squared	0.094366	Mean dependent var	1.010337
Adjusted R-squared	0.091056	S.D. dependent var	2.507305
S.E. of regression	2.390429	Akaike info criterion	4.584974
Sum squared resid	10942.6	Schwarz criterion	4.608112
Log likelihood	-4400.45	F-statistic	28.5059
Durbin-Watson stat	2.003779	Prob(F-statistic)	0

In Table 2, the TR^2 statistic of 181.47 is highly significant with a p-value of 0.00. Thus, the null hypothesis of no ARCH\GARCH effects is emphatically rejected. Many of the lagged squared residuals are statistically significant indicating that volatility clustering is a feature of returns.

As a first try at modeling time-varying volatility in the returns on the NYSE index, we fitted an MA(1)-ARCH(5) model to sr_t. The results are shown in Table 3. Here the model for sr_t can be thought of as comprising a mean equation for sr_t, specified as an MA(1) process, and a variance equation for sr_t, specified as an ARCH(5) process. Both are estimated jointly by the method of maximum likelihood. All the coefficients in the ARCH specification are positive so that the conditional variance is always positive, and the coefficients on the lagged squared residuals sum to 0.731, a number less than one as required for a finite unconditional variance. The estimated coefficients on the lagged squared residuals are all statistically significant. This suggests that even higher-order ARCH models may be required. However, ARCH type models with long lags are not parsimonious and it is usually preferable to estimate a GARCH(1,1) specification for the variance equation.

Table 3: Estimation of MA(1)-ARCH(5) Model for Stock Returns

Dependent Variable: SR
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 1931
Included observations: 1930 after adjustments
Convergence achieved after 14 iterations
MA backcast: 1, Variance backcast: ON
GARCH = C(3) + C(4)*RESID(-1)^2 + C(5)*RESID(-2)^2 + C(6)*RESID(-3)^2 + C(7)*RESID(-4)^2 + C(8)*RESID(-5)^2

Mean Equation

	Coefficient	Std. Error	z-Statistic	Prob.
C	0.073216	0.020863	3.509306	0.0004
MA(1)	0.112154	0.024659	4.548241	0.0000

Variance Equation

	Coefficient	Std. Error	z-Statistic	Prob.
C	0.328317	0.023205	14.1483	0.0000
RESID(-1)^2	0.163148	0.022114	7.377523	0.0000
RESID(-2)^2	0.256414	0.02566	9.992863	0.0000
RESID(-3)^2	0.086334	0.024299	3.553029	0.0004
RESID(-4)^2	0.163758	0.028341	5.778146	0.0000
RESID(-5)^2	0.061118	0.026263	2.327107	0.0200

R-squared	0.002734	Mean dependent var	0.0353
Adjusted R-squared	-0.000898	S.D. dependent var	1.006207
S.E. of regression	1.006659	Akaike info criterion	2.661503
Sum squared resid	1947.682	Schwarz criterion	2.684571
Log likelihood	-2560.35	F-statistic	0.752859
Durbin-Watson stat	2.074987	Prob(F-statistic)	0.627095

Table 4 shows the results of estimating an MA(1)-GARCH(1,1) model for returns (sr_t). The coefficient on the MA(1) term is statistically significant. In the GARCH specification, all the parameter estimates are statistically significant and positive. The ARCH coefficient (α_1) and the GARCH coefficient (β) sum to a value very near unity (0.996), with β (0.886) substantially larger then α_1 (0.109) as is commonly found in financial data.

Table 4: Estimation of MA(1)-GARCH(1,1)

Dependent Variable: SR
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 1931
Included observations: 1930 after adjustments
Convergence achieved after 14 iterations
MA backcast: 1, Variance backcast: ON
GARCH = C(3) + C(4)*RESID(-1)^2 + C(5)*GARCH(-1)

	Mean Equation			
	Coefficient	Std. Error	z-Statistic	Prob.
C	0.075428	0.020434	3.691324	0.0002
MA(1)	0.103656	0.023799	4.355534	0.0000
	Variance Equation			
C	0.012609	0.00267	4.723211	0.0000
RESID(-1)^2	0.109949	0.009086	12.10068	0.0000
GARCH(-1)	0.885863	0.009682	91.49436	0.0000
R-squared	0.003096	Mean dependent var		0.0353
Adjusted R-squared	0.001025	S.D. dependent var		1.006207
S.E. of regression	1.005692	Akaike info criterion		2.613089
Sum squared resid	1946.976	Schwarz criterion		2.627506
Log likelihood	-2516.63	F-statistic		1.49467
Durbin-Watson stat	2.058277	Prob(F-statistic)		0.20121

Table 5 shows the results of performing an LM test for remaining heteroscedasticity (that is, ARCH\GARCH effects) in the standardized MA(1)-GARCH(1,1) residuals. The TR^2 statistic of 3.87 is not statistically significant at conventional significance levels (it has a p-value of 0.57). Furthermore, in the auxiliary regression, none of the coefficients on the lagged squared standardized residuals are significant indicating that there is no remaining heteroscedasticity to be modeled.

Table 5: LM Test for ARCH Effects in Standardized Residuals from MA(1)-GARCH(1,1) Model

ARCH Test:

F-statistic	0.772183	Probability	0.569736
Obs*R-squared	3.86521	Probability	0.568981

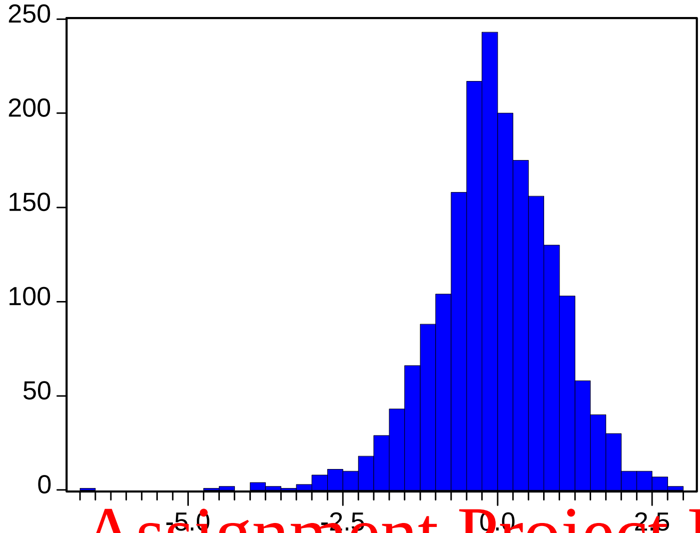
Test Equation:
Dependent Variable: STD_RESID^2
Method: Least Squares
Sample (adjusted): 7 1931
Included observations: 1925 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.001594	0.068388	14.64577	0.0000
STD_RESID^2(-1)	0.002692	0.022824	0.117942	0.9061
STD_RESID^2(-2)	0.037636	0.022823	1.649021	0.0993
STD_RESID^2(-3)	-0.01147	0.022836	-0.50225	0.6156
STD_RESID^2(-4)	-0.01053	0.022821	-0.46149	0.6445
STD_RESID^2(-5)	-0.01843	0.022823	-0.80757	0.4194

R-squared	0.002008	Mean dependent var	1.001504
Adjusted R-squared	-0.00059	S.D. dependent var	2.038313
S.E. of regression	2.038916	Akaike info criterion	4.265826
Sum squared resid	7977.629	Schwarz criterion	4.283164
Log likelihood	-4099.86	F-statistic	0.772183
Durbin-Watson stat	1.999667	Prob(F-statistic)	0.569736

Figure 3 presents a histogram of the standardized MA(1)-GARCH(1,1) residuals and associated summary statistics. The distribution appears symmetric and fat-tailed. The coefficient of kurtosis is quite high and the Jarque-Bera test clearly rejects normality. This suggests that the standard errors reported in Table 4 are not correct and that robust standard errors should be used instead. Table 6 reports estimates of the MA(1)-GARCH(1,1) model for returns with robust standard errors. Notice the coefficient estimates do not change just the standard errors. Nevertheless on the basis of the robust (Bollerslev-Wooldridge) standard errors there is no change in inference in this case.

Figure 3: Histogram of Standardized Residuals from MA(1)-GARCH(1,1) Model



Series: Standardized Residuals	
Sample 2 1931	
Observations 1930	
Mean	-0.050700
Median	-0.044595
Maximum	2.865349
Minimum	-6.741121
Std. Dev.	0.998719
Skewness	-0.547412
Kurtosis	5.047710
Jarque-Bera	433.5869
Probability	0.000000

Table 6: MA(1)-GARCH(1,1) Model for Returns with Robust Standard Errors

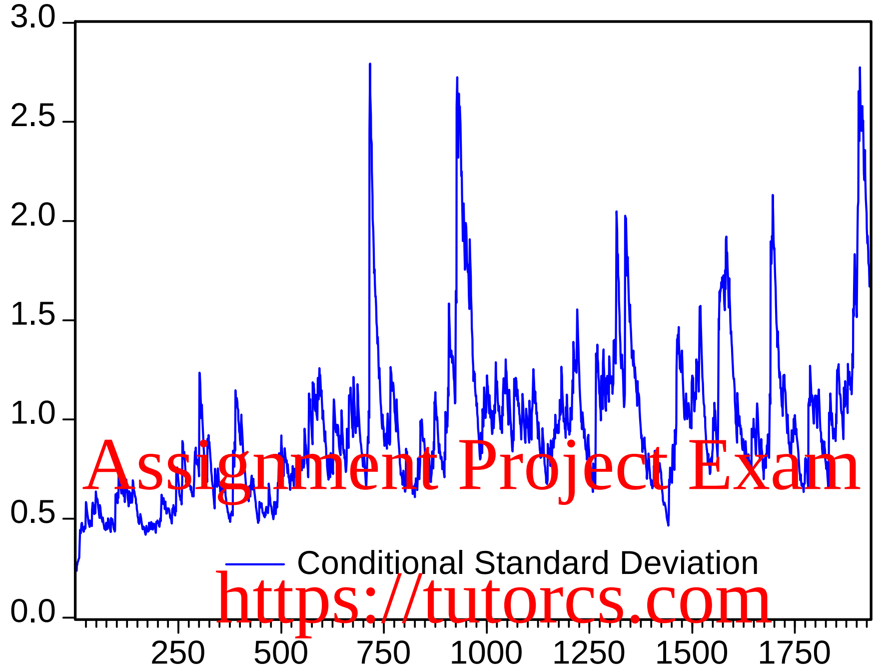
Dependent Variable: SR
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 1931
Included observations: 1930 after adjustments
Convergence achieved after 14 iterations
Bollerslev-Wooldrige robust standard errors & covariance
MA backcast: 1, Variance backcast: ON
GARCH = C(3) + C(4)*RESID(-1)^2 + C(5)*GARCH(-1)

	Mean Equation			
	Coefficient	Std. Error	z-Statistic	Prob.
C	0.075428	0.018001	4.190173	0.0000
MA(1)	0.103656	0.026223	3.952926	0.0001
	Variance Equation			
C	0.012609	0.004079	3.091308	0.0020
RESID(-1)^2	0.109949	0.023523	4.674078	0.0000
GARCH(-1)	0.885863	0.01969	44.99032	0.0000

R-squared	0.003096	Mean dependent var	0.0353
Adjusted R-squared	0.001025	S.D. dependent var	1.006207
S.E. of regression	1.005692	Akaike info criterion	2.613089
Sum squared resid	1946.976	Schwarz criterion	2.627506
Log likelihood	-2516.63	F-statistic	1.49467
Durbin-Watson stat	2.058277	Prob(F-statistic)	0.20121

Figure 4 shows the time series of the estimated conditional standard deviation implied by the estimated MA(1)-GARCH(1,1) model. Clearly volatility fluctuates a great deal and is highly volatile.

Figure 4: Estimated Conditional Standard Deviation, MA(1)-GARCH(1,1) Model



Now we estimate the MA(1)-GARCH(1,1) model for returns using observations 1-1531, leaving 400 observations for use in an out-of-sample forecasting exercise. Specifically, the MA(1)-GARCH(1,1) model for returns is estimated over observations 1-1531 and used to generate a dynamic forecast of the conditional standard deviation for the out-of-sample observations 1532-1931. The results are shown in Figure 5. The forecast begins just following a reduction in volatility (that is, in the conditional standard deviation). The forecast is for a gradual increase in volatility. The long-run volatility forecast is the unconditional standard deviation of sr_t .Letting b_1 denote the MA(1) coefficient, the formula for the unconditional variance is

$$\begin{aligned} \operatorname{var}(sr_t) &= (1+b_1^2)\operatorname{var}(\varepsilon_t) \\ &= \frac{(1+b_1^2)\alpha_0}{1-(\alpha_1+\beta)} = \frac{(1+(0.1037)^2)0.0126}{1-(0.1099+0.8859)} = 3.0323 \end{aligned}$$

Thus, $\sigma = 1.7413$ which is the value the forecast of the conditional standard deviation is approaching in Figure 5.

Figure 5: Estimated and Forecasted Condition Standard Deviation,
MA(1)-GARCH(1,1) Model