Chapter 8

ARCH

8.1 Background

Suppose we begin with the model:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \qquad \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}) \tag{8.1}$$

The familiar least squares estimate for $\boldsymbol{\beta}$ is given by

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}.$$

The above model allows us to state that conditional on an information set Ω_{t-1}, the expectations of y_t and ϵ_t are

$$\begin{aligned} E(y_t|\Omega_{t-1}) &= \boldsymbol{X}_t\boldsymbol{\beta} \\ E(\epsilon_t|\Omega_{t-1}) &= 0 \end{aligned} \tag{8.2}$$

and their conditional variances are

$$V(\epsilon_t|\Omega_{t-1}) = \sigma^2 \tag{8.3}$$
$$V(y_t|\Omega_{t-1}) = \sigma^2. \tag{8.4}$$

- In the conventional classical model the conditional variance of y_t is constant.
- Many time series especially financial data have volatility that clusters
- That is the variance today appears to depend on past volatility
- For instance stock markets experience periods of wide fluctuations in which stock prices both rise and then fall the next day and this pattern of persistent volatility need not accompany a change in the conditional mean

- Suppose that the conditional variance of y_t is assumed to be a function of time, such that
$$V(y_t|\Omega_{t-1}) = \sigma_t^2$$
- What does the presence of the time-varying variance do to the OLS estimation of equation (8.1) ?
- This is just a special case of performing OLS in the presence of heteroskedasticity, so the results from that case will hold:
  - Estimates of $\boldsymbol{\beta}$ to be consistent (as long as the independent variables are asymptotically uncorrelated with the errors),
  - Standard errors are biased
  - Inference on $\boldsymbol{\beta}$ based on these results is invalid.
- One could obtain valid test statistics if the standard errors are calculated to be robust to heteroskedasticity,
- If we examine the shortrate in the data set we see that the first-difference lacks persistence but its square is highly persistent.
- Robert Engle (1982) pioneered this research into Autoregressive Conditional Heteroskedasticity **ARCH** and the filed of empirical finance has been dominated by research along these lines since
- Interestingly as we shall see the errors are only conditionally heteroskedastic but they are in fact unconditionally homoskedastic!
- ARCH models imply time varying second moments and not anything about the first (although it is popular now to place the conditional volatility into the conditional first moment called $ARCH - M$ models)
- There is a HUGE literature on these kinds of models and we will only summarize some standard results and models.

8.2 The Basic ARCH Model

- Following Engle(1982) suppose we have the following specification for an ARCH(1) process:
$$\begin{aligned} y_t &= \epsilon_t \sigma_t^2 \qquad \epsilon_t \sim N(0,1) & (8.5) \\ \sigma_t^2 &= \alpha_0 + \alpha_1 y_{t-1}^2 & (8.6) \end{aligned}$$
- The conditional variance of y_t is given by
$$V(y_t|\Omega_t) = \alpha_0 + \alpha_1 y_{t-1}^2 = \sigma_t^2,$$
while the unconditional variance is constant and equal to $\alpha_0/(1 - \alpha_1)$.

- The conditional distribution of y_t is

$$y_t|\Omega_t \sim N(0, \sigma_t^2)$$

- A more general formulation of the ARCH process is given by

$$\sigma_t^2 = \sigma(y_{t-1}, \dots, y_{t-p}; \alpha_0, \dots, \alpha_p)$$

where p is referred to as the order of the ARCH process.

- Suppose that y_t depends on some (exogenous) independent variables $\boldsymbol{X}_t$ that are in the conditioning information set.

- Then the conditional distribution of y_t is

$$y_t|\Omega_t \sim N(\boldsymbol{X}_t\boldsymbol{\beta}, \sigma_t^2) \tag{8.7}$$

- The full specification of the ARCH model will include Equation 8.7 and the following equations:

$$\begin{aligned}
\sigma_t^2 &= \sigma^2(y_{t-1}, \dots, y_{t-p}; \alpha_0, \dots, \alpha_p) \\
\epsilon_t &= y_t - \boldsymbol{X}_t\boldsymbol{\beta} \quad (\epsilon_t(\boldsymbol{\beta})) && (8.8) \\
\epsilon_t &= \upsilon_t\sigma_t. \qquad \upsilon_t \sim N(0,1) && (8.9)
\end{aligned}$$

- From the specification above we can build up a loglikelihood function as the sum of the conditional densities of the y_t's.

- Letting $\ell(\boldsymbol{y}; \boldsymbol{\beta}, \boldsymbol{\alpha})$ be the average loglikelihood function we have that

$$\ell(\boldsymbol{y}; \boldsymbol{\beta}, \boldsymbol{\alpha}) = \frac{1}{T}\sum_{t=1}^{T} \ell_t(y_t; \boldsymbol{\beta}, \boldsymbol{\alpha})$$

where

$$\ell_t(y_t; \boldsymbol{\beta}, \boldsymbol{\alpha}) = constants - \frac{1}{2} log \sigma_t - \frac{1}{2}\frac{\epsilon_t^2}{\sigma_t^2}$$

- Call ℓ_t the contribution to the average loglikelihood for observation t.

- To estimate the parameters of the model $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ we take the first-order conditions (FOC) of the loglikelihood with respect to the parameters.

- The FOC with respect to $\boldsymbol{\beta}$ are

$$\frac{\partial \ell_t}{\partial \boldsymbol{\beta}} = -\frac{1}{2\sigma_t}\frac{\partial \sigma_t}{\partial \boldsymbol{\beta}} + \left[\frac{\epsilon_t \boldsymbol{X}_t^T}{\sigma_t} + \frac{\partial \sigma_t}{\partial \boldsymbol{\beta}}\frac{\epsilon_t^2}{2\sigma_t^2}\right]$$

$$= -\frac{1}{2\sigma_t}\frac{\partial \sigma_t}{\partial \boldsymbol{\beta}}\left[1 - \frac{\epsilon_t^2}{\sigma_t}\right] + \frac{\epsilon_t \boldsymbol{X}_t^T}{\sigma_t}$$

- To derive these you must note that σ_t is a function of ϵ_t which is itself a function of $\boldsymbol{\beta}$.
- The Hessian can be found by taking the derivatives of the FOC.
- The $\boldsymbol{\beta\beta}$ block of the Hessian will be given by

$$\frac{\partial^2 \ell_t}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = -\frac{\boldsymbol{X}_t^T \boldsymbol{X}_t}{\sigma_t} - \frac{1}{2\sigma_t^2} \frac{\partial \sigma_t}{\partial \boldsymbol{\beta}} \frac{\partial \sigma_t}{\partial \boldsymbol{\beta}^T} \left(\frac{\epsilon_t^2}{\sigma_t} \right)$$

$$-\frac{2\epsilon_t \boldsymbol{X}_t^T}{\sigma_t^2} \frac{\partial \sigma_t}{\partial \boldsymbol{\beta}} + \left(\frac{\epsilon_t^2}{\sigma_t} - 1 \right) \frac{\partial}{\partial \boldsymbol{\beta}^T} \left(\frac{1}{2\sigma_t} \frac{\partial \sigma_t}{\partial \boldsymbol{\beta}} \right).$$

- Taking conditional expectations we have

$$E_{t-1}\left[\frac{2\epsilon_t \boldsymbol{X}_t^T}{\sigma_t^2} \frac{\partial \sigma_t}{\partial \boldsymbol{\beta}} \right] = \mathbf{0}$$

$$E_{t-1}\left[\left(\frac{\epsilon_t^2}{\sigma_t} - 1 \right) \frac{\partial}{\partial \boldsymbol{\beta}^T} \left(\frac{1}{2\sigma_t} \frac{\partial \sigma_t}{\partial \boldsymbol{\beta}} \right) \right] = \mathbf{0}$$

$$E_{t-1}\left[\frac{\partial^2 \ell_t}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right] = -\frac{\boldsymbol{X}_t^T \boldsymbol{X}_t}{\sigma_t} - \frac{1}{2\sigma_t^2} \frac{\partial \sigma_t}{\partial \boldsymbol{\beta}} \frac{\partial \sigma_t}{\partial \boldsymbol{\beta}^T}$$

- The information matrix is the negative of the expectation of the Hessian averaged over all t.
- Thus the $\boldsymbol{\beta\beta}$ block of the information matrix is given by

$$\Im_{\boldsymbol{\beta\beta}} = -\frac{1}{T} \sum_{t=1}^{T} E\left[E_{t-1}\left[\frac{\partial^2 \ell_t}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right] \right]$$

$$\Im_{\boldsymbol{\beta\beta}} = -\frac{1}{T} \sum_{t=1}^{T} E\left[\frac{\boldsymbol{X}_t^T \boldsymbol{X}_t}{\sigma_t} + \frac{1}{2\sigma_t^2} \frac{\partial \sigma_t}{\partial \boldsymbol{\beta}} \frac{\partial \sigma_t}{\partial \boldsymbol{\beta}^T} \right].$$

- We can find the $\boldsymbol{\beta\alpha}$ block of the information matrix by taking first the FOC with respect to $\boldsymbol{\beta}$ and then take the derivative of the resulting expression with respect to $\boldsymbol{\alpha}$.
- The resulting block of the information matrix will have the general form

$$\Im_{\boldsymbol{\beta\alpha}} = \frac{1}{T} \sum_{t=1}^{T} E\left[\frac{1}{2\sigma_t^2} \frac{\partial \sigma_t}{\partial \boldsymbol{\beta}} \frac{\partial \sigma_t}{\partial \boldsymbol{\beta}^T} \right].$$

- Engle(1982) shows that under symmetrical distributions of the y_t, the off-diagonal blocks are zero.

- This implies that the information matrix is block-diagonal between $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, and that we can estimate $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ separately, without asymptotic loss of efficiency

- We can estimate efficiently the conditional mean by using only the estimates of the ARCH process and vice versa

- However, the standard errors of the OLS could be inconsistent

- Block diagonality does not hold for other $ARCH$ processes like $EGARCH$ (exponential)

- Hence the joint maximum likelihood has been common practice from the beginning (even though the estimation is quite non-linear)

- Not surprising, the efficiency gain is largest when the coefficient on the lagged squared errors (α_1) is close to 1.

- From the results above and due to the block diagonality of the information matrix between the parameters of the conditional mean and the conditional variance, the asymptotic distributions of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\alpha}}$ are:

$$\sqrt{T}(\hat{\beta} - \beta_o) \overset{a}{\sim} N(0, \Im_{\beta\beta}^{-1})$$

$$\sqrt{T}(\hat{\alpha} - \alpha_o) \overset{a}{\sim} N(0, \Im_{\alpha\alpha}^{-1})$$

8.3 Testing For ARCH Errors

- The general equation for an ARCH(p) process is given by

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \ldots + \alpha_p \epsilon_{t-p}^2$$

  and the hypothesis we wish to test is whether or not there is an autoregressive component in the variances of the residuals.

- The null hypothesis can then be specified as

$$H_0 : \boldsymbol{\alpha} = 0 \qquad where\ \boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_p]$$

- The model for the conditional variance under the null hypothesis is

$$\sigma_t^2 = \alpha_0$$

  which is to say, that under the null hypothesis the conditional variance is constant and inference based on least squares standard errors is valid.

- The test for simple ARCH errors is extremely easy to perform since it is based on the restricted (homoskedastic) residuals).

- Run the following linear regression

$$\hat{\epsilon}_t^2 = \alpha_0 + \alpha_1 \hat{\epsilon}_{t-1}^2 + \ldots + \alpha_p \hat{\epsilon}_{t-p}^2.$$

  where $\hat{\epsilon}_t$ are the least squares residuals from the estimation of model 8.1.

- This is a form of a Lagrange Multiplier Test as our null hypothesis is that the errors have a constant conditional variance and that OLS of 8.1 allows us to perform valid inference.

- A rejection of a constant variance is usually interpreted as evidence of ARCH errors

- However, as we all know a rejection of the null does not necessarily imply the alternative is the DGP

- Even if the alternative is true, the researcher still does not know the appropriate $ARCH$ model or its order since the test is likely to have power against a variety of ARCH alternatives

8.3.1 Steps in ARCH Testing

1. Regress y_t on the independent variables and save the residuals.
2. Square the residuals from step 1.
3. Set up and estimate the test regression, remembering to skip the first p observations to allow for lags of the squared residuals.
4. Calculate the test statistic as
   (a) $(T-p) \cdot R^2$, where R^2 is the multiple correlation coefficient printed out by the regression package. This statistic will be distributed as χ^2 with p degrees of freedom. Multiply by $(T-p)$ as this is the number of degrees of freedom from the test regression.

8.3.2 Identifying the order of the ARCH process

- Bollerslev(1986) suggests that just as one could use the ACF and PACF to identify the order of an autoregressive process

- Often researchers look at the significance of the $\alpha's$ with various orders of the ARCH test

- In all cases, this is only suggestive and in practice researchers have estimated various ARCH models and done a variety of residual tests to determine if the modelling process has successfully removed the ARCH

- For instance denote the MLE estimates as $\hat{\beta}$ and $\hat{\alpha}$ respectively. One can calculate the standardized residuals

$$\hat{\nu}_t = \frac{\hat{\epsilon}_t}{\hat{\sigma}_t}$$

  and then perform an $ARCH$ test on these to see if there is any residual ARCH present

8.4 Estimation with ARCH Errors in Stata (slight change in notation)

Suppose our model is a general econometric model with ARCH(1) errors that is given by

$$\begin{aligned} y_t &= \boldsymbol{X}_t\boldsymbol{\beta} + \epsilon_t & (8.10) \\ \sigma_t^2 &= \alpha_0 + \alpha_1\epsilon_{t-1}^2 & (8.11) \end{aligned}$$

where $\boldsymbol{X}_t$ is a k-vector $\boldsymbol{X}_t = [1, X_{1t}, \ldots, X_{(k-1)t}]$.

- In estimating an ARCH model one must realize that the equation specifying the ARCH process is estimating a VARIANCE. This will imply that the fitted values from the ARCH process must be non-negative as negative variance terms make no sense. Thus if one is going to estimate an ARCH process using an iterative OLS procedure you must pay particular attention to the estimated parameters of the ARCH process to ensure that they do not violate the non-negativity of the conditional variance.

- To estimate this system by maximum likelihood, one must use a recursive procedure. This is due to the fact that σ_t depends on past squared errors, thus we must perform the $(t-1)^{st}$ iteration in order to obtain the residuals for the t^{th}.

8.4.1 Other forms of ARCH

STATA estimates a wide variety of ARCH models with just the change of a command line

1. *garch* (Generalized ARCH)
2. *saarch* (Simple asymmetric ARCH)
3. *tarch* (Threshold ARCH)
4. *aarch* (Asymmetric ARCH)
5. *narch* (Nonlinear ARCH)

6. *narchk* (Nonlinear arch with single shift)
7. *abarch* (Absolute value ARCH)
8. *atarch* (Absolute threshold ARCH)
9. *sdgarch* (GARCH)
10. *earch* (exponential ARCH)
11. *egarch* (Exponential GARCH)
12. *parch* (Power ARCH)
13. *tparch* (Threshold power ARCH)
14. *aparch* (Asymmetric Power ARCH)
15. *nparch* (Nonlinear power ARCH)
16. *nparchk* (nonlinear power ARCH with single shift)
17. *pgarch* (Power GARCH)
18. *archm* (include ARCH in mean)
19. *archmlags* (GARCH in mean)
20. *arch* exp (exponential ARCH in mean)

- Below we discuss just the GARCH and ARCH-M specifications and the interested reader may pursue the rest!

8.4.2 GARCH

- GARCH differs from ARCH in that past conditional variances will enter the equation for the conditional variance at t.
- The general specification for a GARCH(p,q) model is given by

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \ldots + \alpha_p \epsilon_{t-p}^2 + \gamma_1 \sigma_{t-1}^2 + \ldots + \gamma_q \sigma_{t-q}^2 \tag{8.12}$$

i.e. GARCH(1,1) is given by:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \gamma \sigma_{t-1}^2$$

- Small order GARCH processes perform like larger order ARCH processes, since the right hand side terms involving σ implicitly incorporate all past values of ϵ.

- It is tempting to interpret the $\alpha's$ as capturing the "moving average parts" and $\gamma's$ as the "autoregressevie parts" but this is not the case (see Hamilton p. 666) .

- It can be shown that if ϵ_t is describe by an $GARCH(q,p)$ process then ϵ_t^2 follows a $ARMA(r,p)$ where $r = \max(p,q)$

8.4.3 E-GARCH

The ARCH model implies symmetry from increases or decreases in errors as they are squared. We might want to allow for a more flexible structure which was introduced by Nelson (1991) and is very popular.

Consider the $E - GARCH(1,1)$

$$\ln(\sigma_t) = \alpha_0 + \alpha_1 v_t + \alpha_1(|\nu_{t-1}| - \sqrt{2/\pi}) + \gamma\sigma_{t-1}^2 \qquad v_t = \frac{\epsilon_t}{\sigma_t}$$

8.4.4 ARCH-M

- ARCH-M models differ from ARCH in that the conditional variance enters the expression for the conditional mean.

- With this type of specification the information matrix is no longer block diagonal between the parameters of the conditional mean and the conditional variance.

- OLS of the conditional mean equation no longer provides consistent estimates, and we must use a maximum likelihood procedure to estimate the system.

- Suppose that the ARCH-M model is given by the following

$$\begin{aligned} y_t &= \boldsymbol{X}_t\boldsymbol{\beta} + \delta \cdot \sigma_t^{\frac{1}{2}} + \epsilon_t \\ \sigma_t &= \alpha_0 + \alpha_1\epsilon_{t-1}^2 \end{aligned} \tag{8.13}$$

(the more general form of the model is $y_t = x_t(\beta, \sigma_t^2) + \epsilon_t$).

- Notice how the period t conditional variance enters the equation for the period t conditional mean, and how the equation for the period t conditional variance depends on the squared period $t-1$ error.

- In financial applications, σ_t^2 can be thought of as a proxy for risk.

8.5 Multivariate Garch (see arch.do)

Stata has included some estimation of multivariate GARCH with restrictions. As to what you can feasibly estimate is problem dependent and in the illustrative *arch.do*, some multivariate GARCH are explored

The general set-up is

i

$$
\begin{aligned}
y_t &= Cx_t + \epsilon_t \\
\epsilon_t &= H_t^{\frac{1}{2}} \nu_t \\
vech(H_t) &= S + Avech(\epsilon_{t-i}\epsilon_{t-1}^T) + Bvech(H_{t-1})
\end{aligned}
$$

where

y_t is an $m \times 1$ vector of dependent variables
C is an $m \times k$ matrix of paramters
x_t is a $k \times 1$ vector of independent variables (maybe lags of y_t)
$H^{\frac{1}{2}}$ is the Choloeski factor of the time-varying conditional cov of H_t
ν_t is an $m \times 1$ vector of normal *iid* innovations
S is an mxm symmetric matrix of paramters
A_i is $m \times m$ symmetric (diagonal in vech) matrix of parameters
B_i is a $m \times m$ symmetric (diagonal in vech) matrix

8.5.1 Diagonal vech MGARCH ($DVECH$)

- each condtional variance (covariance) depends on its own past conditonal variance (covariance i, j but not any other r covariances) but not on others (A and B are diagonal)

$$h_{ij,t} = s_{ij} + a_{ij}\epsilon_{i,(t-1)}\,\epsilon_{j,(t-1)} + b_{ij}h_{ij,(t-1)}$$

- This is the first vector model postulated (and estimated) with the parameters grow quadratically in m

- In a $DVECH(1,1)$ numbewr of paramters are $\frac{3m(m+1)}{2}$

- Estimation is difficult because H_t must be positive definite for each t and the off-diagonal elements have to satistfy complicated restirctions

8.5.2 Conditional Correlation MGARCH (CC)

- Conditional correlation (CC) models use the nonlinear combinations of univariate $GARCH$ models to represent the conditional covariances.

- CC models have slower parameter growth than the $DVECH$

$$H_t = D_t^{1/2} R_t D_t^{1/2}$$

where D_t is a diagonal matrix such that

$$D_t = diag(\sigma_{i,t}^2) \qquad i = 1, ..., m$$

- Each of the conditional variances follows a univariate $GARCH$ process and the parameterization of R_t varies across model

$$h_{ij,t} = \rho_{ij}\sigma_{i,t}\sigma_{j,t}$$

where the $\sigma_{i,t}^2$ is a univariate $GARCH$ process

$$\sigma_{i,t}^2 = s_i + \sum_{j=1}^{p_i} \alpha_j \epsilon_{i,t-j}^2 + \sum_{j=1}^{q_i} \beta_j \sigma_{i,t-j}^2$$

- ρ_{ij} governs the amount of condtional covariance from one variable to another

There are 3 other models based on CC

1. **Constant Conditional Correlation** $MGARCH$ models (CCC)

$R_t = 1$ on main diagonal and ρ (same for all off diagonals) $\forall t$

2. **Dynamic Conditional Correlation** MGARCH (DCC) where matrix R_t the "quasi correlation matrix" is time varying and follows a sort of $GARCH$ process (it is weird without a great deal of motivation)

3. **Varying Conditional Correlation** MGARCH (VCC) conditiona correlation matrix is a weighted sum of a time invariant component, a measure of recent correlation among residuals, and last periods conditional correlation. All conditional correlations follow same dynamics (see Stata). I have not got this to work

8.6 Bibliography

Bollerslev, T. ,1986, "Generalized Autoregressive Heteroskedasticity", *Journal of Econometrics,* vol 31 pages 307-327

Engle, R.F, 1982, "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation",*Econometrica*, Vol 50 pages 987-1007

Bera, A.K., and Higgins, M.L., 1993 "Arch Models: Properties, Estimation and Testing," *Jounal of Economic Surveys*, Vol 7 pages 305-362

Engle, R.F., and Rothschild M., 1992 "Special Issue on ARCH models in Finance", *Journal of Econometrics*, Vol, 52 No. 1/2.