MODELLING VOLATILITY IN FINANCIAL TIME SERIES: AN INTRODUCTION TO ARCH

ECONOMETRICS C ⧫ LECTURE NOTE 7

HEINO BOHN NIELSEN

FEBRUARY 6, 2012

So far in this course, we have mainly discussed models for the conditional mean and we have focussed particularly on the structure of time dependence. In many applications within finance, however, there is a major interest also in the conditional variance of a time series, vaguely interpreted as the risk of holding certain assets. In this note we introduce a particular class of so-called *autoregressive conditional heteroskedasticity* (ARCH) models for the conditional variance. In 2003 Robert Engle was awarded the Nobel prize for the ARCH model, and since its introduction in Engle (1982) many interesting refinements have been proposed. In this note we focus on the main ideas, and, although interesting from a theoretical and empirical point of view, we only briefly mention some of the possible extensions. To give you a feeling for the behavior of financial data and the ARCH model we emphasize an empirical example.

OUTLINE

§1 Changing Volatility in Financial Time Series 2

§2 A Misspecification Test for ARCH 3

§3 The ARCH Model Defined 5

§4 Generalized ARCH (GARCH) Models 8

§5 Extensions to the Basic Model 11

§6 Concluding Remarks 16

1 Changing Volatility in Financial Time Series

Applications in financial economics are often interested in both the *mean* and the *variance* of investments, and sometimes in the entire return distribution. An intuitive reason is that an investor typically faces a trade off between the return and risk.

One application where the mean-variance trade-off is obvious is within the area of portfolio management, i.e. how to combine bonds, stocks, etc. in a portfolio, p. A simple setup used in some theoretical work is the so-called mean variance utility function, where the expected utility of the portfolio return, R_p, is defined as being

$$U(R_p) = E[R_p] - \frac{\phi}{2} V(R_p),$$

where ϕ is the degree of risk aversion, measuring the explicit trade-off between the expected return and the associated variance.

Another application where the whole distribution of returns is needed is for calculation the so-called *Value-at-Risk, VaR*. The VaR is used as a stress test and measures the maximum loss of a given portfolio calculated at a given probability level. In otheer words, the 5% VaR of portfolio p is a value so that the probalitity of a larger loss is exactly 5%, and it corresponds to the 5% quantile of the distribution of R_p.

A stylized fact for many financial time series is a tendency for the variance, or *volatility*, to be non-constant. In particular, as noted by Mandelbrot (1963):

> *"...large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes."*

In terms of economics we may think of a large shock to the return at time $t-1$, after which there is a high probability of another large shock at time t. The shock has upset the market and a period of larger uncertainty on the future direction of y_t follows. This idea, often referred to as *volatility clustering*, is just a reflection of high and low market uncertainty. The non-constant variance is known as heteroskedasticity, and this particular form of time varying heteroskedasticity is known as *autoregressive conditional heteroskedasticity* or ARCH. One main insight of the ARCH model is the distinction between *conditional* and *unconditional* variance; where the conditional variance, i.e. the part of the variance that cannot be predicted based on the information set, can be seen as a measure of risk. This is parallel to the distinction in a linear regression between the conditional and unconditional mean.

In many applications the autocorrelation of financial returns is quite low. Note, however, that the ARCH effects will make the *squared returns* highly autocorrelated, and an informal way to examine a data set for ARCH effects is therefore to look at the squared returns and to calculate the autocorrelation function (ACF) for squared returns.

Example 1 (volatility of ibm stock returns): As an example, let y_t denote the percentage return from month to month on the IBM stock on the New York Stock Exchange for the period 1926:1 to 1999:12. The returns are reported in Figure 1 (A). We note that the variance is high is some periods, e.g. in the early 1930'ties and in the 1990'ties, while movements are much smaller in other periods. Notice that the periods of large variation entail both positive and negative returns, so whereas the mean seems by-and-large constant, the variance is non-constant. We say that there are clear ARCH effects.

Figure 1 (B) depicts the squared returns for the IBM stock. In this graph the time dependence is clearly visible reflecting the clusters of high variance. Figure 1 (C) shows the ACF for the returns, y_t, and the squared returns, y_t^2. Whereas the autocorrelations are generally small for the returns they are much larger for squared returns: Again indicating ARCH effects.

We could estimate the variance of returns for the whole sample using the normal formula to estimate the *unconditional variance*,

$$\widehat{s}^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\epsilon}_t^2, \tag{1}$$

where $\hat{\epsilon}_t = y_t - \widehat{\mu}$ is the deviation from the estimated mean of y_t. To illustrate the expected variation in ϵ_t we could draw approximate 95% confidence bands as $\pm 1.96 \cdot \widehat{s}$. These constant confidence bands are reported in Figure 1 (D). Notice that due to the non-constant variance, the constant bands are very poor measures of the uncertainty at a given point in time, and the observations outside the band seem to cluster. As a measure of the uncertainty at a given point in time, t, we could instead calculate the variance using a small window of observations, e.g. the five previous observations:

$$\widehat{s}_t^2 = \frac{1}{5}\sum_{i=1}^{5}\hat{\epsilon}_{t-i}^2. \tag{2}$$

This *conditional variance* measure from (2) is also depicted in graph (D), and seems to be a better measure of the variance at a given point in time. The descriptive measure could be refined by changing the weights in the moving average, and the class of ARCH models we will introduce below is a way of formally embedding this kind of changing confidence bands into the statistical model. ♦

2 A Misspecification Test for ARCH

First, let y_t denote the variable of interest and we think of y_t as a (stationary) time series for e.g. an asset return or an interest rate. We consider the following linear model for the conditional mean,

$$E[y_t \mid x_t] = x_t'\theta, \quad t = 1, 2, ..., T$$

which corresponds to the linear regression model,

$$y_t = x_t'\theta + \epsilon_t, \tag{3}$$

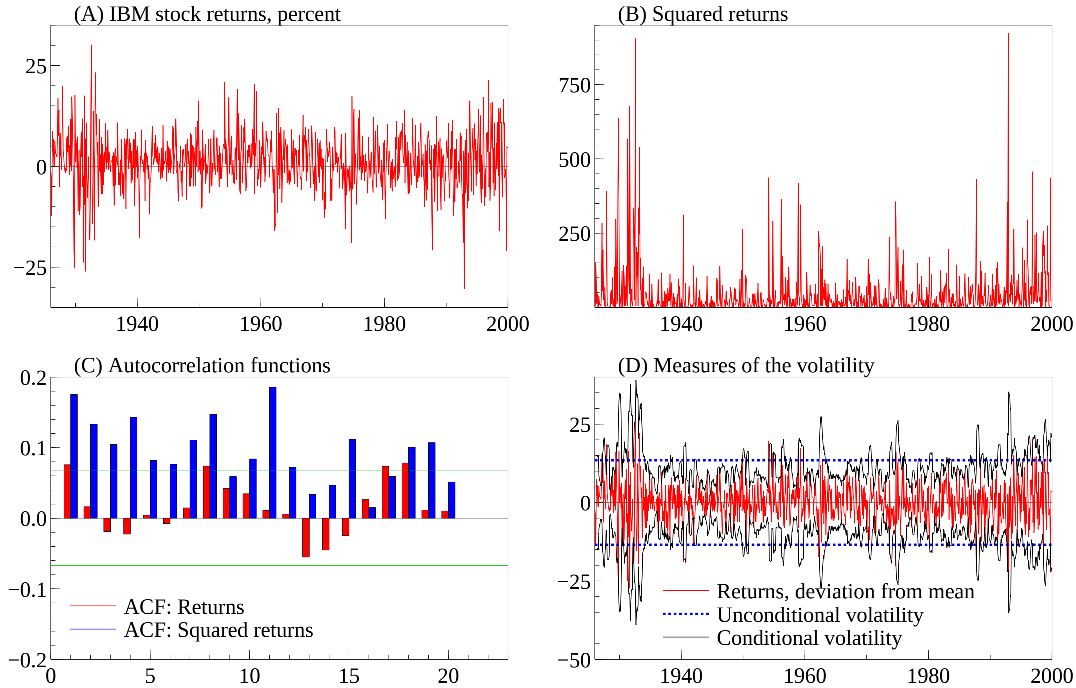

Figure 1: *Returns and squared returns for IBM stock.*

with $E[\epsilon_t \mid x_t] = 0$. In the literature of finance, there is a long history of seeing returns as approximately uncorrelated over time, and it is difficult to predict the conditional mean of an asset returns. This fact is often seen as a weak form of market efficiency, because predictability would typically allow easy speculative gains. It implies that the equation in (3) is often difficult to specify, and $x_t'\theta$ only accounts for a small proportion of the variation in y_t.

In the presence of ARCH effects, OLS estimation of the equation in (3) is consistent but inefficient. The sufficient condition for consistency of the OLS estimator in (3) is that $E[x_t\epsilon_t] = 0$, and ARCH effects will not in general violate that. OLS is no-longer efficient, however, and there exists a non-linear model that takes the ARCH effects into account, and the estimator in this model has a smaller variance.

To test the hypothesis of no-ARCH effects, we can use the standard Breusch-Pagan *LM test* for no-heteroskedasticity and apply it to this particular form of heteroskedasticity. If the null hypothesis is no-ARCH effects up to order p, we consider the auxiliary regression model

$$\hat{\epsilon}_t^2 = \gamma_0 + \gamma_1\hat{\epsilon}_{t-1}^2 + \gamma_2\hat{\epsilon}_{t-2}^2 + ... + \gamma_p\hat{\epsilon}_{t-p}^2 + \text{error}, \tag{4}$$

where $\hat{\epsilon}_t$ is the estimate residual from the regression of interest. The null hypothesis of no ARCH is given by

$$H_0 : \gamma_1 = \gamma_2 = ... = \gamma_p = 0,$$

so that the expected value of $\hat{\epsilon}_t^2$ is a constant for all t. We can use the familiar LM statistic,

$$\hat{\xi}_{ARCH} = T \cdot R^2,$$

which is asymptotically distributed as a $\chi^2(p)$ if the null is true.

It is important to note that the ARCH test has also power against residual autocorrelation. This is because autocorrelation in ϵ_t will imply autocorrelation in ϵ_t^2 (while the opposite is not true in general). Before the ARCH test is applied it is therefore important always to test for no-autocorrelation first. If the residual are *not* autocorrelated but the squared residuals are, that is interpreted as an indication of ARCH effects.

EXAMPLE 2 (MISSPECIFICATION TEST, IBM STOCK RETURN): First consider a linear regression for the conditional mean

$$y_t = \underset{(5.01)}{1.149} + \underset{(2.26)}{0.076} y_{t-1} + \widehat{\epsilon}_t.$$

The lagged dependent variable seems borderline significant, indicating some predictability of returns. The coefficient of determination is low, however, $R^2 = 0.00574$, meaning that the statistical model explains under one percent of the variation in returns. There are no signs of autocorrelation in the residuals, and a test for no-autocorrelation of order 1-2 is accepted with a LM test statistic of 0.228, corresponding to a $p-$value of 0.89 in a $\chi^2(2)$.

To test for the presence of ARCH effects, we consider the auxiliary regression of squared residuals on $p = 5$ lagged squared residuals:

$$\widehat{\epsilon}_t^2 = \underset{(6.74)}{25.804} + \underset{(4.53)}{0.153} \widehat{\epsilon}_{t-1}^2 + \underset{(2.93)}{0.100} \widehat{\epsilon}_{t-2}^2 + \underset{(1.39)}{0.048} \widehat{\epsilon}_{t-3}^2 + \underset{(3.18)}{0.110} \widehat{\epsilon}_{t-4}^2 + \underset{(0.484)}{0.017} \widehat{\epsilon}_{t-5}^2 + \text{error}.$$

We note that many of the lags are significant. The coefficient of determination is $R^2 = 0.06939$ and the LM test statistic for no-ARCH is given by

$$\hat{\xi}_{ARCH} = TR^2 = 882 \cdot 0.06939 = 61.2,$$

which is highly significant in the asymptotic $\chi^2(5)$ distribution. This is a formal indication of ARCH effects. ♦

3 The ARCH Model Defined

To analyze the conditional heteroskedasticity in more details we need a statistical model that allows for ARCH effects. We therefore define the *conditional variance*:

$$\sigma_t^2 = E[\epsilon_t^2 \mid \mathcal{I}_{t-1}],$$

where $\mathcal{I}_{t-1} = \{y_{t-1}, y_{t-2}, ..., x_{t-1}, x_{t-2}, ...\}$ is the information set available at time $t-1$. The class of ARCH models consists of an equation for the conditional mean (3) augmented with an equation for the conditional variance, σ_t^2.

Engle (1982) suggests a statistical model for σ_t^2 which follows directly for the descriptive measure in (2). In particular he suggests the ARCH(p) model for the conditional variance:

$$\sigma_t^2 = \varpi + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + ... + \alpha_p \epsilon_{t-p}^2. \tag{5}$$

Notice that the variance, σ_t^2, is a simple function of p lagged squared residuals. To ensure a consistent model that generates a positive variance, we need to constrain the parameters, $\varpi > 0$, $\alpha_i \geq 0$. The economic interpretation is straightforward: If there is a large shock at time $t-1$, i.e. if ϵ_{t-1}^2 is high, then the variance of the following shocks will be large. Graphically, the width of the confidence bands depends on the magnitude of the past shocks, and the ARCH(p) process has a memory of p periods.

A common way to write the full model is the following

$$\begin{aligned} y_t &= x_t'\theta + \epsilon_t \\ \epsilon_t &= \sigma_t z_t, \end{aligned}$$

where σ_t is defined in (5) and z_t is an IID error term with mean zero and unit variance. At each point in time there is a new independent underlying shock to the system, z_t, which is scaled by σ_t so that the ARCH innovation have a conditional variance of $E[\epsilon_t^2 \mid \mathcal{I}_{t-1}] = \sigma_t^2$, and the conditionally heteroskedastic shock ϵ_t drives the observed return process y_t.

3.1 Interpretation

Another way to understand the model is to decompose ϵ_t^2 into the conditional expectation and the *surprise* in the squared innovations:

$$\epsilon_t^2 = E[\epsilon_t^2 \mid \mathcal{I}_{t-1}] + v_t = \sigma_t^2 + v_t,$$

where it follows that $E[v_t \mid \mathcal{I}_{t-1}] = 0$, so that v_t is an uncorrelated (but not necessarily homoskedastic) sequence. Inserting $\sigma_t^2 = \epsilon_t^2 - v_t$ into the ARCH(p) equation in (5) yields the equation

$$\epsilon_t^2 = \varpi + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + ... + \alpha_p \epsilon_{t-p}^2 + v_t, \tag{6}$$

which shows that the squared innovation, ϵ_t^2, follows an AR(p) process. Note that (6) is exactly the implication we use in the auxiliary regression (4) in the misspecification test.

Note that ϖ is not the unconditional variance. We can illustrate by taking expectations to obtain

$$E[\epsilon_t^2] = \varpi + \alpha_1 E[\epsilon_{t-1}^2] + ... + \alpha_p E[\epsilon_{t-p}^2],$$

which defines a constant unconditional variance of

$$\sigma^2 = \frac{\varpi}{1 - \alpha_1 - ... - \alpha_p},$$

provided that the sum of the coefficients is less than one, $\sum_{i=1}^p \alpha_i < 1$. Note that whereas the ARCH model exhibit conditional heteroskedasticity, the ARCH process is unconditionally homoskedastic.

3.2 Maximum Likelihood Estimation

Before we continue let us briefly discuss the estimation of ARCH models. To make the notation simple we consider an ARCH(1) model with

$$\begin{aligned} y_t &= x_t'\theta + \epsilon_t \\ \sigma_t^2 &= \varpi + \alpha\epsilon_{t-1}^2. \end{aligned}$$

To perform a likelihood analysis we have to specify a distributional shape for ϵ_t. First consider conditional normality:

$$\epsilon_t = \sigma_t z_t, \quad z_t \sim N(0,1),$$

or alternatively that $\epsilon_t \mid \mathcal{I}_{t-1} \sim N(0, \sigma_t^2)$. We can write the likelihood contribution as a function of observed data as

$$\begin{aligned} L_t(\theta, \varpi, \alpha \quad &| \quad y_t, x_t, y_{t-1}, x_{t-1}, ...) \\ &= \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{1}{2}\frac{\epsilon_t^2}{\sigma_t^2}\right) \\ &= \frac{1}{\sqrt{2\pi(\varpi + \alpha(y_{t-1} - x_{t-1}'\theta)^2)}} \exp\left(-\frac{1}{2}\frac{(y_t - x_t'\theta)^2}{\varpi + \alpha(y_{t-1} - x_{t-1}'\theta)^2}\right), \end{aligned}$$

and maximize the sum of the contributions with respect to the parameters θ, ϖ and α. The analytical analysis of the likelihood function is somewhat complicated and we cannot solve the likelihood equations analytically. Instead we use numerical optimization to find the ML estimators.

In place of the normal distribution, other forms of distributions can be used. Many applications use a more fat-tailed distribution to allow a larger proportion of extreme observations. A particularly convenient solution is to use a student $t(v)$ distribution where the degree of freedom, v, determines how fat the tails should be. Here v can be treated as a parameter in the likelihood function and estimated jointly with the remaining parameters.

Example 3 (arch(p) model, ibm stock return): To illustrate the estimation of ARCH(p) models we consider a model given by the two equations

$$\begin{aligned} y_t &= \theta_0 + \theta_1 y_{t-1} + \epsilon_t \\ \sigma_t^2 &= \varpi + \alpha_1\epsilon_{t-1}^2 + \alpha_2\epsilon_{t-2}^2 + ... + \alpha_p\epsilon_{t-p}^2, \end{aligned}$$

where we assume that the error term is normal, $\epsilon_t \mid \mathcal{I}_{t-1} \sim N(0, \sigma_t^2)$. The estimation results for different lag lengths, $p = 1, 2, 3, 5, 7$, are reported in Table 1. The results suggest that the ARCH effects are clearly significant. Unfortunately, it is quite hard to determine the appropriate number of lags in the conditional variance. Many of the coefficients have similar magnitudes and comparable significance. It is an observed weakness of the ARCH model that it is hard to precisely pin down the shape of the memory structure, and the estimated coefficients are often relatively unstable between models. ♦

	ARCH(1)	ARCH(2)	ARCH(3)	ARCH(5)	ARCH(7)
θ_0	1.121 (4.94)	1.181 (5.88)	1.196 (6.11)	1.198 (6.12)	1.194 (6.11)
θ_1	0.113 (2.81)	0.116 (3.05)	0.110 (2.98)	0.102 (2.83)	0.102 (2.85)
ϖ	36.838 (12.2)	30.733 (10.5)	27.260 (9.36)	24.838 (7.80)	21.780 (6.37)
α_1	0.175 (3.35)	0.156 (2.83)	0.155 (2.96)	0.134 (2.61)	0.130 (2.48)
α_2		0.157 (2.32)	0.123 (1.92)	0.098 (1.52)	0.100 (1.47)
α_3			0.118 (2.42)	0.100 (2.13)	0.102 (2.10)
α_4				0.060 (1.26)	0.025 (0.595)
α_5				0.055 (1.67)	0.051 (1.38)
α_6					0.054 (1.26)
α_7					0.055 (1.41)
$\sum_{i=1}^{p} \alpha_i$	0.175	0.313	0.396	0.446	0.517
log-lik	−2929.19	−2916.92	−2912.09	−2909.10	−2904.70
SC	6.635	6.615	6.612	6.621	6.626
HQ	6.622	6.599	6.592	6.594	6.593
AIC	6.614	6.588	6.580	6.577	6.572
ARCH 1-5 test:	[0.00]	[0.08]	[0.63]	[0.98]	[0.98]

Table 1: *Estimation of ARCH(p) models for IBM stock returns.*

4 Generalized ARCH (GARCH) Models

To resolve the problem with many lags in ARCH models Bollerslev (1986) and Taylor (1986) suggested a generalized version of the ARCH model which economize more with the number of parameters. The simplest case is the very popular GARCH(1,1) model defined by the equation

$$\sigma_t^2 = \varpi + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2,$$

where the lagged variance is included along with the squared innovation. A simple interpretation is that the lagged dependent variable allows for a more smooth development in the variance and a longer memory without including many parameters.

To understand the relationship between ARCH and GARCH models we can again use the definition $\sigma_t^2 = \epsilon_t^2 - v_t$, and obtain that

$$\begin{aligned} \sigma_t^2 &= \varpi + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \\ \epsilon_t^2 - v_t &= \varpi + \alpha \epsilon_{t-1}^2 + \beta \left(\epsilon_{t-1}^2 - v_{t-1} \right) \\ \epsilon_t^2 &= \varpi + (\alpha + \beta) \epsilon_{t-1}^2 + v_t - \beta v_{t-1}. \end{aligned}$$

This suggests that the GARCH(1,1) implies an ARMA(1,1) structure for the squared innovation. We recall that that an ARMA model can be seen as a restricted and parsimonious representation of an infinite AR model, and we can think of the GARCH model as a restricted infinite ARCH model. By repeated substitution we can write the GARCH model as

$$\begin{aligned}
\sigma_t^2 &= \varpi + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \\
&= \varpi + \alpha\epsilon_{t-1}^2 + \beta(\varpi + \alpha\epsilon_{t-2}^2 + \beta\sigma_{t-2}^2) \\
&= \varpi(1+\beta) + \alpha\epsilon_{t-1}^2 + \alpha\beta\epsilon_{t-2}^2 + \beta^2\sigma_{t-2}^2 \\
&\vdots \\
&= \varpi(1+\beta+\beta^2+\beta^3+...) + \alpha\epsilon_{t-1}^2 + \alpha\beta\epsilon_{t-2}^2 + \alpha\beta^2\epsilon_{t-2}^2 + ... \\
&= \frac{\varpi}{1-\beta} + \alpha\sum_{j=1}^{\infty}\beta^{j-1}\epsilon_{t-j}^2,
\end{aligned}$$

where we assume that the process started in the infinite past.

Also for the GARCH model we need σ_t^2 to be non-negative and constrain the coefficients to be non-negative. By looking at the ARMA representation we can also read of the condition under which ϵ_t^2 is covariance stationary, namely that $\alpha+\beta<1$. In this case the unconditional variance is given by

$$\sigma^2 = E[\epsilon_t^2] = \frac{\varpi}{1-\alpha-\beta}.$$

The GARCH(1,1) model is extremely popular in applied work, but it can of cause be generalized with more lags as the GARCH(p,q) model:

$$\sigma_t^2 = \varpi + \sum_{j=1}^{p}\alpha_j\epsilon_{t-j}^2 + \sum_{j=1}^{q}\beta_j\sigma_{t-j}^2.$$

The condition for covariance stationarity is $\sum_{j=1}^{p}\alpha_j + \sum_{j=1}^{q}\beta_j < 1$.

4.1 Volatility Forecasts

One important application of ARCH and GARCH models is prediction of future volatility. We use $\sigma_{T+h|T}^2 = E[\epsilon_{T+h}^2 \mid \mathcal{I}_T]$ to denote the forecast of volatility at time $T+h$ given the information set at time T. To construct the forecast from an ARCH(1) we first use the fact $\varpi = \sigma^2(1-\alpha)$ to rewrite the equation in deviations from the mean

$$\begin{aligned}
\sigma_t^2 &= \varpi + \alpha\epsilon_{t-1}^2 \\
\sigma_t^2 &= \sigma^2(1-\alpha) + \alpha\epsilon_{t-1}^2 \\
\sigma_t^2 - \sigma^2 &= \alpha\left(\epsilon_{t-1}^2 - \sigma^2\right).
\end{aligned}$$

To forecast volatility for $T+1$ we find the best prediction

$$\sigma_{T+1|T}^2 = E[\epsilon_{T+1}^2 \mid \mathcal{I}_T] = E\left[\sigma^2 + \alpha\left(\epsilon_T^2 - \sigma^2\right) \mid \mathcal{I}_T\right] = \sigma^2 + \alpha\left(\epsilon_T^2 - \sigma^2\right),$$

where we have used that ϵ_T is in the information set at time T, so that $E\left[\epsilon_T^2 \mid \mathcal{I}_T\right] = \epsilon_T^2$. Forecasts for the next periods are constructed as the recursion

$$\begin{aligned}
\sigma^2_{T+2|T} &= E[\epsilon^2_{T+2} \mid \mathcal{I}_T] \\
&= E\left[\sigma^2 + \alpha\left(\epsilon^2_{T+1} - \sigma^2\right) \mid \mathcal{I}_T\right] \\
&= \sigma^2 + \alpha\left(E\left[\epsilon^2_{T+1} \mid \mathcal{I}_T\right] - \sigma^2\right) \\
&= \sigma^2 + \alpha\left(\sigma^2_{T+1|T} - \sigma^2\right), \\
\sigma^2_{T+3|T} &= E[\epsilon^2_{T+3} \mid \mathcal{I}_T] \\
&= E\left[\sigma^2 + \alpha\left(\epsilon^2_{T+2} - \sigma^2\right) \mid \mathcal{I}_T\right] \\
&= \sigma^2 + \alpha\left(E\left[\epsilon^2_{T+2} \mid \mathcal{I}_T\right] - \sigma^2\right) \\
&= \sigma^2 + \alpha\left(\sigma^2_{T+2|T} - \sigma^2\right),
\end{aligned}$$

etc. Note that the forecast will produce an exponential convergence towards the unconditional variance, σ^2.

For the GARCH(1,1) model we use a similar recursion. First write the model in deviations from mean

$$\begin{aligned}
\sigma_t^2 &= \varpi + \alpha\epsilon^2_{t-1} + \beta\sigma^2_{t-1} \\
\sigma_t^2 &= \sigma^2(1-\alpha-\beta) + \alpha\epsilon^2_{t-1} + \beta\sigma^2_{t-1} \\
\sigma_t^2 - \sigma^2 &= \alpha\left(\epsilon^2_{t-1} - \sigma^2\right) + \beta\left(\sigma^2_{t-1} - \sigma^2\right).
\end{aligned}$$

The first period forecast is given by

$$\begin{aligned}
\sigma^2_{T+1|T} &= E[\epsilon^2_{T+1} \mid \mathcal{I}_T] \\
&= E\left[\sigma^2 + \alpha\left(\epsilon^2_T - \sigma^2\right) + \beta\left(\sigma^2_T - \sigma^2\right) \mid \mathcal{I}_T\right] \\
&= \sigma^2 + \alpha\left(\epsilon^2_T - \sigma^2\right) + \beta\left(\sigma^2_T - \sigma^2\right),
\end{aligned}$$

because ϵ_T is in the information set at time T and σ_T^2 can be calculated from the information set. The next period forecast is

$$\begin{aligned}
\sigma^2_{T+2|T} &= E[\epsilon^2_{T+2} \mid \mathcal{I}_T] \\
&= E\left[\sigma^2 + \alpha\left(\epsilon^2_{T+1} - \sigma^2\right) + \beta\left(\sigma^2_{T+1} - \sigma^2\right) \mid \mathcal{I}_T\right] \\
&= \sigma^2 + \alpha\left(E\left[\epsilon^2_{T+1} \mid \mathcal{I}_T\right] - \sigma^2\right) + \beta\left(E\left[\sigma^2_{T+1} \mid \mathcal{I}_T\right] - \sigma^2\right) \\
&= \sigma^2 + \alpha\left(\sigma^2_{T+1|T} - \sigma^2\right) + \beta\left(\sigma^2_{T+1|T} - \sigma^2\right) \\
&= \sigma^2 + (\alpha+\beta)\left(\sigma^2_{T+1|T} - \sigma^2\right),
\end{aligned}$$

where we have used that $E\left[\epsilon^2_{T+1} \mid \mathcal{I}_T\right] = \sigma^2_{T+1|T}$ and $E\left[\sigma^2_{T+1} \mid \mathcal{I}_T\right] = \sigma^2_{T+1|T}$. For longer horizons we find similarly that

$$\sigma^2_{T+h|T} = \sigma^2 + (\alpha+\beta)\left(\sigma^2_{T+h-1|T} - \sigma^2\right),$$

which is an exponential convergence with speed $\alpha + \beta$.

Example 4 (garch models, ibm stock return): To illustrate, we estimate the AR(1)-GARCH(1,1), i.e.

$$\begin{aligned} y_t &= \theta_0 + \theta_1 y_{t-1} + \epsilon_t \\ \sigma_t^2 &= \varpi + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2. \end{aligned}$$

First column reports the ML estimates obtained by assuming normal innovations, $\frac{\epsilon_t}{\sigma_t} \sim N(0,1)$. We note that the lagged variance is large and significant, and $\alpha + \beta = 0.934$ is quite close to one. The fact that the estimated GARCH models have roots close to unity is a common observation which is discussed in more details in Box 1. To illustrate the results we report in Figure 2 (A) the estimated residuals as well as the confidence bands constructed as $\pm 1.96 \cdot \widehat{\sigma}_t$. In Figure 2 (B) we compare the estimated standard deviation from the GARCH(1,1) with the one obtained in the ARCH(5) model. They share the same general tendencies, but the GARCH(1,1) model is much less erratic and have a longer memory. The difference becomes even more pronounced if we compare the forecasts in Figure 2 (C). The GARCH model suggests a quite high variance for a number of years while the ARCH model converges very fast towards the unconditional variance; this reflects the difference in convergence speed between $\sum_{i=1}^{5} \alpha_i = 0.45$ for the ARCH(5) and $\alpha + \beta = 0.93$ for the GARCH(1,1). Also note that the volatility forecasts converge towards the unconditional variance, which is an average of the low and high variance periods in the past, and not towards the volatility of the low variance periods, although they are often the most common.

We note that the null of normality is strongly rejected ($p-$value of 0.00), due to a large proportion of extreme observations. One refinement is to estimate the model with the assumption of a more general error distribution, and second column reports the results for a student$-t$ distribution, $\frac{\epsilon_t}{\sigma_t} \sim t(\nu)$. The estimated degrees of freedom is $\hat{v} = 8.4$ which gives a much more fat-tailed distribution than the normal. It holds that kurtosis for the normal distribution is $\mathsf{K} = 3$ while it is $\mathsf{K} = 3 + \frac{6}{v-4} = 3 + \frac{6}{8.4-4} = 4.4$ for the $t(8.4)-$distribution. The latter is also close to the kurtosis of the residuals from the basic GARCH(1,1) model. There are also small differences in the estimated parameters, but in a graph the conditional standard errors are almost identical (not shown). ♦

5 Extensions to the Basic Model

Thee are many possible extensions of these basic model. Here we only consider two important cases.

5.1 Asymmetric ARCH Models and the News Impact Curve

The basic models have the feature that the sign of a shock does not matter and positive and negative shocks have the same effects on the conditional variance. The effect of a

Box 1: Persistence in Volatility and the IGARCH

It is a stylized feature of estimated GARCH models that the sum of the coefficients is close to one,

$$\sum_{i=1}^{p} \hat{\alpha}_i + \sum_{i=1}^{q} \hat{\beta}_i \approx 1,$$

known as an *integrated GARCH* or *IGARCH* model. In this case the squared residuals follow a unit root ARMA model, and the unit root has strange implications for the behavior of the model. First it violates what we called the stationarity condition. Secondly it implies that the unconditional variance is not defined, and, finally, a forecasts for the conditional variance will replicate the forecast of a random walk (with drift), i.e. a linear trend!

The latter implication may be a valid characterization of the data, but it is hard to maintain as a behavioral model for investors. As a consequence the IGARCH phenomenon is often discussed as a sign of misspecification of the class of GARCH models. One problem could be that there are structural shifts in the unconditional variance, i.e. in the constant term of the GARCH equation. If this is not modelled it will bias the roots of the ARMA model towards unity, as we have seen for the analysis of unit root data.

The asymptotic analysis of the IGARCH model is given in Nelson (1990). He demonstrates that the IGARCH model is actually strictly stationary, but the variance is infinite (and is not weakly stationary!). A result is that although the analysis looks strange, a test for a unit root in the variance, i.e. a likelihood ratio test for the IGARCH model against the GARCH follows a standard χ^2-distribution.

To illustrate we have replicated the estimates in Table 2 imposing the IGARCH restriction $\alpha+\beta=1$. Based on a likelihood ratio test, the IGARCH model is clearly rejected in this case. For the model with normal errors we get a test statistic of

$$LR(\alpha+\beta=1) = -2\cdot(-2908.03+2901.02) = 14.02,$$

which is clearly significant in a $\chi^2(1)$.

An additional complication in the near-integrated GARCH case is that the unconditional variance is poorly estimated. A simple estimator of the unconditional variance is $\hat{s}^2$ from (1). The unconditional variance in the GARCH model is $\sigma^2 = \varpi/(1-\alpha-\beta)$ and for $\alpha+\beta\approx 1$ the denominator is close to zero and the estimate of the constant term, ϖ, also becomes very small. As a consequence the estimator $\hat{\sigma}^2 = \hat{\varpi}/(1-\hat{\alpha}-\hat{\beta})$ can be very poor, and in the limiting case, $\alpha+\beta\to 1$, it does not exist. In practice, the estimate $\hat{\sigma}^2$ can be far from the estimate $\hat{s}^2$. Since the forecasts for the conditional variance describe a convergence back to the unconditional variance, a poor estimate of σ^2 may render the forecasts useless. A simple solution is to insert $\varpi = \sigma^2(1-\alpha-\beta)$ and write the GARCH equation as

$$\sigma_t^2 = \sigma^2(1-\alpha-\beta) + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2.$$

Instead of estimating $\left(\sigma^2,\alpha,\beta\right)$ (which is equivalent to estimating (ϖ,α,β)), we can fix σ^2 to some robustly estimated value (e.g. s^2) and only estimate the parameters (α,β) in the GARCH model. This is known as *variance targeting.*

	GARCH(1,1) $\frac{\epsilon_t}{\sigma_t} \sim N(0,1)$	GARCH(1,1) $\frac{\epsilon_t}{\sigma_t} \sim t(\nu)$	IGARCH(1,1) $\frac{\epsilon_t}{\sigma_t} \sim N(0,1)$	IGARCH(1,1) $\frac{\epsilon_t}{\sigma_t} \sim t(\nu)$
θ_0	1.179 (6.00)	1.226 (6.17)	1.292 (6.001)	1.312 (6.411)
θ_1	0.104 (2.86)	0.071 (1.93)	0.105 (3.123)	0.070 (1.974)
ϖ	2.932 (2.10)	2.535 (2.60)	0.622 (0.910)	0.822 (1.349)
α	0.097 (3.12)	0.093 (3.88)	0.094 (1.515)	0.107 (2.412)
β	0.837 (16.4)	0.850 (22.8)	0.906 (−−−)	0.893 (−−−)
ν		8.409 (4.31)		6.708 (4.267)
$\alpha+\beta$	0.934	0.944	1.000	1.000
log-lik	-2901.02	-2890.04	-2908.03	-2893.96
SC	6.580	6.562	6.588	6.564
HQ	6.563	6.542	6.574	6.547
AIC	6.553	6.530	6.566	6.537
ARCH 1-5 test:	[0.91]	[0.91]	[0.75]	[0.83]
Normality	[0.00]	...	[0.00]	...

Table 2: *Estimation of GARCH(p,q) models.*

shock ϵ_{t-1} on the conditional variance, σ_t^2, is known as the *news impact curve*; and for the basic model the news impact curve is obviously symmetric.

In some cases it is reasonable to believe that negative shocks have a different impact from positive shocks, and the models can easily be adapted to this situation. In a famous paper, Glosten, Jagannathan, and Runkle (1993) suggest to model the asymmetric effects in the following simple way

$$\sigma_t^2 = \varpi + \alpha\epsilon_{t-1}^2 + \kappa\epsilon_{t-1}^2 I(\epsilon_{t-1}) + \beta\sigma_{t-1}^2,$$

where

$$I(x) = \begin{cases} 1 & \text{if } x < 0 \\ 0 & \text{Otherwise} \end{cases}$$

is the indicator function. We follow PcGive and call this a *threshold* asymmetric model. We note that the news impact curve is now asymmetric, with an impact of squared residuals of α for positive shocks and $\alpha+\kappa$ for negative values.

Example 5 (threshold model, imb stock return): For the IBM stock return we estimate a threshold model given by

$$\begin{aligned} y_t &= \underset{(5.35)}{1.114} + \underset{(3.07)}{0.107} y_{t-1} + \epsilon_t \\ \sigma_t^2 &= \underset{(2.20)}{3.228} + \underset{(2.48)}{0.064}\epsilon_{t-1}^2 + \underset{(1.16)}{0.063}\epsilon_{t-1}^2 I(\epsilon_{t-1}) + \underset{(16.2)}{0.831}\sigma_{t-1}^2 \end{aligned}$$

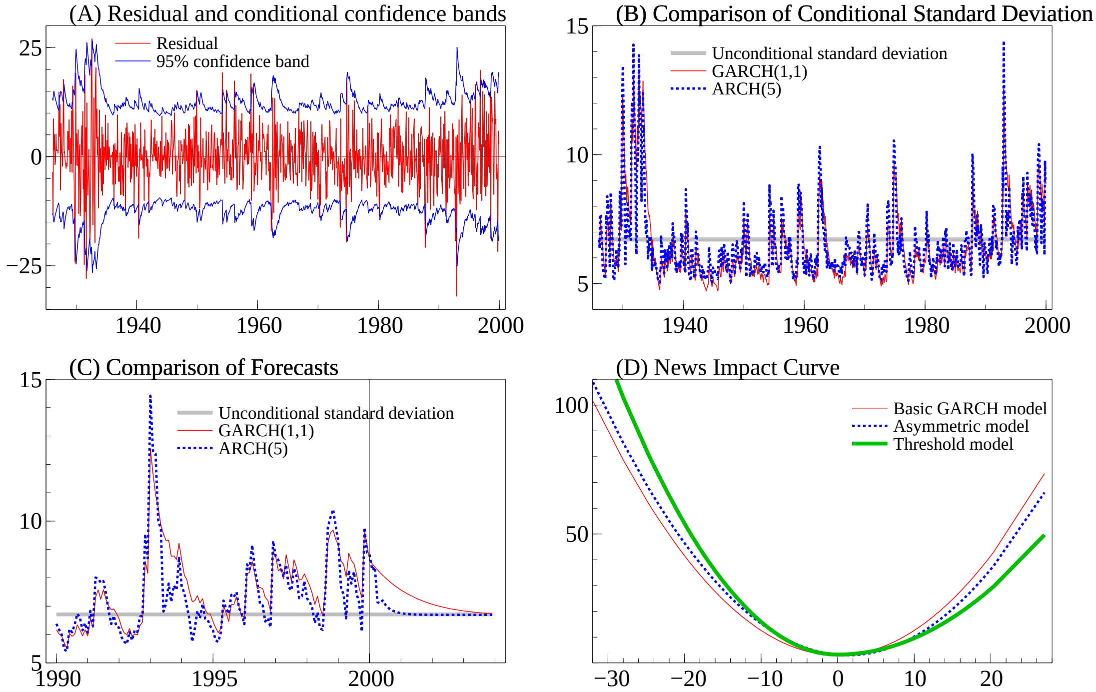

Figure 2: *Results for the analysis of the IBM stock returns.*

with $t-$values in parentheses and a log-likelihood of -2899.308. At face value the estimates suggest an asymmetric effect, with impacts of $\alpha = 0.064$ and $\alpha + \kappa = 0.064 + 0.063 = 0.127$ for positive and negative shocks respectively. This is compared to the basic symmetric news impact curve in Figure 2 (D), and suggests that the market is more upset by negative than positive shocks. The threshold parameter is not statistically significant, however, with a $t-$test statistic of $t_{\kappa=0} = 1.16$. We can also test the significant by a likelihood ratio test, where the statistic $LR(\kappa = 0) = 2 \cdot (-2899.308 + 2901.02) = 3.424$ corresponds to a $p-$value of 0.064. ♦

An alternative form of asymmetry, suggested by Engle and Ng (1993), is

$$\sigma_t^2 = \varpi + \alpha(\epsilon_{t-1} - \gamma)^2 + \beta\sigma_{t-1}^2,$$

where the news impact curve has the same slope for small and large values, but zero is no-longer the neutral shock. We follow PcGive and refer to this as the *asymmetric* model.

Example 6 (asymmetric model, IBM stock return): To illustrate the asymmetric model, we consider the IBM stock return and obtain

$$\begin{aligned} y_t &= \underset{(5.32)}{1.125} + \underset{(3.01)}{0.107} y_{t-1} + \epsilon_t \\ \sigma_t^2 &= \underset{(2.30)}{2.988} + \underset{(3.58)}{0.096}(\epsilon_{t-1} - \underset{(0.298)}{1.332})^2 + \underset{(17.0)}{0.832}\sigma_{t-1}^2. \end{aligned}$$

with $t-$values in parentheses. Based on this specification there is very little evidence for asymmetry. Again we have depicted the news impact curve in Figure 2 (D) and note that the effect is minor. The log-likelihood of this specification is -2899.928. ♦

The two models can easily be combined into a more general specification and a fully asymmetric GARCH(p,q) model could have the form

$$\sigma_t^2 = \varpi + \sum_{i=1}^{p} \alpha_i (\epsilon_{t-i} - \gamma)^2 + \kappa \epsilon_{t-1}^2 I(\epsilon_{t-1} - \gamma) + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2,$$

which allows for a very elaborate specification of the news impact curve.

5.2 ARCH in Mean

Most theoretical models in finance suggest a trade-off between risk and return; and within a given time period investors require a larger expected return from an investment which is riskier, often referred to as a *risk premium*. Similar effects may also work over time, so that investors require a higher return in time periods where the volatility is asserted to be higher. There may also be effects in the opposite direction, see Glosten, Jagannathan, and Runkle (1993) for a discussion.

The argument above suggests that there should be a relationship between the conditional variance and the conditional mean. One way to test the hypothesis is to let the conditional mean depend on the variance, σ_t^2. It is not obvious how the relationship should be, but some suggested examples replaces the simple equation for the conditional mean (3) with

$$\begin{aligned} y_t &= x_t'\theta + \delta\sigma_t^2 + \epsilon_t \\ y_t &= x_t'\theta + \delta\sigma_t + \epsilon_t \\ y_t &= x_t'\theta + \delta \log\left(\sigma_t^2\right) + \epsilon_t. \end{aligned}$$

A positive estimate for δ corresponds to positive risk premium.

Example 7 (garch-in-mean, ibm stock return): We estimate the GARCH-in-mean model for the IBM stock return and obtain

$$\begin{aligned} y_t &= \underset{(1.49)}{0.948} + \underset{(2.92)}{0.104} y_{t-1} + \underset{(0.377)}{0.0060} \sigma_t^2 + \epsilon_t \\ \sigma_t^2 &= \underset{(2.24)}{2.894} + \underset{(3.26)}{0.096} \epsilon_{t-1}^2 + \underset{(16.9)}{0.839} \sigma_{t-1}^2. \end{aligned}$$

The log-likelihood of this specification is -2900.930. The estimated GARCH-in-mean effect is very small and certainly not statistically significant. This seems to suggest that the return has not been driven by changes in the variance. ♦

6 Concluding Remarks

This note has introduced the main ideas of ARCH and GARCH models and some extensions. There is an enormous literature in this field and the number of specific ARCH-type models is exploding. Excellent reviews of the early ARCH literature are given in Bollerslev, Chou, and Kroner (1992), Bollerslev, Engle, and Nelson (1992), and Bera and Higgins (1995); and they also contain many references to specific models and extensions.

References

Bera, A. K., and M. L. Higgins (1995): "On ARCH Models: Properties, Estimation and Testing," in *Surveys in Econometrics*, chapter 8, pp. 215–272. Blackwell, Oxford.

Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31(3), 307–327.

Bollerslev, T., R. Y. Chou, and K. F. Kroner (1992): "ARCH Modelling in Finance - A Review of the Theory and Empirical Evidence," *Journal of Econometrics*, 52, 5–59.

Bollerslev, T., R. F. Engle, and D. B. Nelson (1992): "ARCH Models," in *Handbook of Econometrics, Volume IV*, ed. by R. Engle, and D. McFadden, chapter 49, pp. 2959–3038. North-Holland, Amsterdam.

Engle, R. F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50(4), 987–1008.

Engle, R. F., and V. Ng (1993): "Measuring and Testing the Impact og News on Volatility," *Journal of Finance*, 48, 1749–1777.

Glosten, L., R. Jagannathan, and D. Runkle (1993): "Relationship Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 48(5), 1779–1802.

Mandelbrot, B. (1963): "The Variation of Certain Speculative Prices," *The Journal of Business*, 36(4), 394–419.

Nelson, D. B. (1990): "Stationarity and Persistence in the GARCH(1,1) Model," *Econometric Theory*, 6(3), 318–334.

Taylor, S. J. (1986): *Modelling Financial Time Series*. Wiley, New York.