COMMON MARKET FOR EASTERN AND SOUTHERN AFRICA

MODELLING VOLATILITY IN FINANCIAL MARKETS WITHIN A MULTIVARIATE FRAMEWORK

Training for Experts in Central Banks of COMESA Member States at the COMESA Monetary Institute, Nairobi – Kenya, July 27 – August 07, 2015

Thomas Bwire (PhD)
Bank of Uganda
P.O BOX 7120
Kampala, Uganda
E-mail: tbwire@bou.or.ug

MOTIVATION AND OUTLINE

Understanding financial markets volatility and the ability to robustly forecast the same is an important aspect in investment decisions, portfolio creations, risk management and monetary policy making. This is amplified by the fact that heightened volatility shocks to one variable will most likely trigger volatility of other related variables, the so called spill-over effects. Financial markets volatility generates uncertainty about financial asset returns and increases the associated level of risk. Investors and portfolio managers have threshold levels of risk they can bear, and at policy level, financial market volatility and the associated volatility spillovers are a potential threat to financial stability and can dampen prospects for economic growth.

A good understanding of volatility and a robust forecast of volatility of returns on financial assets over the asset holding period is therefore a good starting point not only for assessing risk levels associated with investment in a financial asset but also importantly, monetary policy makers rely on the estimates of volatility as a barometer for the vulnerability of financial markets and the economy. Given this background, the course is intended to equip Central bank staffs from COMESA member countries with appropriate analytical skills and rigour to be able to adequately measure and forecast volatility in prices of financial market assets, so the decision makers, informed by rigorous robust volatility analysis, may undertake measures to mitigate the adverse effects of uncertainty in financial markets. The two week course will cover:

Day 1 & 2:	Salient features of financial markets volatility and introduction to modelling financial markets volatility.
Day 3 & 4:	Modelling the ARCH and GARCH effects: the ARCH and ARCH-M, GARCH models and Models of Asymmetry (the TGARCH, EGARCH, PARCH & CGARCH models) and departure from the Gaussianity.
Day 5 & 6:	Estimation of the ARCH and GARCH effects and related models.
Day 7:	Modelling Multivariate GARCH
Day 8 & 9:	Forecasting conditional volatility within both univariate and multivariate frameworks and forecast performance evaluation, and commencement of group assigned exercise during day 9.
Day 10:	Presentations on Group exercises.

Each of the theoretical rigour to be covered during the training shall be applied in a practical setting to demonstrate the theoretical econometric issues surrounding volatility. The practical applications of the ARCH and GARCH models as the 'tool of choice' for estimating and forecasting volatility in financial markets shall be demonstrated in a series of supervised hands-on application using the most user friendly E-Views econometric software.

Key reading List

Baba, Y., R.F. Engel, D.F. Kraft and K.F. Kroner (1990), "Multivariate Simultaneous generalized ARCH", *Mimeo*, University of California at San Diego, Department of Economics.

Bollerslev Tim (1986), "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics* **31**, 307-327.

Bollerslev, Tim and Jeffrey M. Wooldridge (1992), "Quasi- maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances", *Econometric Reviews*, 11, 143-172.

Dickey, D.A., and W. A. Fuller (1979), "Distribution of estimators for Autoregressive time series with a unit Root", *Journal of American Statistical Association* 74 (366): 427-31.

Enders Walter, *Applied Econometric Time Series*, John Wiley and Sons Ltd, 2010 (**Chapter 3**)

Engel, F. Robert (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation", *Econometrica* **50**(4), 987-1007

__________ (2004), "Risk and Volatility: Econometric Methods and Financial Practice", *The American Economic Review* **94**(3), 405-420.

Engel, R.F. and Kroner K.F. (1995), "Multivariate Simultaneous generalized ARCH", *Econometric Theory*, 11, 122-150

E-views 7 User's Guide 11, (**Chapter 22-25**)

Glosten L.R., R. Jagannathan and D.E. Runkle (1993), "On the relation between the expected value and the volatility of the nominal excess return on stocks", *Journal of Finance* **48**, 1779-1801

Harris, R and R, Sollis., *Applied Time Series Modelling and Forecasting*, John Wiley and Sons Ltd, 2003 (**Chapters 1 and 8).**

Nelson D.B (1991), "Conditional Heteroskedasticity in asset returns: A new approach", *Econometrica* **59**(2), 347-370.

Phillips, P.C.B., and Pierre Perron (1988), "Testing for a unit root in Time Series Regression", *Biometrika* 75: 335-46.

Poon, S-H and C.W.J Granger (2003), 'Forecasting Volatility in Financial Markets: A Review', *Journal of Economic Literature* XLI, 478-539.

Financial Econometrics Notes by Kevin Sheppard http://www.kevinsheppard.com/images/b/bb/Financial_Econometrics_2013 2014.pdf

Chapter 23-24 of E-views 7 User's Guide 11

DAY 1: Some Preliminaries

1. Volatility definition and measurement

Uncertainty or risk plays an important role in financial analysis and is usually measured with volatility. Volatility is the degree to which financial prices or returns fluctuate. It can also be looked at as the time varying risk associated with the returns on an asset. Obviously, the volatility of an asset is not an observable metric, so its measurement is necessary. In Finance, volatility is often calculated as sample standard deviation, $\hat{\delta}$, or square root of the variance, $\sqrt{\delta^2}$, in this case, $\hat{\delta}$ of returns associated with a financial asset.

Assuming a set of time series observations (N) of returns (Rt) on investing in a financial asset, the standard deviation (δ) is derived as the square root of the variance (δ^2) as in equation (1).

$$\delta = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(R_t - \bar{R}\right)^2} \qquad (1)$$

where $\bar{R}$ is the mean return. $\hat{\delta}$ is a distribution free parameter which depends on the dynamics of the underlying stochastic process and whether or not the parameters are time varying. Very often, when δ is used to measure volatility, the users usually assume a normal distribution (see departure from Gaussianity). However, the relationship between volatility and risk is tenuous. Risk is more often associated with small or negative returns. Volatility, on the other hand, makes no such distinction.

Equation (1) is what is referred to as *historical* or *realized volatility*, i.e. the standard deviation of financial returns on an asset computed over a window of a pre-specified number of past trading periods. However, despite the simplicity with which *historical volatility* is computed, the time frame over which it is computed matters. If the period is too short, it will be too noisy and if the period is too long, it will not be so relevant for today. Moreover, the statistical properties of a sample mean make it a very inaccurate estimate of the true mean especially in small samples because of the influence of extremes, and as noted above, not always that it draws from a normal distribution. Importantly, asset holders are interested in the volatility of returns over the holding period going forward and not over some historical period. This forward looking view of risk requires measures that are able to estimate *conditional volatility* associated with holding a particular asset or a forecast of volatility as well as a measure for today. Moreover, it is logically inconsistent to assume that the variance is constant for a period such as one year ending today and also that it is constant for the year ending on the previous day but with a different value – raising the possibility that volatility is dynamic. ARCH models and their many extensions that we will discuss during this training fill this gap.

Conditional volatility is expected volatility at some future time (say t + h) that is deliberately informed by new information set available at time t (ϖ_t). It means that tomorrow's volatility estimate depends on or is conditional on certain new information available today.

Implied volatility is the volatility that delivers a no-arbitrage option pricing, and is derived using Black-Scholes model. The value of a financial asset is priced on the basis of the assets own price, the exercise price, the time to maturity, the risk-free interest rate and the assets expected standard deviation. Risk-free interest rate is the return on a security whose pay-off is certain.

The approach however is criticized because the assumed geometric Brownian motion for the prices of the underlying asset may not hold in practice. Moreover, the resultant implied volatility of an asset return has been shown to be larger than that obtained by conditional volatility measures. Nonetheless, all these volatility measures are measured in terms of variance (δ^2).

2. Financial market volatility: the stylized facts.

There are several salient features about financial time series and financial market volatility. Financial time series are often available at a higher frequency than macroeconomic time series, often daily, weekly and monthly, and like many of the other macroeconomic time series, on a quarterly and annually basis. Many high frequency financial time series exhibit the property of 'long-memory' i.e. characterized by significant statistical correlation between observations that are distant apart.

Time series data on the returns from investing in a financial asset seem to meander in a fashion that contains periods of high volatility followed by periods of lower volatility (i.e. visually, there are clusters of extreme values in the returns followed by periods in which such extreme values are not present). This feature of a stochastic variable is called heteroskedastic as opposed to homoskedastic i.e. a property of constant variance, and is known as *conditional heteroskedasticity* of asset returns.

3. Empirical properties of financial assets returns

Financial markets literature recognizes several salient characteristics of financial asset returns:

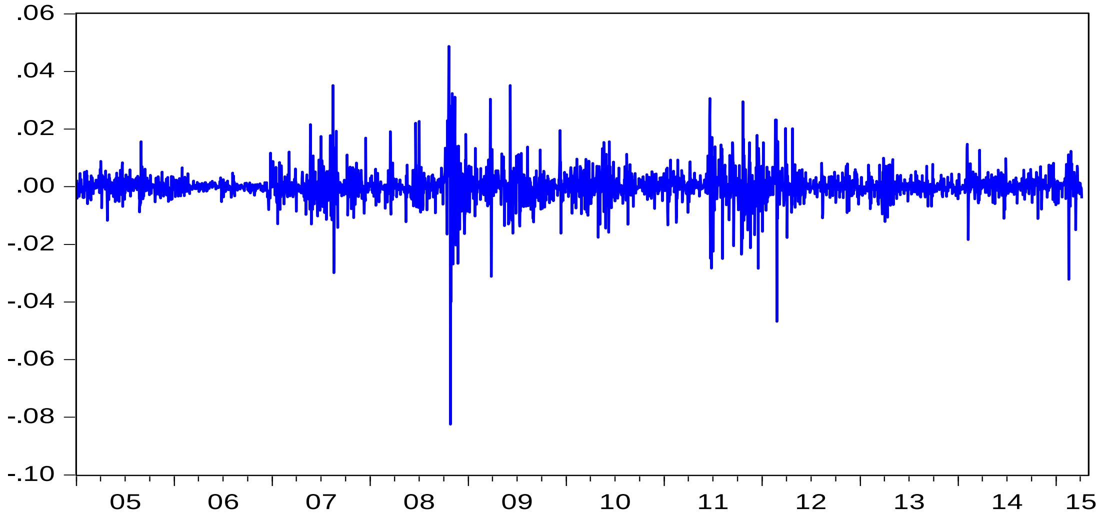
1. Returns on a financial asset evolve over-time in a continuous manner.

Figure 1: Daily Ushs/USD, Jan 2005 – April 2015.

As is shown by the raw data on daily Ushs/USD exchange rate in Figure 1, the rather smooth curve shows what has happened to the exchange rate in Uganda over the last 10 years. We see a stable but somewhat appreciated exchange rate before 2007 and before the financial crisis of 2008 and a steep sustained depreciation and intermittent appreciation thereafter. At the commencement of 2005, one USD was priced at about Ushs. 1,750 and it cost about Ushs 3,000 at the end of April 2015. This implies one dollar invested in the foreign exchange market at the commencement of 2005 had multiplied 1.7 times by the end of April 2015, while the shilling had depreciated by 71.3% over the same period, amounting to an average annual depreciation of 7.1%.

2. Periods of large movements in prices alternate with periods during which prices hardly change. This characteristic feature is commonly referred as *volatility clustering*, and is illustrated in Figure 2.

Figure 2: Intraday % changes in LEXR, Jan 2005 – April 2015.

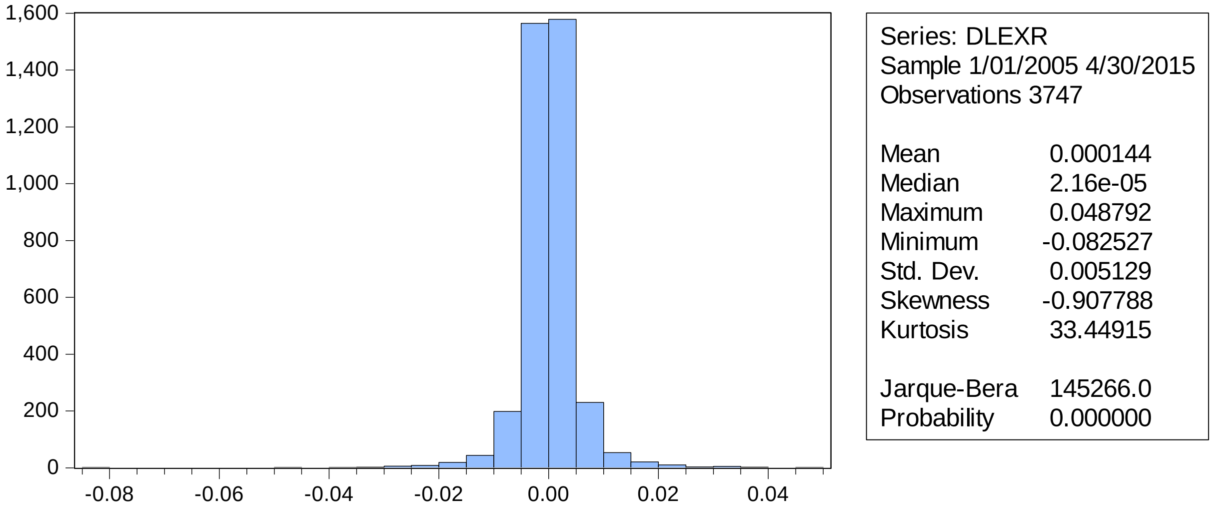


While the picture of the level data is appealing, it is the relative price change over the holding period that really matter. Thus, investors, portfolio and risk managers and monetary policy makers focus attention on the relative intraday percentage change or depreciation or returns as shown above (with exchange rate in natural logarithm). Clearly, the exchange rate return series is centred around zero, with periods of large volatility, followed by periods of relative tranquillity. We can note and relate to the world and domestic events particularly in 2007, 2008-9, 2011/12 and more recently 2014/15 when size of the depreciation and subsequent appreciation dwarfs all the other changes.

Volatile periods are hectic periods and intuitively reflect heightened uncertainty. The usual suspect for such uncertainty is uncertainty about the fundamentals in the economy. But from an econometric point of view, uncertainty about fundamentals only explains a moderate portion of the observed financial market volatility. Comprehensive measurement of volatility, including its forecast requires simple time series models beyond the fundamentals.

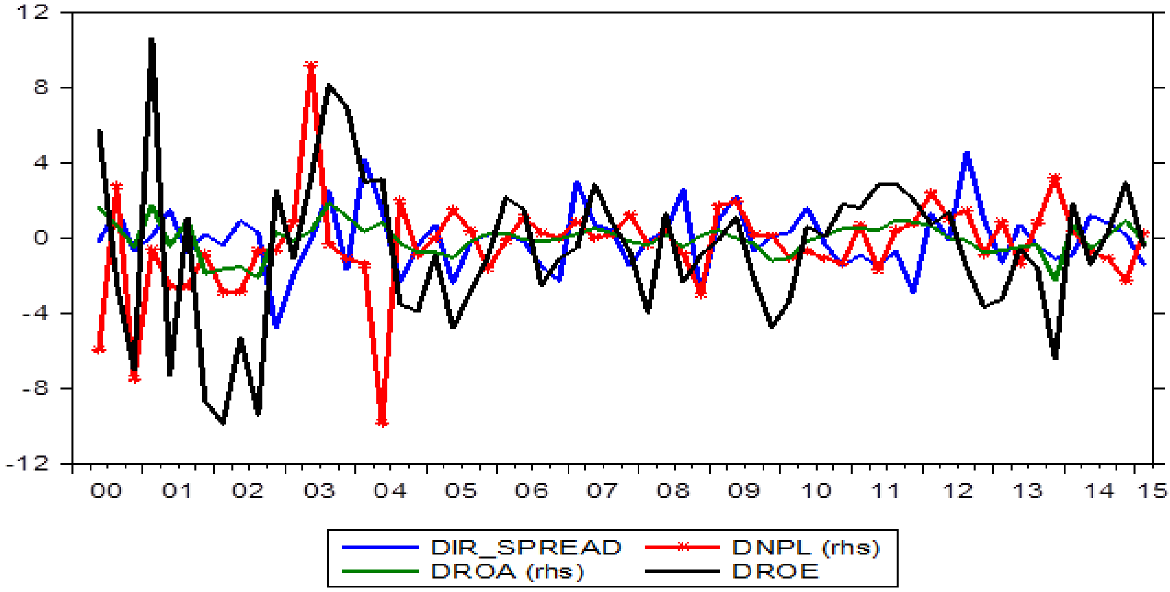
3. Related to the above is the existence of asymmetric movement of the volatility, i.e. a large (small) change in prices is more likely followed by a large (small) price change, i.e. persistence in price movements.
4. Another feature is the non-normal empirical distribution. Usually, the empirical density function has a higher peak around its mean, but fatter tails than that of the corresponding normal distribution or high excess kurtosis (fig 3). In addition, the distribution of returns is typically negatively skewed. This is because large negative movements in financial markets are not usually matched by equally large positive movements.

Figure 3: Frequency distribution of log-Ushs/USD, Jan 2005 – April 2015



5. As shown in fig. 4, quarterly movements in NPL, ROA, ROE and interest rate spread series exhibit co-movements of volatilities across assets and has been found to be true even across financial markets.

Figure 4: Volatility co-movements in multivariate financial time series



The four series do not drift far apart. Moreover, large shocks to ROE appear to be timed similarly with those to ROA. It is this sort of pattern where volatility shocks to one variable triggers volatilities in other related variables that constitute the spill-over effects, the behaviour we later model.

6. As shown in figure 2, the mean of log-return series is close to zero, which is intuitive. To show this, assume the log-return on an asset is defined to take the form:

$$r_t = \log(p_t) - \log(p_{t-1}),$$

where p_t is the price level of the asset at time t, then it follows that:

a) $p_t \sim I(1)$, i.e. is non stationary or is a random walk process around trend, something synonymous with the plot in fig. 1. A unit root in the data generating process (dgp) for p_t or non-stationarity in p_t can be eliminated by transforming the series into its first differences, and is said to be integrated of order one or $I(1)$ if it becomes stationary after first differencing.

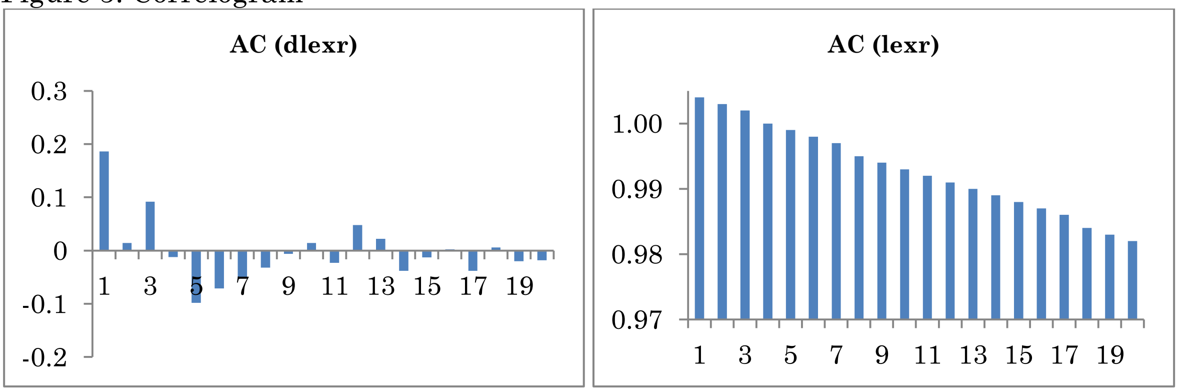
Therefore, iff $p_t \sim I(1)$, then

b) $r_t = \Delta \log p_t \sim I(0)$

That is. is stationary or mean reverting, a behaviour akin to the plot in fig.2.

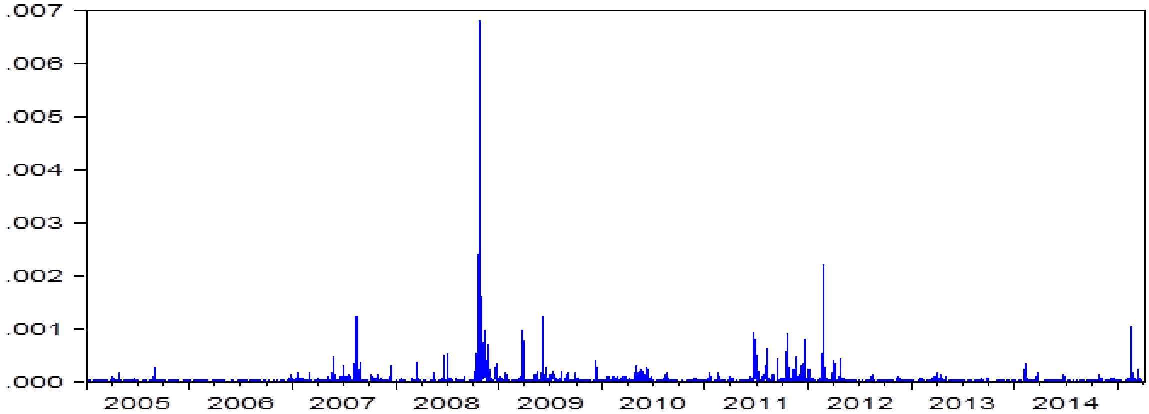
In time series, there are important distinctions between a stationary or $I(0)$ series and a non-stationary or $I(1)$ process. $I(0)$ Series fluctuate around its mean (or is mean reverting) with a finite variance that does not depend on time while its $I(1)$ variant wanders widely. $I(0)$ series has a limited memory of its past behaviour, while $I(1)$ series has infinitely long memory i.e. innovations in the system permanently affects the process. And as shown in figure 5, the autocorrelations of a $I(0)$ series decline rapidly to zero as the lag increases, while they decay to zero very slowly in the case of $I(1)$ series.

Figure 5: Correlogram



7. The other common feature of financial asset returns is that we do not observe autocorrelation in levels (fig. 1), but we do with squares of log-returns. As shown in fig. 6, the autocorrelations of variance, and particularly those of mean absolute deviation stay positive and significantly above zero for all lags.

Figure 6: squares of log-returns



DAY 2: Understanding Heteroskedasticity and Autocorrelation

To underscore these twin concepts in macroeconomic and financial time series data, consider a standard but the simplest stationary univariate model for a random stochastic variable, y_t, observed over the sequence of time $t = 1... T$:

$$y_t = \underbrace{\phi y_{t-1}}_{explained} + \underbrace{\varepsilon_t}_{unexplained} \qquad (2)$$

This simple equation implies that the variable, y, at time t is generated by its own past behaviour, i.e. its own lags – the explained part and a disturbance (or residual) term ε_t -the unknown part. In econometrics, ε_t is assumed to be governed by several standard classical or Gauss-Markov assumptions:

i) $E(\varepsilon_t) = 0 \qquad \forall t$

That is its expected value (or mean) is zero.

ii) $V(\varepsilon_t) = E(\varepsilon_t^2) = \delta^2 \qquad \forall t$

That is it has a constant or time invariant variance, so we say residuals are homoskedastic. Violation of this assumption results in residuals that are heteroskedastic, i.e. varying variance. It is this behaviour that we model in detail in the coming lectures.

Thus, so long as $E(\varepsilon_t) = 0$ and $E(\varepsilon_t^2) = \delta^2$ hold, we say that $\varepsilon_t \sim i.i.d.(0, \delta^2)$, i.e. ε_t are identically and independently distributed as normal with zero mean and constant variance. This is what econometricians call *white noise process*.

iii) $C(\varepsilon_t \varepsilon_{t-s}) = E(\varepsilon_t \varepsilon_{t-s}) = 0; \qquad \forall t \neq s,$

That is the residuals in period t are uncorrelated with those in period s, where period s is own past. This behaviour is particularly important when we come to producing forecasts, since it implies that the correlation structure between y_{t-s} and y_t is the same as that between y_t and y_{t+s}. That is, we can extrapolate backwards-looking relationships into forward-looking relationships to yield forecasts of future values of the series. Note however that violation of this assumption results into serial or autocorrelation – a situation where there is interdependence in the behaviour of residuals over time.

As a fore-mentioned, $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \delta^2$ $E(\varepsilon_t \varepsilon_{t-s}) = 0$; $\forall t \neq s$ are the classical Gauss-Markov simplifying assumptions that ensures a pure white noise process. However, in contrast and in practice, residuals of financial and in as many instances macroeconomic time series data exhibit a behaviour which is such that:

a) $E(\varepsilon_{t-i}^2) \neq \delta^2$, i.e. residuals are heteroskedastic, and
b) $E(\varepsilon_t, \varepsilon_{t-1}) \neq 0$, i.e. residuals are autocorrelated (a problem of serial correlation)

Does this departure from the standard white noise process warrant a worry when modelling financial and macroeconomic time series data?

The answer is yes, and this is why. While developments in econometrics theory suggests that even with heteroskedastic residuals, the popular ordinary least squares (OLS) estimator, ϕ is still consistent (i.e. we get the correct coefficient values especially as the sample becomes reasonably large) and is reasonably efficient (in terms of the variance of $\hat{\phi}$), the real problem is that the standard errors (of $\hat{\phi}$) are not correct. As a result, the *t*- and *F*-tests are no longer valid and inference based on them could be awfully misleading. So we shall, during the great part of the training, be modelling heteroskedasticity using a particular class of models, the Autoregressive Conditional Heteroskedasticity (ARCH) models and related extensions thereof, to obtain more efficient estimates.

Ensuring that the other crucial assumption of time independence of residuals could be problematic but is entirely an art craft of the modeller as it is purely a modelling aspect. It requires that the modeller fits an appropriate autoregressive moving average (ARMA) process. Usually, we want a model with as few lags as possible to get a parsimonious model, but at the same time we want enough lags to remove autocorrelation of the residuals. Below, I describe briefly how this can be achieved.

4.1 Testing for and dealing with serial correlation

Assume for ease of exposition, an autoregressive process of order one, *AR(1)*, defined as:

$$y_t = \phi y_{t-1} + \varepsilon_t; \qquad \varepsilon_t \sim N(0, \delta^2) \text{ and } |\phi| < 1 \text{ - for mean reversion to hold.} \qquad (3)$$

To see the dependence structure of residuals in this random walk process, we have to express the RHS in terms of ε_t only by backward recursive substitution. The equation eventually takes the form:

$$y_t = \varepsilon_t + \phi\varepsilon_{t-1} + \varphi^2\varepsilon_{t-2} + \ldots + \phi^{j-1}\varepsilon_{t-(j-1)}$$ [1]

Serial correlation arises because residuals for periods not too far apart, such as ε_t, ε_{t-1} and so on are related, but this dependence decay gradually to zero as $t \to \infty$ either exponentially if $\phi > 0$, or like a damped sine wave if $\phi < 0$, providing $|\phi| < 1$ holds.

A. Testing for serial correlation (practical issue)

There are two major tests, both of which are inbuilt in E-Views. The Correlogram and Ljung – Box Q-Statistic, and the Breusch – Godfrey Lagrange multiplier test statistic. In both tests, the null hypothesis is that of '*no serial correlation*'.

[1] We will explore this derivation in detail when deriving the conditional mean and variance under the ARCH model.

B. **Dealing with serial correlation**

To rid the stochastic process in (3) of serial correlation, the process is sequentially estimated, up to an order sufficient to remove any serial correlation. The estimation specification is of the form:

$$y_t = c + \beta_i \sum_{i=1}^{\rho} y_{t-i} + \varepsilon_t$$

Noting that ρ here is the appropriate autoregressive order sufficient to remove serial correlation, and y is log returns. In e-views, such a process is simply estimated using OLS by the following user supplied command:

y c ar(1) ... ar(p)

The stationarity condition for general $AR(p)$ processes is for the inverted roots of the lag polynomial to lie inside the unit circle. This result appears at the bottom of the regression output.

In the second step of building an ARMA model, we now introduce the simplest class of the moving average (MA) process, the MA(1) process, which is given by:

$$y_t = \varepsilon_t + \theta\varepsilon_{t-1} \qquad (4)$$

Where $\varepsilon_t \sim N(0,\delta^2)$ and $|\theta| < 1$. For this process, we find that

$$\begin{aligned} E(y_t) &= E(\varepsilon_t + \theta\varepsilon_{t-1}) \\ &= E(\varepsilon_t) - \theta E(\varepsilon_{t-1}) \\ &= 0 \end{aligned}$$

$$\begin{aligned} Var(y_t) &= E(y_t^2) \\ &= E\{(\varepsilon_t + \theta\varepsilon_{t-1})^2\} \\ &= E(\varepsilon_t^2 + \theta^2\varepsilon_{t-1}^2 + 2\theta\varepsilon_t\varepsilon_{t-1}) \\ &= \delta^2(1+\theta^2) \end{aligned}$$

We can also show the $C(y_t, y_{t-s})$.

$$C(y_t, y_{t-s}) = E(y_t y_{t-s})$$

$$E(y_t y_{t-s}) = E\left\{\left(\underbrace{\varepsilon_t + \theta\varepsilon_{t-1}}_{y_t}\right)\left(\underbrace{\varepsilon_{t-s} + \theta\varepsilon_{t-s-1}}_{y_{t-s}}\right)\right\}$$

$$= E\{(\varepsilon_t\varepsilon_{t-s} + \theta^2\varepsilon_{t-1}\varepsilon_{t-s-1} + \theta\varepsilon_t\varepsilon_{t-s-1} + \theta\varepsilon_{t-1}\varepsilon_{t-s})\}$$

$$= E(\varepsilon_t\varepsilon_{t-s}) + \theta^2 E(\varepsilon_{t-1}\varepsilon_{t-s-1}) + \theta E(\varepsilon_t\varepsilon_{t-s-1}) + \theta E(\varepsilon_{t-1}\varepsilon_{t-s})$$

At $s = 1$,

$$C(y_t y_{t-1}) = E(\varepsilon_t\varepsilon_{t-1}) + \theta^2 E(\varepsilon_{t-1}\varepsilon_{t-2}) + \theta E(\varepsilon_t\varepsilon_{t-2}) + \theta E(\varepsilon_{t-1}\varepsilon_{t-1})$$

$$C(y_t y_{t-1}) = \theta\delta^2$$

It should be obvious that for $\forall\ s>1,\ C(y_t y_{t-s})=0$.

Therefore, $$C(y_t, y_{t-s}) = \begin{cases} \delta^2\theta; s=1 \\ 0; s>1 \end{cases}$$

Thus, the mean, variance and auto covariance of the MA(1) process do not depend on *t*, establishing that this process is stationary, for any value of the parameter θ. The autocorrelation function for the MA(1) process is then

$$\rho_s = \begin{cases} \dfrac{\theta}{1+\theta^2}; s=1 \\ 0; s>1 \end{cases}$$

Thus the MA(1) process is 1-*dependent,* and in practice, lower order MA models have been found to be more useful in econometrics than higher order MA models.

It follows thus that the specification of the *ARMA (p,q)* process is given by:

$$y_t = \underbrace{\phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \ldots + \phi^p\varepsilon_{t-p} + \varepsilon_t}_{AR(p)componet} + \underbrace{\theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \ldots + \theta_q\varepsilon_{t-q}}_{MA(q)component};\ \varepsilon_t \sim N(0,\delta^2)$$

The estimation specification is of the form:

y c ar(1) ... ar(p) ma(1) ... ma(q)

As with AR diagnostics, the estimated ARMA process is (covariance) stationary if all the inverted AR and MA roots lie inside the unit circle. In addition, if the ARMA model is correctly specified, the ACF and PACF of the residuals from the model should be nearly white noise.

DAY 3: The Genesis of ARCH Models

As mentioned earlier, uncertainty or risk plays an important role in financial analysis and is usually measured with volatility. The downside is that volatility of an asset is not observable, so it's modelling is necessary. Based on the constructed model, the volatility can be both measured and predicted. Asset holders are interested in the volatility of returns over the holding period, not over some historical period. This forward looking view of risk means that it is important to be able to estimate and forecast the risk associated with holding a particular asset. Although in literature, numerous volatility models have been suggested to capture the characteristics of return for an asset, here we explore the Auto Regressive Conditional Heteroskedasticity (ARCH) and their many extensions, in particular, Generalized Auto Regressive Conditional Heteroskedasticity (GARCH) type models.

Modelling Conditional Volatility

1. The ARCH model

We will extend the AR (1) model previously briefly introduced, in the initial exposition of the ARCH model.

Consider the conventional *AR(1)* model given by:

$$y_t = \phi y_{t-1} + \varepsilon_t; \qquad (5)$$

Where t = 1, 2, ..., T, $\varepsilon_t \sim i.i.d.(0, \delta^2)$ and assume that $|\phi| < 1$. Thus, y_t is a stationary *AR(1)* process. If y_t is generated by the process in (5), then it can be shown that the mean and variance (covariance) of y_t are *constant* or are *unconditional*.

To show this, the *AR(1)* process in (5) involves recursion since y_{t-1} appears on the RHS. The idea is to express the RHS in terms of ε_t only. Thus, given

$$y_t = \phi y_{t-1} + \varepsilon_t$$
$$y_t = \phi(\phi y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t$$
$$y_t = \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2 y_{t-2}$$
$$y_t = \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2(\phi y_{t-3} + \varepsilon_{t-2})$$
$$y_t = \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3 y_{t-3}$$

Repeating the back substitution $j-1$ times, we obtain

$$y_t = \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \ldots + \phi^{j-1}\varepsilon_{t-(j-1)} + \phi^j y_{t-j}$$
$$= \sum_{i=0}^{j-1}\phi^i\varepsilon_{t-i} + \phi^j y_{t-j}$$

Which still depends on y_{t-j}, however, given the restriction that $|\phi| < 1$, then as $j \to \infty$, $\phi^j y_{t-j} \to 0$, such that

$$y_t = \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i} .$$

Based on this representation for the AR(1) process, we find

$$E(y_t) = E\left(\sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}\right)$$
$$= \sum_{i=0}^{\infty} \phi^i E(\varepsilon_{t-i})$$
$$= 0.$$

and

$$V(y_t) = E(y_t^2)$$
$$= E\left\{\left(\sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}\right)^2\right\}$$
$$= E\left\{(\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3\varepsilon_{t-3} + \ldots)^2\right\}$$
$$= E\left\{(\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3\varepsilon_{t-3} + \ldots)(\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3\varepsilon_{t-3} + \ldots)\right\}$$
$$= E\,(\varepsilon_t^2 + \phi^2\varepsilon_{t-1}^2 + \phi^4\varepsilon_{t-2}^2 + \phi^6\varepsilon_{t-3}^2 + \ldots + \{\text{cross terms like } \phi^{i+j}\varepsilon_{t-j} \text{ where } i \neq j\ \})$$
$$= \delta^2 + \phi^2\delta^2 + \phi^4\delta^2 + \phi^6\delta^2 + \ldots + \{0\}$$
$$= \delta^2\,(1 + \phi^2 + \phi^4 + \phi^6 + \ldots)$$
$$= \delta^2 \sum_{i=0}^{\infty} \phi^{2i}$$

Since $|\phi| < 1$, this infinite sum is convergent, so using the well known geometric series result that

$$\sum_{i=0}^{\infty} r^i = \frac{1}{1-r} \quad \text{when } |r| < 1$$

we have

$$V(y_t) = \gamma_0 = \delta^2 \sum_{i=0}^{\infty} \phi^{2i} = \frac{\delta^2}{1-\phi^2} .$$

you may also want to show that

$$C(y_t, y_{t-s}) = \frac{\delta^2 \phi^s}{1-\phi^2}$$

The autocorrelation function for the *AR(1)* process is then

$$\rho_s = \phi^s$$

So the *AR(1)* process has non-zero autocorrelations at all lags, but these decay to 0 as $s \to \infty$ because $|\phi| < 1$ (either exponentially if $\phi > 0$, or like a damped sine wave if $\phi < 0$).

We have thus shown, in particular, that the mean and variance of the *AR(1)* process do not depend on *t provided* $|\phi| < 1$. These are the *unconditional mean* and *the unconditional variance* of y_t .

In reality however, the mean of y_t is conditional on information set available at time $t-i$, hence *conditional mean* of y_t. Denoting this historical information set available at time $t-i$ as ϖ_{t-i}, $i = 1,2,\ldots$, and assuming the same d.g.p for y_t as in (5), the conditional mean of y_t is given by

$$E(y_t \mid \varpi_{t-i}) = \phi y_{t-1}$$

and the *conditional variance* of y_t is given by

$$\left.\begin{aligned} V(y_t \mid \varpi_{t-i}) &= E(\varepsilon_t^2 \mid \varpi_{t-i}) \\ &= \delta_t^2 \end{aligned}\right\}$$

Consistent with figure 2, it is straight forward to see from this expression that the conditional expectation of the variance is time varying – a feature seasoned econometricians call heteroskedasticity. An important question though is how long *t* should be in calculating conditional volatility. As mentioned earlier, if the period is too short, say one past year, it will be too noisy and if it is too long, say past 30 years or more, such a long memory is just not so relevant for today. The solution, according to Engel (2004, p.406, in AER) is *autoregressive conditional heteroskedasticity* (ARCH) process, a process which describes the forecast variance in terms of current observations. That is, the ARCH model, *instead of using short or long sample standard deviations, takes weighted averages of past squared forecast errors, a type of weighted variance which gives more influence to recent information and less to the distant past, making the ARCH model a simple generalization of the sample variance.*

2. Conditional and unconditional mean and variance of ARCH(1) process.

Engel (1982), defines ε_t for the ARCH model as

$$\varepsilon_t = v_t\left(\alpha_0 + \alpha_1\varepsilon_{t-1}^2\right)^{1/2} \tag{6}$$

Where $v_t \sim i.i.d(0,1)$, i.e. white noise process such that $\delta_v^2 = 1$, v_t and ε_{t-1} are independent of each other, i.e. $C(v_t, \varepsilon_{t-1}) = 0$, and α_0 and α_1 are constants such that $\alpha_0 > 0$, $0 < \alpha_1 < 1$.

We can show that ε_t has a mean of zero, and ε_t and v_t are uncorrelated, since v_t is white noise and is independent of ε_{t-1}.

The *unconditional mean*, $E(\varepsilon_t)$ is of the form:

$$\left.\begin{aligned} E(\varepsilon_t) &= \underbrace{E(v_t)}_{0} E\left(\alpha_0 + \alpha_1\varepsilon_{t-1}^2\right)^{\frac{1}{2}} \\ &= 0 \end{aligned}\right\}$$

and it follows from $E(v_t v_{t-i}) = 0$ that

$$E(\varepsilon_t \varepsilon_{t-i}) = 0;\ \forall i \neq 0.$$

The *unconditional variance of* ε_t, $V(\varepsilon_t)$ is of the form:

$$\left.\begin{aligned} V(\varepsilon_t) &= E\left[\left\{\varepsilon_t - \underbrace{E(\varepsilon_t)}_{0}\right\}^2\right] = E(\varepsilon_t^2) \\ &= \underbrace{E(v_t^2)}_{1} E(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2) \\ &= \alpha_0 + \alpha_1 \underbrace{E(\varepsilon_{t-1}^2)}_{1} \\ &= \frac{\alpha_0}{(1-\alpha_1)} \end{aligned}\right\}$$

Noting that the result that $\alpha_0 + \alpha_1 = \dfrac{\alpha_0}{(1-\alpha_1)}$ given $\alpha_0 > 0$, $0 < \alpha_1 < 1$ benefits from the property of geometric series. The unconditional mean and variance are unaffected by the presence of the error process defined by (6).

But we know according to the asset pricing models that the risk premium of holding a financial asset depends on the expected return on the asset and the variance of that return. It is therefore common knowledge that the relevant measure is the risk over the holding period, not the unconditional risk. Thus, the assessment of risk should be determined using the *conditional* distribution of asset returns.

Like the *unconditional mean*, the *conditional mean* of ε_t is zero

$$\left.\begin{aligned} E(\varepsilon_t \mid \varpi_{t-i}) &= \underbrace{E(v_t \mid \varpi_{t-i})}_{0}(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2)^{\frac{1}{2}} \\ &= 0 \end{aligned}\right\}$$

However, since $\delta_v^2 = 1$, the influence of the error process defined by (8) falls entirely on the *conditional variance* of ε_t as shown by (8.4). Thus, the *conditional variance* becomes:

$$\left.\begin{aligned} E(\varepsilon_t^2 \mid \varpi_{t-i}) = h_t &= \underbrace{E(v_t^2 \mid \varpi_{t-i})}_{1}(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2) \\ &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \end{aligned}\right\} \quad (7)$$

Henceforth, h_t is used to refer to the conditional variance of ε_t and is dependent on the realized value of the past squared forecast errors, ε_{t-1}^2. Thus, if the realized value of ε_{t-1}^2 is large, h in period *t* will be large as well. Thus, h_t follows a first-order autoregressive denoted ARCH (1). In an ARCH model, the unconditional and conditional expectations of ε_t are zero. On the contrary, h_t is an autoregressive process resulting in conditional heteroskedastic errors i.e. depends on the squared error term from the last period (ε_{t-1}^2). The coefficients α_0 and α_1 are restricted to be strictly both positive to ensure the conditional variance satisfies certain regularity conditions, including the non-negativity constraint, and the restriction on α_1, i.e. $0 < \alpha_1 < 1$ ensures stability of the process. It therefore follows that since α_1 and ε_{t-1}^2 cannot be negative the minimum value for the h_t is α_0. Thus, the volatility of $\{y_t\}$ is increasing in α_0 and α_1 which implies any unusually large shock in $|v_t|$ will be associated with a persistently large variance in the $\{\varepsilon_t\}$ sequence, i.e. the larger α_1 is, the larger the persistence.

It is also implied that a stochastic variable y_t, generated from a linear model with an ARCH error term, is itself an ARCH process. To show this, substitute (6) into (5) to yield:

$$E(y_t \mid \varpi_{t-i}) = \phi y_{t-1}, \text{ and } h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$$

Thus, h_t of y_t is also a function of the squared error term from the last period (ε_{t-1}^2). Therefore, in terms of *specification*, *ARCH* directly affects the error terms ε_t.

The *ARCH(1)* process in (7) has been extended in several interesting ways, including higher order *ARCH* processes – the Engel's (1982) original *ARCH (q)* model, to other univariate time series models, bivariate as well as multivariate regression models and to systems of equations.

The *AR(p)-ARCH(q)* model form of (5) is given as:

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t \text{ ;} \qquad (8)$$

Where $\varepsilon_t = v_t \left(\alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \right)^{1/2}$; $v_t \sim i.i.d.(0,1)$, and $|\phi| < 1$

The *ARCH(q)* multiple regression model is given as:

$$y_t = \beta_0 + \sum_{i=1}^{k} \beta_i x_{it} + \varepsilon_t \qquad (9)$$

where $\varepsilon_t = v_t \left(\alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \right)^{1/2}$; $v_t \sim i.i.d.(0,1)$, and $|\beta| < 1$

Where x_{it} = lagged exogenous explanatory variables.

Equations (8) and (9) can be extended to include dummy variables in the model for the conditional mean if there is a reason to capture particular features of the market e.g. day-of-the-week' effects.

For brevity, assuming daily data, and if we let $i = 1,\ldots,5$ such that 1 (Monday), ..., and 5 (Friday), then the model adjusts to

$$y_t = \beta_0 + \sum_{i=1}^{k} \beta_i x_{it} + \sum_{l=1}^{5} \xi_l D_{lt} + \varepsilon_t$$

Where $\varepsilon_t = v_t\left(\alpha_0 + \sum_{i=1}^{q}\alpha_i\varepsilon_{t-i}^2 + \sum_{l=1}^{5}\xi_l D_{lt}\right)^{1/2}$; $v_t \sim i.i.d.(0,1)$, $|\beta| < 1$ and $D_{lt} = \begin{cases}1\\0\end{cases}$; if daily feature is observed, zero otherwise, $\alpha_0 > 0$ and $\alpha_1, \alpha_2, ..., \alpha_q \geq 0$. Generalizing the *ARCH* models to systems of equations (the multivariate *ARCH* model) is a natural extension of the original specification.

3. Kurtosis of an ARCH model

Here as already mentioned, it is a good exercise to demonstrate that ARCH models generate excess kurtosis, i.e. the unconditional distribution for ε_t has fatter tails than the normal distribution, even if $v_t \sim i.i.d(0,1)$. The proof of this is quite straight forward.

Under normality, the unconditional fourth moment is given by:

$$k = \frac{E(\varepsilon_t^4)}{E(\varepsilon_t^2)^2}$$

The conditional fourth moment is given by

$$k = \frac{E\left[E_{t-1}(e_t^4\delta_t^4)\right]}{E\left[E_{t-1}(e_t^2\delta_t^2)\right]^2}$$

Where $E(\varepsilon_t^4) = E\left[E_{t-1}(\delta_t^4 e_t^4)\right]$, and $E(\varepsilon_t^2)^2 = E\left[E_{t-1}(\delta_t^2 e_t^2)\right]^2$

So $$k = \frac{E\left[E_{t-1}(e_t^4\delta_t^4)\right]}{E\left[(E_{t-1}e_t^2)^2(\delta_t^2)^2\right]}$$

since $E_{t-1}e_t^2 = 1$ and $E_t(e_t^4) = 3$,

$$k = \frac{3\delta_t^4}{1^2(\delta_t^2)^2}$$ so that

$$k = 3\frac{\delta_t^4}{(\delta_t^2)^2} \geq 3$$

4. The ARCH-M Model

The ARCH in mean (ARCH-M) is extension to the basic ARCH framework to allow the mean of $\{y_t\}$ to depend on its own h_t (**Engle *et al.*, 1987**). The model, relevant in financial applications is used to relate the expected return on an asset to the expected asset risk. A risk-averse agent will require compensation or risk premium for holding a long-term risky asset. Since an asset's *riskiness* is measured by h_t, it follows that the greater the h_t of returns, the greater the risk premium necessary to induce the agent to hold the long-term risky asset, i.e. the risk premium is an increasing function of h_t of returns. Thus, the ARCH-M is of the form:

$$y_t = \underbrace{\beta + \lambda h_t}_{\mu_t} + \varepsilon_t \quad ; \lambda > 0$$

Where: y_t = excess return for holding a long-term asset relative to the risk free one-period Treasury bond

μ_t = risk premium necessary to induce the risk-averse agent to hold the long- term risky asset rather than the risk free one-period Treasury bond.

λ = is a measure of the risk-return trade-off

h_t = the *ARCH(q)* process, defined as: $h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2$, and

ε_t = shock to the excess return on the long-term risky asset.

Noting that holding a risky asset makes sense only if the $E_{t-1} y_t = \mu$. Moreover, providing h_t is constant (i.e., if $\alpha_1 = \alpha_2 = ... = \alpha_q = 0$) the ARCH-M Model degenerates into the traditional case of a constant risk premium.

Other variants of this ARCH-M model are:

$$y_t = \underbrace{\beta + \lambda \sqrt{h_t}}_{\mu_t} + \varepsilon_t \quad ; \lambda > 0 :$$

Where conditional standard deviation is used in place of conditional variance, and

$$y_t = \underbrace{\beta + \lambda \log(h_t)}_{\mu_t} + \varepsilon_t \quad ; \lambda > 0$$

Where the log of conditional variance is used in place of conditional variance

Due to the large persistence in volatility, ARCH models typically require 5-8 lags of ε_t^2 to adequately model h_t or fit the data. In addition, as we have argued, to avoid problems associated with a negative conditional variance it is necessary to impose restrictions on the parameters of the model. Consequently in practice the estimation of ARCH models is not always straight forward. Given this, **Bollerslev (1986)** extends the ARCH model to be an *ARMA* process, i.e. allows for a more lag structure: the generalized *ARCH* (*GARCH*) model.

DAY 4: The GARCH model, GARCH estimation, Models of Asymmetry (the TGARCH, EGARCH, PARCH & CGARCH models) and departure from the Gaussianity.

1. The GARCH Model

Under the GARCH model, h_t is modelled as a function of lagged values of ε_t^2 and lagged values h_t. In general, assuming a d.g.p for the sequence $\{y_t\}$ in (5) and following Engel's (1982) generalized definition of ε_t for the *ARCH* model as in (6), ε_t for the *GARCH* (*p, q*) model is:

$$\varepsilon_t = v_t\left(\alpha_0 + \sum_{i=1}^{q}\alpha_i\varepsilon_{t-i}^2 + \sum_{j=1}^{p}\beta_j h_{t-j}\right)^{0.5} \qquad (10)$$

Where $v_t \sim i.i.d.(0,1)$; $q > 0$; $p \geq 0$; $\alpha_0 > 0$; $\alpha_i \geq 0$, $i = 1,2,\ldots,q$ and $\beta_j \geq 0$, $j = 1,2,\ldots,p$.

The sufficient condition for covariance stationarity of the process $\{\varepsilon_t\}$ is for $\sum_{i=1}^{q}\alpha_i + \sum_{j=1}^{p}\beta_j < 1$.

It is straightforward to show from (10)

That the *unconditional variance* of ε_t becomes:

$$V\left(\varepsilon_t^2\right) = \alpha_0 \Big/ \left(1 - \sum_{i=1}^{q}\alpha_i - \sum_{j=1}^{p}\beta_j\right)$$

And that the *conditional variance* is:

$$h_t = \alpha_0 + \sum_{i=1}^{q}\alpha_i\varepsilon_{t-i}^2 + \sum_{i=1}^{p}\beta_j h_{t-j}$$

And since $\{v_t\}$ is a white noise process, the *conditional* and *unconditional means* of ε_t are equal to zero.

This generalized *GARCH* (*p, q*) model in (10) allows for *ARMA* components in the h_t. It has been shown in applied settings that a *GARCH (p, q)* model with low values of *q* and *p* provides a better fit to the data than an *ARCH (q)* model with a high value of *q*.

In general, if we suppose in (10) that $p = 0$ and so are all values of β_j, and $q = 1$, the *GARCH* (*p, q*) model collapses to an *ARCH(1)* or *GARCH (0,1)* model. Thus, an ordinary ARCH model is a special case of a GARCH specification in which there are no lagged forecast variances in the h_t equation.

Key points to NOTE about GARCH processes:

1. Estimating a GARCH process typically involves an estimation of two interrelated equations:

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t \; ; \; i = 1,\dots, p \qquad \text{(i)}$$

Where
$$\varepsilon_t = v_t \left(\alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \right)^{0.5}$$

and

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_j h_{t-j} \qquad \text{(ii)}$$

Where y_t can be an ARMA process of order (p^m, q^m). The first equation is the model of the mean and the second is the model of the variance. Symbols p^m and q^m implies that the order of the ARMA process for the mean need not equal the order of the *GARCH (p, q)* equation.

2. The two equations are related in that h_t is the conditional variance of ε_t, hence ε_t of the GARCH process is the conditional variance of the mean equation. As a matter of fact ε_t^2 is not h_t itself.

We know that

$$\varepsilon_t = v_t (h_t)^{0.5} \quad \Leftrightarrow \varepsilon_t^2 = v_t^2 h_t .$$

But $\quad Ev_t^2 = E_{t-1} v_t^2 = 1$

Hence $\quad E_{t-1}\varepsilon_t^2 = h_t$

Showing that ε_t^2 is not h_t itself.

3. *GARCH (1, 1)* model is the most popular form of conditional volatility, especially for financial data where volatility shocks are very persistent. It takes the following specification:

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} \qquad (11)$$

As shown, h_t is a function of a constant, α_0, news about volatility from the previous period, measured as the lag of ε_t^2 from the mean equation (the ARCH term) and the previous period's forecast variance: h_{t-1} (the GARCH term).

i) If $\alpha_1 = 0$, there is no volatility clustering
ii) if $\beta_1 = 0$, there is absence of a GARCH term in h_t
iii) *GARCH (1, 1)* process for ε_t is weakly stationary if and only if $\alpha_1 + \beta_1 < 1$.
iv) Volatility persistence is captured by $\alpha_1 + \beta_1$, i.e. the degree of autoregressive decay of the squared residuals.

v) Conditional volatility increases with large values of both α_1 and β_1, but in different ways. The response of h_t to new information is an increasing function of the magnitude of α_1. And the larger the value of β_1, the more is the autoregressive persistence of the h_t.

vi) The value of α_1 must be strictly positive.

Also note that the *GARCH (1, 1)* model can be expressed as an *ARMA (1, 1)* for the ε_t^2. The proof of this is straight forward.

From $$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1},$$

Adding $\varepsilon_t^2 - h_t$ on both sides

$$h_t + \varepsilon_t^2 - h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} + \varepsilon_t^2 - h_t$$

Adding $\beta_1 \varepsilon_{t-1}^2$ and subtracting $\beta_1 \varepsilon_{t-1}^2$ from the RHS.

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} + \beta_1 \varepsilon_{t-1}^2 - \beta_1 \varepsilon_{t-1}^2 + \varepsilon_t^2 - h_t$$

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \varepsilon_{t-1}^2 - \underbrace{\beta_1 \left(\varepsilon_{t-1}^2 - h_{t-1}\right)}_{\omega_{t-1}} + \underbrace{\varepsilon_t^2 - h_t}_{\omega_t}$$

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \varepsilon_{t-1}^2 - \beta_1 \omega_{t-1} + \omega_t$$

Where ω is the volatility surprise, defined as $\varepsilon_t^2 - h_t$.

2. The GARCH-M Model

See ARCH-M model above, but noting that GARCH (1,1) h_t is no more than

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$$

So as shown before, the variance appears in the mean of the sequence $\{ y_t \}$ explicitly. The sequence $\{ y_t \}$ is stationary as long as the variance process is stationary.

3. The Integrated GARCH model

In empirical estimations, $\hat{\alpha}_1$ and $\hat{\beta}_1$ in *GARCH (1,1)* or $\sum_{i=1}^{q} \hat{\alpha}_i$ and $\sum_{i=1}^{p} \hat{\beta}_i$ in a general *GARCH(q, p)* model tend to sum to a value close to one. IGARCH is the limit case, where the sum of these parameters equals exactly one.

Once $\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i = 1$ as is the case for the IGARCH, h_t is not definite anymore, and while the sequence $\{ y_t \}$ is no longer covariance stationary, it remains strictly stationary because the unconditional density of ε_t does not change over time. The *IGARCH (1, 1)* model is written as:

$$h_t = \alpha_0 + (1 - \beta_1) \varepsilon_{t-1}^2 + \beta_1 h_{t-1} \qquad ; \; 0 < \beta_1 < 1.$$

Re-arranging this we get:

$$h_t = \alpha_0 + h_{t-1} + (1 - \beta_1) \left[\varepsilon_{t-1}^2 - h_{t-1} \right].$$

DAY 5 Estimation of the ARCH, GARCH and related models

The appropriate way to obtain a proper order of the GARCH process is to estimate the mean and conditional variance equations simultaneously. As such, GARCH processes are estimated by maximum-likelihood techniques so as to obtain estimates that are fully efficient. E-views and other automated software packages contain built in routines that estimate these models. Other classes of software packages, e.g. *Stata* require the modeller to write a program for the GARCH-type models. In the class of built in routine command packages, e.g. E-views, all that the modeller needs is to specify the order of the process and the assumption about the conditional distribution of ε_t and the computer does the rest. By default, G(ARCH) models in E-views are estimated by the method of M-L under the assumption that the errors are conditionally normally distributed.

DAY 6. Models with Asymmetry and Stochastic Volatility (SV)

1. The Threshold GARCH (TGARCH) Model

An interesting feature of asset prices is that 'bad' news tends to have a more pronounced effect on volatility than does the 'good' news. TGARCH models show how to allow the effects of good and bad news to have different effects on conditional volatility (Glosten *et al.*, 1993). In the model, 'new information' is measured by the size of the shock ε_t, and the good news is when $\varepsilon_{t-i} > 0$ and bad news when $\varepsilon_{t-i} < 0$. The TGARCH process is given by:

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \lambda_1 d_{t-1} \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$$

where d_{t-1} is a dummy variable that is equal to 1 if $\varepsilon_{t-1} < 0$ and is equal to zero if $\varepsilon_{t-1} > 0$. Thus, if

(i) $\varepsilon_{t-1} > 0$, $d_{t-1} = 0$, and the effect of an ε_{t-1} shock on h_t is $\alpha_1 \varepsilon_{t-1}^2$.

(ii) $\varepsilon_{t-1} < 0$, $d_{t-1} = 1$, and the effect of an ε_{t-1} shock on h_t is $(\alpha_1 + \lambda_1)\varepsilon_{t-1}^2$.

a) If $\lambda_1 > 0$, negative shocks will have larger effects on volatility than positive shocks. This is the *leverage effect*, i.e. the tendency for volatility to decline when returns rise and to rise when returns fall (*news impact curve*, Enders, 156).

b) If $\hat{\lambda}_1$ is statistically different from zero, the data contains a threshold effect.

2. The Exponential GARCH (EGARCH) Model

As opposed to the standard GARCH model which necessitates that all estimated coefficients are positive, EGARCH does not require nonnegativity constraint (Nelson, 1991), and is specified as:

$$\ln(h_t) = \alpha_0 + \alpha_1 \left(\frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} \right) + \lambda_1 \left| \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} \right| + \beta_1 \ln(h_{t-1}).$$

Note three interesting features about the EGARCH model:

i) Regardless of the magnitude of $\ln(h_t)$, the implied value of h_t can never be negative so it is permissible for the coefficient to be negative.

ii) The standardized value of ε_{t-1} is used in the place of ε_{t-1}^2, allowing for a more natural interpretation of the size and persistence of shocks. $\frac{\varepsilon_{t-1}}{h_{t-1}^{0.5}}$ is a unit free measure.

iii) The model allows for *leverage effects*.

If $\frac{\varepsilon_{t-1}}{(h_{t-1})^{0.5}} > 0$, the effect of the shock on $\ln(h_t)$ is $\alpha_1 + \lambda_1$, and

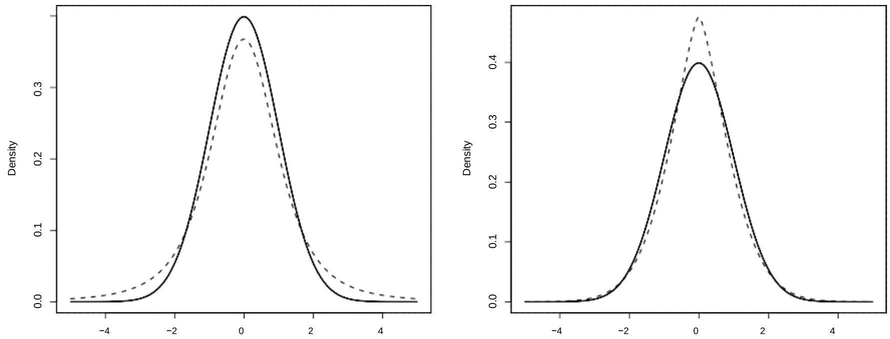
If $\frac{\varepsilon_{t-1}}{(h_{t-1})^{0.5}} < 0$, the effect of the shock on $\ln(h_t)$ is $-\alpha_1 + \lambda_1$. (see Enders 2010: 157).

3 Others include the Power ARCH (PARCH) and the Component GARCH (CGARCH) models.

3. Departure from Gaussianity

A common feature of financial assets is that the distribution function for the rate of return is fat-tailed (have excess kurtosis) than a normal distribution. If there is a good reason to believe that the asset return has a higher probability of a very large loss or gain than indicated by the normal distribution, one might want to estimate M-L using student's *t*–distribution and/or Generalized Error distribution. The comparison to the normal distribution is as follows:

Figure 7: Comparison of the Normal (solid line) and *t*– and *GED* Distributions, respectively.



The *t*-distribution in particular places a greater likelihood on large realizations than does the normal distribution. The good news is E-Views allows one to estimate a GARCH model using any of these distributions.

4. Stochastic Volatility (SV)

A major drawback with the GARCH (1, 1) framework is that it does not allow for a specific error in the dynamics of conditional volatility. Implicitly, the assumption that a large return (in absolute value) is associated with large volatility always holds, even when practically this is far from the statistical feature of financial returns.

SV model (Taylor 1986) extends the GARCH (1, 1) model by introducing the aspect of a specific source of randomness in conditional volatility. It takes the form:

$$\varepsilon_t = \sigma_t z_t = z_t \exp\left(\frac{1}{2}h_t\right) \quad ; \; z_t \sim iidN(0,1)$$

$$h_t = \alpha_0 + \beta_1 h_{t-1} + v_t \quad ; \; v_t \sim iidN\left(0, \sigma_v^2\right)$$

Where $\operatorname{cov}(z_t v_t) = 0$. v_t = source of randomness in the dynamics of conditional volatility, which improves the description of the actual volatility.

Like in the GARCH model, in the SV model, the distribution of the returns has excess kurtosis and persistence is captured by the autoregressive terms β_j. The inclusion of randomness feature makes the SV model much more flexible than the GARCH models, and has been found to fit returns volatility better. However, the possible improvement in capturing the most statistical features of return volatility comes at the price of higher complexity at the estimation level. The difficulty lies with how to evaluate the likelihood function. The error term v_t implies that the SV cannot be estimated directly by M-L. Instead, quasi-maximum likelihood estimation (QMLE) is used.

DAY 7: Modelling Volatility within a multivariate Framework

1.1 Conditional Heteroskedasticity within Multivariate GARCH

Multivariate GARCH is founded on the basis that contemporaneous shocks to variables can be correlated with each other, allowing for volatility spill-overs (positive or negative). That is, volatility shocks to one variable might lead to volatility of other related variables, but to magnitudes that only empirical analysis can reveal. Multivariate GARCH models allow the variances and Covariances to depend on the information set in a vector ARMA manner, the type useful in multivariate financial models, which require the modelling of both variances and Covariances (such as CAPM or dynamic hedging models).

Building from a univariate GARCH model, an *n*-variant model requires allowing the conditional variance-covariance matrix of the *n*-dimensional zero mean random variables ε_t to depend on elements of information set ϖ_{t-1}, that is:

$$\varepsilon_t \sim N(0, \mathbf{H_t}).$$

To put it in context, assume an *n*-dimensional ARCH or GARCH model, which in a compact form is represented as:

$$\mathbf{x}_t = (x_{1t}, x_{2t}, \ldots, x_{Nt})' \qquad (12)$$

Where $\mathbf{x}_t$ = (N x 1). $E(\mathbf{x}_t / \varpi_{t-i}) = \boldsymbol{\mu}$; $\boldsymbol{\mu}$ = (N x 1) & $Var(\mathbf{x}_t / \varpi_{t-i}) = \mathbf{H}_t$; $\mathbf{H}_t$ = (N x N)

The diagonal elements of $\mathbf{H}_t$ are the h_{it} terms and the off-diagonal elements are the h_{ijt} terms. The h_{it} and h_{ijt} are estimated simultaneously by M-L to capture the interactions between volatility of the N time series.

For ease of exposition, take a case of just two variables (*N=2*), x_{1t} and x_{2t} and $p = q = 1$. As in the univariate case, the GARCH (1,1) errors for the two error processes become:

$$\varepsilon_{1t} = v_{1t}(h_{11t})^{0.5}$$
$$\varepsilon_{2t} = v_{2t}(h_{22t})^{0.5}$$

Where h_{11t} and h_{22t} are the conditional variances of ε_{1t} and ε_{2t}, respectively for $\text{var}(v_{1t}) = \text{var}(v_{2t}) = 1$ as before.

To allow for the possibility that ε_{1t} and ε_{2t} are correlated, the conditional variance between ε_{1t} and ε_{2t} becomes h_{12t}, and specifically,

$$h_{12t} = E_{t-1}\varepsilon_{1t}\varepsilon_{2t}$$

Allowing in this framework the interaction of volatility terms with each other, it is possible to construct various representations of the GARCH models available in the literature, including VECH (Engel and Kroner, 1995), diagonal, BEKK (after Baba, Engel, Kraft and Kroner, 1990), and constant correlation. In what follows, we discuss each of these in detail.

The VECH Representation

Under the so-called VECH representation of multivariate GARCH, conditional variance of each variable, h_{11t} and h_{22t}, depends on its own past; the conditional covariance between the two variables, h_{12t}; the lagged squared errors, ε_{1t-1}^2 and ε_{2t-1}^2; and the product of the lagged errors, $\varepsilon_{1t-1}\varepsilon_{2t-1}$. To show this, define expressions for h_{11t}, h_{12t} and h_{22t} as follows:

$$h_{11t} = \xi_{10} + \alpha_{11}\varepsilon_{1t-1}^2 + \alpha_{12}\varepsilon_{1t-1}\varepsilon_{2t-1} + \alpha_{13}\varepsilon_{2t-1}^2 + \beta_{11}h_{11t-1} + \beta_{12}h_{12t-1} + \beta_{13}h_{22t-1} \quad (12.1)$$

$$h_{12t} = \xi_{20} + \alpha_{21}\varepsilon_{1t-1}^2 + \alpha_{22}\varepsilon_{1t-1}\varepsilon_{2t-1} + \alpha_{23}\varepsilon_{2t-1}^2 + \beta_{21}h_{11t-1} + \beta_{22}h_{12t-1} + \beta_{23}h_{22t-1} \quad (12.2)$$

$$h_{22t} = \xi_{30} + \alpha_{31}\varepsilon_{1t-1}^2 + \alpha_{32}\varepsilon_{1t-1}\varepsilon_{2t-1} + \alpha_{33}\varepsilon_{2t-1}^2 + \beta_{31}h_{11t-1} + \beta_{32}h_{12t-1} + \beta_{33}h_{22t-1} \quad (12.3)$$

A compact matrix representation of these equations takes the form:

$$\begin{bmatrix} h_{11,t} \\ h_{12,t} \\ h_{22,t} \end{bmatrix} = \begin{bmatrix} \xi_{10} \\ \xi_{20} \\ \xi_{30} \end{bmatrix} + \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{bmatrix} \begin{pmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}^2 \end{pmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \end{bmatrix} \begin{pmatrix} h_{11,t-1} \\ h_{12,t-1} \\ h_{22,t-1} \end{pmatrix} \quad (12.4)$$

Showing a rich interaction between x_{1t} and x_{2t}.

There are several complications which make it very difficult to estimate the VECH representation of multivariate GARCH models, despite the ease with which it can be conceptualized.

The number of parameters necessary to estimate can be extremely large. Even for a simple two variable GARCH (1,1) model above, there are 21 parameters, and the estimation can be quite complicated with more variables added to the system and if the order of the GARCH process increases. Moreover, as with univariate GARCH models, it is necessary to impose restrictions on the parameters of the product of the lagged errors of the model to ensure the non-negativity of the conditional variances of individual series, which in practice can however be difficult to do. Overcoming these problems involves imposing restrictions on the general model in (12.4) as in the diagonal VECH model.

The Diagonal VECH Representation

This is based on the assumption that each conditional variance is equivalent to that of a univariate GARCH process and the conditional Covariance is quite parsimonious. In the simplest case of *N=2* and $p = q = 1$, the diagonal representation form of the VECH reduces the number of parameters to be estimated to nine from 21, and reduces to the form:

$$\begin{bmatrix} h_{11,t} \\ h_{12,t} \\ h_{22,t} \end{bmatrix} = \begin{bmatrix} \xi_{10} \\ \xi_{20} \\ \xi_{30} \end{bmatrix} + \begin{bmatrix} \alpha_{11} & 0 & 0 \\ 0 & \alpha_{22} & 0 \\ 0 & 0 & \alpha_{33} \end{bmatrix} \begin{pmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}^2 \end{pmatrix} + \begin{bmatrix} \beta_{11} & 0 & 0 \\ 0 & \beta_{22} & 0 \\ 0 & 0 & \beta_{33} \end{bmatrix} \begin{pmatrix} h_{11,t-1} \\ h_{12,t-1} \\ h_{22,t-1} \end{pmatrix} \tag{13}$$

Performing the matrix multiplication yields:

$$\begin{aligned} h_{11,t} &= \xi_{10} + \alpha_{11}\varepsilon_{1,t-1}^2 + \beta_{11}h_{11,t-1} \\ h_{12,t} &= \xi_{20} + \alpha_{22}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + \beta_{22}h_{12,t-1} \\ h_{22,t} &= \xi_{30} + \alpha_{22}\varepsilon_{2,t-1}^2 + \beta_{22}h_{22,t-1} \end{aligned}$$

Showing that variances depend solely on past own squared residuals and past values of itself, and each element of the Covariance matrix, $h_{12,t}$ depends only on past values of itself and past values of $\varepsilon_{1,t}\varepsilon_{2,t}$.

This easy to estimate diagonal VECH model is handy, but setting all $\alpha_{ij} = \beta_{ij} = 0 \ \forall i \neq j$ devoids the model of interactions among the variables – the very essence of multivariate GARCH modelling. Indeed, as can be seen, an $\varepsilon_{1,t-1}$ shock affects $h_{11,t}$ and $h_{12,t}$ but not $h_{22,t}$. Moreover, we require that $h_{12,t}$ be positive definite for all values of ε_{it} in the sample space, a restriction that can be difficult to check, let alone impose during estimation.

The BEKK Representation

The BEKK (Baba, Engel, Kraft and Kroner, 1990) multivariate GARCH model guarantees that that the conditional variances are positive, by forcing all the parameters to enter the model via quadratic forms. It assumes the following model for $\mathbf{H}_t$:

$$\mathbf{H_t} = \boldsymbol{\xi}_\mathbf{0} + \sum_{\mathbf{i=1}}^{\mathbf{q}} \boldsymbol{\alpha}_\mathbf{i}\boldsymbol{\varepsilon}_\mathbf{t-i}\boldsymbol{\varepsilon}'_\mathbf{t-i}\boldsymbol{\alpha}'_\mathbf{i} + \sum_{\mathbf{i=1}}^{\mathbf{p}} \boldsymbol{\beta}_\mathbf{i}\mathbf{H}_\mathbf{t-i}\boldsymbol{\beta}'_\mathbf{i}$$

The BEKK representation in the *N=2* and $p = q = 1$ case in (12.4) becomes:

$$\begin{aligned} \begin{bmatrix} h_{11,t} & h_{12,t} \\ h_{12,t} & h_{22,t} \end{bmatrix} &= \begin{bmatrix} \xi_{11} & \xi_{12} \\ \xi_{12} & \xi_{22} \end{bmatrix} + \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \begin{bmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} & \varepsilon_{2,t-1}^2 \end{bmatrix} \begin{pmatrix} \alpha_{11} & \alpha_{21} \\ \alpha_{12} & \alpha_{22} \end{pmatrix} \\ &+ \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \begin{bmatrix} h_{11,t-1} & h_{12,t-1} \\ h_{12,t-1} & h_{22,t-1} \end{bmatrix} \begin{pmatrix} \beta_{11} & \beta_{21} \\ \beta_{12} & \beta_{22} \end{pmatrix} \end{aligned} \tag{14}$$

Suppressing the time subscripts and the GARCH terms:

$$\begin{aligned} h_{11,t} &= \xi_{11} + \alpha_{11}^2\varepsilon_1^2 + 2\alpha_{11}\alpha_{21}\varepsilon_1\varepsilon_2 + \alpha_{21}^2\varepsilon_2^2 \\ h_{12,t} &= \xi_{12} + \alpha_{11}\alpha_{12}\varepsilon_1^2 + (\alpha_{21}\alpha_{12} + \alpha_{11}\alpha_{22})\varepsilon_1\varepsilon_2 + \alpha_{21}\alpha_{22}\varepsilon_2^2 \\ h_{22,t} &= \xi_{22} + \alpha_{12}^2\varepsilon_1^2 + 2\alpha_{12}\alpha_{22}\varepsilon_1\varepsilon_2 + \alpha_{22}^2\varepsilon_2^2 \end{aligned}$$

As can be seen, BEKK unlike the VECH representation in (12.4), requires only 8 parameters (without the GARCH terms) but 11 parameters (GARCH terms inclusive) to be estimated and allows for the interaction effects that the diagonal representation in (13) does not. As such the model allows for the spill-over effects of shocks to the variance of one of the variables to the others. In principle, the BEKK representation improves on both the VECH and diagonal multivariate GARCH representations. However, despite its improvement over the VECH and diagonal representations, in practice, the BEKK model has been shown to be quite difficult to estimate. A large number of parameters in it are not globally identified, and as such, convergence can be quite difficult to achieve.

The Constant Conditional Correlation (CCC) Representation

CCC representation of the multivariate GARCH model, restricts the conditional correlation coefficients to be equal to the correlation coefficients between the variables, which are simply constants. Thus, as the name suggests, the conditional correlation coefficients are constant over time.

As such, in the simplest case of *N=2* and $p = q = 1$, the CCC model assume:

$$h_{12,t} = \rho_{12}\left(h_{11,t}h_{22,t}\right)^{0.5}$$

The covariance equation entails only one parameter, and the variance terms need not be diagonalized and the covariance terms are proportional to $\left(h_{11,t}h_{22,t}\right)^{0.5}$.

2.2 Conditional Heteroskedasticity, Unit roots and Cointegration

Conditional heteroskedasticity is a common feature of many financial time series, but as discussed earlier, an assumption of many time series models is that the error terms are zero mean, homoskedastic, *iid* random variables. This makes testing for a unit or testing for cointegrating relationships, in the presence of conditional heteroskedasticity an important issue. However, developments in time series econometrics have shown that, asymptotically, the Dickey Fuller (Dickey and Fuller, 1979) tests are robust to the presence of conditional heteroskedasticity, such as ARCH and GARCH (see e.g., among others, Phillips and Perron, 1988). Consequently, in applied work on testing for unit roots or cointegration in financial time series, it is extremely rare that potential difficulties caused by presence of conditional heteroskedasticity are considered to be a problem.

DAY 8: Forecasting the Conditional Variance, Forecast performance evaluation

3. Forecasting the Conditional Variance

Consider the GARCH (1, 1) model:

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} \qquad \text{(Given in (11))}$$

One-step-ahead forecast

Simply update h_t by one period and since ε_t^2 and h_t are known in period t, one–step-ahead forecast is:

$$h_{t+1} = \alpha_0 + \alpha_1 \varepsilon_t^2 + \beta_1 h_t .$$

The *j-step-ahead forecasts*

First, we know that:

$$\varepsilon_t^2 = v_t^2 h_t$$

Updating this by j periods gives:

$$\varepsilon_{t+j}^2 = v_{t+j}^2 h_{t+j}$$

Taking the conditional expectations gives:

$$E_t \varepsilon_{t+j}^2 = E_t (v_{t+j}^2 h_{t+j})$$

The fact that $\operatorname{cov}(v_{t+j}, h_{t+j}) = 0$ & $E_t v_{t+j}^2 = 1$, it follows that:

$$E_t \varepsilon_{t+j}^2 = E_t h_{t+j} \qquad (15)$$

Update eq. 11 by j-periods and take the conditional expectation to obtain:

$$E_t h_{t+j} = \alpha_0 + \alpha_1 E_t \varepsilon_{t+j-1}^2 + \beta_1 E_t h_{t+j-1}$$

Combine this with eq. (15) to get:

$$E_t h_{t+j} = \alpha_0 + (\alpha_1 + \beta_1) E_t h_{t+j-1} \qquad (16)$$

Using eq. (16) and given h_{t+1}, it is possible to obtain the j-step-ahead forecast of the conditional variance recursively.

$$\underset{j\to\infty}{E_t h_{t+j}} = \alpha_0 \left[1 + (\alpha_1 + \beta_1) + (\alpha_1 + \beta_1)^2 + \ldots + (\alpha_1 + \beta_1)^{j-1}\right] + (\alpha_1 + \beta_1)^j h_t$$

Providing $\alpha_1 + \beta_1 < 1$, the conditional forecast of h_{t+j} will converge to the long-run value

$$Eh_t = \frac{\alpha_0}{(1 - \alpha_1 - \beta_1)}$$

Similarly, consider the ARCH (q) model:

Further extensions should consider conditional heteroskedasticity, unit roots and cointegration to the estimation of multivariate GARCH.

4. Forecast Evaluation

Once an appropriate model (all ideal conditions must hold) has been chosen, and all the diagnostic tests are done, the next step is to test for accuracy of the forecasts. Several tests are used to determine the accuracy of the forecasting model, including;

The Mean Absolute Error (MAE) = $\frac{1}{n}\sum_{t=1}^{n}|\varepsilon_t|$

The Mean Square Error (MSE) = $\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^2$.

The lower the values of MAE and MSE, the better the forecasts

Root Mean Squared Error (RMSE) = $\sqrt{\frac{1}{T}\sum_{j=1}^{T-1}\left(F_{t+j}-A_{t+j}\right)^2}$

Where A represents actual values and F represents the forecasts. The RMSE is representative of the size of the error because it is measured in the same units as the data. A lower RMSE signifies better/more accurate forecasts

Theil's Forecast Accuracy (U) = $\frac{\left[\sum_{i=1}^{n}(P_i - A_i)^2\right]^{0.5}}{\left[\sum_{i=1}^{n}A_i^2\right]^{0.5}}$

Where P and A stand for a pair of predicted and observable values. If the theil Coefficient is 0 then, P=A, therefore perfect forecasts. If the theil coefficient is greater than 1, then the forecasts are not good. In other words, the closer the theil coefficient is to zero, the better the forecasts.