ARCH, GARCH, and IGARCH for Unequal Variances

Engle(1982) introduced a model in which the variance at time t is modeled as a linear combination of past squared residuals and called it an ARCH (autoregressive conditionally heteroscedastic) process. Bolerslev (1986) introduced a more general structure in which the variance model looks more like an ARMA than an AR and called this a GARCH (generalized ARCH) process. These proceedures model an error term ϵ_t in terms of a standard white noise $e_t \sim N(0,1)$ as $\epsilon_t = \sqrt{h_t}\, e_t$ where $h_t$satisfies the type of recursion used in an ARMA model . For the original ARCH approach we have

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

Bolerslev added lags of h to the model to produce GARCH(p,q) models where $h_t$satisfies the type of recursion used in a more general ARMA model

$$h_t = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \gamma_j \, h_{t-j}$$

In this way, the error term has a conditional variance that is a function of the magnitudes of past errors ϵ_{t-i}. Engle's original ARCH structure has γ_j=0. Because h_t is the variance rather than its logarithm, certain restrictions must be placed on the α_i, γ_j, and ω to ensure positive variances. For example if these are all restricted to be positive then positive initial values of h_t will ensure all h_t are positive. For this reason, Nelson (1991) suggested replacing h_t by $\log(h_t)$ and an additional modification, with the resulting process being called EGARCH. These approaches allow the standard deviation to change with each observation. Nelson and Cao give constraints on the α and γ values that ensure nonnegative estimates of h_t. These are the default in PROC AUTOREG. More details are given in the PROC AUTOREG documentation and in Hamilton (1994) which is a quite detailed reference for time series.

Recall that SAS PROC AUTOREG will fit a regression model with autoregressive errors using the maximum likelihood method based on a normal distribution. In place of the white noise shocks in the autoregressive error model you can specify a GARCH(p,q) process. If it appears, as suggested by your analysis of the Dow standard deviations, that the process describing the error variances is a unit root process, then the resulting model is referred to as integrated GARCH or IGARCH. If the usual stationarity conditions are satisfied, for a GARCH process, forecasts of h_t will revert to a long run mean. In an IGARCH model, mean reversion is no longer a property of h_t so that forecasts of h_t will tend to reflect the most recent variation rather than the average historic variation. You would expect the variation during the depression to have little effect on future h_t values in an IGARCH model for the Dow Jones industrial average, for example.

We can look at a long series of historic Dow Jones industrial average closing prices.

demo: LongDow.sas

Notice that the overnight percent change in price P_t from yesterday to today would be P_t/P_{t-1}, and if this is 1.04 then you have a 4% overnight gain or "return". FOr number near 1, like 1.04, the logarithms are very close to the ratio minus 1 (ln(1.04) is approximately 0.04) and the log of the price ratio is often called the return. The dataset MORE is created with the Dow Jones data and future dates appended with missin values of DDOW. Here DDOW is the difference of log transformed prices, that is, DDOW is a column of returns.

A normallity test is provided, as the theory rests heavily on the assumption of normality. The normality test used here is the that of Jarque and Bera (1982). This is a general test of normality based on a measurement of skewness b_1 and one of kurtosis b_2-3 using residuals r_t where

$$b_1 = \frac{\sum_{t=1}^{n} r_t^3/n}{(\sum_{t=1}^{n} r_t^2/n)^{3/2}} \qquad \text{and} \qquad b_2-3 = \frac{\sum_{t=1}^{n} r_t^4/n}{(\sum_{t=1}^{n} r_t^2/n)^2} - 3$$

The expression $\sum_{t=1}^{n} r_t^j/n$ is sometimes called the (raw) j^{th} moment of r. The fractions involve third and fourth moments scaled by the sample variance. The numerators are sums of approximately independent terms and thus satisfy a central limit theorem. Both have, approximately, mean 0 when the true errors are normally distributed. Approximate variances of the skewness and kurtosis are 6/n and 24/n. Odd and even powers of normal errors are uncorrelated, so squaring each of these approximately normal variates and dividing by its variance produces a pair of squares of approximately independent N(0,1) variates. The sum of these squared variates, therefore, follows a Chi-square distribution with 2 degrees of freedom under the normality null hypothesis. The Jarque-Bera test

$$JB = N(b_1^2/6 + (b_2-3)^2/24)$$

has a Chi-square distribution with 2 degrees of freedom under the null hypothesis.

Example: Dow Jones

To investigate models for the daily percentage change in the Dow Jones Industrial Average Y_t you will use $D_t=\log(Y_t)-\log(Y_{t-1})$. Calling this variable DDOW, you issue this code:

```
proc autoreg data=MORE;
 model ddow = / nlag=2
 garch=(p=2,q=1,type=integ,noint);
 output out=out2 ht=ht PREDICTED=f LCLI=l UCLI=u;
 run;
```

PROC AUTOREG allows the use of regression inputs, however here there is no apparent time trend or seasonality and no other regressors are readily available. The model statement ddow = (with no inputs) specifies that the regression part of your model is only a mean. Note the way in which the h_t sequence, predicted values and default upper and lower forecast limits have been requested in the data set called out2.

In the output, the estimate of the mean is seen to be 0.000363. Since ddow is a difference, a mean is interpreted as a drift in the data and since the data are log differences, the number $e^{0.000363} = 1.0003631$ is an estimate of the long run daily growth over this time period. With 8892 days in the study period, the number $e^{(0.000363)(8891)} = 25$ represents a 25 fold increase, roughly an 11.3% yearly growth rate! This is not remotely like the rate of growth seen, except in certain portions of the graph. PROC AUTOREG starts with OLS estimates so that the average ddow over the period is the OLS intercept 0.0000702 from the output. This gives $e^{(0.0000702)(8891)} = 1.87$ indicating 87% growth for the full 30 year period, was realized. This has to be more in line with the graph because, as you saw earlier, except for rounding error it is Y_n/Y_1.

```
                     The AUTOREG Procedure
                    Dependent Variable    ddow

                 Ordinary Least Squares Estimates

   SSE                     1.542981    DFE                    8891
   MSE                    0.0001735    Root MSE            0.01317
   SBC                   -51754.031    AIC              -51761.124
   Regress R-Square          0.0000    Total R-Square       0.0000
   Durbin-Watson             1.9427


                                     Standard                Approx
Variable        DF     Estimate         Error    t Value    Pr > |t|

Intercept        1    0.0000702      0.000140       0.50      0.6155
```

Estimates of Autocorrelations

Lag	Covariance	Correlation
0	0.000174	1.000000
1	4.956E-6	0.028561
2	-6.3E-6	-0.036278

Estimates of Autocorrelations

```
Lag    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

  0    |                    |********************|
  1    |                    |*                   |
  2    |                   *|                    |
```

Preliminary MSE 0.000173

Estimates of Autoregressive Parameters

Lag	Coefficient	Standard Error	t Value
1	-0.029621	0.010599	-2.79
2	0.037124	0.010599	3.50

Algorithm converged.

Integrated GARCH Estimates

SSE	1.54537859	Observations	8892
MSE	0.0001738	Uncond Var	.
Log Likelihood	28466.2335	Total R-Square	.
SBC	-56887.003	AIC	-56922.467
Normality Test	3886.0299	Pr > ChiSq	<.0001

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	0.000363	0.0000748	4.85	<.0001
AR1	1	-0.0868	0.009731	-8.92	<.0001
AR2	1	0.0323	0.009576	3.37	0.0008
ARCH1	1	0.0698	0.003963	17.60	<.0001
GARCH1	1	0.7078	0.0609	11.63	<.0001
GARCH2	1	0.2224	0.0573	3.88	0.0001

Why is the IGARCH model giving 25 fold increase? It seems unreasonable. The model indicates, and the data displays, large variability during periods where there were steep drops in the Dow. A method that accounts for different variances tends to downweight observations with high variability. In fact there are some periods in which the 11.3% annual rate required for a 25 fold increase (1.113^{20}=25) was actually exceeded, such as in the periods leading up to the great depression, after FDR assumed office, and toward the end of WWII. The extremely large variances associated with periods of decrease or slow growth give them low weight, and that would tend to increase the estimated growth rate, but it is still not quite enough to explain the results.

Perhaps more importantly, the rejection of normality by the Jarque-Bera test introduces the possibility of bias in the estimated mean. In an ordinary least squares (OLS) regression of a column **Y** of responses on a matrix **X** of explanatory variables, the model is $\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{e}$ and the estimated parameter $\hat{\boldsymbol{\beta}}=(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y}) = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{e})$ is unbiased whenever the random vector **e** has mean 0. In regard to bias, it does not matter if the variances are unequal or even if there is correlation among the errors. These features only affect the variance of the estimates, causing biases in the standard errors for $\hat{\boldsymbol{\beta}}$ but not in the estimates of $\boldsymbol{\beta}$ themselves. In contrast to OLS, GARCH and IGARCH models are fit by maximum likelihood assuming a normal distribution. Failure to meet this assumption could affect bias in parameter estimates such as the estimated mean.

As a check to see if bias can be induced by nonnormal errors, data from a model having the same h_t sequence as that estimated for the Dow Jones log differences data were generated for innovations $e_t \sim N(0,1)$ and again for innovations $(e_t^2-1)/\sqrt{2}$ so this second set of innovations used the same normal variables in a way that gave a skewed distribution still having mean 0 and variance 1. The mean was set at 0.00007 for the simulation and 50 such data sets were created. For each dataset, IGARCH models were fit for each of the two generated series and the estimated means were output to a dataset. The overall mean and standard deviation of each set of 50 means were:

	Mean	Standard Deviation
Normal Errors	0.000071	0.0000907
Skewed Errors	0.000358	0.0001496

Thus it seems that finding a factor of 5 bias in the estimate of the mean (of the differenced logs) could be simply a result of error skewness, in fact the factor of 5 is almost exactly what the simulation shows. The simulation results also show that if the errors had been normal, good estimates of the true value, known in the simulation to be 0.00007, would have resulted.

Using type=integ specifies an IGARCH model for h_t which, as with any unit root model, will have a linearly changing forecast if an intercept is present. You thus use noint to suppress the intercept. Using p=2 and q=1 your model has the form

$$h_t = h_{t-1} + 0.7078\,(h_{t-1}-h_{t-2}) + 0.2224(\,h_{t-2}-h_{t-3}) + 0.0698\,\epsilon_{t-1}^2$$

You can look at h_t as a smoothed local estimate of the variance, computed by adding to the previous smoothed value (h_{t-1}) a weighted average of the most recent 2 changes in these smoothed values and the square of the most recent shock.

By default, PROC AUTOREG uses a constant variance to compute prediction limits, however you can output the h_t values in a dataset as shown and then, recalling that h_t is a local variance, add and subtract $t_{0.975}\sqrt{h_t}$ from your forecast to produce forecast intervals that incorporate the changing variance. Both kinds of prediction intervals are shown in the bottom right corner of the Dow Jones graph where the more or less horizontal bands are the AUTOREG defaults and the bands based on h_t form what looks like a border to the data. The dataset has the historic data and 500 additional days with dates but no values of D_t. PROC AUTOREG will produce h_t values and default prediction limits for these. In general, future values of all inputs need to be included for this to work, but here the only input is the intercept.

The default prediction intervals completely miss the local features of the data and come off the end of the data with a fairly wide spread. Since the last few data periods were relatively stable, the h_t based intervals are appropriately narrower. It appears that (h_t-h_{t-1}) , $(\,h_{t-1}-h_{t-2})$ and ϵ_t^2 were fairly small at the end of the series, contributing very little to h_{t+1} so that h_{n+1} is approximately h_n as are all h_{n+j} for $j>0$. The forecast intervals coming off the end of the series thus have about the same width as the last forecast interval in the historic data. They are almost, but not exactly, two horizontal lines.

The autoregressive error model is seen to be

$$Z_t = 0.0868\,Z_{t-1} - 0.0323\,Z_{t-2} + \epsilon_t^2$$

where $\epsilon_t = \sqrt{h_t}\,e_t$ and although the lag Z coefficients are statistically significant, they are small so that their contribution to forecasts and to the width of prediction intervals into the future is imperceptible in the graph.

Clearly the IGARCH estimated mean 0.000363 is unacceptable in light of the nonnormality, the resulting danger of bias, and its failure to represent the observed growth over the period. The ordinary mean 0.00007 is an unbiased estimate and exactly reproduces the observed growth. The usual conditions leading to the (OLS) formula for the standard error of a mean do not hold here, but more will be said about this shortly. The problem is not with IGARCH versus GARCH, in fact a GARCH(2,1) model also fits the series quite nicely but still gives an unacceptable estimate of the mean of D_t. Note that the average of n independent values of ϵ_t has variance $n^{-2}\sum_{t=1}^{n} h_t$ if e_t has mean 0 and variance 1.

The AR(2) error series

$$Z_t = \alpha_1 Z_{t-1} + \alpha_2 Z_{t-2} + \epsilon_t$$

can be summed from 1 to n on both sides and divided by n to get $(1-\alpha_1-\alpha_2)\bar{Z}$ approximately equal to $\bar{\epsilon}$. From $\epsilon_t = \sqrt{h_t} e_t$ it follows that the variance of $(1-\alpha_1-\alpha_2)\bar{Z}$ is $n^{-2}\sum_{t=1}^{n} h_t$ and that of $\bar{Z}$ is thus $(1-\alpha_1-\alpha_2)^{-2} n^{-2}\sum_{t=1}^{n} h_t$. Hamilton (1994 page 663) indicates that maximum likelihood estimates of h_t are reasonable under rather mild assumptions for ARCH models even when the errors are not normal. Also the graphical evidence indicates that the estimated h_t series has captured the variability in the data nicely. Proceeding on that basis, you sum the estimated h_t series and use estimated autoregressive coefficients to estimate the standard deviation of the mean

$$n^{-1}\sqrt{(1-\alpha_1-\alpha_2)^{-2}\sum_{t=1}^{n} h_t} \text{ as } 8892^{-1}\sqrt{(1-.0868+0.0323)^{-2} 1.55846} = 0.0001485.$$

In this way you get t=0.00007/0.0001485 = 0.5 which is not significant at any reasonable level.

Interestingly, and despite the comments above, a simple t test on the D_t data, ignoring all of the variance structure, gives about the same t. A little thought shows that this could be anticipated for the special case of this model. The summing of h_t and division by n yields what might be thought of as an average variance over the period. Because the αs are small here, the average of h_t divided by n is a reasonable approximation to the variance of $\bar{Z}$ and thus of $\bar{D}$. To the extent that the squared residuals $(D_t - \bar{D})^2$ provide approximate estimates of the corresponding conditional variances h_t the usual OLS formula for the standard error of a mean, $\sqrt{n^{-1}\sum_{t=1}^{n}(D_t - \bar{D})^2/(n-1)}$ gives an estimate of the standard error of the mean. Additional care would be required, such as consideration of the assumed unit root structure for h_t and the error introduced by ignoring the αs, to make this into a rigorous argument. However this line of reasoning does suggest that the naive t test, produced for example by PROC MEANS, might be reasonable for this particular data. There is no reason to expect the naive approach to work well in general.

This example serves to illustrate several important points. One is that careful checking of model implications against what happens in the data is a crucial component of proper analysis. This would typically involve some graphics. Another is that failure to meet assumptions is sometimes not so important but at other times can render estimates meaningless. Careful thinking and a knowledge of statistical principles is crucial here. The naive use of statistical methods without understanding the underlying assumptions and limitations can lead to ridiculous claims. Computational software is not a replacement for knowledge.

Bolerslev, Tim (1986). "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 31:307-27

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

Hamilton, J. D. (1994) *Time Series Analysis*, Princeton University Press

Jarque, C.M. and Bera, A.K. (1980), "Efficient Tests for Normality, Homoskedasticity and Serial Independence of Regression Residuals," *Economics Letters,* 6, 255-259

Said, S.E. and Dickey, D. A. (1984). "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," *Biometrika* 71, 599-607.