POSSIBLE SPECIFICATION ERRORS

Consider the following regression

$$\underset{(Nx1)}{Y} = \underset{(Nxk)}{X}\underset{(kx1)}{b} + \underset{(Nx1)}{e}$$

Recall the assumptions regarding the errors: $e_i \sim iidN(0, \sigma^2)$

Possible specification errors include:

i) Non-normality: The estimator β is still BLUE (best linear unbiased estimator; with the minimum variance)

but the inferences procedures are now only asymptotically valid

HETEROSCEDASTICITY

Recall that homoscedasticity implies:

$$\underset{(NxN)}{E(ee')} = E\begin{bmatrix} e_1^2 & e_1e_2 & \cdots & e_1e_N \\ e_2e_1 & e_2^2 & \cdots & e_2e_N \\ \vdots & \vdots & \cdots & \vdots \\ e_Ne_1 & e_Ne_2 & \cdots & e_N^2 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 I$$

Instead, when there is heterocedasticity we have

$$\underset{(NxN)}{E(ee')} = E\begin{bmatrix} e_1^2 & e_1e_2 & \cdots & e_1e_N \\ e_2e_1 & e_2^2 & \cdots & e_2e_N \\ \vdots & \vdots & \cdots & \vdots \\ e_Ne_1 & e_Ne_2 & \cdots & e_N^2 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \sigma_N^2 \end{bmatrix}$$

AUTOCORRELATED DISTURBANCES

Recall that serially uncorrelated errors imply that

$$\underset{(NxN)}{E(ee')} = E\begin{bmatrix} e_1^2 & e_1e_2 & \cdots & e_1e_N \\ e_2e_1 & e_2^2 & \cdots & e_2e_N \\ \vdots & \vdots & \cdots & \vdots \\ e_Ne_1 & e_Ne_2 & \cdots & e_N^2 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 I$$

In other words $E(e_ie_j) = 0$, for $i \neq j$.

When there are autocorrelated disturbances, we have

$$\underset{(NxN)}{E(ee')} = E\begin{bmatrix} e_1^2 & e_1e_2 & \cdots & e_1e_N \\ e_2e_1 & e_2^2 & \cdots & e_2e_N \\ \vdots & \vdots & \cdots & \vdots \\ e_Ne_1 & e_Ne_2 & \cdots & e_N^2 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2 & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma^2 & \cdots & \sigma_{2N} \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma^2 \end{bmatrix}$$

HETEROSCEDASTICITY AND THE $Var(\beta)$

Recall that the variance-covariance matrix of the $kx1$ vector of the estimators β is given by

$$\underset{(kxk)}{Var(\beta)} = E[\underset{(kx1)}{(\beta - b)}\underset{(1xk)}{(\beta - b)'}]$$

Recall also that the LS vector estimator β is

$$\beta = (X'X)^{-1}X'Y$$

and that under homoscedastic errors

$$\underset{(kxk)}{\widehat{Var}(\beta)} = E[\underset{(kx1)}{(\beta - b)}\underset{(1xk)}{(\beta - b)'}] = \sigma^2(X'X)^{-1}$$

Next recall that in the presence of heteroscedasticity

$E(ee') \neq \sigma^2 I$ but

$$E(ee') = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \sigma_N^2 \end{bmatrix} = \Omega$$

It can be shwon that the variance-covariance matrix of β is given by

$$E[\underset{(kx1)}{(\beta - b)}\underset{(1xk)}{(\beta - b)'}] = [(X'X)^{-1}X'\Omega X(X'X)^{-1}]$$

Of course Ω is unobserved since σ_i^2, $i = 1, \ldots, \sigma_N^2$ are unobserved

White proposed to replace σ_i^2 by the residual $\hat{e}_i^2$

Thus the estimated variance-covariance matrix of β, $\underset{(kxk)}{\widehat{Var}(\beta)}$, is given by

$$\underset{(kxk)}{\widehat{Var}(\beta)} = [\underset{(kxk)}{(X'X)^{-1}} \underset{(kxN)}{X'} \underset{(NxN)}{\hat{\Omega}} \underset{(Nxk)}{X} \underset{(kxk)}{(X'X)^{-1}}]$$

where $\hat{\Omega} = diag\{\hat{e}_1^2, \hat{e}_2^2, \cdots, \hat{e}_N^2\}$.

TESTS FOR HETEROSCEDASTICITY

1. WHITE's test

1st step. Run the following regression

$$Y = Xb + e$$

and save the vector of the residuals: $\widehat{e} = Y - X\beta = Y - \widehat{Y}$

2nd step. Regress the squared residuals, $\widehat{e}_i^2$, on a constant, all the regressors from step 1, their squares and their cross products

(the auxiliary regression)

Example: Suppose we have three regressors: $1, X_{2i}$ and X_{3i}, that is $k = 3$

In the 2nd step we run the auxiliary regression

$$\hat{e}_i^2 = c_1 + c_2 X_{2i} + c_3 X_{3i} + c_4 X_{2i}^2 + c_5 X_{3i}^2 + c_6 X_{2i} X_{3i} + u_i$$

The null hypothesis is: $H_0 : c_2 = c_3 = \cdots = c_6 = 0$ (homoskedasticity)

The alternative hypothesis is $H_0 : c_2, c_3, \cdots, c_6 \neq 0$ (heteroscedasticity)

White show that Nr^2 from the auxiliary regression under the null hypothesis follows (asymptotically)

the χ^2 distribution with q degrees of freedom where q is the number of the regressors (apart from the constant) in the auxiliary regression

In other words,

$$nr^2 \overset{a}{\sim} \chi^2(q)$$

In our example $q = 5$.

If homoscedasticity is rejected there is no indication of the form of heteroscedasticity, so one has to compute White's standard errors

$$\begin{aligned}
\widehat{Var}(\beta) &\neq s^2(X'X)^{-1}, \text{ but} \\
\underset{(kxk)}{\widehat{Var}(\beta)} &= [\underset{(kxk)}{(X'X)^{-1}} \underset{(kxN)}{X'} \underset{(NxN)}{\widehat{\Omega}} \underset{(Nxk)}{X} \underset{(kxk)}{(X'X)^{-1}}]
\end{aligned}$$

where $\widehat{\Omega} = diag\{\widehat{e}_1^2, \widehat{e}_2^2, \cdots, \widehat{e}_N^2\}$

One problem with White's test is that q might be very large. For example, if $k = 10$ then $q = k(k+1)/2 - 1 = 54$

2. BREUSCH-PAGAN or GODFREY Test

The Breusch-Pagan or Godfrey test has two forms. Here we will examine the simplest one

1st step. Run the following regression

$$Y = Xb + e$$

and save the vector of the residuals: $\widehat{e} = Y - X\beta = Y - \widehat{Y}$

Consider a $px1$ vector of variables, Z_i

$$Z_i' = [1, Z_{2i}, \ldots, Z_{pi}]$$

2nd step. Run the auxiliary regression

$$\hat{e}_i^2 = \underset{(1xp)}{Z_i'} \underset{(px1)}{c} + u_i$$

where c is a $px1$ vector of parameters $c' = [c_1, c_2, \ldots, c_p]$

The null hypothesis of homoscedasticity is:

$H_0 : c_2 = c_3 = \cdots c_p = 0$

It can be shown that Nr^2 from the above regression is asymptotically distributed as $\chi^2(p-1)$:

$$Nr^2 \overset{a}{\sim} \chi^2(p-1)$$

AUTOCORRELATED DISTURBANCES

Another possible source of misspecification is serial correlation in the residuals

$$E(u_t u_{t-s}) = \gamma_s \neq 0,$$

for $s = \pm 1, \pm 2, \ldots,$

When we run the following regression

$$Y = Xb + e$$

we hope to include all relevant variables in the X matrix

One possible reason for autocorrelated disturbances is the omission of some important variables from the X matrix

Example:

$$\begin{aligned} \text{Real Relation} &: \quad Y_t = b_1 + b_2 X_t + b_3 Y_{t-1} + e_t \\ \text{Estimated one} &: \quad Y_t = b_1 + b_2 X_t + v_t \end{aligned}$$

Notice that the relation between e_t and v_t is

$$v_t = b_3 Y_{t-1} + e_t$$

which is autocorrelated because Y_t is autocorrelated

DURBIN-WATSON TEST FOR SERIAL CORRELATION

The Durbin-Watson test statistic is computed from the OLS residuals: $\widehat{e} = Y - X\beta$

It has the following form

$$d = \frac{\sum_{t=2}^{T}(\widehat{e}_t - \widehat{e}_{t-1})^2}{\sum_{t=1}^{T}\widehat{e}_t^2}$$

i) If the e's are positively autocorrelated, then successive values will be tend to be close to each other

$\rightarrow |e_t - e_{t-1}| < |e_t|$ In this case d should be small

ii) If the e's are negatively autcorrelated, then successive observations will be on the opposite sides of the horizontal axis

$\rightarrow |e_t - e_{t-1}| > |e_t|$ In this case d should be big

iii) If the e's are uncorrelated then the d should be in between

Note that

$$\widehat{\rho} \simeq \frac{\sum_{t=2}^{T} \widehat{e}_t \widehat{e}_{t-1}}{\sum_{t=2}^{T} \widehat{e}_{t-1}^2}$$

which is the estimated slope coefficient from the following regression

$$\widehat{e}_t = a + b\widehat{e}_{t-1} + u_t$$

It can be shown that

$$d = 2(1 - \widehat{\rho})$$

Thus, the above equation shows that the range of d is from 0 (when $\widehat{\rho} = 1$) to 4 (when $\widehat{\rho} = -1$).

Moreover,

$$\begin{aligned} d &< 2\text{, for positive autocor of the } e'\text{s,} \\ d &> 2\text{, for negative autocor of the } e'\text{s,} \\ d &\simeq 2\text{, for zero autocor of the } e'\text{s.} \end{aligned}$$

Durbin and Watson established upper (d_u) and lower (d_l) bounds for the critical values.

These bounds depend only on the sample size and the number of regressors.

They are used to test the hypothesis of zero autocorrelation against the alternative of positive first order autocorrelation.

The testing procedure is as follows:

1. If $d < d_l$, reject the hypothesis of nonautocorrelated errors in favor of the hypothesis of positive first-order autocorrelation.

2. If $d > d_u$ do not reject the null hypothesis

3. If $d_l < d < d_u$ the test is inconclusive

If the value of d exceeds 2, one may wish to test the null hypothesis against the alternative of negative first-order autocorrelation.

The test is done by calculating $4 - d$ and comparing tis statistic with the tabulated critical values as if one were testing for positive autocorrelation.

There are two important qualifications to the use of the Durbin-Watson test. First, it is necessary to include a constant term in the regression.

Second, it is strictly valid only for a nonstochastic X matrix.

Thus it is not applicable when lagged values of the dependent variable appear among the regressors.

BREUSCH-GODFREY test

These two authors independently built on the work of Durbin to develop LM tests against general autoregressive or moving average disturbance processes.

Suppose the specified equation is

$$y_i = a + bx_i + u_i \tag{1}$$

with

$$u_i = cu_{i-1} + e_i, \text{ or } (1 - cL)u_i = e_i \tag{2}$$

where it is assumed that $|c| < 1$ and that the e's are independently and identically distributed normal variables with zero mean and variance σ_e^2.

Multiplying both sides of equation (1) by $1 - cL$ and substituting equation (2) gives

$$\begin{aligned} (1 - cL)y_i &= a(1 - c) + (1 - cL)bx_i + \underbrace{e_i}_{(1-cL)u_i} \Rightarrow \\ y_i &= a(1 - c) + cy_{i-1} + bx_i - cbx_{i-1} + e_i \end{aligned} \tag{3}$$

$$y_i = a(1-c) + cy_{i-1} + bx_i - cbx_{i-1} + e_i$$

We want to test the hypothesis that $c = 0$. Equation (3) is nonlinear in the parameters.

The test is obtained in two steps:

First, apply OLS to equation (1)

$$y_i = a + bx_i + u_i$$

to obtain the residuals $\widehat{u}_i$.

Then regress $\widehat{u}_i$ on a constant, x_i and $\widehat{u}_{i-1}$.

Under the null H_0, nr^2 is asymptotically $\chi^2(1)$.

BOX–PIERCE LJUNG STATISTIC

The Box-Pierce Q statistic is based on the squares of the first p autocorrelation coefficients (r_j) of the OLS residuals.

The statistic is defined as

$$Q = N \sum_{j=1}^{p} r_j^2$$

where

$$r_j = \frac{\sum_{i=j+1}^{N} e_i e_{i-j}}{\sum_{i=1}^{N} e_i^2}$$

Under the hypothesis of zero autocorrelations for the residuals, Q will have an asymptotic χ^2 distribution with degrees of freedom equal to p minus the number of parameters estimated in fitting an ARMA model.

An improved small-sample performance is expected from the revised Lung-Box statistic

$$Q' = N(N+2)\sum_{j=1}^{p} r_j^2/N - j$$

These statistics are sometimes used to test for autocorrelated disturbances in the type of regression equation that we have been considering,

but this application is inappropriate because equations such as equation (3) are not pure AR schemes but have exogenous variables as well:

$$y_i = a(1-c) + cy_{i-1} + bx - cdx_{i-1} + e_i$$

The effect on the distribution of Q or Q' is unknown.

PROOFS FROM WEEK 7 ONWARDS

(ONLY FOR EC5501)

HETEROSCEDASTICITY AND THE $Var(\beta)$

Recall that the variance-covariance matrix of the $kx1$ vector of the estimators β is given by

$$\underset{(kxk)}{Var(\beta)} = E[\underset{(kx1)}{(\beta - b)}\underset{(1xk)}{(\beta - b)'}]$$

since $E(\beta) = b$ (see below).

Recall also that the LS vector estimator β is

$$\beta = (X'X)^{-1}X'Y$$

Since $Y = Xb + e$, it follows that

$$\begin{aligned} \beta &= (X'X)^{-1}X'(Xb + e) \\ &= (X'X)^{-1}X'Xb + (X'X)^{-1}X'e \\ &= b + (X'X)^{-1}X'e \qquad (4) \end{aligned}$$

From equation (4) we have $E(\beta) = b$ and

$$\beta - b = (X'X)^{-1}X'e$$

Hence,

$$\begin{aligned} E[(\underset{(kx1)}{\beta - b})(\underset{(1xk)}{\beta - b})'] &= E\{(X'X)^{-1}X'e[(X'X)^{-1}X'e]'\} \\ &= E[(X'X)^{-1}X'ee'X(X'X)^{-1}] \end{aligned}$$

since $[(X'X)^{-1}]' = (X'X)^{-1}$ and $(X')' = X$

Next, we have

$$E[(\underset{(kx1)}{\beta - b})(\underset{(1xk)}{\beta - b})'] = [(X'X)^{-1}X'E(ee')X(X'X)^{-1}]$$

Next recall that in the presence of heteroscedasticity

$E(ee') \neq \sigma^2 I$ but

$$E(ee') = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \sigma_N^2 \end{bmatrix} = \Omega$$

Thus the variance-covariance matrix of β is given by

$$E[(\underset{(kx1)}{\beta - b})(\underset{(1xk)}{\beta - b})'] = [(X'X)^{-1}X'\Omega X(X'X)^{-1}]$$

Of course Ω is unobserved since σ_i^2, $i = 1, \ldots, \sigma_N^2$ are unobserved

White proposed to replace σ_i^2 by the residual $\hat{e}_i^2$

Thus the estimated variance-covariance matrix of β, $\underset{(kxk)}{\widehat{Var}(\beta)}$, is given by

$$\underset{(kxk)}{\widehat{Var}(\beta)} = [\underset{(kxk)}{(X'X)^{-1}} \underset{(kxN)}{X'} \underset{(NxN)}{\hat{\Omega}} \underset{(Nxk)}{X} \underset{(kxk)}{(X'X)^{-1}}]$$

where $\hat{\Omega} = diag\{\hat{e}_1^2, \hat{e}_2^2, \cdots, \hat{e}_N^2\}$.

PROOF OF $d = 2(1 - \widehat{\rho})$

The correlation between e_t and e_{t-1} is given by

$$\rho = \frac{E(e_t e_{t-1})}{\sqrt{Var(e_t^2)}\sqrt{Var(e_{t-1}^2)}}$$

The sample correlation coefficient is given by

$$\widehat{\rho} = \frac{\sum_{t=2}^{T} \widehat{e}_t \widehat{e}_{t-1}}{\sqrt{\sum_{t=2}^{T} \widehat{e}_t^2}\sqrt{\sum_{t=2}^{T} \widehat{e}_{t-1}^2}}$$

Notice that for a large T

$$\sum_{t=1}^{T} \widehat{e}_t^2 \simeq \sum_{t=2}^{T} \widehat{e}_t^2 \simeq \sum_{t=2}^{T} \widehat{e}_{t-1}^2$$

(ignoring end point differences)

Thus

$$\widehat{\rho} \simeq \frac{\sum_{t=2}^{T} \widehat{e}_t \widehat{e}_{t-1}}{\sum_{t=2}^{T} \widehat{e}_{t-1}^2}$$

which is the estimated slope coefficient from the following regression

$$\widehat{e}_t = a + b\widehat{e}_{t-1} + u_t$$

Thus

$$\begin{aligned}
d &= \frac{\sum_{t=2}^{T}(\widehat{e}_t - \widehat{e}_{t-1})^2}{\sum_{t=1}^{T}\widehat{e}_t^2} = \\
&= \frac{\sum_{t=2}^{T}\widehat{e}_t^2 + \sum_{t=2}^{T}\widehat{e}_{t-1}^2 - 2\sum_{t=2}^{T}\widehat{e}_t\widehat{e}_{t-1}}{\sum_{t=1}^{T}\widehat{e}_t^2} \\
&\simeq \frac{2\sum_{t=2}^{T}\widehat{e}_{t-1}^2 - 2\sum_{t=2}^{T}\widehat{e}_t\widehat{e}_{t-1}}{\sum_{t=2}^{T}\widehat{e}_{t-1}^2} = \\
&= 2 - 2\underbrace{\frac{\sum_{t=2}^{T}\widehat{e}_t\widehat{e}_{t-1}}{\sum_{t=2}^{T}\widehat{e}_{t-1}^2}}_{\widehat{\rho}} \\
&= 2(1 - \widehat{\rho})
\end{aligned}$$

DURBIN TESTS FOR A REGRESSION CONTAINING LAGGED VALUES OF THE DEPENDENT VARIABLE

As has been pointed out, the Durbin-Watson test procedure was derived under the assumption of a nonstochastic X matrix, which is violated by the presence of the dependent variable among the regressors.

Consider the relation

$$y_t = b_1 y_{t-1} + \cdots + b_r y_{t-r} + b_{r+1} x_{1t} + \cdots + b_{r+s} x_{st} + u_t, \tag{5}$$

where

$$u_t = \varphi u_{t-1} + e_t$$

with $|\varphi| < 1$, and $e_t \sim N(0, \sigma^2 I)$.

For this general case Durbin's basic result is that under the null hypothesis, $H_0 : \varphi = 0$, the statistic

$$h = \phi \sqrt{\frac{N}{1 - N\widehat{Var}(\beta_1)}} \overset{a}{\sim} N(0, 1), \qquad (6)$$

where $\widehat{Var}(\beta_1)$ is the estimated variance of β_1 in the OLS fit of equation (5)

and $\phi = \sum_{t=2}^{N} \widehat{u}_t \widehat{u}_{t-1} / \sum_{t=2}^{N} \widehat{u}_{t-1}^2$ is the estimate of φ from the regression of $\widehat{u}_t$ on $\widehat{u}_{t-1}$

the $\widehat{u}_t$'s being the residuals from the OLS regression of equation (5).

The test procedure is as follows. Compute h in equation (6) and if $h > 1.645$ (the 10% critical value for the normal distribution)

reject the null hypothesis at the 5% level in favor of the hypothesis of positive, first-order autocorrelation.

For negative h a similar one sided test for negative autocorrelation can be carried out.

The test breaks down if $N\widehat{Var}(\beta_1) > 1$.

Durbin showed that an asymptotically equivalent procedure is the following:

1. Estimate the OLS regression in equation (5) and obtain the residuals $\widehat{u}$'s

2. Estimate the OLS regression of

$$\widehat{u}_t \text{ on } \widehat{u}_{t-1},\, y_{t-1},\, \ldots, y_{t-r},\, x_{1t},\, \ldots, x_{st}$$

3. If the coefficient of $\widehat{u}_{t-1}$ in this regression is significantly different from zero by the usual t test,

reject the null hypothesis $H_0 : \varphi = 0$.

Durbin indicates that this last procedure can be extended to test for an AR(p) disturbance rather than an AR(1) process by simply adding additional lagged $\widehat{e}$'s to the second regression and testing the joint significance of the coefficients of the lagged residuals.

ESTIMATION WITH AUTOCORRELATED DISTRURBANCES:

If one or more of the tests described in the previous analysis suggest autocorrelated disturbances, what should be done?

One procedure is to start by checking whether the autocorrelation may not be a sign of misspecification in the original relationship.

Suppose that the correct relationship is

$$y_t = \gamma_1 + \gamma_2 x_t + \gamma_3 x_{t-1} + \gamma_4 y_{t-1} + e_t, \qquad (7)$$

where e_t are white noise.

A researcher's economic theory, however, delivers the proposition that y_t is influenced only by x_t.

When this model is fitted to the data, it is not surprising that significant autocorrelation is found in the errors.

To take this autocorrelation into account in the estimation of the relationship, the researcher now specifies

$$y_t = b_1 + b_2 x_t + u_t \tag{8}$$

with

$$\begin{aligned} u_t &= \varphi u_{t-1} + e_t \text{ or} \\ (1 - \varphi L) u_t &= e_t \end{aligned} \tag{9}$$

and proceeds to estimate it by, say, GLS.

The correct model in equation (7) involves five parameters, namely four parameters and a variance,

whereas our reseacher's specification has just four parameters.

The researcher is thus imposing a possibly invalid restriction on the parameters of the true model.

The nature of this restriction may be seen by rewriting equation (8) in the form

$$(1-\varphi L)y_t = b_1(1-\varphi)+b_2(1-\varphi L)x_t+\underbrace{e_t}_{(1-\varphi L)u_t} \qquad (10)$$

(We multiply both sides of equation (8) by $(1-\varphi L)$ and use equation (9).

Equation (7)

$$y_t = \gamma_1 + \gamma_2 x_t + \gamma_3 x_{t-1} + \gamma_4 y_{t-1} + e_t,$$

can be written as

$$\begin{aligned}(1-\gamma_4 L)y_t &= \gamma_1 + (\gamma_2 + \gamma_3 L)x_t \; + e_t \text{ or} \\ (1-\gamma_4 L)y_t &= \gamma_1 + \gamma_2(1+\frac{\gamma_3}{\gamma_2}L)x_t \; + e_t \quad (11)\end{aligned}$$

Comparison of the parameters in equations (10) and (11) shows that the restriction involved in moving from the former to the latter is

$$\gamma_4 = -\frac{\gamma_3}{\gamma_2} \qquad (12)$$

This is known as common factor restriction.

There are two alternative estimation procedures:

(i) Non-linear least squares

(ii) COCHRANE-ORCUTT

Rewriting equation (10) we have

$$(1-\varphi L)y_t = b_1(1-\varphi) + b_2(1-\varphi L)x_t + e_t \quad (13)$$

The above equation can also be written as

$$(y_t - b_1 - b_2 x_t) = \varphi(y_{t-1} - b_1 - b_2 x_{t-1}) + e_t. \quad (14)$$

If φ were known in equation (13) the b's could be estimated by straightforward OLS.

Similarly, if the b's were known in equation (14) φ could be estimated by an OLS regression with the intercept suppressed.

Cochrane and Orcutt suggest an iterative estimation procedure based on this pair of relations.

Start say, with an estimate of guess $\widehat{\varphi}^1$ of the autocorrelation parameter and use it to compute the differences $(1 - \widehat{\varphi}^1 L)y_t$ and $(1 - \widehat{\varphi}^1 L)x_t$.

These transformed variables are then used in the OLS regression (13):

$$(1 - \varphi L)y_t = b_1(1 - \varphi) + b_2(1 - \varphi L)x_t + e_t$$

yielding estimated coefficients $\beta_1^{(1)}$ and $\beta_2^{(2)}$.

These in turn are used to compute the variables in equation (14):

$$(y_t - b_1 - b_2 x_t) = \varphi(y_{t-1} - b_1 - b_2 x_{t-1}) + e_t.$$

and an OLS regression yields a new estimate $\widehat{\varphi}^2$.

The iterations continue until a satisfactory degree of convergence is reached.

The concern with iterative procedures is that they may converge to a local minimum and not necessarily to the global minimum.

A precaution is to fit equation (13) for a grid of φ values in steps of 0.1 from, say, -0.9 to 0.9

and then iterate from the regression with the smallest RSS.

GARCH HETEROSCEDASTICITY

In our analysis up to now we assume that the error term ε_t has zero mean:

That is

$$\begin{aligned} E(\varepsilon_t) &= E(\varepsilon_t|\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots,) \\ &= E_{t-1}(\varepsilon_t) = 0 \end{aligned}$$

constant unconditional variance:

$$Var(\varepsilon_t) = E(\varepsilon_t^2) = \sigma^2$$

and constant conditional (on time $t-1$) variance:

$$\begin{aligned} h_t &= Var(\varepsilon_t|\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots,) \\ &= E(\varepsilon_t^2|\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots) \\ &= E_{t-1}(\varepsilon_t^2) = E(\varepsilon_t^2) = \sigma^2 \end{aligned}$$

However, in many cases this is an unrealsistic.assumption. The conditional variance is not constant but changes with time. It depends on information up to to time $t-1$.

ARCH and GARCH MODELS

The next question is what kind of formulation one should use for h_t.

Naturally, h_t should be a function of q past squared errors. This is an ARCH(p) process:

$$\begin{aligned} h_t &= \omega + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-q}^2 \\ &= \omega + \Sigma_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2. \end{aligned}$$

ARCH: Autoregressive conditional heteroscedasticity.

One can include lagged values of the conditional variance in the above expression. This is the GARCH(p, q) process:

$$\begin{aligned} h_t &= \omega + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-q}^2 + \\ &\quad + \beta_1 h_{t-1} + \cdots + \beta_p h_{t-p} \\ &= \omega + \Sigma_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \Sigma_{j=1}^{p} \beta_j h_{t-j}. \end{aligned}$$

GARCH: Generalized ARCH.

The conditional variance h_t must be positive at all time t.

POSITIVITY OF THE CONDITIONAL VARIANCE

Consider the GARCH(p, q) process:

$$h_t = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \Sigma_{j=1}^{p} \beta_j h_{t-j}.$$

Sufficient conditions for the positivity of the conditional variance are

$$\begin{aligned} \omega &> 0, \ \alpha_i \geq 0, \text{ for } i = 1, \ldots, q \\ \beta_j &\geq 0, \ \text{for } j = 1, \ldots, p. \end{aligned}$$

One can write the error term ε_t as

$$\varepsilon_t = \sqrt{h_t} e_t$$

where e_t are independently and identically distributed random variables with zero mean and unit variance:

$E(e_t) = E_{t-1}(e_t) = 0$ and $E(e_t^2) = E_{t-1}(e_t^2) = 1$.

This implies that

$$E_{t-1}(\varepsilon_t) = \sqrt{h_t} E_{t-1}(e_t) = 0$$

and

$$E_{t-1}(\varepsilon_t^2) = h_t E_{t-1}(e_t^2) = h_t.$$

Note that using the law of the iterated expectations we have

$E(\varepsilon_t^2) = E[E_{t-1}(\varepsilon_t^2)] = E(h_t)$.

Consider the ARCH(1) process: $h_t = \omega + \alpha_1 \varepsilon_{t-1}^2$. Taking expectations from both of this equation gives

$$\begin{aligned} E(h_t) &= \omega + \alpha_1 E(\varepsilon_{t-1}^2) \\ &= \omega + \alpha_1 E(h_{t-1}). \end{aligned}$$

Next if $0 \leq \alpha_1 < 1$ the ARCH(1) process is covariance stationary and $E(h_t) = E(h_{t-1})$. $E(h_t)$ is then

$$E(h_t) = \frac{\omega}{1 - a_1}.$$

Similarly, for the ARCH(q) process,

assuming that $\Sigma_{i=1}^{q}\alpha_i < 1$, we have

$$\begin{aligned} E(h_t) &= \omega + \Sigma_{i=1}^{q}\alpha_i E(\varepsilon_{t-i}^2) \\ &= \omega + \Sigma_{i=1}^{q}\alpha_i E(h_{t-i}) \\ &= \omega + \Sigma_{i=1}^{q}\alpha_i E(h_t) \Rightarrow \\ E(h_t) &= \frac{\omega}{1 - \Sigma_{i=1}^{q}\alpha_i}. \end{aligned}$$

Finally, consider the GARCH(p, q) process.

If $\Sigma_{i=1}^{q}\alpha_i + \Sigma_{j=1}^{p}\beta_j = \alpha^* < 1$, then

$$\begin{aligned} E(h_t) &= \omega + \Sigma_{i=1}^{q}\alpha_i E(\varepsilon_{t-i}^2) + \Sigma_{j=1}^{p}\beta_j E(h_{t-j}) \\ &= \omega + \Sigma_{i=1}^{q}\alpha_i E(h_{t-i}) + \Sigma_{j=1}^{p}\beta_j E(h_{t-j}) \\ &= \omega + [\Sigma_{i=1}^{q}\alpha_i + \Sigma_{j=1}^{p}\beta_j] E(h_t) \Rightarrow \\ E(h_t) &= \frac{\omega}{1 - \alpha^*}. \end{aligned}$$

since $E(\varepsilon_{t-i}^2) = E(h_{t-i}) = E(h_t)$

TESTING FOR ARCH EFFECTS:

The obvious test implies the following steps:

(i) Fit y to X by OLS and obtain the residuals $\{\widehat{\varepsilon}_t\}$.

(ii) Compute the OLS regression: $\widehat{\varepsilon}_t^2 = \mathrm{a}_0 + a_1 \widehat{\varepsilon}_{t-1}^2 + \cdots + a_p \widehat{\varepsilon}_{t-p}^2 + \mathrm{u}_t$

If these coefficients are significantly different from zero, the assumption of conditionally homoscedastic disturbances is rejected in favor of ARCH disturbances.

One should remember, however, that various specification errors in the original relation can give false indications of ARCH disturbances.