**See ARCH-README-FORECASTING.docx, ARCH-ForecastInvestigate-Compare.R, arch1-ForecastInvestigate-SAVE.R**

**ARCH Models - Introduction**

Nobel Prize in Economic Sciences:

<https://www.nobelprize.org/prizes/economic-sciences/2003/summary>

ARCH models are time series models for hetereoscedastic error terms, i.e., models for when  depends on t. In this section,  will be denoted by 

ARCH stands for autoregressive conditionally hetereoscedastic. The generalization of these models to be discussed later are GARCH models with the G meaning “generalized”.

When to use this models?

Sometimes in a time series plot, the variation is small, then large for a small period of time, then the variation goes back to being small again. Examples of where this happens:

1) stock or bond prices

2) money exchange rates

In the ARIMA modeling framework, what is usually done is the following:

* Suppose xt denotes the series that appears to have non-constant variance.
* A common method to handle this is to use the natural log of the series, log(xt).
* This series will often appear to be nonstationary in the mean, so a solution to that problem is first differences: yt = log(xt) – log(xt-1). Notice that yt has a mean of 0.
* This yt series ACF and PACF will often look like the ACF and PACF from a white noise process leaving the model to be ARIMA(0,1,0) for yt! This corresponds to the “efficient market hypothesis” discussed in finance that stock prices follow a random walk.

A closer look at yt reveals the following,

yt = log(xt) – log(xt-1)

= log(xt/xt-1)

= log(current value / past value)

This is somewhat similar to a “return” for an investment, defined as

rt = (xt - xt-1)/xt-1

For example, if a stock price at the end of trading the previous day is 50 and at the end of today it was 60, the return would be:

rt = (60 - 50)/50 = 0.2

Thus, the stock went up 20%.

Other types of structures are often still present in yt which would lead to an additional model:

1. The distribution of yt has “heavier” tails than a normal distribution (remember the relationship between a t-distribution with say 5 degrees of freedom and a standard normal distribution).
2. The  are correlated and often the correlation is non-negative.
3. The changes in yt tend to be clustered. Because of this clustering, one could say there is dependence in the variability or “volatility” of observed values.

Example: Monthly returns of value-weighted S&P 500 Index from 1926 to 1991 (SP500.R, sp500.txt)

This data set is taken from Pena, Tiao, and Tsay’s textbook. The data set is already in the returns rt format. Notice the changes in the variability (and their corresponding dates).

> sp500 <- read.table(file = "sp500.txt",

 header = FALSE, col.names = "x", sep = "")

> head(sp500)

 x

1 0.0225

2 -0.0440

3 -0.0591

4 0.0227

5 0.0077

6 0.0432

> x <- ts(data = sp500$x, start = 1926, deltat = 1/12)

> x

 Jan Feb Mar Apr May Jun

1926 0.0225 -0.0440 -0.0591 0.0227 0.0077 0.0432

1927 -0.0208 0.0477 0.0065 0.0172 0.0522 -0.0094

 Jul Aug Sep Oct Nov Dec

1926 0.0455 0.0171 0.0229 -0.0313 0.0223 0.0166

1927 0.0650 0.0445 0.0432 -0.0531 0.0678 0.0190

<EDITED>

> plot(x = x, ylab = expression(x[t]), xlab = "t", type =

 "l", col = "red",main = "S&P 500 data series")

> points(x = x, pch = 20, col = "blue")

> abline(v = c(1930, 1940, 1950, 1960, 1970, 1980, 1990),

 lty = "dotted", col = "gray")

> abline(h = c(-0.2, 0, 0.2, 0.4), lty = "dotted", col =

 "gray")



Example: Simulated ARCH(2) data with α0 = 0.1, α1 = 0.5, α2 = 0.2, n = 10,000 (arch2.R)

Data is simulated from the model above. Don’t worry about the name yet! For now, just understand that this series is example of where there is dependence in the variability. See the program for the code (after we do a model fitting example).

Suppose the series is denoted by xt.



Around time ≈ 2000 and various other places, there is more variability than which appears in the rest of the series.



According to the ACF and PACF of xt, the data appear to be white noise! However, take a look at the ACF and PACF of . It looks like an AR(3) model would be appropriate to model ! (Note: Actually it should have shown up to be AR(2))

ARCH(1) model

Understanding the variability is important in finance because investors expect higher returns as compensation for higher degrees of volatility (think of this as risk).

Consider a model that allows dependence in the variances of yt, denoted by . Below is the ARCH(1) model:

yt = σtεt where  = α0 + α1 and εt ~ independent (0,1)

After recording the video: The reason for E(yt) = 0 is a little more complicated than what was described at this point in the video. The notes and video will examine the derivation later.

Notes:

* yt has a mean of 0. Thus, the time series of interest could be xt with yt = xt - μ. Alternatively, possible values of yt are yt = (1-B)xt, yt = (1-B)log(xt), or yt = (1-B)xt / xt = rt.
* Relate this model to what we would have for an ARMA(0,0) model for yt before: yt = wt where wt ~ ind. (0,σ2). We could have instead used yt = σwt where wt ~ ind. (0,1).
* We could even state the model as

xt = μ + wt

where wt ~ ind. (0, ),  = α0 + α1, and yt = xt - μ.

* Because the variance changes, this is where the H part of the ARCH name comes about.
* For now, εt will be taken as having a normal distribution.
* The α0 and α1 parameters have constraints on possible values they can take on so that  > 0. For example, α1 > 0 to make sure  > 0 for any α0. More specific constraints are given later.
* One can think of this as yt is white noise with variance depending on the past.
* Also, one may need to find an ARMA model for the original series xt itself if it has autocorrelation in it. We can then work with the residuals as yt to find an ARCH model. We will discuss finding an ARMA model and ARCH model together later in the notes.
* You could rewrite the model as 
* Conditional on yt-1, yt has a normal distribution, i.e.,

yt|yt-1 ~ N(0, α0+α1)

Thus, if you know the value of yt-1, yt has a normal distribution with mean 0 and variance α0 + α1.

* The above representation means that Var(ytIyt-1) =  = α0 + α1. Also, because E(yt|yt-1) = 0,



So,  = α0 + α1. Thus, the conditional variance of yt comes about through previous values of  like an AR(1) model! This is where the AR and C parts of the ARCH name come about!

One could also express the model as

=  +  - 

= (α0+α1) + (σtεt)2 - 

= α0+α1 + (-1)

= α0+α1 + νt where νt = (-1)

Again, you can think of the ARCH(1) model as an AR(1) model for  with νt as the error term. This error term is different from the usual error term found in an AR(1) model.

* The “unconditional” mean of yt, E(yt), is found using the following property:

Let A and B be two random variables. Then EA(A), the expected value of A using the marginal distribution of A, can be found using EB[EA|B(A|B)] where the second expectation is with respect to the conditional distribution of A given B is known. More generally,

EA[g(A)] = EB{EA|B[g(A)|B]}

where g() is a function of A. For a proof, see Casella and Berger’s textbook.

Using this result,



* One way to find the “unconditional” variance of yt is the following:





where the last step is the result of  must be constant through time since  = α0+α1 + νt is a causal AR(1) process when 0 ≤ α1 < 1 (not future dependent).

* The above representation of the variance leads to the following:

 = α0/(1 - α1)

because

 

* The kurtosis for a random variable A is defined to be:

= 

It measures the peakedness or flatness of a probability distribution function. The larger the value, the more spread out the distribution is. For a standard normal distribution, this becomes

 = 3

One can use the moment generating function of a standard normal distribution to figure this out.

The fourth moment for yt can be shown to be  provided that  < 1 because the fourth moment must be positive (a random variable to the 4th power is always positive). Because we already have the constraint of 0 ≤ α1 < 1, this means 0 ≤ α1 < .

The kurtosis for yt then becomes  > 3. Thus, the distribution of yt will always have “fatter” tails than a standard normal. Using the “usual” method of finding outliers will find more of them. Pena, Tiao, and Tsay’s textbook says “this is in agreement with the empirical finding that outliers appear more often in asset returns than that implied by an iid sequence of normal random variates.”

* Because we have an AR(1) structure for , then the autocorrelation between  and  is  ≥ 0. The autocorrelation is always greater than 0 because of the constraints on α1. We can look for this behavior in an ACF plot! This positive autocorrelation allows us to model the phenomenon that changes in yt tend to be clustered (i.e., volatile yt values lead to more volatile values – think in terms of a stock return).

Parameter estimation

The likelihood can be written out using yt|yt-1 ~ independent N(0, α0+α1). See Shumway and Stoffer’s textbook’s for more on this likelihood function and notice how the first observation is conditioned upon. This function involves writing out a joint probability distribution as a product of conditional probability distributions and one marginal probability distribution.

Given this likelihood, “conditional” maximum likelihood estimation can proceed in a similar manner as maximum likelihood estimation has done before. Standard likelihood methods can be used to find the covariance matrix for parameter estimates.

ARCH(m) model

yt = σtεt where  = α0 + α1 + α2 +…+ αm and εt ~ independent N(0,1)

Conditions on the parameters are αi ≥ 0 for all i = 1, …, m and α1 + + αm < 1.

Weaknesses of ARCH models

1. The model treats positive and negative returns (yt) in the same manner.
2. The model is restrictive with regard to what values the αi can take on – see the ARCH(1) example.
3. The model does not provide new insight to understanding financial time series; only a mechanical way to describe the variance.
4. The model often over predicts the volatility because it responds slowly to isolated large shocks to the return series.

GARCH(m,r)

The “G” stands for “Generalized”

yt = σtεt where

 = α0 + α1 + … + αm +  + … + 

and εt ~ independent N(0,1)

The additional parameters help incorporate past variances. More will be discussed about this model later.