Topic 8: GLS, Heteroskedasticity and Autocorrelation

Dr. Ani Dasgupta
Department of International Business, Mass. Maritime Academy
and
Department of Economics, Boston University

1 OLS with Non-spherical Disturbances

What happens to the OLS estimates if $E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}'|\mathbf{X}) = \sigma^{\mathbf{2}}\boldsymbol{\Omega}$ with $\boldsymbol{\Omega}$ being a symmetric p.d. matrix but $\boldsymbol{\Omega} \neq \mathbf{I}$? Two standard instances where this might occur are: heteroskedasticity and autocorrelation. In the first case, the disturbances have zero covariance among them, but their second moment may depend on the observation number (one reason for that could be that the second moment depends on the right hand side variables and since they differ from observation to observation the variance of the disturbance term is non-constant. In this case, the covariance matrix is of the form:

$$\begin{bmatrix} \sigma_1^2 & & & & \\ & \sigma_2^2 & & 0 & \\ & & \cdot & & \\ & & & \cdot & \\ & 0 & & & \cdot \\ & & & & \sigma_n^2 \end{bmatrix}$$

In the second case covariance among different disturbance terms are allowed to be nonzro. In this case, the exact nature of the covariance matrix depends on what is assumed about the process $\{\varepsilon_t\}$. Typically, $\{\varepsilon_t\}$ is assumed to follow some stationary ergodic process often specifically modeled as an ARMA(p,q) process the governing equation of which is

$$\varepsilon_t - \phi_1\varepsilon_{t-1} - \cdots - \phi_p\varepsilon_{t-p} = v_t - \theta_1 v_{t-1} - \cdots - \theta_q v_{t-q}$$

where v_t's are i.i.d. zero mean random variables (often referred to as 'white noise'). Thus today's disturbance depends on past disturbances as well as a stream of past white noise terms. A special case of this is an $AR(1)$ process, where for instance i.e. $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$ (v_t's are i.i.d., $E(v_t) = 0,\ |\rho| < 1$), when the covariance matrix has the following structure:

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & . & . & \rho^{n-1} \\ \rho & 1 & \rho & . & . & \rho^{n-2} \\ . & . & . & & & . \\ . & . & . & . & & . \\ . & . & . & & . & . \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & & & 1 \end{bmatrix}$$

where $\sigma^2 =$Var(ϵ_t) (You can check this by showing $E(\epsilon_t\epsilon_s) = \rho^{|t-s|}$).

We will shortly discuss efficient ways of estimating the β coefficients under such scenarios. However, note that under such 'nonspherical error structure' (a name given to the violation of $E(\boldsymbol{\epsilon}\epsilon'|\mathbf{X}) = \sigma^{\mathbf{2}}\mathbf{\Omega}$ condition) the following facts are true:

Fact 1. The OLS coefficients are still unbiased (since for this to hold all we require is $E(\boldsymbol{\epsilon}|\mathbf{X}) = \mathbf{0}$).

Fact 2. They are consistent under the assumptions made in the chapter on large sample theory (plim $\frac{1}{n}(\mathbf{X'X})$ is a p.d. matrix, plim $\frac{1}{n}\mathbf{X'}\boldsymbol{\epsilon} = \mathbf{0}$)

Fact 3. They are not efficient any more, in the Gauss-Markov sense (since we make use of $E(\boldsymbol{\epsilon}\epsilon'|\mathbf{X}) = \sigma^{\mathbf{2}}\mathbf{I}$ in the proof of the Gauss-Markov Theorem)

Fact 4. They are still *typically* asymptotically normal under certain special conditions. The versions of the Central Limit theorems needed to prove this claim are much more sophisicated than the Lindberg-Levy version we have seen before (for heteroskedasticity we need the so-called Lindberg-Feller Theorem and for autocorrelation we need the so called Gordin's Theorem (see Greene for instance, for their statements)). We will not pursue this topic. However, we should mention that even if asymptotic normality holds, the standard t-statistics reported for the coefficients are no longer valid as the (asymptotic) variance covariance matrix of $\mathbf{b_{OLS}}$ is no more $\sigma^2\mathbf{\Sigma_x}^{-\mathbf{1}}$. It is $plim(\frac{\mathbf{X'X}}{n})^{-1}\, plim(\frac{\mathbf{X'}\varepsilon\varepsilon'\mathbf{X}}{n})\, plim(\frac{\mathbf{X'X}}{n})^{-1}$.

When we have heteroskedasticity but the observations are independent, it is still possible to (consistently) estimate $Var(\mathbf{b})$. The estimated variance matrix was given by White in a famous 1980 paper and its expression is $n(\mathbf{X'X})^{-1}\left(\frac{1}{n}\sum_{i=1}^n e_i^2\mathbf{x_i x_i}'\right)(\mathbf{X'X})^{-1}$ (thus $A\hat{v}ar(\mathbf{b}) = (\frac{\mathbf{X'X}}{n})^{-1}\left(\frac{1}{n}\sum_{i=1}^n e_i^2\mathbf{x_i x_i}'\right)(\frac{\mathbf{X'X}}{n})^{-1}$) where e_i's are the residuals obtained after the OLS regression.[1] However, the White estimator often behaves poorly in small

[1] Recall that we had already discussed this in the context of large sample theory. We had also

samples. For better small sample performance it is sometime recommended that instead of using ${e_i}^2$ as a proxy of the variance of $\boldsymbol{\epsilon}_i$ one should use $\frac{{e_i}^2}{m_{ii}}$ where m_{ii} is the $i-th$ diagonal entry of the $\mathbf{M}$ matrix.

If the observations are not independent, the standard errors are typically calculated using the so-called Newey-West estimator. Recall that in order to calculate the asymptotic variance of the OLS estimator, we need two things: plim $\frac{1}{n}(\mathbf{X}'\mathbf{X})$ and plim $(\frac{\mathbf{X}'\varepsilon\varepsilon'\mathbf{X}}{n})$. As long as the regressors are stationary and ergodic, the first probability limit is consistently estimated using the sample averages. However, if we try to mimic the idea of the White estimator for the other probability limit by using $\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} e_i e_j \mathbf{x_i}\mathbf{x_j}'$ as an estimator, we run into the difficulty that the 'sandwich' matrix need not be positive definite and we may end up with negative values for the standard errors of some coefficients. Newey and West proposed the following estimator which gets around this problem:

$$\frac{1}{n}\sum_{i=1}^{n} e_i^2 \mathbf{x_i}\mathbf{x_i}' + \frac{1}{n}\sum_{i=1}^{p}\sum_{j=i+1}^{n} w_i e_j e_{j-i}(\mathbf{x_i}\mathbf{x_{j-i}}' + \mathbf{x_{j-i}}\mathbf{x_i}')$$

where the weights $w_i = 1 - \frac{i}{p+1}$ and the 'lag' p is typically chosen roughly as $n^{.25}$. There are other competing estimators for this matrix, but this one is the most standard estimator used and it is available as a Stata command (search under 'Newey').

2 Generalized Least Squares

If we know $\mathbf{\Omega}$, it is however trivial to obtain an unbiased efficient estimator of $\boldsymbol{\beta}$ (under some mild assumptions which is also consistent and asymptotically normal). The trick is to transform the present model to a model where all the OLS assumptions apply.

So consider the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \epsilon \tag{1}$$

where

$$E(\boldsymbol{\epsilon}\epsilon'|\mathbf{X}) = \sigma^{\mathbf{2}}\mathbf{\Omega} \tag{2}$$

One reason we write the right-hand side as $\sigma^2\mathbf{\Omega}$ rather than simply $\mathbf{\Omega}$, is that often in applications, the structure of a variance-covariance matrix is known except for a few unknown (parameters). Another reason is that the estimates of the coefficients will turn out to be invariant to any multiplicative constant in tha variance-covariance matrix, as we will shortly see.

mentioned that standard errors calculated this way are called 'robust' or White standard errors and they can be implemented in Stata by using the 'robust' option after your regression command (followed by a comma).

We now look for a n.s. matrix $\mathbf{P}$ such that when (**??**) is premultiplied by $\mathbf{P}$, we get a classical model where OLS assumptions apply.[2]

The new model is

$$(\mathbf{Py}) = (\mathbf{PX})\boldsymbol{\beta} + (\mathbf{P}\boldsymbol{\epsilon}) \tag{3}$$

Let $\tilde{\mathbf{y}} = \mathbf{Py}$ = new y, $\tilde{\mathbf{X}} = \mathbf{PX}$ = new $\mathbf{X}$, $\tilde{\boldsymbol{\epsilon}} = \ \mathbf{P}\boldsymbol{\epsilon}$ = new disturbance. For the new model to conform to the classical linear model, we need to ensure that $E(\mathbf{P}\epsilon\epsilon'\mathbf{P}'|\mathbf{X}) = \sigma^{\mathbf{2}}\mathbf{I}$. Since the left hand side is $\mathbf{PE}(\epsilon\epsilon'|\mathbf{X})\mathbf{P}' = \sigma^{\mathbf{2}}\mathbf{P\Omega P}'$, we need to find $\mathbf{P}$, such that $\mathbf{P\Omega P}' = \mathbf{I}$.

However, such a $\mathbf{P}$ is easily seen to be $\boldsymbol{\Omega}^{-\mathbf{1/2}}$ (Since $\boldsymbol{\Omega}$ is symmetric p.d., it has a symmetric inverse which has a symmetric square root). Hence, an unbiased and efficient (by virtue of Gauss-Markov) estimator of $\boldsymbol{\beta}$ is

$$\mathbf{b}_{GLS} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}}$$

$$= (\mathbf{X}'\mathbf{P}'\mathbf{PX})^{-\mathbf{1}}(\mathbf{X}'\mathbf{P}'\mathbf{Py})$$

$$= (\mathbf{X}'\boldsymbol{\Omega}^{-\mathbf{1}}\mathbf{X})^{-\mathbf{1}}(\mathbf{X}'\boldsymbol{\Omega}^{-\mathbf{1}}\mathbf{y}) \text{ where we use } \mathbf{P}' = \mathbf{P} = \boldsymbol{\Omega}^{-\mathbf{1/2}} \tag{4}$$

To prove consistency and asymptotic normality, one needs to assume $plim\,\frac{1}{n}\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$ is an invertible matrix (in addition to orthogonality between the new regressors and the disturbance); once this is assumed the results follow using the techniques in the chapter on large sample theory. You should also be able to show that $\text{Var}(\mathbf{b_{GLS}}) = \sigma^{\mathbf{2}}(\mathbf{X}'\boldsymbol{\Omega}^{-\mathbf{1}}\mathbf{X})^{-\mathbf{1}}$

So far, we haven't said anything about estimating σ^2 and $\boldsymbol{\Omega}$ both of which may be unknown. If $\boldsymbol{\Omega}$ is known and only σ^2 is to be estimated, OLS ideas apply - an estimate is given by

$$\hat{\sigma}^2 = \frac{(\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\mathbf{b}_{GLS})'(\tilde{\mathbf{y}} - \tilde{\mathbf{X}}b_{GLS})}{n-k}$$

$$= \frac{(\mathbf{y} - \mathbf{X}\mathbf{b}_{GLS})'\boldsymbol{\Omega}^{-1}(\mathbf{y} - \mathbf{X}\mathbf{b}_{GLS})}{n-k} \tag{5}$$

If $\boldsymbol{\Omega}$ is to be estimated, some structure must be imposed on it (otherwise, there are too many parameters that need to be determined!). If $\boldsymbol{\Omega} = \boldsymbol{\Omega}(\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is a 'reduced set

[2] $\mathbf{P}$ is a matrix of constants, i.e. it should not have X's or y's in it.

of parameters' (such as ρ in the $AR(1)$ in the example), and if a consistent estimator of $\hat{\boldsymbol{\theta}}$ is available, we simply replace $\boldsymbol{\Omega}$ by $\boldsymbol{\Omega}(\hat{\boldsymbol{\theta}})$ in (**??**) and (**??**). This method is called FGLS (Feasible Generalized Least Squares). Often, this is done in two stages. First, one uses OLS to find a consistent estimator. Then, one uses the implied errors to estimate θ. Then the GLS estimator is calculated by using equation **??** with the estimated Ω in there.

3 Heteroskedasticity

From our previous discussion about GLS we had learnt that presence of nonspherical disturbances (i.e. $E(\boldsymbol{\epsilon\epsilon}'|\mathbf{X}) \neq \sigma^2\mathbf{I}$) causes inefficiency for OLS estimates. In this topic, we discuss how to detect the presence of such disturbances and what remedial action(s) may be taken to correct for them. We concentrate on heteroskedasticity first which posits that $E(\boldsymbol{\epsilon\epsilon}'|\mathbf{X})$ is a diagonal matrix. The implied assumption that non-diagonal terms are zero is often deemed to be valid for cross-sectional data.

The econometrician should start out by asking himself whether his model is likely to admit the possibility of heteroskedasticity and if so through what channels. For instance, if one is using price of houses as dependent variable, it is conceivable that a priori, if the mean or expected price of a house is small then its variance will also be small. Consider the following example due to Sen and Srivastava (1990)[3]: It would appear less likely that a house worth $\$250,000$ would sell for $\$300,000$ than that a million dollar one would sell for $\$1050,000$. Also, it appears to be more likely that the less expensive house would sell for $\$300,000$ than the more expensive one for $1.2 million. Thus, the standard deviation of the selling price is not constant, nor does it vary in proportion to intrinsic value. Rather, it is something in between. But whatever it is, one might suspect that $var(y)$ could be some function of $E(y)$.

The variance of the left hand side may depend on specific variables which may or may not be part of the explanatory variables for the expected value of the left hand side variable. For example, suppose you have collected data on money patients have spent towards medical charges in a certain hospital. You might have data on the patient's age, sex, identity of physician in charge, severity of illness (categorized in a scale of $1-4$). etc. While you may have reasons to believe why and how expected charges relate to these variables, could you identify variables which affect the dispersion of the dependent variable? Two such variables are age and the identity of physician (why?).

Once you are suspicious about heteroskedasticity, it is time to do some preliminary checking. A good idea is to examine the plot of the residuals both against $\hat{y}$ and against

[3] Regression Analysis: A. Sen and M. Srivastava (1990) Springer Verlag

other explanatory variables. If the observed scatter 'fans out' or 'fans in' we may wish to examine the issue further by conducting a formal test. Note, however, that eyeballing the scatter is never a full-proof way to be aware heteroskedasticity because theoretically e_i and e_j (error terms corresponding to two observations) may have different variances even when ϵ_i and ϵ_j don't and they may have equal (roughly) variances even when ϵ_i and ϵ_j have different variances - it all depends on the $\mathbf{X}$ matrix. This should make sense, as $\mathbf{e} = \mathbf{M}\epsilon$.

3.1 Tests for Heteroskedasticity

We will discuss 3 tests of heteroskedasticity which vary according to generality of the setup as well as power. Recall that power of a test is probability of rejecting H_0 when it is indeed false ($=$ $1-$Prob(type II error)).

3.1.1 White's Test

This is the most general and least powerful of the 3 tests. It has the following steps:

Step 1: Estimate model using OLS; save residuals e_i.

Step 2: Regress the squared residuals against a constant , the regressors used in Step 1 and all squared and cross products of the regressors, e.g. if the model used in Step 1 is

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i,$$

then the regression in the second step will be of the form:

$$e_i^2 = \alpha_1 + \alpha_2 X_{i2} + \alpha_3 X_{i3} + \alpha_4 X_{i2}^2 + \alpha_5 X_{i3}^2 + \alpha_6 X_{i2} X_{i3} + u_i$$

Step 3: Reject H_0 at $\alpha\%$ if nR^2 (where R^2 is the coefficient of determination from the last regression and n sample size) is $>$ $\chi^2_\alpha(k-1)$ where k is the # of regressors (including 1) in the Step 2 (auxiliary) regression (Thus, $k-1=5$ in the example).

Note that the White test is nonconstructive, i.e. it gives us no indication as to which variables are causing heteroskedasticity.

3.1.2 Goldfeld-Quandt Test

This is less general than White's test, but more powerful. To perform this test, we must be able to identify a certain variable which is 'causing' heteroskedasticity (this variable may or may not be part of the regressors). The steps of this test are:

Step 1: Order the observations by the size of Z, the variable that is assumed to be related to the size of the error variance.

Step 2: Leave out C 'middle' values and thereby create 2 subsamples of sizes $\frac{n-c}{2}$. The subsample sizes should exceed k and at the same time each subsample should have enough observations (dropping more than a third of the sample is discouraged).

Step 3: Run OLS models on both subsamples. Let RSS_1 be the residual sum of squares associated w/ the subsample of larger variances. Let RSS_2 be the residual sum of squares associated w/ the subsample of smaller variances.

The test statistic is

$$\frac{RSS_1}{RSS_2} \text{ and it is distributed (under null) as } F_{\frac{n-c}{2}-k,\ \frac{n-c}{2}-k}$$

3.1.3 Breusch-Pagan Test

This is the most powerful and the least general of the 3 tests. The premise here is that $\sigma_i^2 = \sigma^2 f(\alpha' Z)$ where Z_i is a vector of independent variables supposedly causing heteroskedasticity.

Step 1: Estimate regression model

$$Y_i = \beta_1 + \beta_2 X_{i2} + ... + \beta_k X_{ik} + u_i$$

and obtain OLS residuals e_i.

Step 2: Form ML estimate of error variance $\tilde{\sigma}^2$

$$\begin{aligned} \text{Define } \tilde{\sigma}^2 &= \frac{e'e}{n} \\ \text{Generate } V_i &= \frac{e_i^2}{\tilde{\sigma}^2} \end{aligned}$$

Step 3: Regress V_i on Z_i (a vector). Obtain $(1/2)ESS$. Under null, $(1/2)ESS \overset{a}{\sim} \chi^2(p-1)$ where p is the number of vars. in Z_i.

3.2 Estimation

If one has no clue about what is causing heteroskedasticity, it is best to use OLS estimates (recall they are unbiased and consistent!) and for testing purposes only estimate $\sigma^2(X'\Omega X)^{-1}$ matrix using the so-called White heteroskedasticity estimator as has been discussed before.

If there is strong suspicion that $var(y_i)$ depends only on $E(y_i)$, some suitable variance-stabilizing transformation may do the trick. Otherwise, two methods are suggested: Weighted Least Squares (via FGLS) and ML estimation.

3.2.1 Weighted Least Squares

Suppose we have a model $y_i = \mathbf{x}_i'\boldsymbol{\beta} + \epsilon_i$ and while $E(\epsilon_i|\mathbf{x}_i) = 0$, $\mathrm{Var}(\epsilon_i|\mathbf{x}_i) = \sigma_i^2$ with $\sigma_i^2 \neq \sigma_j^2$. Assume that it is possible to obtain consistent estimates of these variances. Then we can rewrite our model as $\frac{y_i}{\sigma_i} = \frac{\mathbf{x}_i'}{\sigma_i}\beta + \frac{\epsilon_i}{\sigma_i}$. This model now obeys all OLS conditions and hence estimating it by OLS is not only unbiased and consistent, but also efficient. Incidentally, this procedure is just equivalent to using the GLS estimator $(\mathbf{X'\Omega^{-1}X})^{-1}(\mathbf{X'\Omega^{-1}Y})$ where

$$\boldsymbol{\Omega} = \begin{bmatrix} \sigma_1^2 & & & & \\ & \sigma_2^2 & & 0 & \\ & & \cdot & & \\ & & & \cdot & \\ & 0 & & & \cdot \\ & & & & \sigma_n^2 \end{bmatrix}$$

This is also called weighted least squares (WLS) where $(\sigma_i^2)^{-1}$'s are the 'weights'. Once you specify the weighting variable, STATA automatically takes care of the appropriate estimation the right way; all you have to do is to append the option [aweight = varname] where varname $\propto (\sigma_i^2)^{-1}$.

If a model fails one or more tests of heteroskedasticity a standard practice is to carry out FGLS as follows:

Step 1: Run OLS, collect residuals e_i, generate e_i^2, $\ln(e_i^2)$.

Step 2: Regress $\ln e_i^2$ on Z_i (Z may include variables from X, if dealing with non-negative variables, Z often includes log of X variables, which can capture $\sigma_i^2 = X_i^\alpha$ type of relation)

Step 3: Obtain predicted values of e_i^2, use their reciprocals as weights in doing a WLS on the original regression.

3.2.2 ML Estimation

First we write the program for likelihood generator (assuming $\epsilon_i \sim N(0, \sigma_i^2)$).

```
Capture program drop heteromle
program define heteromle
args lnf xbeta zalpha
quietly replace ‘lnf’ =-‘zalpha’-0.5* ((\$ML_y1-‘xbeta’)/(exp(‘zalpha’)))^2$
end
```

[Recall $f(y|x,z) = \frac{1}{\sqrt{2\pi}e^{z_i\alpha}} \exp\left\{-\frac{1}{2}\left(\frac{y_i - x_i'\beta}{\exp(z_i'\alpha)}\right)^2\right\}$. Hence, ignoring the constant $\sqrt{2\pi}$, $\ln\, L_i = -z_i'\alpha - 0.5\left(\frac{y_i - x_i'\beta}{\exp(z_i'\alpha)}\right)^2$]

Next, we issue a command of the type: “ml model lf heteromle (xbeta: avgexp=age income incsq ownrent) (zalpha : lninc)”

After doing an “ml check”, issuing “ml maximize” produces the needed estimates.

4 Autocorrelation

If $E(\epsilon_i\epsilon_j|\mathbf{X}) \neq 0$ for $i \neq j$, we say that error terms are autocorrelated. As I have mentioned before, time series data are notorious for autocorrelation, but sometime one can also have 'spatial autocorrelation'. This occurs when observational units are in geographical proximity and neighborhood units tend to mimic one another. Autocorrelation patterns can sometime be eyeballed from residual patterns, with negative autocorrelation associated with lots of crossing over the zero line by the residuals, and positive autocorrelations associated with residuals staying negative (or positive) for a while. But beware though that this last pattern can also be caused by nonlinear relationsship between y and one of the regressor if the data are sorted by ascending or descending values of that regressor.

4.1 Tests

4.1.1 Durbin-Watson

The classic test for autocorrelation is the Durbin-Watson test, which is out of favor these days. The DW statistic test

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

is supposed to be small for positive autocorrelation and large for negative autocorrelation (which makes intuitive sense, if errors are positively correlated when e_{t-1} is large and positive you expect e_t to be also large and positive).

Expanding the above,

$$d = \frac{\sum_{t=2}^{n} e_t^2 + \sum_{t=2}^{n} e_{t-1}^2 - 2\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2} \simeq 2(1-\phi)$$

where ϕ is the correlation coefficient between $\epsilon_t, \epsilon_{t-1}$: Hence, d ranges from 0 to 4, if d is much greater than 2 that indicates negative autocorrelation and if it is much less than 2, that indicates positive autocorrelation.

So far as actual testing is concerned, unfortunately the actual distribution depends on the x's. So DW offered lower and upper bounds for the critical regions (d_L & d_U), which can be obtained from tables.

if $d < d_L$ the test rejects null in favor of +ve autocorrelation.

if $d > d_U$ the test doesn't reject

if $d_L < d < d_U$, the test is inconclusive.

To test for negative autocorrelation, one calculates $4 - d$ and proceeds similarly as if one is testing for +ve autocorrelation.

NOTE: For DW test

a. A constant term must be included.

b. The test is strictly valid only for nonstochastic regressors. Also, if lagged dependent variables are included, the test is biased.

4.1.2 Durbin's h (appropriate when the rhs contains lagged variables)

The steps for this test (for testing for AR(1) errors) are:

1. Estimate OLS regression

$$y_t = \beta_1 y_{t-1} + ... + \beta_\gamma y_{t-\gamma} + \beta_{\gamma+1} x_{1t} + ... + \beta_{\gamma+s} x_{st} + \epsilon_t$$

Obtain e_t.

2. Regress e_t on $e_{t-1},\ y_{t-1},\ ...,\ y_{t-\gamma},\ x_{1t},\ ...,\ x_{st}$. Reject H_0 if the coeff. on e_{t-1} is significant at the appropriate level.

This can be extended to test $p-$th order autocorrelation. Simply include $e_{t-1},\ e_{t-2},\ ...,\ e_{t-p}$ as explanatory regressors in the second regression, and test the <u>joint</u> significance of the lagged error coefficients.

4.1.3 Breusch-Godfrey

This test is appropriate for testing the null hypothesis of no autocorrelation versus the alternative of an AR(p) disturbance process (it also works for MA processes).

Let $\mathbf{e} = \mathbf{y} - \mathbf{Xb}$, be the $n \times 1$ OLS residuals, $\hat{\sigma}^2 = \frac{\mathbf{e'e}}{n}$ the ML estimate of $\sigma^2_{\epsilon_i}$. Define

$$\mathbf{E}_p = \begin{bmatrix} 0 & 0 & . & . & . & 0 \\ e_1 & 0 & . & . & . & 0 \\ e_2 & e_1 & . & . & . & 0 \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ e_{n-1} & e_{n-2} & . & . & . & e_{n-p} \end{bmatrix}$$

The test statistic is

$$t = \mathbf{e'E_p}\left[\mathbf{E'_pE_p} - \mathbf{E'_pX(X'X)^{-1}X'E_p}\right]^{-1}\mathbf{E'_pe}/\hat{\sigma}^2$$

and the test statistic is χ^2 distributed with p degrees of freedom.

Operationally these are the steps:

1. Fit the OLS regression of Y on X, obtain e.

2. Regress e on $[E_p,\ X]$; obtain R^2.

3. Refer nR^2 to $\chi^2(p)$.

STATA: In Stata, use 'estat dwatson', 'estat durbinalt' and 'estat bgodfrey' to obtain the test statistics for these three tests.

4.2 Correcting For Autocorrelation

Before one goes about correcting for autocorrelation,.one must be warned that after often autocorrelation in estimated models (particularly w/ time series data) appear not because the disturbance is not white noise, but because the model is misspecified. For example, if the true model is

$$y_t = \gamma_1 + \gamma_2 x_t + \gamma_3 x_{t-1} + \gamma_4 y_{t-1} + \epsilon_t \tag{6}$$

then specifying a model such as

$$y_t = \beta_1 + \beta_2 x_t + u_t \tag{7}$$

will obviously result in signs of autocorrelation (since intuitively a high u_t in model (**??**) will lead to high y_t which'll be 'represented' in u_{t+1}, which (assuming $\gamma_4 > 0$) could cause u_{t+1} to be positively correlated with u_t).

Appending a relation such as

$$u_t = \varphi u_{t-1} + \epsilon_t \tag{8}$$

is not a solution since (**??**) & (**??**) together imply

$$y_t = \beta_1 + \beta_2 x_t + \varphi(y_{t-1} - \beta_1 - \beta_2 x_{t-1}) + \epsilon_t$$

$$= \beta(1-\varphi) + \beta_2 x_t - \varphi\beta_2 x_{t-1} + \varphi y_{t-1} + \epsilon_t \tag{9}$$

This is equivalent to imposing the (false) restriction $\gamma_3 + \gamma_2\gamma_4 = 0$ on the true model (**??**).

4.3 Estimation

If one needs a quick and dirty solution with $AR(1)$ errors, one should employ some version of iterative FGLS. Under the assumption that the error term is $AR(1)$the covariance matrix is

$$\sigma^2\mathbf{\Omega} = {\sigma_\varepsilon}^2 \begin{bmatrix} 1 & \rho & \rho^2 & . & . & \rho^{n-1} \\ \rho & 1 & \rho & . & . & \rho^{n-2} \\ . & . & . & & & . \\ . & . & . & . & & . \\ . & . & . & & . & . \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & & & 1 \end{bmatrix}$$

Writing the process as $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$ with u_t being iid with variance σ^2, we have (exploiting the stationarity of the process), ${\sigma_\varepsilon}^2 = \frac{{\sigma_u}^2}{1-\rho^2}$, when the above can be written as ${\sigma_u}^2\mathbf{\Gamma}$, where

$$\mathbf{\Gamma} = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & . & . & \rho^{n-1} \\ \rho & 1 & \rho & . & . & \rho^{n-2} \\ . & . & . & & & . \\ . & . & . & . & & . \\ . & . & . & & . & . \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & & & 1 \end{bmatrix}$$

It may be verified that in this case

$$\mathbf{\Gamma}^{-1} = \begin{bmatrix} 1 & -\rho & 0 & . & . & . & 0 & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & . & . & . & 0 & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & . & . & . & 0 & 0 & 0 \\ . & . & . & & & & . & . & . \\ . & . & . & & & & . & . & . \\ . & . & . & & & & . & . & . \\ 0 & 0 & 0 & . & . & . & -\rho & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & . & . & . & 0 & -\rho & 1 \end{bmatrix}$$

If one had an estimate of ρ in $\hat{\rho}$ we could use FGLS: $\mathbf{b}_{FGLS} = (\mathbf{X'\hat{\Omega}^{-1}X})^{-\mathbf{1}}(\mathbf{X'\hat{\Omega}^{-1}Y})$. (Note that given the expression of the estimator, the fact that we don't know σ_u will not matter). Practically, the way an estimate of ρ is obtained is by using the OLS errors to run a regression of the form $e_t = \rho e_{t-1} + \epsilon_t$, use the estimated $\hat{\rho}$ to get FGLS and then iterate the whole procedure using the new errors. This is the **Prais-Winsten** procedure.

The Cochrane-Orcutt procedure is a simpler version of the Prais-Winsten procedure.

Rewrite the model

$$y_t = x_t'\beta + u_t,\ u_t = \rho u_{t-1} + \epsilon_t$$

as

$$y_t - \rho y_{t-1} = \beta(x_t - \rho x_{t-1}) + \epsilon_t \tag{10}$$

If we have an estimate of ρ ($\hat{\rho}$), we can construct the variables $y_t - \hat{\rho} y_{t-1}$ & $x_t - \hat{\rho} x_{t-1}$ and OLS can be used to determine estimates of β. In practice, we start with a guess and reiterate the procedure using residuals obtained from the previous step.

STATA: To perform the prais-winsten procedure, you first create a time variable. If you already have one, its fine, otherwise issue the command: "gen time=_n"; and then issue a tsset command "tsset time". Next, you issue "prais depvar $x1$ $x2$ ... xn" etc.

To perform Cochran-Orcutt, append the option 'corc' to the 'prais' command. Note that these tests are testing for first-order autocorrelation only.