**ARCH Models – ARIMA**

We previously defined an ARCH(m) model as

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.

We could also incorporate ARIMA models with this too. For example, we could have

ϕ(B)(1-B)dxt = θ(B)wt

where wt ~ ind. (0, ), , and yt represents the residuals from the ARIMA model.

Thus, non-stationarity in the variance can be taken care of by using ARCH models with an ARIMA model.

Model Building

1. Build an ARIMA model for the observed time series to remove any autocorrelation in the data. Refer to the residuals as yt.
2. Examine the squared series, , to check for heteroscedasticity. This can be done by doing an ACF and PACF plot of the  values. Remember, we are constructing an AR-like model for . What would you expect the PACF to show if an ARCH model is needed? The Ljung-Box-Pierce test can also be performed on the  values as well.
3. Decide on the order of the ARCH model for  and perform maximum likelihood estimation of all parameters.

Example: U.S. GNP (GNP.R)

Shumway and Stoffer used ARIMA and ARCH models to examine a U.S. GNP data set with quarterly given values. An AR(1) model to the first differenced, log-transformed data was recommended originally by the authors prior to ARCH models being introduced.

We will estimate the AR(1) and ARCH model all at once using the garchFit() and ugarchfit() functions.

Let xt = GNP at time t for this problem.

> library(package = astsa)

> head(gnp)

 Qtr1 Qtr2 Qtr3 Qtr4

1947 1488.9 1496.9 1500.5 1524.3

1948 1546.6 1571.1

> tail(gnp)

 Qtr1 Qtr2 Qtr3 Qtr4

2001 9224.3 9199.8 9283.5

2002 9367.5 9376.7 9477.9

> x <- gnp

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

 "l", col = "red", main = "GNP data")

> grid(col = "gray", lty = "dotted")



There is non-stationarity with respect to the mean. Below are the first differences.

> plot(x = diff(x = x, lag = 1, differences = 1), ylab =

 expression(x[t] - x[t-1]), xlab = "t", type = "l", col

 = "red", main = "First differences of GNP data")

> grid(col = "gray", lty = "dotted")



Shumway and Stoffer also work with the log transformation and make this transformation prior to the first differences. This corresponds to our earlier discussion of

yt = log(xt) – log (xt-1)

= log(xt/xt-1)

= log(current value / past value)

being close to a “return” in an investment (although GNP is not an investment). Because I would like to replicate their example, I chose to do the same here for the remainder of this example.

Below is the estimate of the an AR(1) model along with the usual examinations of model fit.

> gnpgr <- diff(x = log(gnp$x), lag = 1, differences = 1)

> mod.fit.ar <- arima(x = gnpgr, order = c(1, 0, 0),

 include.mean = TRUE)

> mod.fit.ar

Call:

arima(x = gnpgr, order = c(1, 0, 0), include.mean = TRUE)

Coefficients:

 ar1 intercept

 0.3467 0.0083

s.e. 0.0627 0.0010

sigma^2 estimated as 9.03e-05: log likelihood = 718.61, aic = -1431.22

> source(file = "examine.mod.R")

> examine.mod(mod.fit.obj = mod.fit.ar, mod.name =

 "ARIMA(1,1,0)")

$z

 ar1 intercept

 5.525479 8.539823

$p.value

 ar1 intercept

3.285877e-08 0.000000e+00





The estimated ARIMA model for the series is

(1 – 0.3467B)(1 – B)log(xt) = 0.0083 + wt

After recording the video: The constant term in the model was changed from 0.0083(1 – 0.3467) to 0.0083 because this represents the drift term.

Notice the ACF and PACF plots of the residuals look like the corresponding plots for white noise. Also, notice the normal Q-Q shows the “fat tails” of the residual distribution.

The same model could be estimated using

 mod.fit.ar <- arima(x = log(x), order = c(1, 1, 0), xreg

 = 1:length(x))

However, we will not be able to include a value for d in garchFit() or ugarchfit() when we include the ARCH model component. So, the code used here allows us to see the model fitting process without d.

Let yt denote the residuals from the ARIMA model’s fit.

> #Examine ACF and PACF of the squared residuals

> y <- as.numeric(mod.fit.ar$residuals) #Without the

 #as.numeric() the values are not spaced correctly

 #on the plots

> par(mfrow = c(1,2))

> acf(x = y^2, type = "correlation", lag.max = 20, xlim =

 c(1,20), ylim = c(-1,1), xlab = "h", main =

 expression(paste("Estimated ACF for ", y[t]^2)))

> pacf(x = y^2, lag.max = 20, ylim = c(-1,1), xlim =

 c(1,20), xlab = "h", main = expression(paste(

 "Estimated PACF for ", y[t]^2)))

> par(mfrow = c(1,1))



There are only marginally significant values in the ACF and PACF for the squared residuals. Shumway and Stoffer say, “it appears there may be some dependence, albeit small, left in the residuals.” I say, “Maybe… “. Following their example, we can fit an ARMA(1,0) AND ARCH(1) model simultaneously using the garchFit() function of the fGarch package. Below is the code and output from garchFit():

> library(fGarch)

> mod.fit <- garchFit(formula = ~ arma(1,0) + garch(1, 0),

 data = gnpgr)

Series Initialization:

 ARMA Model: arma

 Formula Mean: ~ arma(1, 0)

 GARCH Model: garch

 Formula Variance: ~ garch(1, 0)

<EDITED>

> summary(mod.fit)

Title:

 GARCH Modelling

Call:

 garchFit(formula = ~arma(1, 0) + garch(1, 0), data =

 gnpgr)

Mean and Variance Equation:

 data ~ arma(1, 0) + garch(1, 0)

<environment: 0x000000003c53fea8>

 [data = gnpgr]

Conditional Distribution:

 norm

Coefficient(s):

 mu ar1 omega alpha1

0.00527795 0.36656255 0.00007331 0.19447134

Std. Errors:

 based on Hessian

Error Analysis:

 Estimate Std. Error t value Pr(>|t|)

mu 5.278e-03 8.996e-04 5.867 4.44e-09 \*\*\*

ar1 3.666e-01 7.514e-02 4.878 1.07e-06 \*\*\*

omega 7.331e-05 9.011e-06 8.135 4.44e-16 \*\*\*

alpha1 1.945e-01 9.554e-02 2.035 0.0418 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log Likelihood:

 722.2849 normalized: 3.253536

Standardised Residuals Tests:

 Statistic p-Value

 Jarque-Bera Test R Chi^2 9.118036 0.01047234

 Shapiro-Wilk Test R W 0.9842406 0.01433658

 Ljung-Box Test R Q(10) 9.874326 0.4515875

 Ljung-Box Test R Q(15) 17.55855 0.2865844

 Ljung-Box Test R Q(20) 23.41363 0.2689437

 Ljung-Box Test R^2 Q(10) 19.2821 0.03682246

 Ljung-Box Test R^2 Q(15) 33.23648 0.004352736

 Ljung-Box Test R^2 Q(20) 37.74259 0.00951899

 LM Arch Test R TR^2 25.41625 0.01296901

Information Criterion Statistics:

 AIC BIC SIC HQIC

-6.471035 -6.409726 -6.471669 -6.446282

> par(mfrow = c(1,1))

> plot(mod.fit, which = 13)



Notes:

* The estimated model is

(1 – 0.3666B)(1 – B)log(xt) = 0.0053 + wt

After recording the video: The constant term in the model was changed from 0.0053(1 – 0.3666) to 0.0053 because this represents the drift term.

with



* The hypothesis test for α1 = 0 vs. ≠ 0 has a p-value of 0.0418 indicating that there is marginal evidence that α1 ≠ 0.
* What do the Jarque Bera Test and the Ljung-Box test suggest about the model? The QQ-plot for the ARCH model residuals suggests a similar conclusion. Shumway and Stoffer mention the problems, but do not explore any resolutions.
* Why doesn’t garchFit() or ugarchfit() allow for a value of d? The reason may be due to an investment return being more important that the actual value of an investment on a per share/unit basis. Of course, GNP for this example does not correspond to an investment.