ARCH MODELS AND CONDITIONAL VOLATILITY

A drawback of linear stationary models is their failure to account for changing volatility: The width of the forecast intervals remains constant even as new data become available, unless the parameters of the model are changed. To see this for $h = 1$, write the series as

$$x_t = \sum_{k=0}^{\infty} a_k e_{t-k} \quad , \quad (a_0 = 1) \quad ,$$

where $\{e_t\}$ is independent white noise. The width of the forecast interval is proportional to the square root of the one-step forecast error variance, $var\,[x_{n+1} - f_{n,1}] = var\,[e_{n+1}] = \sigma_e^2$, a constant. On the other hand, actual financial time series often show sudden bursts of high volatility. For example, if a recent innovation was strongly negative (indicating a crash, etc.), a period of high volatility will often follow. A clear example of this is provided by the daily returns for General Motors from Sept 1’st to Nov 30’th, 1987. The volatility increases markedly after the crash, and stays high for quite some time. Nevertheless, the forecast intervals based on a given Gaussian ARMA model would have exactly the same width after the crash as before. This would destroy the validity of post-crash forecast intervals. It would have equally devastating consequences for any methodology which requires a good assessment of current volatility, such as the Black-Scholes method of options pricing. In the ARCH model, which we discuss here, the forecast intervals are able to widen immediately to account for sudden changes in volatility, without changing the parameters of the model. Because of this feature, ARCH (and other related) models have become a very important element in the analysis of economic time series.

The acronym ARCH stands for AutoRegressive Conditional Heteroscedasticity. The term "heteroscedasticity" refers to changing volatility (i.e., variance). But it is not the variance itself which changes with time according to an ARCH model; rather, it is the *conditional* variance which changes, in a specific way, depending on the available data. The conditional variance quantifies our uncertainty about the future observation, given everything we have seen so far. This is of more practical interest to the forecaster than the volatility of the series considered as a whole.

To provide a context for ARCH models, let’s first consider the conditional aspects of the linear $AR(1)$ model $x_t = a x_{t-1} + e_t$, where the $\{e_t\}$ are independent with zero mean and equal variances. If we want to predict x_t from x_{t-1}, the best predictor is the conditional mean, $E[x_t \mid x_{t-1}] = ax_{t-1}$. The success of the $AR(1)$ model for forecasting purposes arises from the fact that this conditional mean is allowed to depend on the available data, and evolve with time. The conditional variance, however, is simply $var[x_t \mid x_{t-1}] = var[e_t] = \sigma_e^2$, which remains constant regardless of the given data. The novelty of the ARCH model is that it allows the conditional variance to depend on the data.

The concept of conditional probability (and therefore conditional mean and variance) plays a key role in the construction of forecast intervals. It could be argued that a reasonable definition of a 95% forecast interval is one for which the future observation has a 95% probability of falling in the interval, **conditionally on the observed data.** That is, of all possible realizations (paths) which coincide with the data available up to now, 95% of them should have a future value x_{n+h} which lies within the forecast interval. For the purposes of forecasting, we do not care too much about realizations which are not consistent with the available information. So we do not demand that x_{n+h} must lie within this particular prediction interval in 95% of all *possible* realizations. The ARMA forecast intervals we derived in an earlier handout have conditional coverage rate of $1-\alpha$, if the model is correct and assuming that that the innovations are normally distributed. However, these intervals are not valid if the data are actually generated by an ARCH model.

Definition of ARCH Model

The $ARCH(q)$ model for the series $\{\varepsilon_t\}$ is defined by specifying the conditional distribution of ε_t, given the information available up to time $t-1$. Let ψ_{t-1} denote this information. It consists of the knowledge of all available values of the series, and anything which can be computed from these values, e.g., innovations, squared observations, etc. In principle, it may can even include the knowledge of the values of other related time series, and anything else which might be useful for forecasting and is available by time $t-1$.

We say that the process $\{\varepsilon_t\}$ is $ARCH(q)$ if the conditional distribution of ε_t given the available information ψ_{t-1} is

$$\varepsilon_t \mid \psi_{t-1} \sim N(0, h_t) \ , \tag{1}$$

$$h_t = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \ , \tag{2}$$

with $\omega > 0$, $\alpha_i \geq 0$ for all i, and $\sum_{i=1}^{q} \alpha_i < 1$.

Equation (1) says that the *conditional distribution* of ε_t *given* ψ_{t-1} is normal, $N(0, h_t)$. In other words, given the available information ψ_{t-1} the next observation ε_t has a normal distribution with a (conditional) mean of $E[\varepsilon_t \mid \psi_{t-1}] = 0$, and a (conditional) variance of $var[\varepsilon_t \mid \psi_{t-1}] = h_t$. We can think of these as the mean and variance of of ε_t computed over all paths which agree with ψ_{t-1}.

Equation (2) specifies the way in which the conditional variance h_t is determined by the available information. Note that h_t is defined in terms of *squares* of past innovations. This, together with the assumptions that $\omega > 0$ and $\alpha_i \geq 0$, guarantees that h_t is positive, as it must be since it is a conditional variance.

Properties of ARCH Models

Perhaps it is surprising that if, instead of restricting to paths which agree with the available information ψ_{t-1} we consider all *possible* paths, **the $\{\varepsilon_t\}$ are a zero mean white noise process.** That is, unconditionally, considering all possible paths, we have $E[\varepsilon_t] = 0$, $var[\varepsilon_t] = \omega/(1 - \sum_{i=1}^{q} \alpha_i)$, a finite constant, and $cov(\varepsilon_i, \varepsilon_j) = 0$ if $i \neq j$. So unconditionally, the process is stationary as long as $\sum_{i=1}^{q} \alpha_i < 1$, which we assumed in the definition of the model. It is only the conditional volatility which changes with time, not the overall volatility.

In spite of its name, **the ARCH model is not autoregressive.** However, if we add $\eta_t = \varepsilon_t^2 - h_t$ to both sides of Equation (2), we get

$$\varepsilon_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \eta_t \ .$$

It can be shown that $\{\eta_t\}$ is zero mean white noise. Therefore, **the squared process $\{\varepsilon_t^2\}$ is an autoregression** $[AR(q)]$ with nonzero mean, and AR parameters $\alpha_1, \ldots, \alpha_q$.

The ARCH(q) model is nonlinear, since if the ε_t could be expressed as $\varepsilon_t = \sum_{k=0}^{\infty} a_k e_{t-k}$, then we would have $var[\varepsilon_t \mid \psi_{t-1}] = var[\varepsilon_t \mid e_{t-1}, e_{t-2}, \cdots] = var[e_t]$, a constant. This contradicts Equations (1) and (2), so $\{\varepsilon_t\}$ must not be a linear process.

The observations $\{\varepsilon_t\}$ of an ARCH(q) model are non Gaussian, since the model is nonlinear. The distribution of the $\{\varepsilon_t\}$ tends to be more long-tailed than normal. Thus, outliers may occur relatively often. This is a useful feature of the model, since it reflects the leptokurtosis which is often observed in practice. Moreover, once an outlier does occur, it will increase the conditional volatility for some time to come. (See Equation (2).) Once again, this reflects a pattern often found in real data. It may seem odd that, while the conditional distribution of ε_t given ψ_{t-1} *is* Gaussian, the unconditional distribution is not. Roughly speaking, the reason for this is that the unconditional distribution is an average of the conditional distributions for each possible path up to time $t-1$. Although each of these conditional distributions is Gaussian, the variances h_t are unequal. So we get a *mixture* of normal distributions with unequal variances, which is not Gaussian.

Although they are uncorrelated, **the $\{\varepsilon_t\}$ are not independent.** This is easy to see, since if $\{\varepsilon_t\}$ were independent, then they would be a linear process; but we have already shown that the $ARCH(q)$ process is in fact nonlinear. One might hope, then, that as in the case of threshold AR and bilinear models, it might be possible to find some nonlinear predictor of ε_t based on ψ_{t-1}. But since $E[\varepsilon_t \mid \psi_{t-1}] = 0$, we see that **the $\{\varepsilon_t\}$ are a Martingale difference.** Thus, the very best (linear or nonlinear) predictor of ε_t based on the available information is simply the trivial predictor, namely the series mean, 0. In terms of point forecasting of the series itself, then, the ARCH models offer no advantages over the linear ARMA models. The advantage of the ARCH models lies in their ability to describe the time-varying stochastic conditional volatility, which can then be used to improve the reliability of interval forecasts and to help us in understanding the process.

Although the series $\{\varepsilon_t\}$ itself is not forecastable, the squared series $\{\varepsilon_t^2\}$ is forecastable: The best forecast of ε_t^2 is

$$E[\varepsilon_t^2 \mid \psi_{t-1}] = var[\varepsilon_t \mid \psi_{t-1}] = h_t = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \quad .$$

Note that we would get the same forecast by using the $AR(q)$ representation for the squared process $\{\varepsilon_t^2\}$ discussed earlier.

ARMA Models With ARCH Errors

Because it is Martingale difference and therefore unpredictable, the $ARCH(q)$ model is not usually used by itself to describe a time series data set. Instead, we can model our data $\{x_t\}$ as an $ARMA(k, l)$ process where the *innovations* $\{\varepsilon_t\}$ are $ARCH(q)$. That is,

$$x_t = \sum_{j=1}^{k} a_j x_{t-j} + \sum_{j=1}^{l} b_j \varepsilon_{t-j} + \varepsilon_t \quad , \tag{3}$$

where $\{\varepsilon_t\}$ is $ARCH(q)$. Because $\{\varepsilon_t\}$ is white noise, the model (3) is in many respects just an ordinary ARMA model. In fact, it *is* just an ARMA model, since our original definition of *ARMA* models simply required that the innovations be white noise. So the ACF and PACF of the $\{x_t\}$ given by model (3) are the same as usual, i.e., the ACF will cut off beyond lag l if $k = 0$, and the PACF will cut off beyond lag k if $l = 0$, etc. In addition, the best linear forecast of x_{n+h} based on $x_n, x_{n-1}, \cdots$ will be no different than what we have studied earlier, once again because $\{\varepsilon_t\}$ is a white noise process. In fact, because the $\{\varepsilon_t\}$ are a Martingale difference, it is not hard to show that the best *possible* predictor of x_{n+h} is simply the best *linear* predictor. So much of the theory and practice of Box-Jenkins modeling still applies to Model (3). Perhaps the most notable exception to this is forecast intervals. These were derived (e.g., in our earlier handout) under the additional assumption that the innovations are Gaussian. Thus, these intervals will not be valid for model (3), in which the innovations are non Gaussian.

One-Step Forecast Intervals for ARMA Models with ARCH Errors

Suppose $\{x_t\}$ is $ARMA(k, l)$ with $ARCH(q)$ errors, $\{\varepsilon_t\}$. Then

$$x_{n+1} = \sum_{j=1}^{k} a_j x_{n+1-j} + \sum_{j=1}^{l} b_j \varepsilon_{n+1-j} + \varepsilon_{n+1} \quad ,$$

where $\{\varepsilon_t\}$ is $ARCH(q)$. Assume that all parameter values and model orders are known. The best forecast of x_{n+1} based on ψ_n is

$$f_{n,1} = E[x_{n+1} | \psi_n] = \sum_{j=1}^{k} a_j x_{n+1-j} + \sum_{j=1}^{l} b_j \varepsilon_{n+1-j} \quad ,$$

since $E[\varepsilon_{n+1} | \psi_n] = 0$. The corresponding one-step forecast error is $x_{n+1} - f_{n,1} = \varepsilon_{n+1}$. We would like to construct a forecast interval for x_{n+1} which has a *conditional* coverage rate of $1-\alpha$, given the observed information ψ_n. To do this, we use our knowledge about the properties of the forecast error, given ψ_n. Since $\{\varepsilon_t\}$ is an ARCH process, we know from Equation (1) that, conditionally on ψ_n, the forecast error ε_{n+1} is distributed as $N(0, h_{n+1})$. Thus,

$$1-\alpha = prob\{(-z_{\alpha/2}\sqrt{h_{n+1}} < \varepsilon_{n+1} < z_{\alpha/2}\sqrt{h_{n+1}}) | \psi_n\} \quad .$$

Adding $f_{n,1}$ to all sides of the inequality above, we conclude that

$$1-\alpha = prob\{(f_{n,1} - z_{\alpha/2}\sqrt{h_{n+1}} < x_{n+1} < f_{n,1} + z_{\alpha/2}\sqrt{h_{n+1}}) | \psi_n\} \quad .$$

Thus, we have just shown that a one-step forecast interval, with a conditional coverage rate of $1-\alpha$, is given by

$$f_{n,1} \pm z_{\alpha/2}\sqrt{h_{n+1}} \quad .$$

The interval can actually be constructed since h_{n+1} depends only on the information available at time n. The width of the one-step forecast intervals is $2z_{\alpha/2}\sqrt{h_{n+1}}$. This width will increase immediately in the event of a sudden large fluctuation in the series. (See Equation (2)).

Simulation of ARCH Models

The definition of an ARCH model (Equations (1) and (2)) is stated in terms of the conditional distribution of ε_t given ψ_{t-1}. This does not tell us explicitly how to generate an ARCH process. Here is one way to actually construct an ARCH process in terms of independent standard normals, $\{e_t\}$: First, set $\varepsilon_t = e_t$ for $t = -q+1, \ldots, 0$. These are just "start-up" values which allow us to compute $h_0, \ldots, h_q$. Next, for $t = 1, \ldots, n$, simply set $\varepsilon_t = e_t\sqrt{h_t}$ where h_t is given by equation (2). Then for $t = 1$ to n, if we condition on ψ_{t-1}, then h_t can be treated as a constant, so that ε_t has a normal distribution with mean 0 and variance h_t. Thus, $\{\varepsilon_t\}_{t=1}^{n}$ is indeed $ARCH(q)$.

Using this method, I simulated an ARCH(1) data set with $n = 100$, $\omega = .25$, $\alpha = .5$, and an ARCH(2) data set with $n = 100$, $\omega = .25$, $\alpha_1 = .6$, $\alpha_2 = .35$. For the ARCH(1) data, the volatility increases strongly in the second half of the series. Integrating this series (i.e., taking partial sums) gives a "random walk" with ARCH errors. The integrated series has a "crash" around $t = 50$. For the ARCH(2) and there are two positive outliers at $t = 61$ and $t = 63$, and two more at $t = 80$ and $t = 81$. These cause sudden strong upswings in the integrated series.

Graphical Methods For Selecting q in $ARCH(q)$ Models.

Because the $ARCH(q)$ process ε_t is white noise, we would typically expect to find that the *ACF* and *PACF* of the raw data are not significant at most lags. Thus, they are not useful for selecting a value of the unknown order q of an *ARCH* process. Examination of the *ACF* and *PACF* of the *squared* series, $\{\varepsilon_t^2\}$ will often be more fruitful. Since, as we have seen, the squared series obeys an $AR(q)$ model, we would hope to find that the *PACF* of the squared series cuts off beyond lag q, while the *ACF* of the squared series "dies down".

For the simulated $ARCH(1)$ series, the *ACF* and *PACF* for the raw data are nonsignificant at all lags, the *ACF* of the squares dies down, and the *PACF* of the squares is highly significant at lag 1, and just barely significant at lag 2. This suggests an $ARCH(1)$, or perhaps an $ARCH(2)$.

For the simulated $ARCH(2)$ series, the ACF of the raw data is almost significant at lag 1, barely significant at lag 4, but nonsignificant at all other low lags. As we will explain, it is sensible to ignore the (marginally) significant lags of the ACF of the raw data, and declare the series a white noise, though perhaps not *strict* white noise. The *PACF* of the squares exhibits a clear cutoff beyond lag 1. Thus, using the graphical method we would (incorrectly) identify the process as $ARCH(1)$.

Since any *ARCH* process is also a white noise process, its true *ACF* and *PACF* must be zero at all lags. Of course, the *sample ACF* and *PACF* will all differ from zero, due to sampling variation. We have been assuming throughout the course that for any white noise process, the sample *ACF* and *PACF* will have a standard error of $1/\sqrt{n}$. Actually, this is not necessarily true unless we have *strict* (independent) white noise. For $ARCH(1)$ processes, it can be shown that the standard error of the sample ACF at lag k is

$$var\,[r_k] \approx \frac{1}{n}[1+2\,\alpha_1^k/(1-3\alpha_1^2)] \ .$$

Thus, the variance of r_k will be *greater* than $1/n$ in this case. For example, if $\alpha_1 = .4$, then the variance of r_1 is approximately $2.5/n$, compared to $1/n$ if no ARCH effects are present. So if our data were generated by this $ARCH(1)$ model, we would find that r_k is significantly different from zero much more than 5% of the time, even though the series is white noise. It is clear, then, that using the ACF and $PACF$ of the raw data to decide whether the series is white noise remains more of an art than a science. On the other hand, using the ACF and $PACF$ of the *squares* to decide whether the raw series is *independent* is a justifiable procedure. This is so because if $\{\varepsilon_t\}$ is strict white noise and n is large, the sample ACF and $PACF$ values for the squared series are (essentially) normally distributed with zero mean and variance $1/n$. Thus any significant lags of the ACF or $PACF$ of the squares can be taken as evidence that the raw series is not independent.

Instead of using graphical methods, which may be problematic, it is possible to use AIC_C to select q automatically for the $ARCH(q)$ model, but this requires that we estimate the ARCH parameters themselves.