HANDOUT 02

AUTOCORRELATION

Outline of autocorrelation lecture:

A Simple Message to
Autocorrelation Correctors: Don't!
*Mizon (1995)

The model with a correction for autocorrelation is a restriction on a more general model with lagged values of both dependent and independent variables. We considered a means of testing this specification as an alternative to "fixing" the problem of autocorrelation.
**Greene (2012, p.979)

…Indeed, most econometricians would now totally reject the use of the GLS procedure. Serial correlation can arise from a number of problems which may be readily solved in some cases.
***Cameron (2005, p.252)

In fact, with time series data, autocorrelated residuals are, much more often than not, an indication of some error in the way we have specified the regression equation rather than genuine autocorrelation in disturbances.
****Thomas (1997, p.307)

I. Introduction

The term autocorrelation can be defined as correlation between values of a series in *time* or *space*.[1]

Recall the "no autocorrelation" assumption of classical linear regression model:

- **Assumption 5 (A5)**: *No autocorrelation*[2] or *zero covariance* between u_t and u_s [i.e., $Cov(u_t, u_s) = 0$]. This assumption can be made stronger by assuming that the values of u_t are all statistically *independent*[3].

[1] This handout is heavily based on the lecture notes of Haluk Erlat (Erlat, 1997).
[2] "No autocorrelation" means that the *correlation* between any u_t and u_s ($t \neq s$) is zero.
[3] If the disturbance terms are statistically independent, the value which the disturbance term takes in one period does not depend on the value which it takes in any other period. However, remember that zero correlation does not always imply independence. Suppose u_t is a normally-distributed (so it has a symmetric distribution) random variable with zero mean. Hence $E(u_t) = 0$ and $E(u_t^3) = 0$ since $E(u_t^3)$ is the third moment about the mean and if any distribution is symmetric the third moment about the mean must be zero. Now let $u_s = u_t^2$.

- This condition states that there should be no systematic association between the values of the disturbance term in any two observations.
- If this condition is not satisfied, OLS will again give ***inefficient*** estimates.
- Recall that *correlation* between any random variables *G* and *H* is given by:

$$\rho = \frac{Cov(G,H)}{\sigma_G \sigma_H}$$

where σ_G and σ_H are standard deviations of *G* and *H*, respectively. Therefore, if the correlation between *G* and *H* is zero ($\rho = 0$), it implies that $Cov(G,H)$.

As a result, the "*no autocorrelation*" assumption implies that $Cov(u_t, u_s) = 0$ where $t \neq s$.[4]

- Note that covariance between X and Y is given by

$$Cov(X,Y) = E\left\{\left[X - E(X)\right]\left[Y - E(Y)\right]\right\}$$

Therefore, *no autocorrelation* implies *zero covariance*:

Clearly u_t and u_s are not independent: if you know u_t, you also know u_s. And if you know u_s, you know the absolute value of u_t. The covariance of u_t and u_s is

$$Cov(u_t, u_s) = E(u_t.u_s) - E(u_t)E(u_s) = E(u_t^3) - 0.E(u_s) = 0 - 0.E(u_s) = 0$$

Here the correlation $\rho = 0$ since $\rho = \frac{Cov(u_t, u_s)}{\sqrt{Var(u_t)Var(u_S)}}$. This situation is an example where the variables are not independent, yet they have zero (linear) correlation, $\rho = 0$. If u_t and u_s are statistically independent then, one can always write this identity: $E(u_t.u_s) = E(u_t)E(u_s)$, which always implies zero covariance and hence zero correlation. However, as can be seen from our example that zero covariance or correlation do not always imply statistical independence.

[4] Recall that a zero value of the covariance indicates no linear dependence between *G* and *H*.

$$Cov(u_t,u_s)=E\{[u_t-E(u_t)][u_s-E(u_s)]\}=0$$

From *Assumption 3*, we know that $E(u_t)=0$ and $E(u_s)=0$. Thus, no autocorrelation assumption implies that

$$Cov(u_t,u_s)=E\left\{\left[u_t-\underbrace{E(u_t)}_{0}\right]\left[u_s-\underbrace{E(u_s)}_{0}\right]\right\}=0\Rightarrow$$

$$E\left(u_tu_s\right)=0$$

II. Sources of Autocorrelation

The sources of autocorrelation can be distinguished into two broad categories (Erlat, 1997, pp.1-6). The first set of reasons is related to the nature of the data being used. The second set of reasons concern the specification of the systematic part of the regression equation (whether or not there is a misspecification in the model).

A. Genuine (Pure) Autocorrelation Reasons (Data-Based Reasons)

Following Thomas (1997, p.308) we call this type of autocorrelation as *Genuine Autocorrelation*.

Autocorrelation is usually encountered in the use of time series data. Following Erlat (1997), Gujarati (2011) and Kennedy (1998), one can list following sources which may result to autocorrelation within this context.

Inertia or business cycles One source is the behavior of most economic time series. Series such as GDP, Price indices, production, employment, unemployment exhibit business cycles. Starting at the bottom of the recession, when economic recovery starts, most of these series start moving upward. In this upward movement the value of a series at one point in time is greater than its previous value. Thus there is a "momentum" built into them and it continues until something happens (e.g., increase in interest rate, taxes, exchange rate, etc.) Therefore, in regression using time series data, successive observations are likely to be dependent (Gujarati, 2011, p.414).

Prolonged influence of shocks In time series data, random shocks (disturbances) have effects that often persist over more than one time period. An earthquake, flood, strike or war, global economic crisis, policy shock, will probably affect the economy's operation in periods following the period in which it occurs (Kennedy, 19998, pp.121-122)

To see how prolonged influence of shocks result to autocorrelation suppose that we have an equation that relates the aggregate demand for money in the economy to a number of explanatory variables. Any policy shock that occurs will have an impact on money demanded through the error term. Also, a shock usually takes several periods to work through the system. This means that, in any one period, the current error term contains not only the effects of current shocks but also the carryover from the previous shocks. This carryover will be related to, or *correlated with*, the effects of earlier shocks (Hill, Griffiths and Judge, 2001, p.258).

Another simple example can be as follows: A major road improvement project might reduce traffic accidents not only in the year of completion but also in future years. Such impacts which persist over multiple years, produce autocorrelated disturbances (Stock and Watson, 2012, p.406)

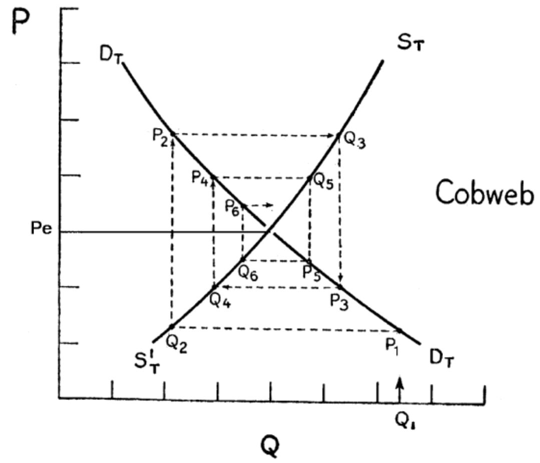
Spatial autocorrelation[5] In regional cross-sectional data, a random shock affecting economic activity in one region may cause economic activity in the near (or adjacent) region to change due to close economic relations between the regions. Shocks due to weather

[5] **Autocorrelation in cross-section data** If we are using a cross-section data which is ordered and carry out a Durbin-Watson autocorrelation test then it will be a test for heteroscedasticity rather than for autocorrelation. We can only have (genuine) autocorrelation in cross section data if it has a spatial dimension. One example of this would be if we were to try to model burglary rates across regions of a city. We might find that one region is "exporting" or "importing" burglaries from a contiguous region. It is possible to construct tests by comparing pairs of residuals from contiguous regions to give us a test of *spatial autocorrelation* (Cameron, 2005, p.252)

similarities might also be tend to cause the error terms between near (adjacent) regions to be related (Kennedy, 1998, p.121).

Cobweb Phenomenon Another source is the *Cobweb phenomenon*. This is usually encountered in the agricultural sector where the production decisions react to changes in price with a lag of production season. For example, if price at period *t* is less than the price at *t-1*, then the production in period *t+1* is expected to decrease. In this case the disturbance in period *t+1* will be affected by the disturbance at period *t*.

This process will lead to fluctuations in price around an equilibrium or expected value [P_e or $E(P)$] so that for a given point in time *t*, the disturbances will alternate in sign.



Source: Adapted from Ezekiel (1938)

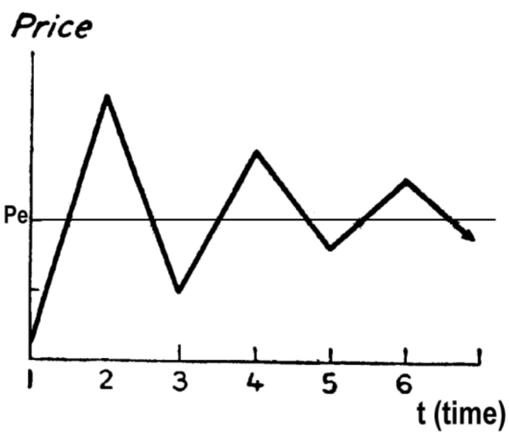


Figure 1 Cobweb Theorem (Convergent Cobweb)

u_t

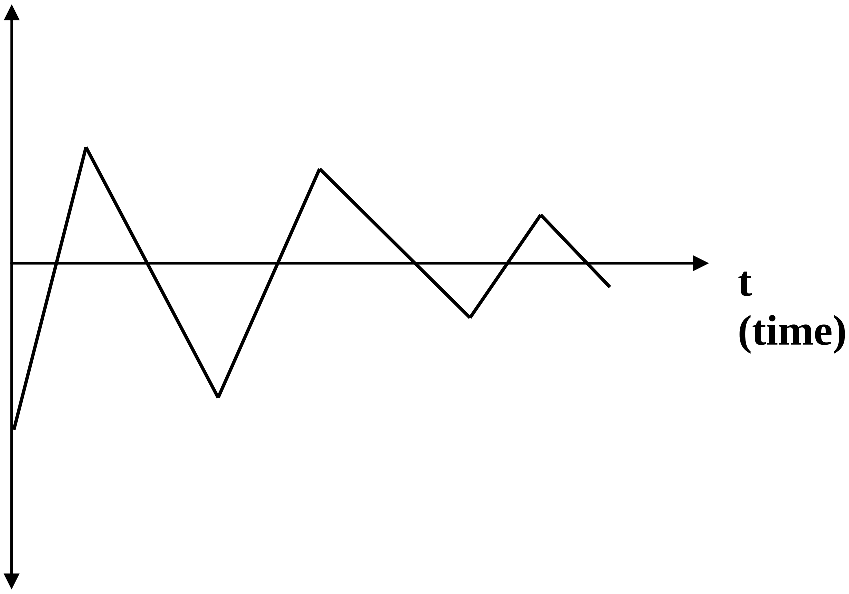


Figure 2 Disturbances [$u_t=P_t-E(P_t)$] in the Case of Covergent Cobweb

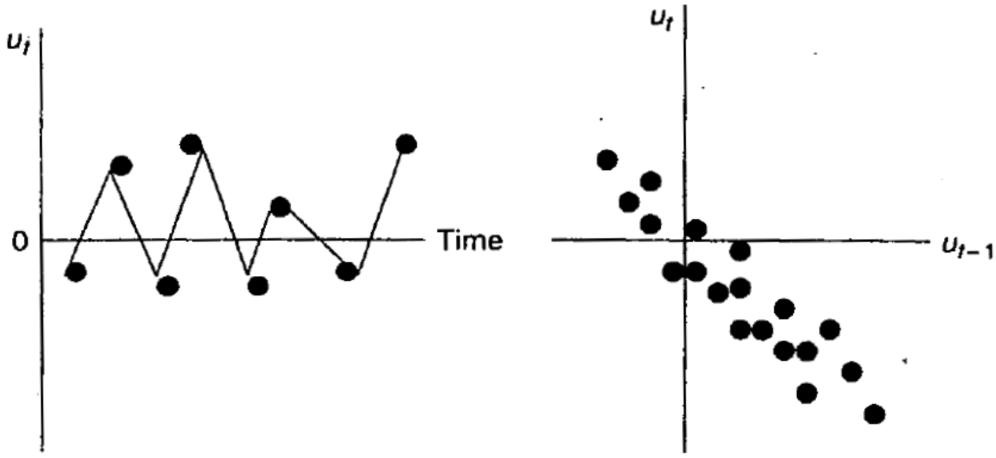


Figure Y. Negative Autocorrelation (Asteriou and Hall, p.151)

Data Manipulation or Transformation The last data-based source of autocorrelation can be stated as the manipulation of the raw data. For example, if quarterly data are generated by summing monthly observations (and sometimes by dividing by three); then this operation, by dampening the fluctuations in the monthly data, will generate a smoother quarterly data compared to the monthly data. Therefore, the graph plotting the quarterly data will look much smoother (showing less fluctuations) than the monthly data. This smoothness itself may lead to a systematic pattern in the disturbances and thereby autocorrelation. The examples can be seen in the following figures.

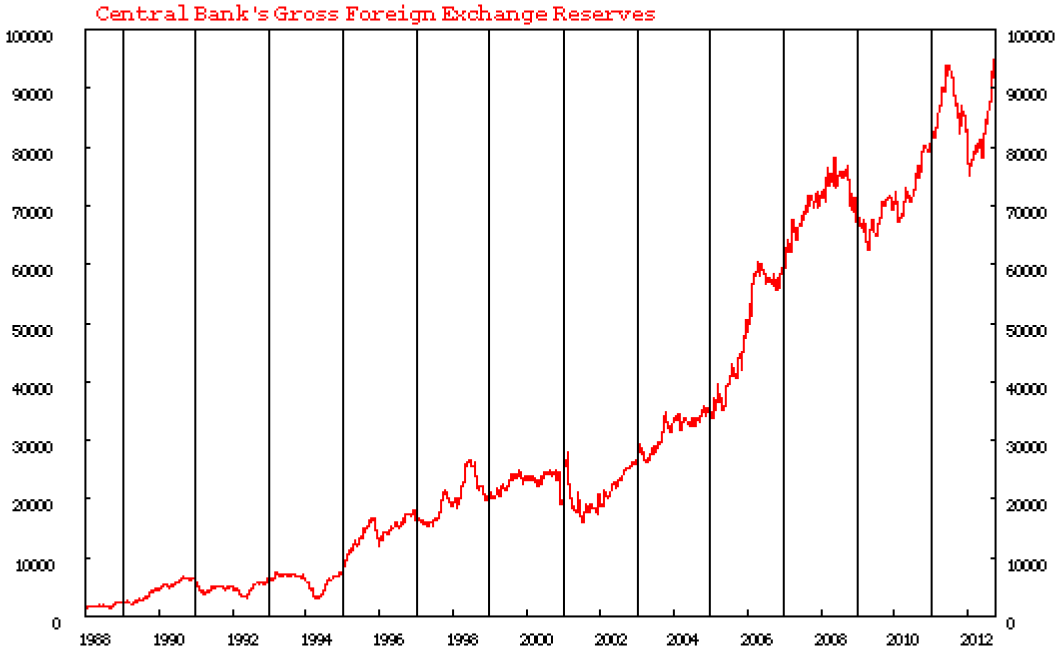


Figure 3a Central Bank Reserves (Monthly Data, Million USD)

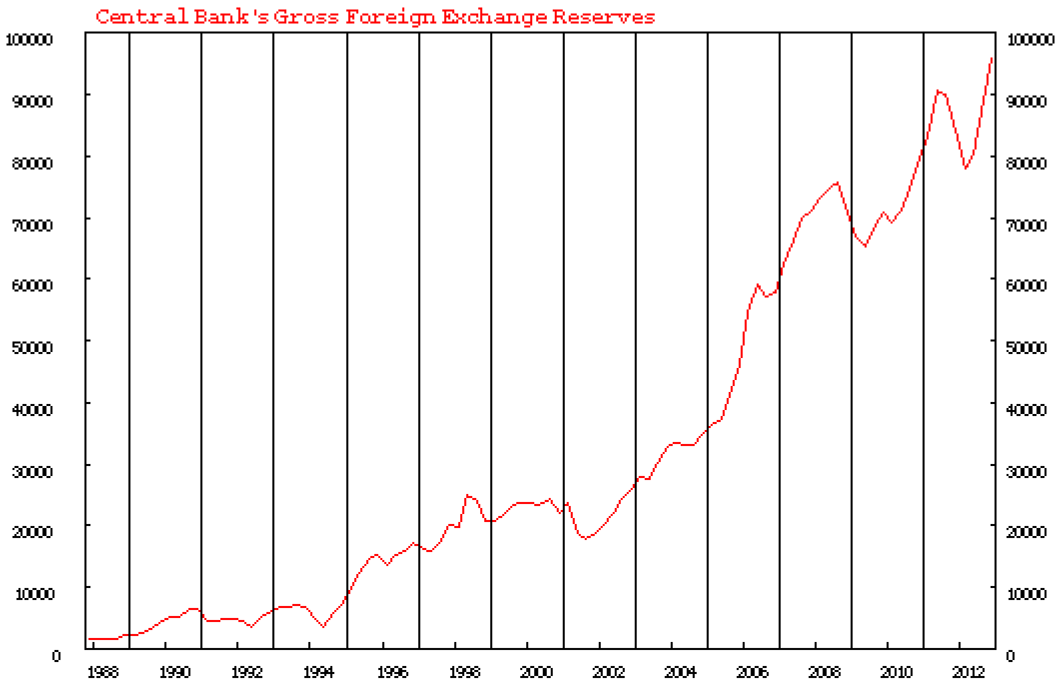


Figure 3b Central Bank Reserves (Quarterly Data, Million USD)

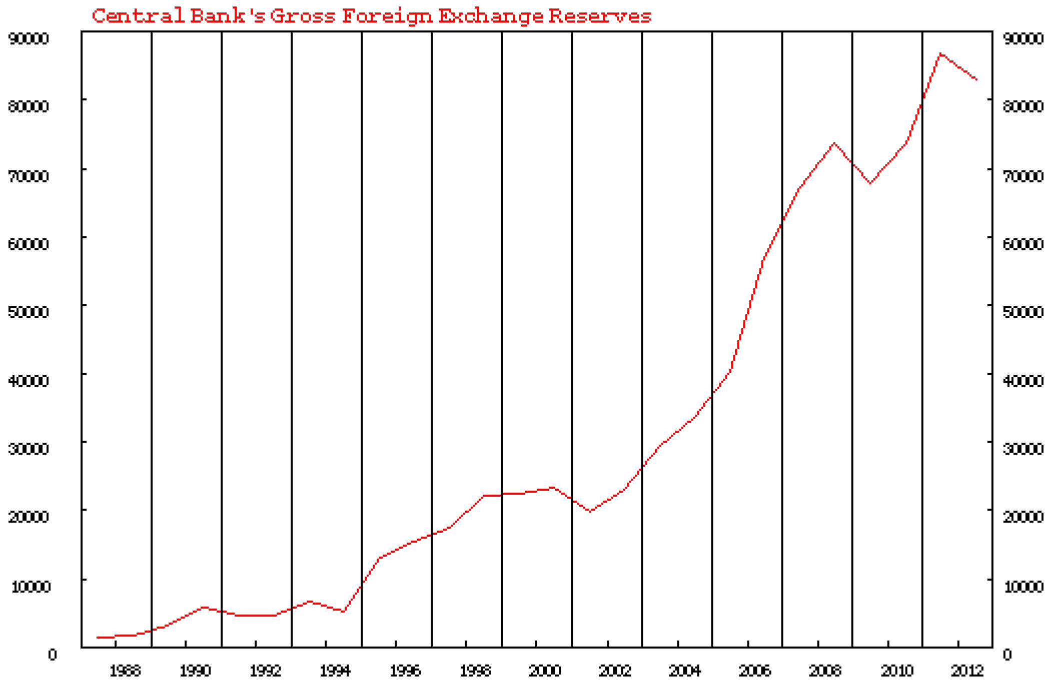


Figure 3c Central Bank Reserves (Annual Data, Million USD)

Two sorts of data problems can be categorized: (1) those that were in the data collected and (2) those arising from choices made by the researcher over variable definitions (Cameron, 2005, p.254).

- Perhaps surprisingly, problems may be caused by attempts to improve the accuracy of data by the recording agencies (such as Turkish Statistical Institute, TurkStat). Let us suppose that the government decides to make an effort to improve the accuracy of data on the variable we have chosen as our dependent variable. If this is a gradual process whereby, for example, data gets closer to the true value by say 5 percent of the gap at the start of each year, this will induce a false dynamic process in the Y variable.

- In addition, definitions of the rates of change may induce serial correlation in a series that was not originally serially correlated due to the form of differencing used. For example, defining rates of inflation on a "year on year" basis rather than a continuous basis or using data already adjusted by moving average methods may induce serial correlation.

B. Apparent (Impure) Autocorrelation Reasons (Misspecification-Based Reasons)

If the systematic part of the regression is misspecified this may result to autocorrelation in the disturbances. Following Dougherty (2007, p.372) we call this type of autocorrelation as *Apparent Autocorrelation (or Impure Autocorrelation)*.

- One example is omitted-variable case situation in which we leave out a relevant explanatory variable. This omitted explanatory variable may be a time series exhibiting a high degree of autocorrelation among its observations and this may be reflected as autocorrelation in the disturbances.

- Second example can be the specification error by ignoring a structural break in the dependent variable and this implies leaving out the dummy variables which account for this structural shift.

- Another common example is leaving out lagged values of the dependent variable when the dependent variable is a function of these lagged values together with other explanatory variables.

  - This systematic relationship between a dependent variable and its lagged values will operate through the disturbance term and will result to autocorrelation.

- Fourth misspecification may result from the use of an inappropriate functional form for the systematic part of the Population Regression Function (PRF).
  - For example suppose that the correct functional form is as follows:

$$Y_t = b_0 + b_1 X_t + b_1 X_t^2 + u_t$$

But we estimate this model:

$$Y_t = b_0 + b_1 X_t + v_t$$

where obviously $v_t = b_1 X_t^2 + u_t$ so the disturbances v_t may be autocorrelated. This can be seen in the following figures. The disturbances which are defined as $v_t = Y_t - E(Y_t)$ exhibits a definite pattern. They first take on positive values then become negative and then again positive.

- The pattern of autocorrelation plotted in Figure 4b where the *disturbances show few sign changes*, *is called positive autocorrelation*.

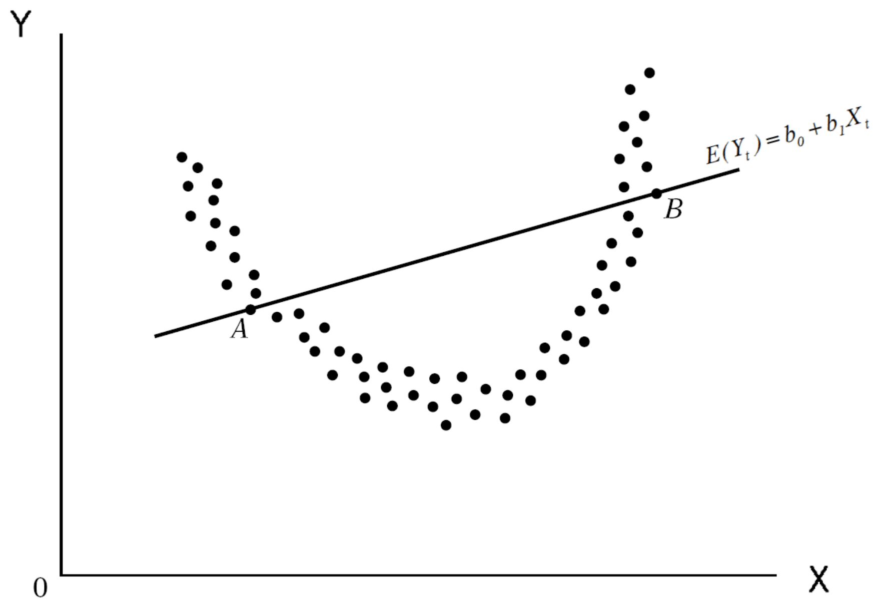


Figure 4a Incorrect Functional Form

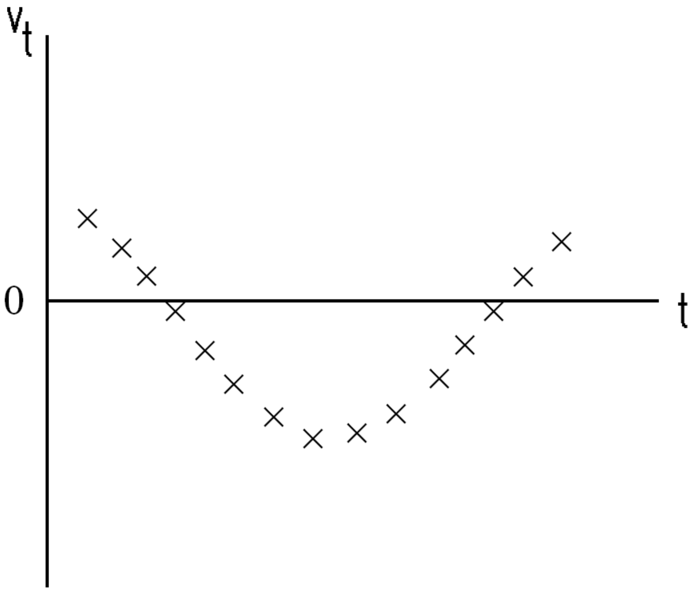


Figure 3b Incorrect Functional Form (Disturbance terms)

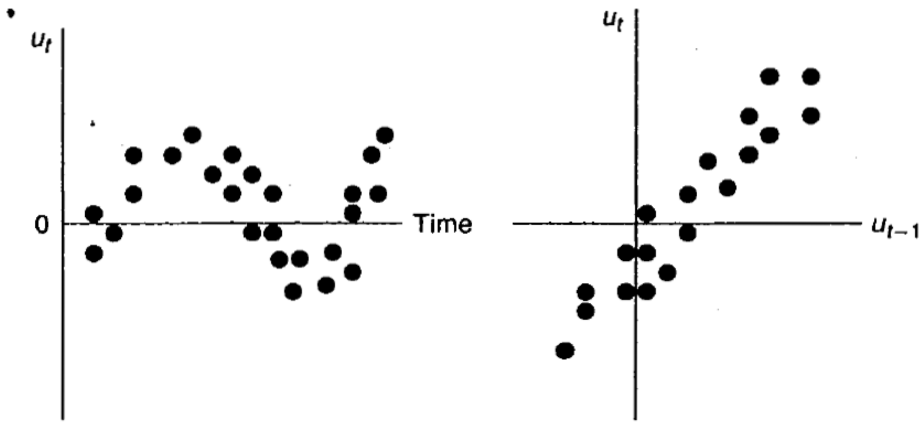


Figure X. Positive Autocorrelation (Asteriou and Hall, p.151)

- If the disturbances exhibit *frequent sign changes, as in Figure 2, then it is said that there is negative autocorrelation*.

Note that both forms of autocorrelation (positive or negative) can be observed whatever the sources of autocorrelation.

However, note here that, the cure of autocorrelation will be radically different depending upon the source of autocorrelation. If the source belongs to the second category (misspecification of systematic part of regression) then the solution to the autocorrelation problem is correct the specification problem.

III. Consequences of Autocorrelation

The consequences of autocorrelation on the OLS estimates can be summarized as follows:

(1) *The OLS estimators are still unbiased and consistent.*
- This is because both unbiasedness and consistency do not depend on assumption of $Cov(u_t, u_s) = 0, t \neq s$ which is in case of autocorrelation is violated.

(2) *The OLS estimators will be inefficient and therefore no longer BLUE.*
- If we are able to correctly model the autocorrelated errors, then there exists an alternative estimator with a lower variance. Having a lower variance means there is a higher probability of obtaining a coefficient estimate close to its true value. It also means that hypothesis tests have greater power and a lower probability of a Type II error.

(3) *The formulas for the standard errors usually computed for the OLS estimator are no longer correct, and hence confidence intervals and hypothesis tests that use these standard errors may be misleading.*
- The estimated variances of the regression coefficients based on OLS estimates will be biased and inconsistent, and therefore hypothesis testing is no longer valid. In most of the cases, R^2 will be overestimated (indicating a better fit than the one that truly exists) and the t-statistics will tend to be higher (indicating higher significance of our estimates than the correct one).

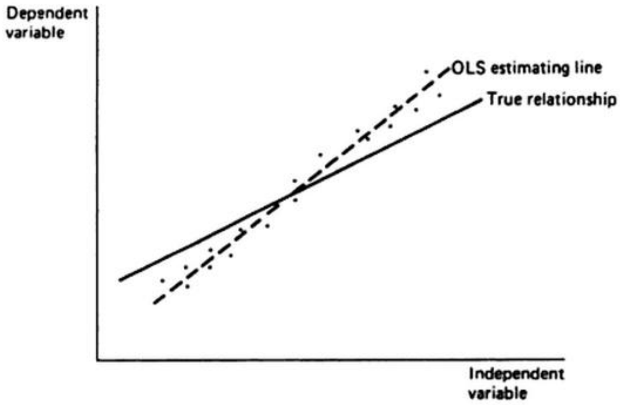


Figure Z. Better fit than the true one in AC case (Kennedy)

IV. What to Do When We Find Autocorrelation

Suppose that we find autocorrelation in our model. What will we do?

We have 3 options:

1) Try to find out if the autocorrelation is genuine (or pure autocorrelation) autocorrelation, that is, not the result of any model misspecification.

2) If you have been convinced that it is genuine (pure) autocorrelation,

    1) In large samples ($T \geq 30,\ \textit{preferably}\ T \geq 50$), you may use the Newey-West method to obtain standard errors of OLS estimators that are corrected for autocorrelation. That is, you can still use OLS estimates with Newey-West standard errors.

    2) One can use appropriate transformation of the original model so that in the transformed model we do not have the problem of genuine AC. As in the case of heteroscedasticity, we will use some type of GLS methodology: use GLS estimators!

        a. If we know ρ → GLS estimator
        b. If we do not know ρ → EGLS estimator

3) In some situations we can continue to use OLS method (in small samples such as $T \le 20$ and when coefficient of correlation $\rho \le 0.3$) with OLS standard errors.

V. Forms of Autocorrelation

We will mostly deal with Markov first-order autoregressive autocorrelation, denoted by AR(1), where the disturbance term u_t in the model $Y_t = \beta_0 + \beta_1 X_t + u_t$ is generated by the process

$$u_t = \rho u_{t-1} + \varepsilon_t$$

where ε_t is a random variable whose values in any observation is independent of its value in all other observations. Here note ε_t is a *white noise*:

$$E(\varepsilon_t) = 0$$
$$E(\varepsilon_t^2) = \sigma^2$$
$$E(\varepsilon_t \varepsilon_s) = 0, t \ne s$$

This type of AC is described as autoregressive because u_t is being determined by lagged values of itself plus a random component ε_t, sometimes named as *innovation* (white noise error term). It is described as first order because u_t depends only on u_{t-1} and the innovation. A process of the type:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \rho_4 u_{t-4} + \rho_5 u_{t-5} + \varepsilon_t$$

can be described as fifth-order AR process, denoted as AR(5).

Another alternative to autoregressive AC is moving-average (MA) autocorrelation, where u_t is determined as a weighted sum of current and previous values of ε_t. For example, the process:

$$u_t = \lambda_0\varepsilon_t + \lambda_1\varepsilon_{t-1} + \lambda_2\varepsilon_{t-2} + \lambda_3\varepsilon_{t-3}$$

would be described as MA(3). In general:

$$u_t = \lambda_0\varepsilon_t + \lambda_1\varepsilon_{t-1} + \lambda_2\varepsilon_{t-2} + \ldots + \lambda_m\varepsilon_{t-m}$$

Is called a moving-average process of order m and is denoted by MA(m).

We will focus on AR(1) autocorrelation because it seems to be the most common type in studies:

$$u_t = \rho u_{t-1} + \varepsilon_t$$

We can define 3 different cases:

(1) If $\rho = 0$ then $u_t = \varepsilon_t$, so that u_t like ε_t obeys all the classical assumptions, including non-autocorrelation. This is the case of no-autocorrelation.

(2) If $\rho > 0$, this case is referred as positive AC. It implies that positive values of u_{t-1} will tend to be followed by positive values of u_t and negative values of u_{t-1} will tend to be followed by negative values of u_t. In practice autocorrelation is in most cases positive. The main reasons of this are economic growth and cyclical movements of the economy.

(3) If $\rho < 0$, this is called negative AC. It implies that positive values of u_{t-1} will tend to be followed by negative values of u_t and vice versa. In such a case successive disturbances will tend to alternate in sign over time.

A. The First-Order Autoregressive Scheme Properties

(1) $u_t = \rho u_{t-1} + \varepsilon_t$ with $|\rho| < 1,$

where $E(\varepsilon_t) = 0,\ E(\varepsilon_t^2) = \sigma_\varepsilon^2,\ E(\varepsilon_t \varepsilon_s) = 0,\ t \neq s$

we can write:

(2) $u_{t-1} = \rho u_{t-2} + \varepsilon_{t-1}$

Substituting (2) in (1) yields

$$u_t = \rho\left[\rho u_{t-2} + \varepsilon_{t-1}\right] + \varepsilon_t$$

(3) $u_t = \rho^2 u_{t-2} + \rho\varepsilon_{t-1} + \varepsilon_t$

We can also write:

(4) $u_{t-2} = \rho u_{t-3} + \varepsilon_{t-2}$

Substituting (4) in (3) produces

$$u_t = \rho^2\left[\rho u_{t-3} + \varepsilon_{t-2}\right] + \rho\varepsilon_{t-1} + \varepsilon_t$$

(5) $u_t = \rho^3 u_{t-3} + \rho^2\varepsilon_{t-2} + \rho\varepsilon_{t-1} + \varepsilon_t$

If we continue substituting i periods (when i is large, $i \to \infty$) we have

(6) $u_t = \varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \rho^3\varepsilon_{t-3} + ...$

Then,

$$u_t = \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i}$$

This is the value of disturbance term when it is autocorrelated with a first-order autoregressive scheme(AR1).

<u>Mean of Autocorrelated u_t's</u>

$$E(u_t) = \sum_{i=0}^{\infty} \rho^i \underbrace{E(\varepsilon_{t-i})}_{=0},$$

so $E(u_t)=0$

<u>Variance of Autocorrelated u_t's</u>

$$u_t = \rho u_{t-1} + \varepsilon_t$$

$$Var(u_t) = Var\left[\rho u_{t-1}\right] + Var(\varepsilon_t)$$

$$Var(u_t) = \rho^2 \underbrace{Var\left[u_{t-1}\right]}_{Var(u_t)} + \underbrace{Var(\varepsilon_t)}_{\sigma_\varepsilon^2}$$

$$(1-\rho^2)Var(u_t) = \sigma_\varepsilon^2$$

$$Var(u_t) = \frac{\sigma_\varepsilon^2}{(1-\rho^2)}$$

<u>Covariance of Autocorrelated u_t's</u>

(1) $u_t = \rho u_{t-1} + \varepsilon_t$

$$Cov(u_t, u_{t-1}) = E\left[u_t - \underbrace{E(u_t)}_{=0}\right]\left[u_{t-1} - \underbrace{E(u_{t-1})}_{=0}\right]$$

$Cov(u_t, u_{t-1}) = E[u_t u_{t-1}]$

From (1)

$u_t u_{t-1} = \rho u_{t-1}^2 + \varepsilon_t u_{t-1}$

Taking expectation of both sides

$$E[u_t u_{t-1}] = \rho \underbrace{E(u_{t-1}^2)}_{=\sigma_u^2} + \underbrace{E(\varepsilon_t u_{t-1})}_{=0}$$

$$E[u_t u_{t-1}] = \rho.\sigma_u^2$$

Similarly,

$$E[u_t u_{t-2}] = \rho^2.\sigma_u^2$$

$E[u_t u_{t-i}] = \rho^i.\sigma_u^2$ (When the number of lags increases (i) covariances will decrease)

The correlation coefficient between u_t and u_{t-1}

$$\rho_{u_t u_{t-1}} = \frac{Cov(u_t, u_{t-1})}{Var(u_t)} = \frac{\rho \sigma_u^2}{\sigma_u^2} = \rho$$

The correlation coefficient between u_t and u_{t-i}

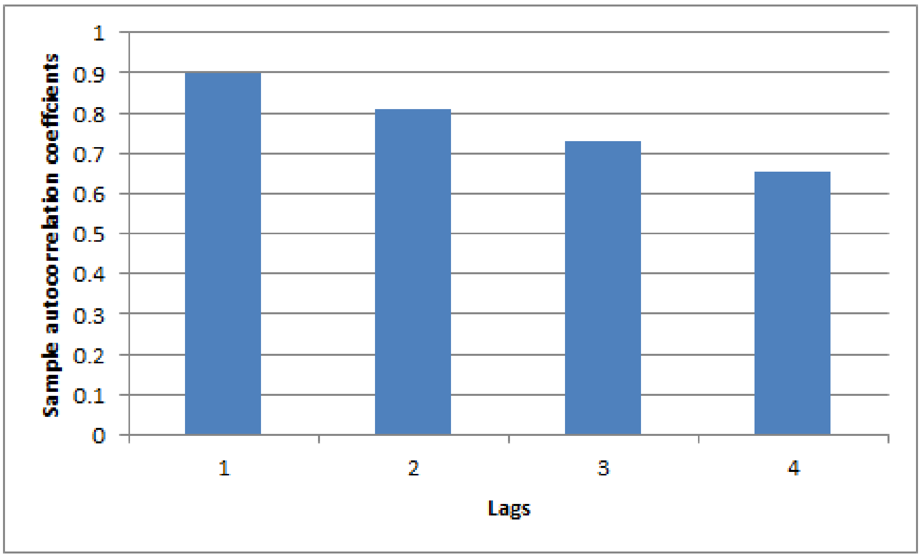
$$\rho_{u_t u_{t-i}} = \frac{Cov(u_t, u_{t-i})}{\sqrt{\underbrace{Var(u_t)}_{\sigma_u^2} \underbrace{Var(u_{t-i})}_{\sigma_u^2}}} = \frac{\rho^i \sigma_u^2}{\sigma_u^2} = \rho^i$$

Hence the autocorrelated coefficients:

$$\rho_{u_t u_{t-i}} = \rho^i$$

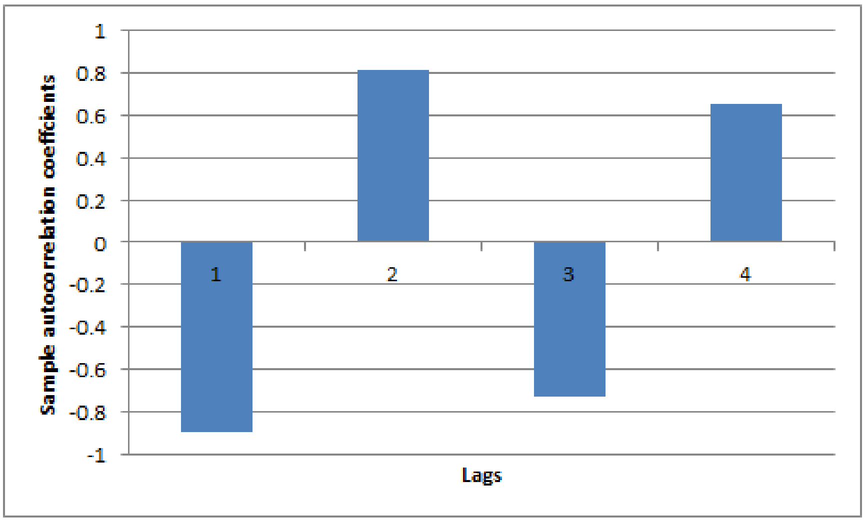
As can be seen as the lag increases $(i \uparrow)$ since $|\rho| < 1$, the autocorrelation coefficient will decrease gradually.

Example 1 $\hat{\rho} = 0.9$ → $\hat{\rho}_1$=0.9, $\hat{\rho}_2$=0.81, $\hat{\rho}_3$=0.729, $\hat{\rho}_4$=0.6561

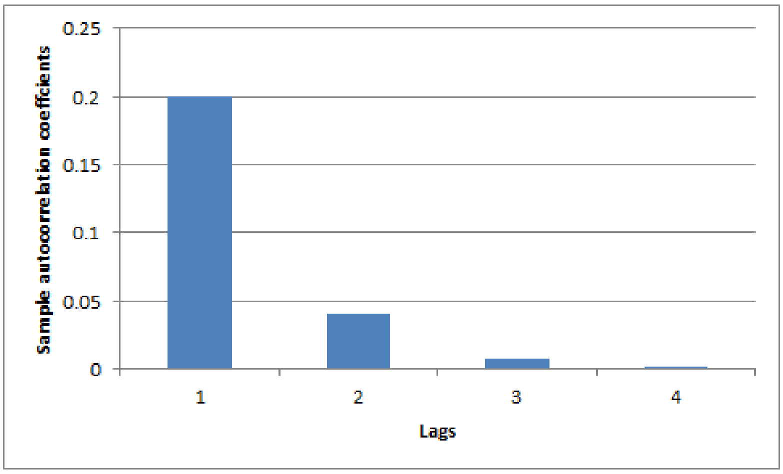


In the graph a gradual decrease is observed.

Example 2 $\hat{\rho} = -0.9$ → $\hat{\rho}_1$=-0.9, $\hat{\rho}_2$=0.81, $\hat{\rho}_3$=-0.729, $\hat{\rho}_4$=0.6561



Example 3 $\hat{\rho} = 0.2$ → $\hat{\rho}_1$=0.2, $\hat{\rho}_2$=0.04, $\hat{\rho}_3$=0,008, $\hat{\rho}_4$=00016

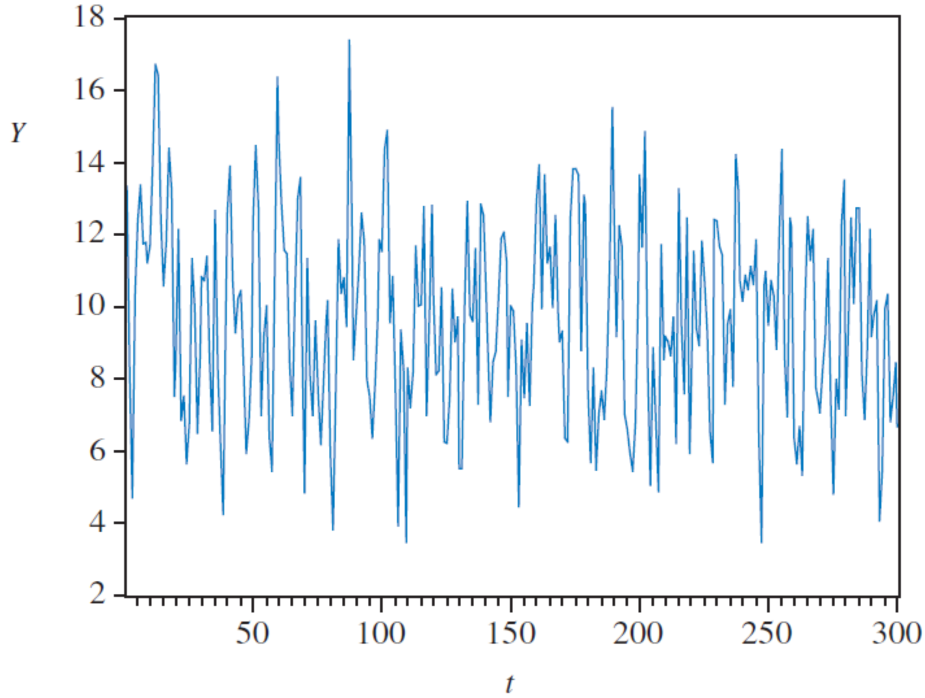


Hence, when the correlations between the current disturbance term and previous period disturbance is weaker (like $\rho = 0.2$) *the correlations between the current disturbance and the disturbances at more distant lags die out relatively quickly*.

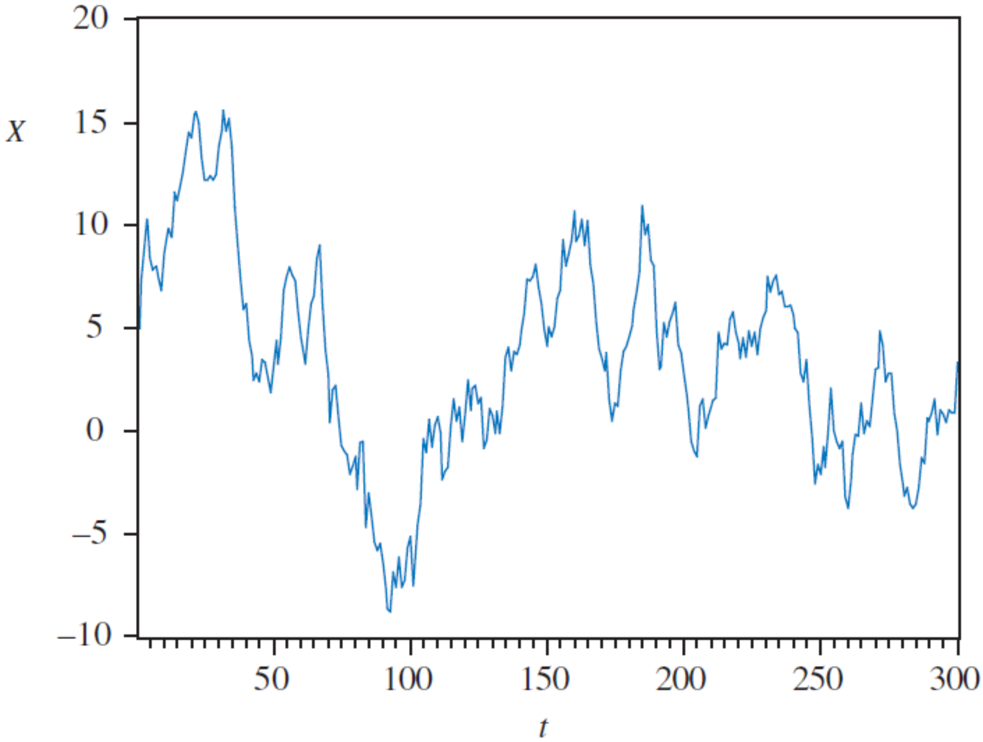
VI. Tests of Autocorrelation

A. Importance of Stationarity

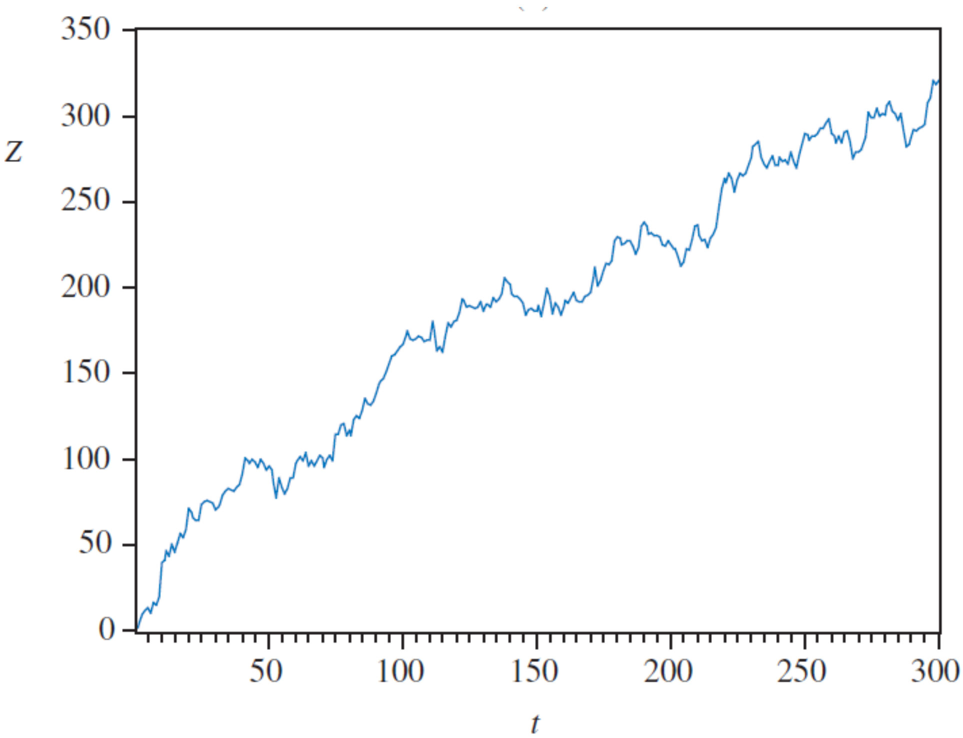
An assumption that we maintain throughout this lecture note dealing with autocorrelation is that the variables in our equations are *stationary*. This assumption will take on more when we deal with time series analysis. For the moment we note that a stationary variable is one that is not explosive, nor trending, and nor wandering aimlessly without returning to its mean. These features can be illustrated with some graphs. Figures 1(a), 1(b) and 1(c) contain graphs of the observations on three different variables, plotted against time. Plots of this kind are routinely considered when examining time-series variables. The variable Y that appears in Figure 1(a) is considered stationary because it tends to fluctuate around a constant mean without wandering or trending. On the other hand, X and Z in Figures 1(b) and 1(c) possess characteristics of nonstationary variables. In Figure 1(b) X tends to wander, or is ‘‘slow turning,’’ while Z in Figure 1(c) is trending. For now the important thing to remember is that this lecture note dealing with autocorrelaion issue is concerned with modeling and estimating dynamic relationships between stationary variables whose time series have similar characteristics to those of Y. That is, they neither wander nor trend (HGL, pp.339-340).


18
16
14
12
10
8
6
4
2
Y
50
100
150
200
250
300
t

(*a*)


20
15
10
5
0
–5
–10
X
50
100
150
200
250
300
t

(*b*)



(c)

Figure 1 (a) Time series of a stationary variable; (b) time series of a nonstationary variable that is ‘‘slow-turning’’ or ‘‘wandering’’; (c) time series of a nonstationary variable that ‘‘trends.’’

B. Autocorrelation Coefficients

The correlation coefficient between u_t and u_{t-1} can be written as:

$$\rho_{u_t u_{t-1}} = \frac{Cov(u_t, u_{t-1})}{\sqrt{Var(u_t)Var(u_{t-1})}}$$

$$\rho_{u_t u_{t-1}} = \frac{E[u_t - E(u_t)]E[u_{t-1} - E(u_{t-1})]}{\sqrt{E[u_t - E(u_t)]^2 E[u_{t-1} - E(u_{t-1})]^2}}$$

$$\rho_{u_t u_{t-1}} = \frac{E[u_t]E[u_{t-1}]}{\sqrt{E[u_t^2]E[u_t^2]}}$$

For a population of sample size N this can be calculated as:

$$\rho_{u_t u_{t-1}} = \frac{\dfrac{\sum_{t=2}^{N} u_t u_{t-1}}{N}}{\sqrt{\dfrac{\sum_{t=1}^{N} u_t^2}{N} \dfrac{\sum_{t=2}^{N} u_{t-1}^2}{N}}}$$

$$\rho_{u_t u_{t-1}} = \frac{\sum_{t=2}^{N} u_t u_{t-1}}{\sqrt{\sum_{t=1}^{N} u_t^2 \sum_{t=2}^{N} u_{t-1}^2}}$$

For infinite (or large) populations $\sum_{t=1}^{N} u_t^2 \approx \sum_{t=2}^{N} u_{t-1}^2$. Therefore we can obtain two different expressions by (1) substituting $\sum_{t=2}^{N} u_{t-1}^2$ for $\sum_{t=1}^{N} u_t^2$ and, (2) substituting $\sum_{t=1}^{N} u_t^2$ for $\sum_{t=2}^{N} u_{t-1}^2$.

(1) Let us substitute $\sum_{t=2}^{N} u_{t-1}^2$ for $\sum_{t=1}^{N} u_t^2$. Then, we get

$$\rho_{u_t u_{t-1}} = \frac{\sum_{t=2}^{N} u_t u_{t-1}}{\sum_{t=2}^{N} u_{t-1}^2}$$

Now recall the AR(1) process:

$$u_t = a.u_{t-1} + \varepsilon_t \qquad t = 1, \ldots, N$$

If we apply OLS to this AR(1) process for the whole population:

$$a = \frac{\sum_{t=2}^{N} u_t u_{t-1}}{\sum_{t=2}^{N} u_{t-1}^2}$$

which is clearly the same with ρ for large populations. This is the reason why in most books the AR(1) model is given in the form:

$$u_t = \rho u_{t-1} + \varepsilon_t$$

where ρ is first-order autocorrelation coefficient.

(1) Let us substitute $\sum_{t=1}^{N} u_t^2$ for $\sum_{t=2}^{N} u_{t-1}^2$. Then, we get

$$\rho_{u_t u_{t-1}} = \frac{\sum_{t=2}^{N} u_t u_{t-1}}{\sum_{t=1}^{N} u_t^2}$$

which is the population autocorrelation coefficient. This expression can not be calculated due to two reasons: (1) Most of the cases we do not know population values, and (2) the u_t's are not observable.

The solution to first problem is easy: we can use sample instead of population. If we would observe u_t's, then we would estimate ρ over a sample of size T as (r is the autocorrelation coefficient calculated over a sample):

$$r_{u_t u_{t-1}} = \frac{\sum_{t=2}^{T} u_t u_{t-1}}{\sum_{t=1}^{T} u_t^2} \qquad \text{where } t=1,2,\ldots,T$$

However, since u_t's are unobservable, all we can do is to use its estimated counterparts namely $\hat{u}_t$'s, that is, residuals. Hence the *first-order sample autocorrelation coefficient*, $\hat{\rho}$, will be:

$$\hat{\rho} = \frac{\sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^{T} \hat{u}_t^2} \qquad \text{where t=1,2,\ldots,}T$$

More generally, the *k-th order sample autocorrelation* for u_t that gives the correlation between observations that are *k* periods apart (the correlation between u_t and u_{t-k}) is given by

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{T} \hat{u}_t \hat{u}_{t-k}}{\sum_{t=1}^{T} \hat{u}_t^2} \qquad \text{where t=1,2,\ldots,}T$$

1. Individual Significance of Autocorrelation Coefficients and Correlogram

Suppose that applying the formula above to the series $\hat{u}_t$ yields, for the first four sample autocorrelations:

$$\hat{\rho}_1 = 0.494,\ \hat{\rho}_2 = 0.411,\ \hat{\rho}_3 = 0.154,\ \hat{\rho}_4 = 0.200$$

The autocorrelations at lags one and two are moderately high: those at lags three and four are much smaller – less than half the magnitude of the earlier ones. How do we test whether an autocorrelation is significantly different from zero? Let the *k*th order population autocorrelation be denoted by ρ_k. Then, when the null hypothesis $H_0: \rho_k = 0$ is true, it turns out that $\hat{\rho}_k$ has an approximate normal distribution with mean zero and variance 1/T (Hill, Griffths and Lim, 2011, p.349). Thus a suitable test statistic is:

$$Z = \frac{\hat{\rho}_k - 0}{\sqrt{1/T}} = \sqrt{T}\hat{\rho}_k \sim N(0,1)$$

At a 5% significance level, we reject $H_0 : \rho_k = 0$ when $\sqrt{T}\hat{\rho}_k \geq 1.96$ or $\sqrt{T}\hat{\rho}_k \leq -1.96$.

Let us assume that for the series $\hat{u}_t$, T=98, then the values of the test statistic Z for the first four lags are:

$$Z_1 = \sqrt{98}.(0.494) = 4.89$$
$$Z_2 = \sqrt{98}.(0.414) = 4.10$$
$$Z_3 = \sqrt{98}.(0.154) = 1.52$$
$$Z_4 = \sqrt{98}.(0.200) = 1.98$$

Thus, we reject the hypothesis $H_0 : \rho_1 = 0$ and $H_0 : \rho_2 = 0$, we have insufficient evidence to reject $H_0 : \rho_3 = 0$, and $\hat{\rho}_4$ is on the borderline of being significant. We conclude that, the residuals exhibits significant serial correlation (autocorrelation) at lags one and two.

Correlogram

A useful device for assessing the significance of autocorrelations is a diagrammatic representation of the correlogram (Hill, Griffiths and Lim, 2011, pp.349-350). A plot of the autocorrelations $\hat{\rho}_k$ against the lag k is called the correlogram. This plot provides a first idea of possible serial correlation. The correlogram, also called the sample autocorrelation function, is the sequence of autocorrelations $\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3, \ldots$. It shows the correlation between observations that are one period apart, two periods apart, three periods apart, and so on. We indicated that an autocorrelation $\hat{\rho}_k$ will be significantly different from zero at a 5% significance level if $\sqrt{T}\hat{\rho}_k \geq 1.96$ or if $\sqrt{T}\hat{\rho}_k \leq -1.96$. Alternatively, we can say that $\hat{\rho}_k$ will be significantly different from zero if $\hat{\rho}_k \geq \frac{1.96}{\sqrt{T}}$ or if $\hat{\rho}_k \leq \frac{-1.96}{\sqrt{T}}$. By drawing the values $\pm\frac{1.96}{\sqrt{T}}$ as bounds on a graph that illustrates the magnitude of each of the $\hat{\rho}_k$, we can see at a glance which correlations are significant.

A graph of the correlogram for $\hat{u}_t$ for the first 12 lags appears in Figure K below. The heights of the bars represent the correlations and the horizontal lines drawn at $\pm\frac{2}{98}=\pm 0.202$ are the significance bounds. We have used 2 rather than 1.96 as a convenient approximation. We can see at a glance that $\hat{\rho}_1$ and $\hat{\rho}_2$ are significantly different from zero, that $\hat{\rho}_4$ and $\hat{\rho}_{12}$ are bordering on significance, and the remainder of the autocorrelations are not significantly different from zero.

Note that your software may not produce a correlogram that is exactly the same as Figure Z. It could use spikes instead of bars to denote the correlations, it might provide a host of additional information, and the width of its significance bounds might vary with different lags. If the significance bounds vary, it is because they use a refinement of the large sample approximation.

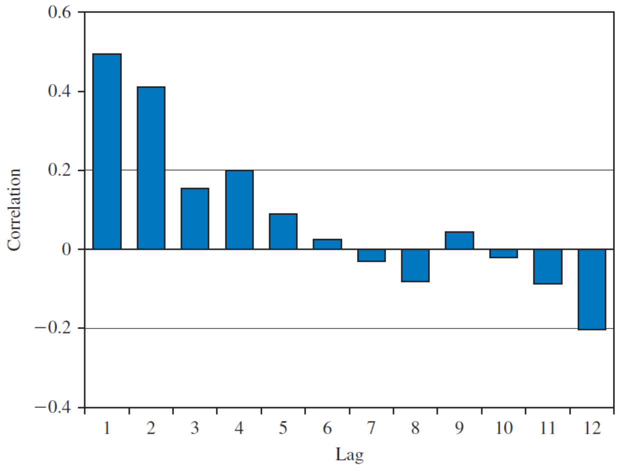


Figure K. Corellogram for $\hat{u}_t$

2. Joint Significance of Autocorrelation Coefficients

Box–Pierce Portmanteau test (or simply *Box–Pierce test*) for the joint significance of the first *m* autocorrelation coefficients is given by:

$$BP_m = T\sum_{k=1}^{m}\hat{\rho}_k \qquad (8)$$

and $BP \sim \chi^2_m(\alpha)$ under the null hypothesis of no serial autocorrelation.

Sometimes the correlations in (8) are weighted because higher order autocorrelations are based on fewer observations. This gives the *Ljung–Box test* (also denoted as the *Q-test*):

$$LB = T\sum_{k=1}^{m}\frac{T+2}{T-k}\hat{\rho}_k$$

and $LB \sim \chi^2_p(\alpha)$ under the null hypothesis of no serial autocorrelation.

The Box–Pierce test and the Ljung–Box test require that the regressors X_t *in the model are non-stochastic (fixed)*: for example, if the model has lagged terms of dependent variable as explanatory variable (such as Y_{t-1}), this tests cannot be used.

C. Durbin-Watson Test

The most frequently used statistical test for the presence of serial correlation is the Durbin-Watson (DW) test (Durbin and Watson, 1950), which is valid when the following assumptions are met:

(1) The regression model involves an intercept term
(2) Serial correlation is assumed to be of first-order,
(3) The equation does not include a lagged dependent variable as an explanatory variable (i.e. Y_{t-1})

Consider the model:

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + ... + \beta_k X_{tk} + u_t$$

where $u_t = \rho u_{t-1} + \varepsilon_t$ with $|\rho| < 1$

Then under the null hypothesis $H_0 : \rho = 0$ the DW test involves the following steps:

Step 1 Estimate the model by OLS and obtain the residuals, $\hat{u}_t$'s.
Step 2 Calculate the DW test statistic as follows:

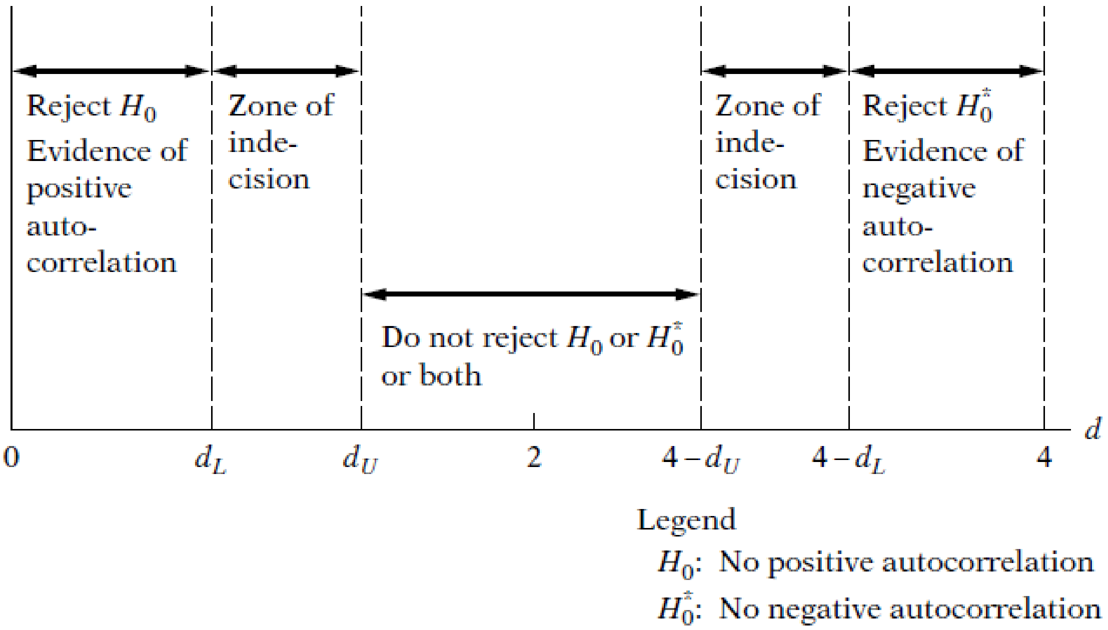
$$DW = \frac{\sum_{t=2}^{T} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{T} \hat{u}_t^2}$$

Step 3 Construct table given below, substituting with your DW critical values d_U, d_L. Note that table of critical values according to $k^/$ which is the number of explanatory variables excluding the intercept term.
Step 4a To test for positive serial correlation the hypotheses are:

$H_0 : \rho = 0$ no autocorrelation
$H_A : \rho > 0$ positive autocorrelation

Step 4b To test for negative serial correlation the hypotheses are:

$H_0 : \rho = 0$ no autocorrelation
$H_A : \rho < 0$ positive autocorrelation



A Rule of Thumb for the DW Test

$$DW = \frac{\sum_{t=2}^{T}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{T}\hat{u}_t^2}$$

$$DW = \frac{\sum_{t=2}^{T}(\hat{u}_t^2 + \hat{u}_{t-1}^2 - 2\hat{u}_t\hat{u}_{t-1})}{\sum_{t=1}^{T}\hat{u}_t^2}$$

$$DW = \frac{\sum_{t=2}^{T}\hat{u}_t^2 + \sum_{t=2}^{T}\hat{u}_{t-1}^2 - 2\sum_{t=2}^{T}\hat{u}_t\hat{u}_{t-1}}{\sum_{t=1}^{T}\hat{u}_t^2}$$

$$DW = \frac{\sum_{t=2}^{T}\hat{u}_t^2}{\sum_{t=1}^{T}\hat{u}_t^2} + \frac{\sum_{t=2}^{T}\hat{u}_{t-1}^2}{\sum_{t=1}^{T}\hat{u}_t^2} - 2\frac{\sum_{t=2}^{T}\hat{u}_t\hat{u}_{t-1}}{\sum_{t=1}^{T}\hat{u}_t^2}$$

For large samples, $\frac{\sum_{t=2}^{T}\hat{u}_t^2}{\sum_{t=1}^{T}\hat{u}_t^2} \approx 1$ and $\frac{\sum_{t=2}^{T}\hat{u}_{t-1}^2}{\sum_{t=1}^{T}\hat{u}_t^2} \approx 1$

$$DW \cong 2 - 2\underbrace{\frac{\sum_{t=2}^{T}\hat{u}_t\hat{u}_{t-1}}{\sum_{t=1}^{T}\hat{u}_t^2}}_{\hat{\rho}}$$

$$DW \cong 2 - 2\hat{\rho}$$

$$DW \cong 2(1 - \hat{\rho})$$

where recall that $\hat{\rho}$ is the estimated correlation coefficient between u_t and u_{t-1}.

The implications of this expression are as follows:

(1) If there is no AC, then ρ=0 and DW=2.

- Hence a value of DW near to 2 indicates that there is no evidence of serial correlation.

(2) If there is exact positive AC, then ρ=1 and DW=0

- Strong positive autocorrelation means that ρ will be close to +1 and DW will get very low values (close to zero) for positive autocorrelation.

(3) If there is exact negative AC, then ρ=-1 and DW=4

- When ρ is close to -1 then DW will be close to 4 indicating strong negative serial correlation.

Hence, we can see that, as a rule of thumb, when DW statistic is very close to 2 then we do not have serial correlation.

Example 1

You are given the following estimation results:

$\hat{Y}_t$ = 36.1882 + 0.2928 X_{t2} + 0.0023 X_{t3} $\quad$ t = 1, …, 30

R^2 = 0.98 DW = 0.34435

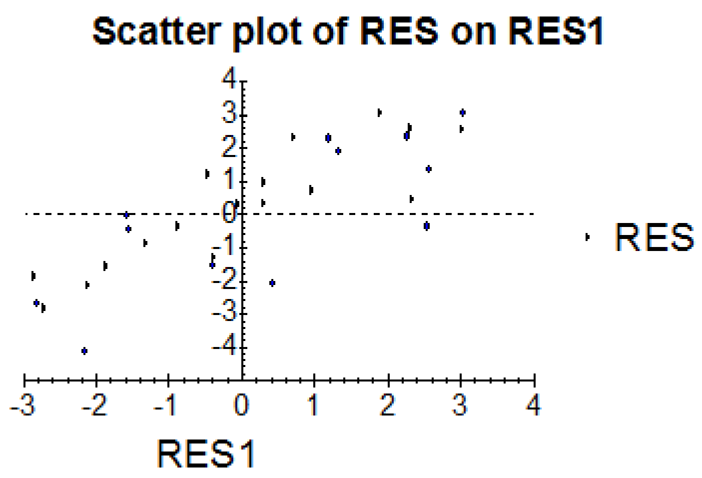
where;

Y_t is real consumption expenditure
X_{t2} is real disposable income
X_{t3} is real wealth

Test for the presence of first-order positive autocorrelation.

Solution

First let us look at the residual plots. The plots implies positive AC.



DW = 0.34435
From Tables d_L = 1.284; d_U = 1.567
So we can reject the null of zero autocorrelation against the alternative that $\rho > 0$

Example 2

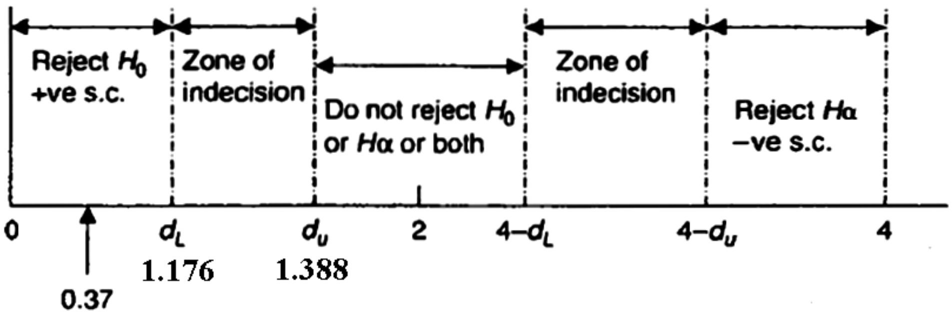
DW=0.1380, T=32, $k^{/}$=1
d_L=1.373 and d_u=1.502 at 5% significance level.
Since DW=0.1380<d_L, reject H_0 of no positive correlation at 0.05 level of significance, so there is positive AC.

Example 3

DW=1.43, T=50, k'=4
d_L=1.378, and d_u=1.721 at 5% significance level.
Since d_L<DW<d_u, inconclusive regarding whether there is positive autocorrelation.

Example 4

DW=0.37, T=38, k'=2
d_L=1.176, and d_u=1.388 at 5% significance level.
Since DW<d_L there is strong evidence that there is positive autocorrelation at 0.05 level of significance.



D. Breusch-Godfrey LM Test for Serial Correlation

The DW test has several drawbacks that make its use inappropriate in various cases. For instance (a) it may give inconclusive results, (b) it is not applicable when a lagged dependent variable is used, and (c) it can't take into account higher orders of serial correlation.

For these reasons Breusch (1978) and Godfrey (1978) developed an LM test which can accommodate all the above cases.

Consider the model:

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \ldots + \beta_k X_{tk} + u_t \qquad (1)$$

where t=1,2,…,T and

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + ... + \rho_m u_{t-m} + \varepsilon_t \qquad \text{with } |\rho| < 1 \tag{2}$$

The Breusch-Godfrey LM test combines these two equations by putting (2) into (1):

$$\begin{aligned} Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + ... + \beta_k X_{tk} \\ + \rho_1 u_{t-1} + \rho_2 u_{t-2} + ... + \rho_m u_{t-m} + \varepsilon_t \end{aligned} \tag{3}$$

Since u's are unobservable, we use the residuals, $\hat{u}$'s. Hence the auxiliary regression of LM test becomes:

$$\begin{aligned} Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + ... + \beta_k X_{tk} \\ + \rho_1 \hat{u}_{t-1} + \rho_2 \hat{u}_{t-2} + ... + \rho_m \hat{u}_{t-m} + \varepsilon_t \end{aligned} \tag{3*}$$

On the other hand, note that:

$$Y_t = \hat{\beta}_0 + \hat{\beta}_1 X_{t1} + \hat{\beta}_2 X_{t2} + ... + \hat{\beta}_k X_{tk} + \hat{u}_t$$

hence we can rewrite (3*) as:

$$\begin{aligned} \hat{\beta}_0 + \hat{\beta}_1 X_{t1} + \hat{\beta}_2 X_{t2} + ... + \hat{\beta}_k X_{tk} + \hat{u}_t = \\ \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + ... + \beta_k X_{tk} \\ + \rho_1 \hat{u}_{t-1} + \rho_2 \hat{u}_{t-2} + ... + \rho_m \hat{u}_{t-m} + \varepsilon_t \end{aligned}$$

Rearranging this equation yields:

$$\begin{aligned} \hat{u}_t = \overbrace{(\beta_0 - \hat{\beta}_0)}^{\alpha_0} + \overbrace{(\beta_1 - \hat{\beta}_1)}^{\alpha_1} X_{t1} + \overbrace{(\beta_2 - \hat{\beta}_2)}^{\alpha_2} X_{t2} + ... + \overbrace{(\beta_k - \hat{\beta}_k)}^{\alpha_k} X_{tk} \\ + \rho_1 \hat{u}_{t-1} + \rho_2 \hat{u}_{t-2} + ... + \rho_m \hat{u}_{t-m} + \varepsilon_t \end{aligned}$$

Then finally we can write:

$$\hat{u}_t = \alpha_0 + \alpha_1 X_{t1} + \alpha_2 X_{t2} + ... + \alpha_k X_{tk} + \rho_1 \hat{u}_{t-1} + \rho_2 \hat{u}_{t-2} + ... + \rho_m \hat{u}_{t-m} + \varepsilon_t \quad (4)$$

and the null and the alternative hypotheses can be stated as:

$H_0 : \rho_1 = \rho_2 = ... = \rho_m = 0$, *hence no autocorrelation*
H_A*: at least one of the* ρ *'s is not zero, thus, serial correlation.*

The steps for carrying out the test are the following:

Step 1 Estimate the original model (1) by OLS and obtain the residuals, $\hat{u}_t$.

Step 2 Run the following regression model with the number of lags used (*m*) being determined according to the order of serial correlation you are willing to test. Get the R^2 of equation (4).

$$\hat{u}_t = \alpha_0 + \alpha_1 X_{t1} + \alpha_2 X_{t2} + ... + \alpha_k X_{tk} + \rho_1 \hat{u}_{t-1} + \rho_2 \hat{u}_{t-2} + ... + \rho_m \hat{u}_{t-m} + \varepsilon_t \quad (4)$$

where the usable sample size to estimate the equation above will be $T - m$ due to m lagged residual terms. Because $\beta_k - \hat{\beta}_k$ are centered around zero, if equation (4) is a regression with significant explanatory power, that power will come from $\hat{u}_{t-1}$, $\hat{u}_{t-2}$ and $\hat{u}_{t-m}$.

Step 3 Compute the LM statistic

$$BG_{LM} = (T - m)R^2$$

from the OLS regression run in step 2.

Note that $BG_{LM} \sim \chi^2_m$ hence if this *LM* statistic is bigger than the χ^2_m critical value for a given level of significance, then we reject the null of serial correlation and conclude that serial correlation is present.

Choice of lag order Note that the choice of *m* is arbitrary in LM test. However, the periodicity of the data (quarterly, monthly, weekly etc.) will often give us a suggestion for the size of *m*. Some possible decision rules are as follows: (1) you may test for several orders, (2) you can use autocorrelation coefficients to decide for the order, (3) you may check the significance of the auxiliary regression's estimated coefficients for residuals, (4) one can use the so-called Akaike and Schwarz information criteria to select the lag length (5) for annual data 1st order, for quarterly data 4th order, and for monthly data 12th order, etc., can be chosen.

Example of Breusch-Godfrey LM Test

Consider the model:

$$\ln C_t = a_0 + a_1 \ln DI_t + a_2 \ln P_t + u_t$$

where C_t is consumption, DI_t is disposable income and P_t is price, and suppose that we have a quarterly data for them. We proceed by testing for fourth-order serial correlation due to the fact that we have quarterly data. In order to test for serial correlation of fourth order we use the Breusch- Godfrey LM test. The results of this test are shown in Table K1.

We can see from the first columns that the values of both the LM statistic and the F statistic are quite high, suggesting the rejection of the null of no serial correlation. It is also evident that this is so due to the fact that the p-values are very small (smaller than 0.05 for a 95% confidence interval). So, serial correlation is definitely present. However, if we observe the regression results, we see that only the first lagged residual term is statistically significant, indicating, most probably, that the serial correlation is of first order. Rerunning the test for a first-order serial correlation the results are as shown in Table K2. This time the LM statistic is much higher, as well as the t statistic of the lagged residual term. So, the autocorrelation is definitely of first order.

Table K1 Results of Breusch-Godfrey LM Test (4^{th} order AC)

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	17.25931	Probability	0.000000
Obs**R*-squared	**26.22439**	Probability	**0.000029**

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 02/12/04 Time: 22:51

Variable	*Coefficient*	*Std. Error*	*t-Statistic*	*Prob.*
C	−0.483704	0.489336	−0.988491	0.3306
LDISP	0.178048	0.185788	0.958341	0.3453
LPRICE	−0.071428	0.093945	−0.760322	0.4528
RESID(−1)	0.840743	0.176658	**4.759155**	0.0000
RESID(−2)	−0.340727	0.233486	−1.459306	0.1545
RESID(−3)	0.256762	0.231219	1.110471	0.2753
RESID(−4)	0.196959	0.186608	1.055465	0.2994

R-squared	0.690115	Mean dependent var	1.28E−15
Adjusted *R*-squared	0.630138	S.D. dependent var	0.044987
S.E. of regression	0.027359	Akaike info criterion	−4.194685
Sum squared resid	0.023205	Schwarz criterion	−3.893024
Log likelihood	86.69901	*F*-statistic	11.50621
Durbin–Watson stat	1.554119	Prob(*F*-statistic)	0.000001

Table K2 Results of Breusch-Godfrey LM Test (1^{st} order AC)

Breusch–Godfrey Serial Correlation LM Test:

F-statistic	53.47468	Probability	0.000000
Obs**R*-squared	**23.23001**	Probability	**0.000001**

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 02/12/04 Time: 22:55

Variable	*Coefficient*	*Std. Error*	*t-Statistic*	*Prob.*
C	−0.585980	0.505065	−1.160208	0.2540
LDISP	0.245740	0.187940	1.307546	0.1998
LPRICE	−0.116819	0.094039	−1.242247	0.2226
RESID(−1)	0.828094	0.113241	**7.312638**	0.0000

R-squared	0.611316	Mean dependent var	1.28*E* − 15
Adjusted *R*-squared	0.577020	S.D. dependent var	0.044987
S.E. of regression	0.029258	Akaike info criterion	−4.126013
Sum squared resid	0.029105	Schwarz criterion	−3.953636
Log likelihood	82.39425	*F*-statistic	17.82489
Durbin–Watson stat	1.549850	Prob(*F*-statistic)	0.000000

Example for LM Test [for AR(1) Case]

*In this example, we show that for AR(1) case, the BG -LM test can be done using t-table values. This BG-LM test is known as **Durbin's M** test.*

The following equations are estimated for the 1990.1-1993.12 period:

(1) $\hat{GTM}_t = 250.2 - 89.65 \text{ MPR}_t$ $R^2 = 0.598$

se (240.92) (33.59)

(2) $\hat{u}_t$ =-48.65 - 2.82 MPR$_t$ + 2.942 $\hat{u}_{t-1}$
se (35.49) (6.85) (0.68)

Test for the presence of first-order autocorrelation in (1)

Solution

H_0: $\rho_1=0$
H_A: $\rho_1\neq 0$

t=2.942/0.68=4.33

$t_{\alpha/2,T-k-1}=t_{0.025,\ 46}=2.021 \Rightarrow$ $t>t_{table}$, so reject H_0 at α=0.05 level of significance. There is autocorrelation.

E. Durbin's h Test in the Presence of Lagged Dependent Variables

We mentioned before in the assumptions of the DW test, that the DW test is not applicable when our regression model includes lagged dependent variables as explanatory variables. Therefore, if the model under examination has the form:

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + ... + \beta_k X_{tk} + \gamma Y_{t-1} + u_t \qquad (1)$$

the DW test is no longer valid. Durbin (1970) devised a test statistic that can be used for such models, and this *h* statistic has the form:

$$h = \left(1 - \frac{DW}{2}\right)\sqrt{\frac{T^{/}}{1 - \hat{\sigma}^2_{\hat{\gamma}} T^{/}}}$$

where $T^{/}$=number of observations in (1), (=T-1), and $\hat{\sigma}^2_{\hat{\gamma}}$= estimated variance of the OLS estimated coefficient associated with lagged variable Y_{t-1}.

For "*large samples*" the test statistic (*h*) has a standard normal distribution. Therefore, for a test of the null hypothesis of no autocorrelation against the 2-sided alternative of autocorrelated errors, at a 5% level, the decision rule is if $-1.96<h<1.96$ do not reject the null hypothesis. By applying this decision rule it can be seen that there is no evidence for autocorrelation in the residuals. As a cautionary note; users should consider that this test may not be accurate in "*small samples*". This test is not as powerful, in a statistical sense, as the Breusch–Godfrey LM test.

Steps

1. Estimate (1). Obtain $\hat{u}_t$ and DW. Obtain the estimated variance of the OLS estimated coefficient associated with lagged variable Y_{t-1}, i.e., $\hat{\sigma}^2_{\hat{\gamma}}$.

2. Compute $h=\left(1-\frac{DW}{2}\right)\sqrt{\frac{T^{/}}{1-\hat{\sigma}^2_{\hat{\gamma}}T^{/}}}$

3. Test;

    H_0: $\rho=0$ (no AC)
    H_A: $\rho\neq 0$ (AC)

$h \sim N(0,1)$ asymptotically. So if the sample size is reasonably large and $|h|>1.96\ (=Z_{0.025}=Z_{\alpha/2})$, we reject H_0 at 0.05 level of significance so there is AC.

Remarks

a. It does not matter how many X variables or how many lagged values of Y are included in the regression model. To compute *h*, we need consider only the variance of the coefficient of lagged Y_{t-1}
b. The test is not applicable if ($\hat{\sigma}^2_{\hat{\gamma}}T^{/}$) exceeds 1. In practice, however, this does not usually happen.

Example

You are given time series data for the period 1977.1-1991.2 on the estimation of aggregate consumption by disposable income.

(OLS)(1) $C_t=13.2+0.88Y_t$ $R^2=0.988$ DW=1.11
(3.38) (0.01)

(OLS)(2) $C_t=5.08+0.64C_{t-1}+0.33Y,$ $R^2=0.993$ DW=2.12
se (3.00) (0.10) (0.09)

and test the model (2) for serial correlation (take $\alpha=0.05$).

Solution

There is lagged dependent variable, so we can use Durbin-h test (or LM test).

H_0: $\rho=0$ (no AC)
H_A: $\rho\neq 0$ (AC)

$$h=\left(1-\frac{DW}{2}\right)\sqrt{\frac{T^{/}}{1-\hat{\sigma}_{\hat{\gamma}}^{2}T^{/}}}$$

where $T^{/}$=T–1, $\hat{\sigma}_{\hat{\gamma}}^{2}$= estimated variance of C_{t-1}.

$$h = (-0.06).\sqrt{\frac{57}{1-(0.10)^2(57)}} = (-0.06).\sqrt{\frac{57}{0.43}} = -0.69$$

0.69<1.96 (=$Z_{0.025}$) $\Rightarrow$ Do not reject H_0 at $\alpha=0.05$ level of significance. There is no autocorrelation.

VII. Correcting for Genuine (Pure) Autocorrelation

Consider the following AR(1) process:

(1) $$Y_t = \beta_0 + \sum_{i=1}^{k} \beta_i X_{ti} + u_t \qquad u_t = \rho u_{t-1} + \varepsilon_t$$

where ε_t satisfies all GM assumptions

Then;

(2) $$Y_t = \beta_0 + \sum_{i=1}^{k} \beta_i X_{ti} + \rho u_{t-1} + \varepsilon_t$$

If ρ is known → GLS procedure will be applied.
If ρ is unknown → EGLS procedure will be applied.

A. GLS and EGSL Methods

1. Generalized Least Squares (GLS) (ρ is known)

Recall

$$Y_t = \beta_0 + \sum_{i=1}^{k} \beta_i X_{ti} + u_t \, ,$$

then,

$$Y_{t-1} = \beta_0 + \sum_{i=1}^{k} \beta_i X_{t-1,i} + u_{t-1}$$

$$u_{t-1} = Y_{t-1} - \beta_0 - \sum_{i=1}^{k} \beta_i X_{t-1,i} \qquad t=2\ldots..T$$

Multiply both sides by ρ

$$\rho u_{t-1} = \rho Y_{t-1} - \rho\beta_0 - \rho\sum_{i=1}^{k}\beta_i X_{t-1,i}$$

and put it in equation (2):

$$(3)\ Y_t = \beta_0 + \sum_{i=1}^{k}\beta_i X_{ti} + \rho Y_{t-1} - \rho\beta_0 - \rho\sum_{i=1}^{k}\beta_i X_{t-1,i} + \varepsilon_t$$

$$\underbrace{Y_t - \rho Y_{t-1}}_{Y_t^*} = \underbrace{(1-\rho)}_{X_{t0}^*}\beta_0 + \sum_{i=1}^{k}\beta_i X_{ti} - \rho\sum_{i=1}^{k}\beta_i X_{t-1,i} + \varepsilon_t$$

$$Y_t^* = X_{t0}^*.\beta_0 + \sum_{i=1}^{k}\beta_i \underbrace{\left[X_{ti} - \rho X_{t-1,i}\right]}_{X_{ti}^*} + \varepsilon_t$$

$$(4)\ Y_t^* = \beta_0.X_{t0}^* + \sum_{i=1}^{k}\beta_i X_{ti}^* + \varepsilon_t \qquad t=2.....T$$

where $\varepsilon_t \rightarrow$white noise residual. There is no problem to apply OLS to model (4).

Now we can apply OLS to the transformed model (4) and obtain GLS estimator (for equation 1) of our original model.

Example Suppose we know that ρ=0.7

observation	Y_t	Y_{t-1}	$Y_t - \rho Y_{t-1}$
t=1	30	-	-
t=2	38	30	38-(0.7).30
t=3	45	38	45-(0.7).38
t=4	62	45	62-(0.7).45
.	.	.	.
.	.	.	.
.	.	.	.

Recall: *GLS is nothing but OLS applied to the transformed model that satisfies the Gauss-Markov assumptions*.

As can be seen from the example, in this procedure we lose one observation because the first observation has no legged counterpart. This loss of one observation can make a substantial difference in the results particularly in small samples. In addition, without transforming the first observation, the error variance will not be homoscedastic (it can be shown).

To avoid this loss of one observation, the first observation on Y and X is transformed as follows:

$$Y_1^* = \sqrt{1-\rho^2} \qquad \text{and} \qquad X_1^* = X_1\sqrt{1-\rho^2}$$

This transformation is known as *Prais-Winsten transformation*.

2. Estimated Generalized Least Squares (GLS) (ρ is unknown)

The GLS procedure is difficult to implement since ρ is rarely known is practice. Therefore, we need to find ways of estimating ρ.

We have several possibilities:

1) First-Difference Method
2) ρ estimated from residuals
3) Iterative methods to estimate ρ
    - *a) Cochrane-Orcutt (CO) iterative procedure*
    - b) CO two-step procedure
    - c) Durbin two-step procedure
    - d) Hildreth-Lu search procedure

All of them estimate ρ and then apply transformation. Most common is CO iterative procedure. Eviews uses CO iterative procedure. One advantage of CO iterative method is that it can be used to estimate not only AR(1) scheme but also higher order autoregresive schemes, such as AR(2).

B. AC Correction and Common Factor (COMFAC) Test

Suppose that we have genuine (pure) auto correlation such as:

(1) $$Y_t = \beta_0 + \beta_1 X_t + u_t$$

where $$u_t = \rho u_{t-1} + \varepsilon_t$$

(1') $$Y_t = \beta_0 + \beta_1 X_t + \rho u_{t-1} + \varepsilon_t$$

Then we can write

(2) $$Y_{t-1} = \beta_0 + \beta_1 X_{t-1} + u_{t-1}$$

Multiply both sides by ρ;

$$\rho Y_{t-1} = \rho\beta_0 + \rho\beta_1 X_{t-1} + \rho u_{t-1}$$

(3) $$\rho u_{t-1} = \rho Y_{t-1} - \rho\beta_0 - \rho\beta_1 X_{t-1}$$

Putting (3) into (1') yields;

$$Y_t = \beta_0 + \beta_1 X_t + \rho Y_{t-1} - \rho\beta_0 - \rho\beta_1 X_{t-1} + \varepsilon_t$$

Rearranging we have

(4) $$Y_t = \underbrace{\beta_0(1-\rho)}_{\lambda_1} + \underbrace{\rho}_{\lambda_2} Y_{t-1} + \underbrace{\beta_1}_{\lambda_3} X_t \underbrace{-\beta_1\rho}_{\lambda_4} X_{t-1} + \varepsilon_t$$

which is $ARDL(1,1)$[6]

(4') $$Y_t = \lambda_1 + \lambda_2 Y_{t-1} + \lambda_3 X_t + \lambda_4 X_{t-1} + \varepsilon_t$$

with the restriction,

$$\lambda_4 = -\beta_1\rho \quad \text{or} \quad \lambda_4 = -\lambda_2\lambda_3$$

The presence of this implicit restriction provides us with an opportunity to perform a test of the validity of the model specification of genuine AC based on AR(1) scheme. This test is known as the "*Common Factor test (COMFAC Test)*".

The test helps us to discriminate between cases where DW statistic is low because the disturbance term is genuinely subject to a AR(1) process and cases where it is low for other reasons such as model misspecification etc.

The usual F test of a restriction is not appropriate since the restriction $\lambda_4 = -\lambda_2\lambda_3$ is not linear. Instead we calculate the *Likelihood Ratio (LR)* statistic.

$$LR[COMFAC] = T.\ln\left(\frac{SSR_R}{SSR_U}\right)$$

$SSR_R \Rightarrow SSR$ of Model 4
$SSR_U \Rightarrow SSR$ of Model 4'

We can compare: $LR[COMFAC] \sim \chi^2_p(\alpha)$

[6] *ARDL(maximum lag of Y, maximum lag of X)*

For our example, the COMFAC test is as follows:

H_0: $\lambda_4 + \lambda_2\lambda_3 = 0$
H_A: $\lambda_4 + \lambda_2\lambda_3 \neq 0$

If $LR\left[COMFAC\right] > \chi_1^2(0.05) \Rightarrow RH_0$ of COMFAC restriction is valid. Hence the source of AC is not $u_t = \rho u_{t-1} + \varepsilon_t$. Thus GLS or EGLS *must not be applied* to correct AC!

If we have 2 variables in the original model;

(5) $Y_t = \beta_0 + \beta_1 X_t + \beta_2 Z_t + u_t$ where $u_t = \rho u_{t-1} + \varepsilon_t$

Then we can write

(5') $Y_t = \beta_0 + \beta_1 X_t + \beta_2 Z_t + \rho u_{t-1} + \varepsilon_t$

From (5) we can also write

$$Y_{t-1} = \beta_0 + \beta_1 X_{t-1} + \beta_2 Z_{t-1} + u_{t-1}$$

Multiplying both sides by ρ produces;

$$\rho Y_{t-1} = \rho\beta_0 + \rho\beta_1 X_{t-1} + \rho\beta_2 Z_{t-1} + \rho u_{t-1}$$

(6) $\rho u_{t-1} = \rho Y_{t-1} - \rho\beta_0 - \rho\beta_1 X_{t-1} - \rho\beta_2 Z_{t-1}$

Putting (6) into (5') we get;

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Z_t + \rho Y_{t-1} - \rho\beta_0 - \rho\beta_1 X_{t-1} - \rho\beta_2 Z_{t-1} + \varepsilon_t$$

(7) $Y_t = \underbrace{(1-\rho)\beta_0}_{\lambda_1} + \underbrace{\rho}_{\lambda_2} Y_{t-1} + \underbrace{\beta_1}_{\lambda_3} X_t + \underbrace{\beta_2}_{\lambda_4} Z_t \underbrace{-\rho\beta_1}_{\lambda_5} X_{t-1} \underbrace{-\rho\beta_2}_{\lambda_6} Z_{t-1} + \varepsilon_t$

with the restrictions

$$\lambda_5 = -\rho\beta_1 \quad \text{or} \quad \lambda_5 = -\lambda_2\lambda_3$$
$$\lambda_6 = -\rho\beta_2 \quad \text{or} \quad \lambda_6 = -\lambda_2\lambda_4$$

(7') $Y_t = \lambda_1 + \lambda_2 Y_{t-1} + \lambda_3 X_t + \lambda_4 Z_t + \lambda_5 X_{t-1} + \lambda_6 Z_{t-1} + \varepsilon_t$

To test for COMFAC restrictions we follow the following steps:

(1) H_0: $\lambda_5 + \lambda_2\lambda_3 = 0$ and $\lambda_6 + \lambda_2\lambda_4 = 0$

H_A: $\lambda_5 + \lambda_2\lambda_3 \neq 0$ and $\lambda_6 + \lambda_2\lambda_4 \neq 0$

(2) Run model (7) which is our restricted model and get its SSR as SSR_R. Run model (7') which is our unrestricted model and get its SSR as SSR_U.

(3) Calculate

$$LM\left[COMFAC\right] = T.\ln\left(\frac{SSR_R}{SSR_U}\right)$$

and compare to $\chi_2^2(0.05)$ since

$$LM\left[COMFAC\right] \sim \chi_p^2(\alpha)$$

$p \rightarrow$ number of COMFAC restrictions. Here p=2

If $LM\left[COMFAC\right] > \chi_2^2(0.05) \Rightarrow$ RH_0, so we do not use GLS or EGLS to fix AC since AR(1) model does not seem to be an adequate specification for AC.

- *If not rejected the coefficient of Y_{t-1} may be interpreted as an estimate of ρ.*

C. The Newey-West Method of Correcting the OLS Standard Errors

Instead of using the EGLS methods, we can still use OLS but correct the standard errors for auto correlation by a procedure developed by *Newey and West*. This is an extension of White's Heteroscedasticity-consistent standard errors (HC Standard errors).

The corrected Standard errors are known as *HAC (heteroscedasticity and autocorrelation consistent) standard errors*, or simply *Newey-West standard errors*.

Most modern computer packages now calculate the HAC standard errors. It is important to note that the HAC procedure is valid in large samples and may not be appropriate in small samples.

Consequently, in large samples we have HAC standard errors so we do not have to worry about the EGLS transformation.

Therefore, if a sample is reasonably large ($T \geq 50$) one should use the Newey-West standard errors not only in situations of autocorrelation but also in cases of heteroscedasticity since HAC standard errors can handle both.

Example

$$\widehat{Y}_t = 32.7419 + 0.6704 X_t$$

se→ (1.3940) (0.0157)
t→ (23.4874) (42.7813)

$R^2=0.9765$, DW=0.1739

The estimation results with HAC standard errors:

$$\widehat{Y}_t = 32.7419 + 0.6704 X_t$$

HAC se→ (2.9162) (0.0302)
HAC t→ (11.227) (22.199)

$R^2=0.9765$, DW=0.1739

Note that the HAC standard errors are much greater than the OLS standard errors and therefore the HAC *t* ratios are much smaller than the OLS *t* ratios.

This shows that OLS had in fact underestimated the true standard errors. This can also be seen as a sign for autocorrelation and heteroscedasticity. But we do not need to worry about it since HAC standard errors have taken this into account in correcting OLS standard errors.

D. OLS versus EGLS and HAC in Genuine Autocorrelation

The practical problem facing the researcher is this: In the presence of autocorrelation, OLS estimators, although unbiased, consistent and asymptotically normally distributed, are not efficient. Therefore, the usual inference procedure based on the t, F and χ^2 tests is no longer appropriate. On the other hand, EGLS and HAC procedure estimators that are efficient, but the finite or small-sample, properties of these estimators are not certain. This means that in small samples, the EGLS and HAC might actually do worse than OLS.

As a matter of fact, in a Monte Carlo study Griliches and Rao (1969) found that *if the sample is relatively small (T=15-20)* and the *coefficient of autocorrelation, ρ, is less than 0.3*, OLS is as good or better than EGLS. Hence, one may use OLS in small samples in which the estimated ρ is less than 0.3.

VIII. Resolving Apparent (Impure) Autocorrelation and Dynamic Econometric Models

A. Finite Distributed Lags

For psychological, technological, and institutional reasons, a regressand may respond to a regressor(s) with a time lag. Regression models that take into account time lags are known as *dynamic* or *lagged* regression models (Gujarati). There are two types of *lagged models*: *distributed-lag* and *autoregressive*. If the regression model includes not only the current but also the lagged (past) values of the explanatory variables (the X's), it is called a *distributed-lag model*. If the model includes one or more lagged values of the dependent variable among its explanatory variables, it is called an *autoregressive model*. Thus, $Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + u_t$ represents a *distributed-lag model*, whereas $Y_t = \alpha + \beta X_t + \gamma Y_{t-1} + u_t$ is an example of an *autoregressive model*.

More generally we may write:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + ... + \beta_q X_{t-q} + u_t \qquad (1)$$

which is a *distributed-lag model* with a *finite lag* of *q* time periods: it is called a *finite distributed lag model of order q*[7]. The coefficient β_0 is known as the **short-run**, or **impact**, **multiplier** because it gives the change in the mean value of Y following a unit change in X in the same time period. If the change in X is maintained at the same level thereafter, then, $(\beta_0 + \beta_1)$ gives the change in (the mean value of) Y in the next period, $(\beta_0 + \beta_1 + \beta_2)$ in the following period, and so on. These partial sums are called **interim**, or **intermediate**, **multipliers**. Finally, after *q* periods we obtain

[7] It is called a finite distributed lag model of order q because it is assumed that after a finite number of periods *q*, changes in X no longer have an impact on Y.

$$\sum_{s=0}^{q} \beta_s = \beta_0 + \beta_1 + \beta_2 + \ldots + \beta_q = \beta$$

which is known as the **long-run**, or **total**, **distributed-lag multiplier**, provided the sum β exists. The *total multiplier* is the final effect on Y of the sustained increase after *q* or more periods have elapsed.

The *mean lag* is simply the weighted average of the lags, in which the lag *s* periods enters with weight β_s, or with relative weight $\dfrac{\beta_s}{\sum_{s=0}^{q} \beta_s}$, thus:

$$\text{Mean lag} = \sum_{s=0}^{q} s \underbrace{\left(\frac{\beta_s}{\sum_{s=0}^{q} \beta_s} \right)}_{weight} = \frac{\sum_{s=0}^{q} s\beta_s}{\sum_{s=0}^{q} \beta_s}$$

Similarly the *median lag* is the number of periods required for the long-run adjustment to be one-half complete.

A purely distributed-lag model can be estimated by OLS, but in that case there is the problem of multicollinearity since successive lagged values of a regressor tend to be correlated. As a result, some shortcut methods have been devised. These include the Koyck, the adaptive expectations, and partial adjustment mechanisms and Almon polynomial distributed-lag model. However we will not go into details of these methods here.

On the other hand, autoregressiveness poses estimation challenges; if the lagged regressand (dependent variable) is correlated with the error term, OLS estimators of such models are not only biased but also are inconsistent[8].

[8] Bias and inconsistency are the case with the Koyck and the adaptive expectations models; the partial adjustment model is different in that it can be consistently estimated by OLS despite the presence of the lagged regressand (G, p.702). To estimate the Koyck and adaptive expectations models consistently, the most popular method is the method of instrumental variable. The instrumental variable is a proxy variable for the lagged regressand but with the property that it is uncorrelated with the error term. An alternative to the lagged regression models just discussed is the Almon polynomial distributed-lag model, which avoids the estimation

Despite the estimation problems, which can be surmounted, the distributed and autoregressive models have proved extremely useful in empirical economics because they make the otherwise static economic theory a dynamic one by taking into account explicitly the role of time. Such models help us to distinguish between short- and long-run response of the dependent variable to a unit change in the value of the explanatory variable(s). Thus, for estimating short- and long-run price, income, substitution, and other elasticities these models have proved to be highly useful (G, p.703)

1. Assumptions

As generally the *time-series variables are random*, it is useful to revise the necessary assumptions under which we can consider the properties of least squares and other estimators.

In distributed lag models due to their time series characteristics both *Y* and *X* are usually random. Consider the example of unemployment (Y_t) and output growth (X_t). They are both *random*. *They are observed at the same time; we do not know their values prior to ''sampling.'' We do not ''set'' output growth and then observe the resulting level of unemployment.*

To accommodate this randomness we assume that the X's are random and that u_t is independent of all X's in the sample—past, current, and future. This assumption, in conjunction with the other multiple regression assumptions, is sufficient for the least squares estimator to be unbiased and to be best linear unbiased conditional on the X's in the sample. With the added assumption of normally distributed error terms, our usual *t* and *F* tests have finite sample justification. Accordingly, the multiple regression assumptions[9] can be modified

problems associated with the autoregressive models. The major problem with the Almon approach, however, is that one must prespecify both the lag length and the degree of the polynomial. There are both formal and informal methods of resolving the choice of the lag length and the degree of the polynomial.

[9] Remember that for multiple regression model of $Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t2} + ... + \beta_K X_{tK} + u_t$ where t=1,…,T.

1) $E(u_t) = 0$

for the distributed lag model as follows (assumptions of the distributed lag model):

Consider $Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + ... + \beta_q X_{t-q} + u_t$ where t=q+1,…,T.

1) $E(u_t) = 0$
2) $Var(u_t) = \sigma^2$
3) $Cov(u_t, u_s) = 0,\ t \neq s$
4) Y and X are stationary *random* variables, and u_t is independent of *current*, *past* and *future* values of X.
5) $u_t \sim N(0, \sigma^2)$ (required in finite samples for hypothesis testing)

2. Example (Okun's Law)

To illustrate and expand on the various distributed lag concepts, we introduce an economic model known as Okun's Law. In this model the change in the unemployment rate from one period to the next depends on the rate of growth of output in the economy:

$$U_t - U_{t-1} = -\gamma(G_t - G_N) \qquad (2)$$

where U_t is the unemployment rate in period *t*, G_t is the growth rate of output in period *t*, and G_N is the ''normal'' growth rate, which we assume is constant over time[10].

The parameter γ is positive, implying that when the growth of output is above the normal rate, unemployment falls; a growth rate below the

---

2) $Var(u_t) = \sigma^2$
3) $Cov(u_t, u_s) = 0,\ t \neq s$
4) The values of each X_{tk} are not random (fixed in repeated sampling).
5) $u_t \sim N(0, \sigma^2)$ (required in finite samples for hypothesis testing)

[10] The normal growth rate G_N is the rate of output growth needed to maintain a constant unemployment rate. It is equal to the sum of labor force growth and labor productivity growth.

normal rate leads to an increase in unemployment. We expect $0<\gamma<1$, reflecting that output growth leads to less than one-to-one adjustments in unemployment.

Denoting the change in unemployment by $DU_t = \Delta U_t = U_t - U_{t-1}$, setting $\beta_0 = -\gamma$, $\alpha = \gamma G_N$ and including an error term then yields:

$$DU_t = \alpha + \beta_0 G_t + u_t \qquad (3)$$

Recognizing that changes in output are likely to have a distributed-lag effect on unemployment—not all of the effect will take place instantaneously—we expand (3) to include lags of G_t:

$$DU_t = \alpha + \beta_0 G_t + \beta_1 G_{t-1} + \beta_2 G_{t-2} + \ldots + \beta_q G_{t-q} + u_t \qquad (4)$$

To estimate this relationship we use quarterly U.S. data on unemployment and the percentage change in gross domestic product (GDP) from quarter 2, 1985, to quarter 3, 2009. Output growth is defined as $G_t = \dfrac{GDP_t - GDP_{t-1}}{GDP_{t-1}}.100$

These data are stored in the *okun.rar* file that can be reached from online.metu.edu.tr service of the ECON302.

Table 9.1 Spreadsheet of Observations for Distributed Lag Model

t	Quarter	U_t	U_{t-1}	DU_t	G_t	G_{t-1}	G_{t-2}	G_{t-3}
1	1985Q2	7.3	•	•	1.4	•	•	•
2	1985Q3	7.2	7.3	−0.1	2.0	1.4	•	•
3	1985Q4	7.0	7.2	−0.2	1.4	2.0	1.4	•
4	1986Q1	7.0	7.0	0.0	1.5	1.4	2.0	1.4
5	1986Q2	7.2	7.0	0.2	0.9	1.5	1.4	2.0
96	2009Q1	8.1	6.9	1.2	−1.2	−1.4	0.3	0.9
97	2009Q2	9.3	8.1	1.2	−0.2	−1.2	−1.4	0.3
98	2009Q3	9.6	9.3	0.3	0.8	−0.2	−1.2	−1.4

The OLS estimation results are as follows:

Table 9.2 Estimates for Okun's Law Finite Distributed Lag Model

Lag Length $q = 3$				
Variable	Coefficient	Std. Error	t-value	p-value
Constant	0.5810	0.0539	10.781	0.0000
G_t	−0.2021	0.0330	6.120	0.0000
G_{t-1}	−0.1645	0.0358	−4.549	0.0000
G_{t-2}	−0.0716	0.0353	−2.027	0.0456
G_{t-3}	0.0033	0.0363	0.091	0.9276
Observations = 95	$R^2 = 0.652$		$\hat{\sigma} = 0.1743$	
Lag Length $q = 2$				
Variable	Coefficient	Std. Error	t-value	p-value
Constant	0.5836	0.0472	12.360	0.0000
G_t	−0.2020	0.0324	−6.238	0.0000
G_{t-1}	−0.1653	0.0335	−4.930	0.0000
G_{t-2}	−0.0700	0.0331	−2.115	0.0371
Observations = 96	$R^2 = 0.654$		$\hat{\sigma} = 0.1726$	

Least squares estimates of the coefficients and related statistics for are reported in table above for lag lengths q=2 and q=3. Note that all coefficients of G and its lags have the expected negative sign and are significantly different from zero at a 5% significance level, with the exception of that for G_{t-3} when q=3. A variety of measures are available for choosing q. In this case we drop G_{t-3} and settle on a model of order 2 because β_3 is insignificant and has the wrong sign, and β_0; β_1, and β_2 all have the expected negative signs and are significantly different from zero. The information criteria AIC and SC are another set of measures that can be used for assessing lag length.

B. Autoregressive Distributed Lag Models (ARDL Models)

An autoregressive distributed lag (ARDL) model is one that contains both lagged X_t's and lagged Y_t's. In its general form, with p lags of Y and q lags of X, an ARDL(p, q) model can be written as

$$Y_t = \delta + \theta_1 Y_{t-1} + \ldots + \theta_p Y_{t-p} + \delta_0 X_t + \delta_1 X_{t-1} + \ldots + \delta_q X_{t-q} + u_t \qquad (5)$$

In the context of the above equation, the assumption 4 introduced in previous section is no longer valid since it says that u_t is not correlated with current, past, and future values of Y_{t-1}; X_t and X_{t-q}. Since Y_t is a future value of Y_{t-1} and Y_t depends directly on u_t, the assumption will be violated. We can, however, replace it with a weaker, more tenable assumption—namely, that u_t is uncorrelated with *current and past* values of the right-hand-side variables. Under this assumption, the OLS estimator is *no longer unbiased*, but it does *have the desirable large sample property of consistency*, and, if the errors are normally distributed, it is best in a large sample sense. Thus, we replace the assumption 4 as we said:

Consider $Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t2} + ... + \beta_K X_{tK} + u_t$ where some of the X_{tK} may be *lagged values of Y*;

1) $E(u_t) = 0$
2) $Var(u_t) = \sigma^2$
3) $Cov(u_t, u_s) = 0, t \neq s$
4) Y and X are stationary *random* variables, and u_t *is uncorrelated with all* X_{tk} *and their past values.*
5) $u_t \sim N(0, \sigma^2)$ (required in finite samples for hypothesis testing)

The ARDL model has several advantages. It captures dynamic effects from lagged X's and lagged Y's, and by including a sufficient number of lags of Y and X, we can eliminate serial correlation in the errors. Moreover, an ARDL model can be transformed into one with only lagged X's which go back into the infinite past:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + ... + u_t \qquad (6)$$

$$Y_t = \alpha + \sum_{s=0}^{\infty} \beta_s X_{t-s} + u_t$$

Because it does not have a finite cut off point, this model is called an infinite distributed lag model. It contrasts with the finite distributed lag model, where the effect of the lagged x's was assumed to cut off to zero after q lags. Like before, the parameter β_s is the distributed lag weight or the s-period delay multiplier showing the effect of a change in X_t on Y_{t+s}. The total or long-run multiplier showing the long-run effect of a sustained change in X_t is $\sum_{s=0}^{\infty}\beta_s$. For the transformation from (5) to (6) to be valid, the effect of a change must gradually die out. Thus, the values of β_s for large *s* will be small and decreasing, a property that is necessary for the infinite sum $\sum_{s=0}^{\infty}\beta_s$ to be finite.

OLS is an appropriate estimation technique under assumptions of the previous section, but the main concern for estimation is choice of the lag lengths p and q.

There are a number of different criteria for choosing *p* and *q*. Because they all do not necessarily lead to the same choice, there is a degree of subjective judgment that must be used. Four possible criteria are

1) Has serial correlation in the errors been eliminated? If not, then least squares will be biased in small and large samples. It is important to include sufficient lags, especially of Y, to ensure that serial correlation does not remain. It can be checked using the correlogram or Lagrange multiplier tests.

2) Are the signs and magnitudes of the estimates consistent with our expectations from economic theory? Estimates which are poor in this sense may be a consequence of poor choices for p and q, but they could also be symptomatic of a more general modeling problem.

3) Are the estimates significantly different from zero, particularly those at the longest lags?

4) What values for p and q minimize information criteria such as the AIC and SBC?

1. Example (Phillips Curve)

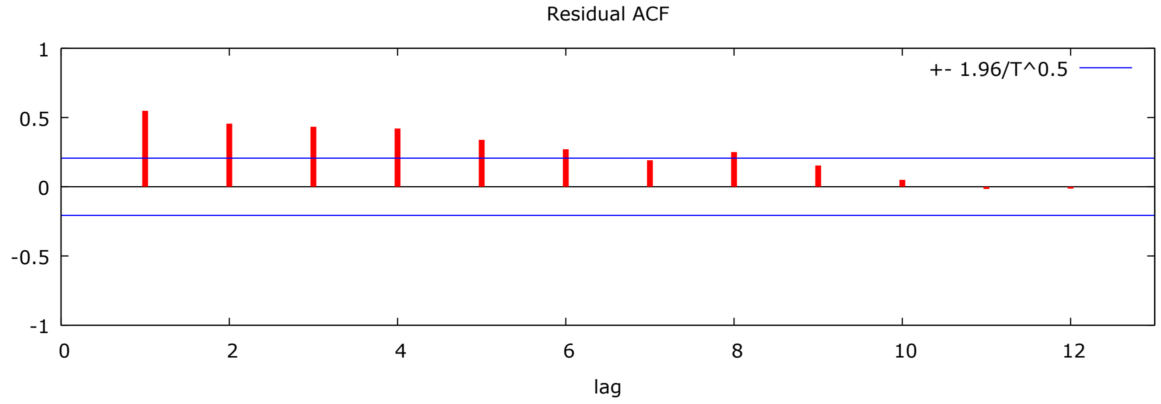
Model 11: OLS, using observations 1987:2-2009:3 (T = 90)
Dependent variable: inf

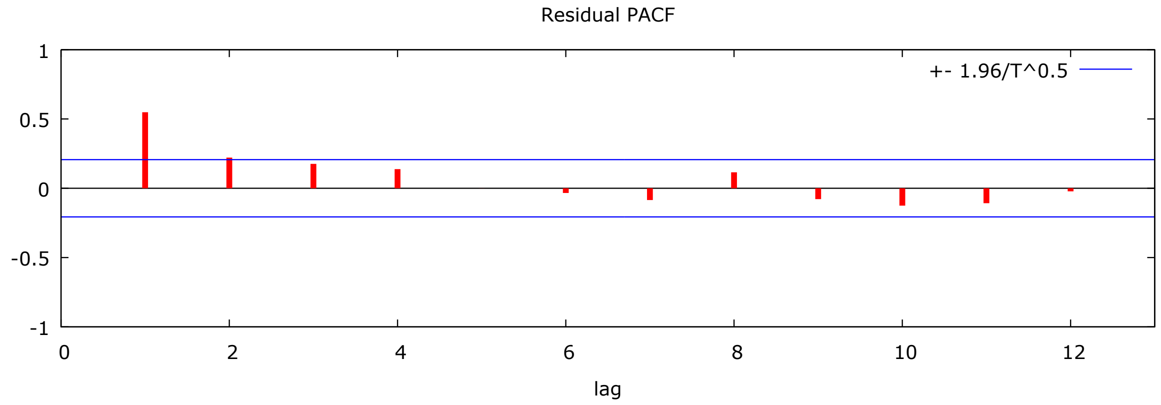
	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	0.777621	0.0658249	11.8135	<0.00001	***
d_u	-0.527864	0.229405	-2.3010	0.02375	**

Mean dependent var	0.791111	S.D. dependent var	0.636819
Sum squared resid	34.04454	S.E. of regression	0.621989
R-squared	0.056752	Adjusted R-squared	0.046033
F(1, 88)	5.294666	P-value(F)	0.023754
Log-likelihood	-83.95817	Akaike criterion	171.9163
Schwarz criterion	176.9160	Hannan-Quinn	173.9325
rho	0.549882	Durbin-Watson	0.887289

Corelogram:

```
Residual autocorrelation function

  LAG      ACF          PACF         Q-stat. [p-value]

    1   0.5487  ***   0.5487 ***     28.0056  [0.000]
    2   0.4557  ***   0.2213 **      47.5475  [0.000]
    3   0.4332  ***   0.1761 *       65.4091  [0.000]
    4   0.4205  ***   0.1383         82.4327  [0.000]
    5   0.3390  ***  -0.0003         93.6296  [0.000]
    6   0.2710  **   -0.0338        100.8674  [0.000]
    7   0.1912  *    -0.0850        104.5151  [0.000]
    8   0.2507  **    0.1155        110.8611  [0.000]
    9   0.1534       -0.0784        113.2669  [0.000]
   10   0.0500       -0.1252        113.5256  [0.000]
   11  -0.0157       -0.1076        113.5516  [0.000]
   12  -0.0132       -0.0218        113.5700  [0.000]
```





There is AC in this model. Let us try ARDL model for estimation.

OLS, using observations 1988:2-2009:3 (T = 86)
Dependent variable: inf

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	0.130792	0.106368	1.2296	0.22264	
d_u	-0.811945	0.269261	-3.0155	0.00349	***
d_u_1	0.145599	0.289896	0.5022	0.61695	
d_u_2	0.0349051	0.317024	0.1101	0.91262	
d_u_3	0.0409242	0.303401	0.1349	0.89306	
d_u_4	-0.327582	0.287826	-1.1381	0.25864	
inf_1	0.245415	0.110832	2.2143	0.02981	**
inf_2	0.103233	0.114222	0.9038	0.36896	
inf_3	0.166993	0.114691	1.4560	0.14950	
inf_4	0.242344	0.116723	2.0762	0.04125	**

Mean dependent var	0.748837	S.D. dependent var	0.619059
Sum squared resid	17.82454	S.E. of regression	0.484286
R-squared	0.452813	Adjusted R-squared	0.388015
F(9, 76)	6.988033	P-value(F)	2.78e-07
Log-likelihood	-54.35655	Akaike criterion	128.7131
Schwarz criterion	153.2566	Hannan-Quinn	138.5907
rho	-0.042260	Durbin-Watson	2.064445

OLS, using observations 1988:1-2009:3 (T = 87)
Dependent variable: inf

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	0.1001	0.0982599	1.0187	0.31137	
d_u	-0.790172	0.188533	-4.1912	0.00007	***
inf_1	0.23544	0.101556	2.3183	0.02295	**
inf_2	0.121328	0.103757	1.1693	0.24569	
inf_3	0.16769	0.10496	1.5977	0.11401	
inf_4	0.281916	0.10138	2.7808	0.00674	***

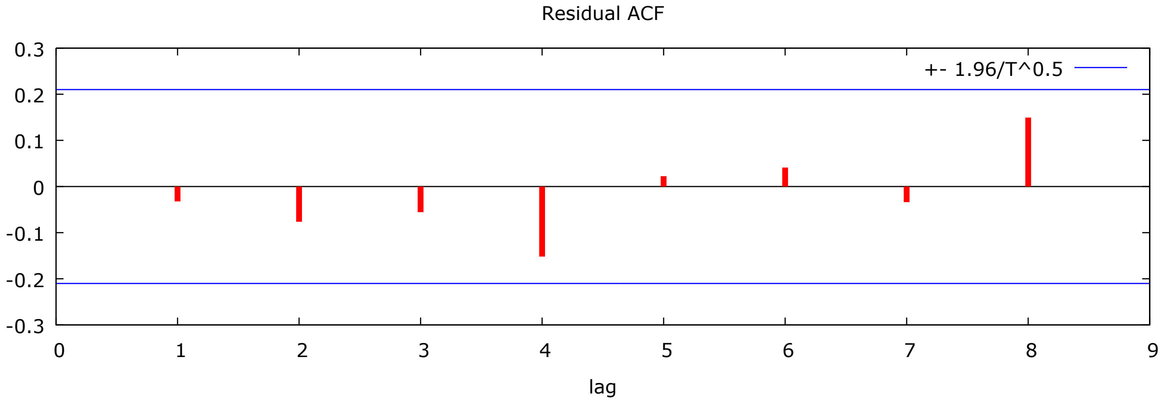
Mean dependent var	0.760920	S.D. dependent var	0.625682
Sum squared resid	18.23336	S.E. of regression	0.474450
R-squared	0.458422	Adjusted R-squared	0.424992
F(5, 81)	13.71262	P-value(F)	1.07e-09
Log-likelihood	-55.47215	Akaike criterion	122.9443
Schwarz criterion	137.7397	Hannan-Quinn	128.9020
rho	-0.032772	Durbin's h	-0.903935

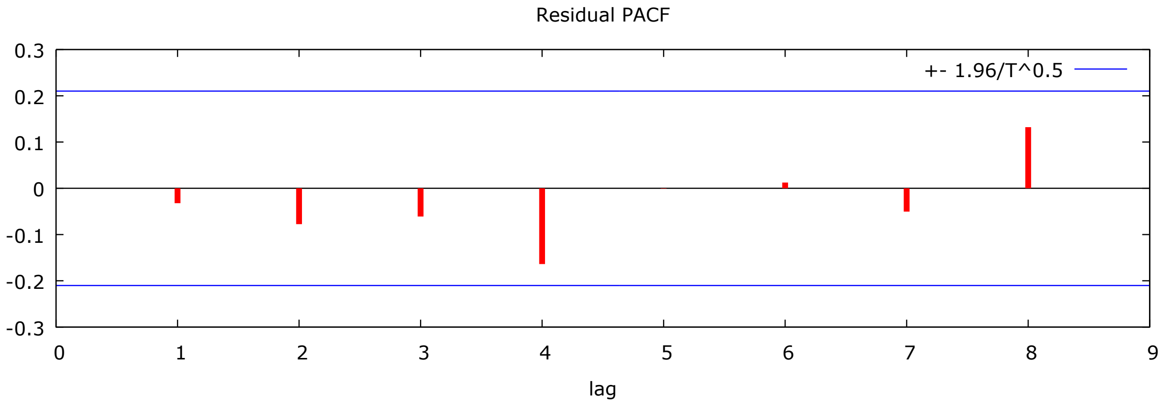
AC test:

```
Breusch-Godfrey test for autocorrelation up to order 4
OLS, using observations 1988:1-2009:3 (T = 87)
Dependent variable: uhat

             coefficient   std. error   t-ratio    p-value
  --------------------------------------------------------
  const      -0.109089      0.112161    -0.9726    0.3338
  d_u        -0.0178808     0.214627    -0.08331   0.9338
  inf_1      -0.135065      0.268527    -0.5030    0.6164
  inf_2       0.109993      0.289074     0.3805    0.7046
  inf_3      -0.168853      0.286317    -0.5897    0.5571
  inf_4       0.319178      0.250648     1.273     0.2067
  uhat_1      0.0840957     0.284311     0.2958    0.7682
  uhat_2     -0.191709      0.285707    -0.6710    0.5042
  uhat_3      0.0763031     0.279820     0.2727    0.7858
  uhat_4     -0.458110      0.236134    -1.940     0.0560   *

  Unadjusted R-squared = 0.077256

Test statistic: LMF = 1.611691,
with p-value = P(F(4,77) > 1.61169) = 0.18

Alternative statistic: TR^2 = 6.721272,
with p-value = P(Chi-square(4) > 6.72127) = 0.151

Ljung-Box Q' = 3.05947,
with p-value = P(Chi-square(4) > 3.05947) = 0.548
```

Result: NO AC

```
Residual autocorrelation function

 LAG      ACF          PACF         Q-stat. [p-value]

   1  -0.0322       -0.0322         0.0934  [0.760]
   2  -0.0764       -0.0775         0.6245  [0.732]
   3  -0.0555       -0.0611         0.9086  [0.823]
   4  -0.1518       -0.1638         3.0595  [0.548]
   5   0.0224       -0.0007         3.1070  [0.683]
   6   0.0411        0.0123         3.2685  [0.774]
   7  -0.0338       -0.0504         3.3793  [0.848]
   8   0.1494        0.1320         5.5664  [0.696]
```





Result: NO AC

Table 9.4 AIC and SC Values for Phillips Curve ARDL Models

p	q	AIC	SC	p	q	AIC	SC
1	0	−1.247	−1.160	1	1	−1.242	−1.128
2	0	−1.290	−1.176	2	1	−1.286	−1.142
3	0	−1.335	−1.192	3	1	−1.323	−1.151
4	0	−1.402	−1.230	4	1	−1.380	−1.178
5	0	−1.396	−1.195	5	1	−1.373	−1.143
6	0	−1.378	−1.148	6	1	−1.354	−1.096

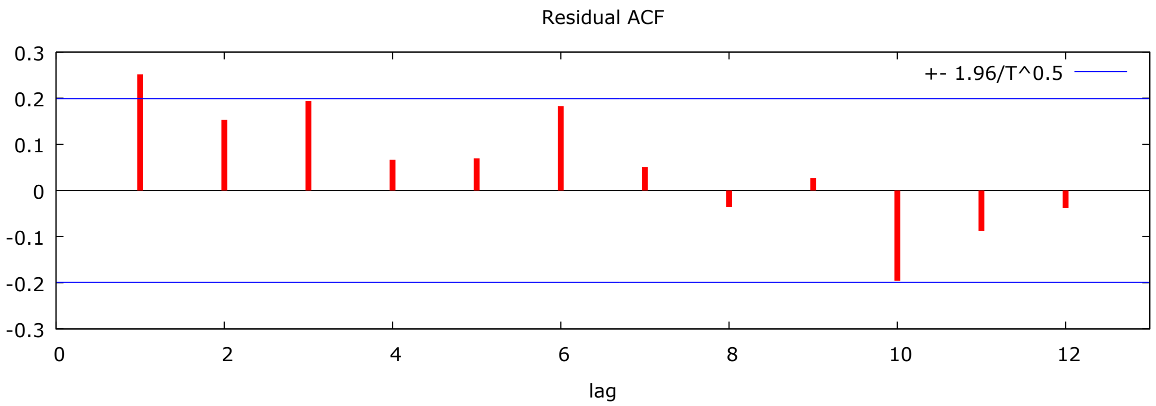
ARDL(4,0) is pointed out by AIC and SBC.

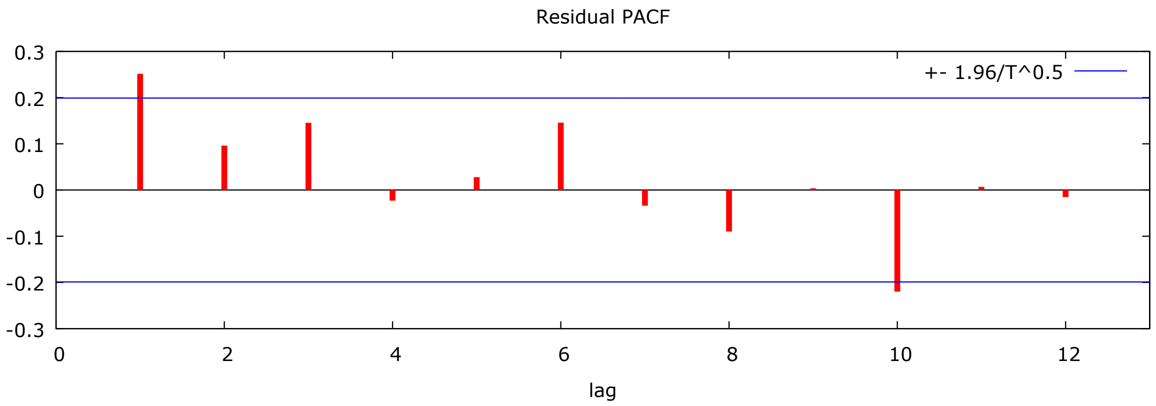
2. Example (Okun's Law)

Model 1: OLS, using observations 1985:3-2009:3 (T = 97)
Dependent variable: d_u

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	0.421338	0.0460029	9.1590	<0.00001	***
g	-0.311801	0.0321719	-9.6917	<0.00001	***

Mean dependent var	0.023711	S.D. dependent var	0.287508
Sum squared resid	3.990212	S.E. of regression	0.204944
R-squared	0.497167	Adjusted R-squared	0.491874
F(1, 95)	93.92957	P-value(F)	7.53e-16
Log-likelihood	17.11999	Akaike criterion	-30.23998
Schwarz criterion	-25.09056	Hannan-Quinn	-28.15781
rho	0.252546	Durbin-Watson	1.490251





```
Breusch-Godfrey test for autocorrelation up to order 4
OLS, using observations 1985:2-2009:3 (T = 98)
Dependent variable: uhat

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const      -0.0444271    0.0458515    -0.9689   0.3351
  g           0.0377440    0.0327434     1.153    0.2520
  uhat_1      0.228771     0.104912      2.181    0.0318   **
  uhat_2      0.108799     0.111587      0.9750   0.3321
  uhat_3      0.213257     0.114245      1.867    0.0651   *
  uhat_4     -0.0164596    0.113101     -0.1455   0.8846

  Unadjusted R-squared = 0.114500

Test statistic: LMF = 2.974013,
with p-value = P(F(4,92) > 2.97401) = 0.0233

Alternative statistic: TR^2 = 11.220956,
with p-value = P(Chi-square(4) > 11.221) = 0.0242

Ljung-Box Q' = 13.0098,
with p-value = P(Chi-square(4) > 13.0098) = 0.0112
```

There is AC in the model. Let us try ARDL by starting from ARDL(4,4).

Model 2: OLS, using observations 1986:3-2009:3 (T = 93)
Dependent variable: d_u

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	0.320122	0.105019	3.0482	0.00309	***
g	-0.167614	0.0326717	-5.1302	<0.00001	***
g_1	-0.0993119	0.0384564	-2.5825	0.01156	**
g_2	-0.00996099	0.0393811	-0.2529	0.80094	
g_3	0.0675639	0.0401257	1.6838	0.09598	*
g_4	-0.032178	0.0417702	-0.7704	0.44328	
d_u_1	0.347286	0.112357	3.0909	0.00272	***
d_u_2	0.0812942	0.121125	0.6712	0.50398	
d_u_3	0.177623	0.121783	1.4585	0.14847	
d_u_4	-0.206892	0.110192	-1.8776	0.06395	*

Mean dependent var	0.025806	S.D. dependent var	0.291884
Sum squared resid	2.203405	S.E. of regression	0.162933
R-squared	0.718884	Adjusted R-squared	0.688402
F(9, 83)	23.58353	P-value(F)	2.08e-19
Log-likelihood	42.06941	Akaike criterion	-64.13881
Schwarz criterion	-38.81282	Hannan-Quinn	-53.91290
rho	0.037408	Durbin-Watson	1.852595

Model 3: OLS, using observations 1986:3-2009:3 (T = 93)
Dependent variable: d_u

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	0.283362	0.0933251	3.0363	0.00319	***
g	-0.170854	0.0323213	-5.2861	<0.00001	***
g_1	-0.0952262	0.0379967	-2.5062	0.01413	**
g_2	-0.0144314	0.0388568	-0.3714	0.71128	
g_3	0.0664671	0.0400033	1.6615	0.10033	
d_u_1	0.336489	0.11121	3.0257	0.00329	***
d_u_2	0.0857541	0.120693	0.7105	0.47935	
d_u_3	0.196722	0.118944	1.6539	0.10188	
d_u_4	-0.180828	0.104615	-1.7285	0.08757	*

Mean dependent var	0.025806	S.D. dependent var	0.291884
Sum squared resid	2.219160	S.E. of regression	0.162538
R-squared	0.716874	Adjusted R-squared	0.689910
F(8, 84)	26.58597	P-value(F)	5.17e-20
Log-likelihood	41.73812	Akaike criterion	-65.47623
Schwarz criterion	-42.68284	Hannan-Quinn	-56.27291
rho	0.050136	Durbin-Watson	1.826062

Model 4: OLS, using observations 1986:3-2009:3 (T = 93)
Dependent variable: d_u

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	0.375684	0.075754	4.9593	<0.00001	***
g	-0.182942	0.0318165	-5.7499	<0.00001	***
g_1	-0.0887159	0.0381835	-2.3234	0.02255	**
g_2	-0.0121405	0.0392325	-0.3095	0.75774	
d_u_1	0.321584	0.111989	2.8716	0.00516	***
d_u_2	0.0397725	0.118688	0.3351	0.73837	
d_u_3	0.139566	0.115034	1.2133	0.22839	
d_u_4	-0.1753	0.105639	-1.6594	0.10072	

Mean dependent var	0.025806	S.D. dependent var	0.291884
Sum squared resid	2.292094	S.E. of regression	0.164213
R-squared	0.707569	Adjusted R-squared	0.683486
F(7, 85)	29.38097	P-value(F)	3.42e-20
Log-likelihood	40.23444	Akaike criterion	-64.46888
Schwarz criterion	-44.20808	Hannan-Quinn	-56.28815
rho	0.038560	Durbin-Watson	1.834593

Model 5: OLS, using observations 1986:3-2009:3 (T = 93)
Dependent variable: d_u

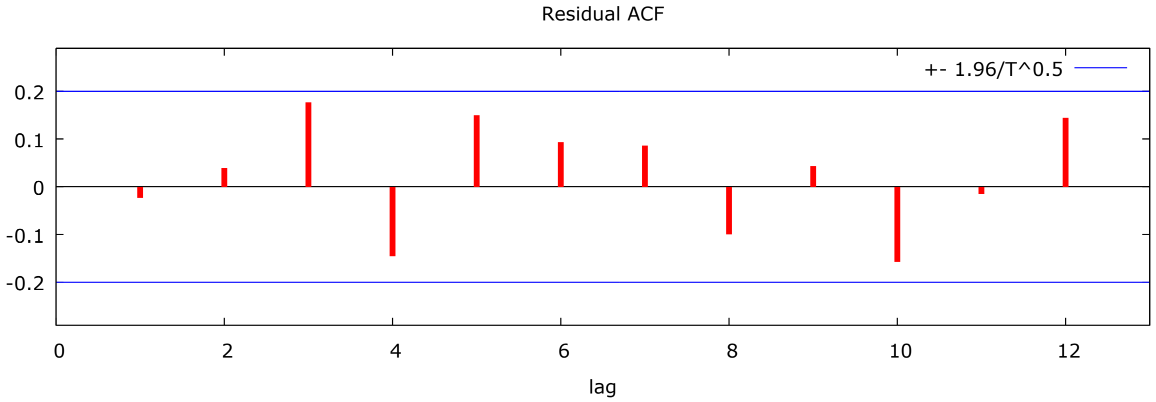
	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	0.361054	0.058877	6.1323	<0.00001	***
g	-0.183488	0.0316	-5.8066	<0.00001	***
g_1	-0.0892011	0.0379502	-2.3505	0.02104	**
d_u_1	0.329875	0.108164	3.0498	0.00304	***
d_u_2	0.0518018	0.11155	0.4644	0.64355	
d_u_3	0.137002	0.114131	1.2004	0.23328	
d_u_4	-0.171953	0.10453	-1.6450	0.10362	

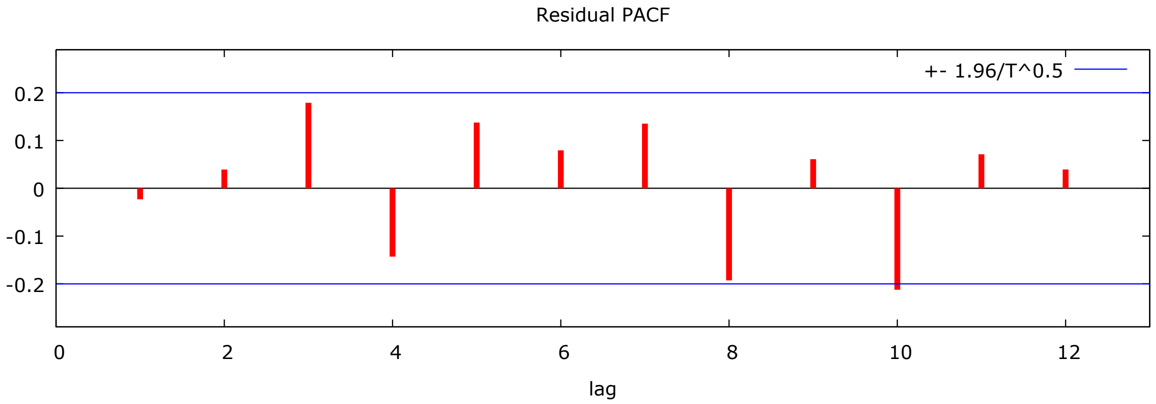
Mean dependent var	0.025806	S.D. dependent var	0.291884
Sum squared resid	2.294676	S.E. of regression	0.163347
R-squared	0.707239	Adjusted R-squared	0.686814
F(6, 86)	34.62591	P-value(F)	5.79e-21
Log-likelihood	40.18208	Akaike criterion	-66.36416
Schwarz criterion	-48.63597	Hannan-Quinn	-59.20602
rho	0.023259	Durbin-Watson	1.863929

Model 6: OLS, using observations 1985:4-2009:3 (T = 96)
Dependent variable: d_u

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	0.37801	0.0578398	6.5355	<0.00001	***
g	-0.184084	0.0306984	-5.9965	<0.00001	***
g_1	-0.0991552	0.0368244	-2.6926	0.00842	***
d_u_1	0.350116	0.084573	4.1398	0.00008	***

Mean dependent var	0.025000	S.D. dependent var	0.288736
Sum squared resid	2.422724	S.E. of regression	0.162277
R-squared	0.694101	Adjusted R-squared	0.684126
F(3, 92)	69.58413	P-value(F)	1.40e-23
Log-likelihood	40.39577	Akaike criterion	-72.79155
Schwarz criterion	-62.53415	Hannan-Quinn	-68.64534
rho	-0.024372	Durbin's h	-0.419594






```
Breusch-Godfrey test for autocorrelation up to order 4
OLS, using observations 1985:4-2009:3 (T = 96)
Dependent variable: uhat

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const      -0.0404068    0.0796285    -0.5074   0.6131
  g           0.00724623   0.0310951     0.2330   0.8163
  g_1         0.0235344    0.0524849     0.4484   0.6550
  d_u_1       0.0697011    0.160284      0.4349   0.6647
  uhat_1     -0.0806161    0.197649     -0.4079   0.6844
  uhat_2      0.0460614    0.117952      0.3905   0.6971
  uhat_3      0.200825     0.112315      1.788    0.0772  *
  uhat_4     -0.170182     0.114794     -1.483    0.1418

  Unadjusted R-squared = 0.063965

Test statistic: LMF = 1.503393,
with p-value = P(F(4,88) > 1.50339) = 0.208

Alternative statistic: TR^2 = 6.140635,
with p-value = P(Chi-square(4) > 6.14064) = 0.189

Ljung-Box Q' = 5.53979,
with p-value = P(Chi-square(4) > 5.53979) = 0.236
```

So no AC! ARDL(1,1) is chosen.

IX. Autoregressive Conditional Heteroscedasticity (ARCH)

With time series data, it is possible for serial correlation to ocur in the variance of the disturbance rather than in the disturbance itself.

A "large" disturbance in one period, resulting in an unusual value for Y in that period, is likely to result in greater uncertainty (which is measured by σ^2) in the next period.

We can capture this possibility by making σ^2 vary with the disturbance in the previous period:

(1) $$\sigma_t^2 = \sigma^2 + \gamma u_{t-1}^2$$

Equation (1) is known as a first-order autoregressive conditional heteroscedasticity (ARCH) process.

Many studies, particularly those involving speculative prices, have encountered ARCH effects, and it is important to be aware of such possibilities.

The p-th-order ARCH process $\left[ARCH(p)\right]$ can be written as:

$$\sigma_t^2 = \sigma^2 + \gamma_1 u_{t-1}^2 + \gamma_2 u_{t-2}^2 + ... + \gamma_p u_{t-p}^2$$

Testing for ARCH(p)

Consider the model

(X) $$Y_t = \beta_0 + \sum_{i=1}^{k} \beta_i X_{ti} + u_t$$ where t=1…..T

Steps:

1. Estimate model (X) by OLS and obtain residuals $\widehat{u}_t$
2. Estimate the following model by OLS and get R^2

$$\widehat{u}_t^2 = \alpha_0 + \alpha_1 \widehat{u}_{t-1}^2 + \alpha_2 \widehat{u}_{t-2}^2 + ... + \alpha_p \widehat{u}_{t-p}^2 + e_t \qquad \text{where} \quad t=1,\ldots,T-p$$

and calculate $LM\left[ARCH(p)\right]$ statistic.

$$LM\left[ARCH(p)\right] = T'.R^2$$

where $T'=T-p$

3. Test

$H_o = \alpha_1 = \alpha_2 = ... = \alpha_p = 0$
$H_A =$ at least one $\alpha_i \neq 0$

If $LM\left[ARCH(p)\right] > \chi_p^2(\alpha) \Rightarrow$ RH_0. So there is ARCH effects!

- *If ARCH effects were believed to be present, then estimation would have to be by Maximum Likelihood (ML) not OLS!*

Appendix 1 Cobweb Theorem

The Cobweb Theorem was developed by Henry Schultz, Jan Tingerman and Arthur Hanai in 1930's. However the name Cobweb was first coined by Nicholas Kaldor in 1934. The Cobweb phenomenon is explained below by a figure taken from Chauhan (2009, p.60).

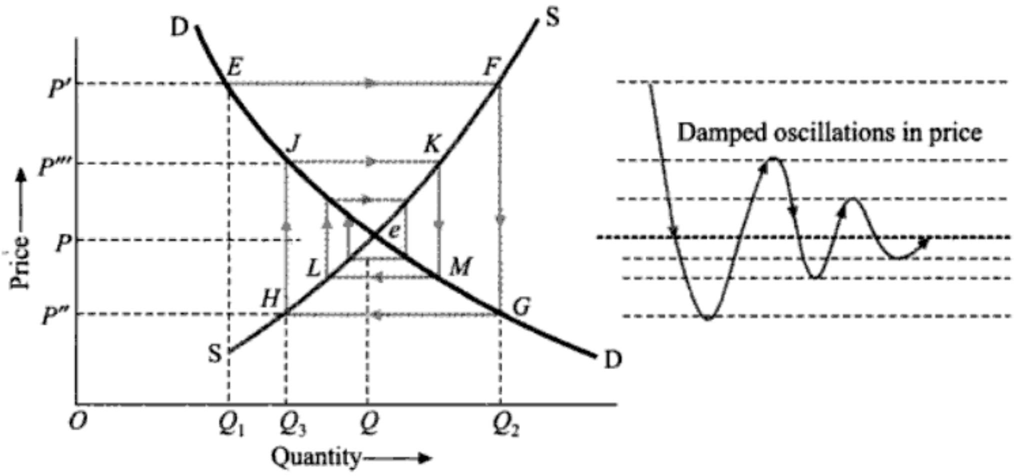


FIGURE 2.29 The figure explains how the previous year's price governs the current year's supply by cultivators. Let the corn price in year 1 be *OP′*. The price is higher than the equilibrium price, *OP*, because supply of corn by the cultivators in year 1 is only OQ_1. Encouraged by the high price of year 1, the cultivators come up with matching supply, OQ_2 in year 2. This being higher than the corresponding level of demand at *OP′*, an excess supply measuring Q_1Q_2 (*EF*) results to depress price down to *OP″* (See the movement along *FG*). Against the expected price *OP′*, the cultivators fetched a much lower price (*OP″*) in year 2. Governed by the price they fetched in year 2, the cultivators produce and supply the matching quantity, OQ_3, creating in the process an excess demand of Q_3Q_2 (HG) in the year 3. This exerts an upward pressure on price, which rises to *OP‴* in year 3. Prompted by this price they would supply more than OQ_3, (OQ_3 + *JK*), creating an excess supply (*JK*) in year 4. The cultivators never learn from their past experiences and continue to be guided by the previous year's price, year after year, until they eventually hit the equilibrium at point *e*. The lagged variations in price–quantity, in this figure, lead to a cobweb that ends up at the equilibrium point *e*. The cobweb is called the **convergent cobweb,** which results from the **damped oscillations** in price (right panel) and leads to **a stable equilibrium** in the long-run.

Appendix 2 GSL Versus OLS

Suppose that the disturbances are autocorrelated, but this fact is ignored and the coefficients of the model are estimated by OLS and the usual t and F tests are carried out. What are the consequences of such a practice?

Our true model is

(1) $Y_t = \beta_0 + \sum_{i=1}^{k} \beta_i X_{ti} + u_t$ t=1…T

where $E(u_t) = 0$ and $Var(u_t) = \sigma^2$

(2) $Cov(u_t,u_s) = E(u_tu_s) = \sigma_{ts} \neq 0$ for $t \neq s$, while we "falsely" estimate (1) under the assumption that $Cov(u_t,u_s) = 0$ for $t \neq s$.

Recall from your ECON301 lectures that the OLS estimator can be written as:

(3) $$\tilde{\beta}_i = \beta_i + \sum_{t=1}^{T} a_{ti}u_t$$

where $a_{ti} = \dfrac{\hat{v}_{ti}}{\sum \hat{v}_{ti}^2}$ where $\hat{v}_{ti}$ being the residuals from the regression of the X_{ti} on the remaining explanatory variables.

The variance of $\tilde{\beta}_i$ is

$$\text{Var}(\tilde{\beta}_i) = \text{E}\left[\tilde{\beta}_i - E(\tilde{\beta}_i)\right]^2$$

Since $E(\tilde{\beta}_i) = \beta_i$, we have:

$$Var(\tilde{\beta}_i) = E\left[\tilde{\beta}_i - \beta_i\right]^2$$

From (3) we can write as usual :

$$Var(\tilde{\beta}_i) = E\left[\sum_{t=1}^{T} a_{ti}u_t\right]^2$$

Expanding yields:

$$Var(\tilde{\beta}_i) = E\Big[a_{1i}^2u_1^2 + a_{2i}^2u_2^2 + a_{3i}^2u_3^2 + \ldots\ldots + a_{Ti}^2u_T^2$$

$$+ 2a_{1i}u_1a_{2i}u_2 + 2a_{2i}u_2a_{3i}u_3 + \ldots.. + 2a_{T-1i}u_{T-1}a_{Ti}u_T + \ldots. + 2a_{Ti}\underset{T\neq S}{u_T}a_{Si}u_S\Big]$$

$$Var(\tilde{\beta}_i) = \sum_{t=1}^{T} a_{ti}^2E(u_t^2) + 2\sum a_{ti}a_{t-1,i}E(u_{t-1}u_t)$$

$$Var(\tilde{\beta}_i) = \sigma^2\sum a_{ti}^2 + 2\sum a_{ti}a_{t-1,i}E(u_{t-1}u_t)$$

$$Var(\tilde{\beta}_i) = \sigma^2 \sum a_{ti}^2 + 2\sum_{\substack{t=1 \\ t \neq s}}^{T} \sum_{s=1}^{T} a_{ti} a_{si} \underbrace{E(u_t u_s)}_{\sigma_{ts}}$$

(4) $$\underbrace{Var(\tilde{\beta}_i)}_{\sigma^2_{\tilde{\beta}_i}} = \sigma^2 \sum a_{ti}^2 + 2 \sum_{\substack{t \\ t \neq s}} \sum_{s} a_{ti} a_{si} \sigma_{ts}$$

If we ignore $E(u_t u_s) = \sigma_{ts}$ we will falsely have:

$$Var(\tilde{\beta}_i)^F = \tilde{\sigma}^2 \sum a_{ti}^2$$

If we use $\tilde{\sigma}^2 = \dfrac{\sum \tilde{u}_t^2}{T-k-1}$, then we will estimate $\tilde{V}ar(\tilde{\beta}_i)^F = \tilde{\sigma}^2 \sum a_{ti}^2$

or

(5) $${}^F\tilde{\sigma}^2_{\tilde{\beta}_i} = \tilde{\sigma}^2 \sum a_{ti}^2$$

If we take expectation of both sides of (5), we can not obtain (4) even if $E(\tilde{\sigma}^2) = \sigma^2$ holds;

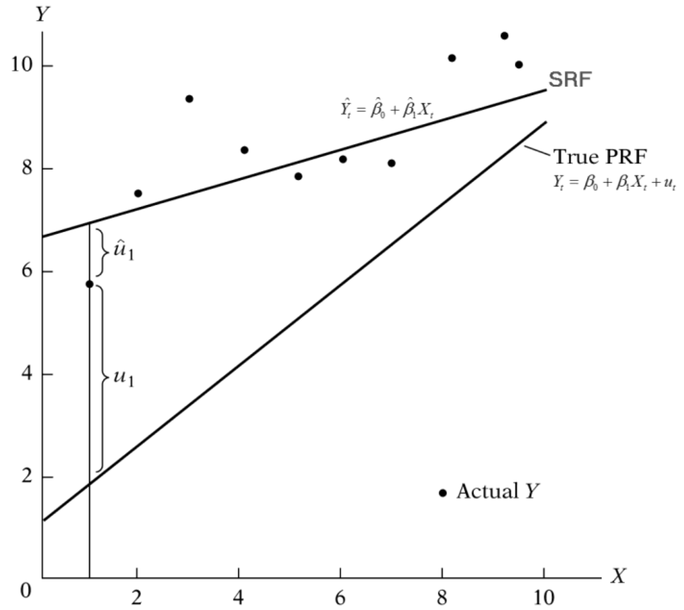
$$E\left[{}^F\tilde{\sigma}^2_{\tilde{\beta}_i}\right] = E(\tilde{\sigma}^2) \sum a_{ti}^2$$

$$E\left[{}^F\tilde{\sigma}^2_{\tilde{\beta}_i}\right] = \sigma^2 \sum a_{ti}^2$$

which is obviously different from $\sigma^2 \sum a_{ti}^2 + 2\sum_{\substack{t \\ t \neq s}} \sum_{s} a_{ti} a_{si} \sigma_{ts}$ by $2\sum_{\substack{t \\ t \neq s}} \sum_{s} a_{ti} a_{si} \sigma_{ts}$. Hence ${}^F\tilde{\sigma}^2_{\tilde{\beta}_i}$ is a biased estimator of $\sigma^2_{\tilde{\beta}_i}$. This bias would further increase when $\tilde{\sigma}^2$ is not an unbiased estimator of σ^2 :

i.e. when $E\left[\tilde{\sigma}^2\right] \neq \sigma^2$. In fact, $\tilde{\sigma}^2$ is indeed not an unbiased estimator of σ^2 under autocorrelation!

Under autocorrelation, σ^2 may be underestimated since $\tilde{\sigma}^2 = \dfrac{\sum \hat{u}_t^2}{T-k-1}$ has $\sum \hat{u}_t^2$ in the numerator. For example, in negative autocorrelation situation shown in the figure below, it can be seen that $\hat{u}_t$'s are not good representatives of u_t's and they are generally smaller.



Hence, we may conclude that there are two sources of underestimation of the variances of the OLS estimators ($\sigma^2_{\tilde{\beta}_i}$)

(1) the term of $2\sum_{t}\sum_{s} a_{ti} a_{si} \sigma_{ts}$ ($t \neq s$) is ignored

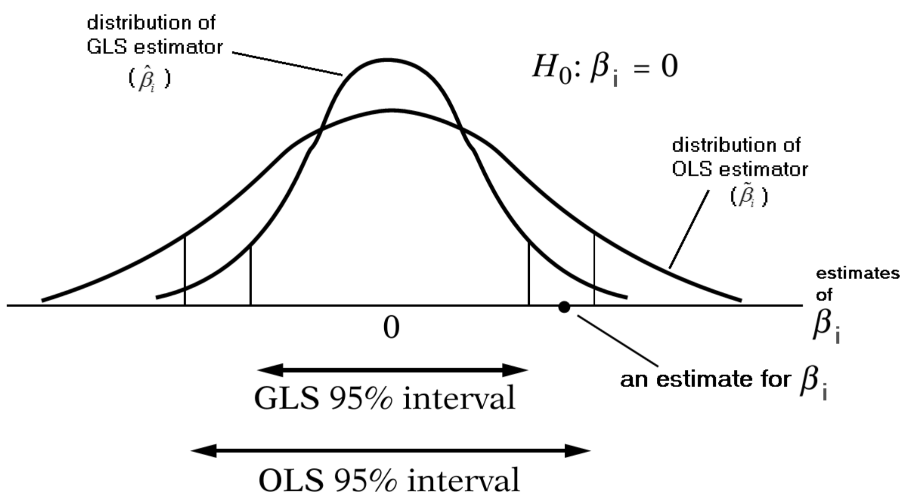
(2) the estimate of σ^2 in most cases has a downward bias due to positive autocorrelation in the u_t's.

Now suppose that we use OLS estimators $\tilde{\beta}_i$ but use the variance formula given by (4)

$$Var(\tilde{\beta}_i) = \sigma^2 \sum a_{ti}^2 + 2 \sum_{\substack{t \\ t \neq s}} \sum_{s} a_{ti} a_{si} \sigma_{ts}$$

Is this the solution of autocorrelation problem? Unfortunately not, even if we use (4) and even if we know σ^2, there is stil a problem of using OLS under autocorelation. It can be shown that OLS estimators are stil unbiased but they have no longer minimum variance.

In other words they are not efficient. It is possible to find another estimator with a smaller variance compared to OLS estimators. This estimator is GLS estimator: $\hat{\beta}_i$



Since OLS does not have minimum varience, the confidence imntervals derived from there are likely to be wider than those based on the GLS estimator. The implication of this finding for hypothesis testing is clear:

> *We are likely to declare a coefficient statistically insignificant even though in fact (based on correct GLS procedure) it may be.* In other words, instead of "rejecting" null hypothesis we can conclude "do not reject" null hypothesis.

This can be seen from the figure. The point given for any estimate of β_i in figure will be declared as insignificant if we calculate OLS confidence intervals using OLS estimates and variences. But if we were to use the (correct) GLS confidence interval, we could reject $H_0 : \beta_i = 0$ since the point (estimate) lies in the region of rejection for GLS estimator distribution.

<u>The message is</u><u>: </u>to establish confidence intervals and to test hypothesis, one should use GLS not OLS even though the estimators derived from OLS are unbiased and consisent.

Summary: Results of OLS Estimation Disregarding Autocorrelation

1) the residual variance $\tilde{\sigma}^2 = \dfrac{\sum \hat{u}_t^2}{T-k-1}$ is likely to underestimate the true σ^2.

2) As a result, we are likely to overestimate R^2 (and we can be very happy, although we must not be!)

3) Even if σ^2 is not underestimated, $Var(\tilde{\beta}_i)^F$ may under estimate $Var(\tilde{\beta}_i)$ even though $Var(\tilde{\beta}_i)$ is also inefficient compared to $Var(\hat{\beta}_i^{GLS})$

where

$$Var(\tilde{\beta}_i)^F = \tilde{\sigma}^2 \sum a_{ti}^2$$
$$Var(\tilde{\beta}_i) = \sigma^2 \sum a_{ti}^2 + 2\sum_{\substack{t \\ t \neq s}} \sum_{s} a_{ti} a_{si} \sigma_{ts}$$

4) Therefore, the usual t and F tests of significance are no longer valid, and if applied, are likely to give seriously misleading conclusions about the statistical significance of the estimated regression coefficients.

References

- Cameron, Samuel (2005) *Econometrics*, McGraw Hill, Berkshire.
- Chauhan, S.P.S., (2009) *Microeconomics: An Advanced Treatise*, Eastern Economy Edition, New Delhi.
- Dougherty, Christopher (2007) *Introduction to Econometrics*, Oxford, New York.
- Erlat, Haluk (1997) *Introduction to Econometrics*, *Chapter 6: Autocorrelation*, Draft (corrected for misprints), Ankara.
- Ezekiel, Mordecai (1938) The Cobweb Theorem, *The Quarterly Journal of Economics*, Vol. 52, No. 2, pp. 255-280.
- Gujarati, D., and Porter (2011) *Basic Econometrics*, McGraw Hill, New York.
- Hill, R. C., Griffiths, W. E., and Judge, G. G., (2001) *Undergraduate Econometrics*, Second Edition, Wiley, New York.
- Kennedy, Peter (1998) *A Guide to Econometrics*, Fourth Edition, Blackwell, New york.
- Stock, J., and Watson, M. M., (2012) *Introduction to Econometrics*, Third Edition, Pearson, Boston.