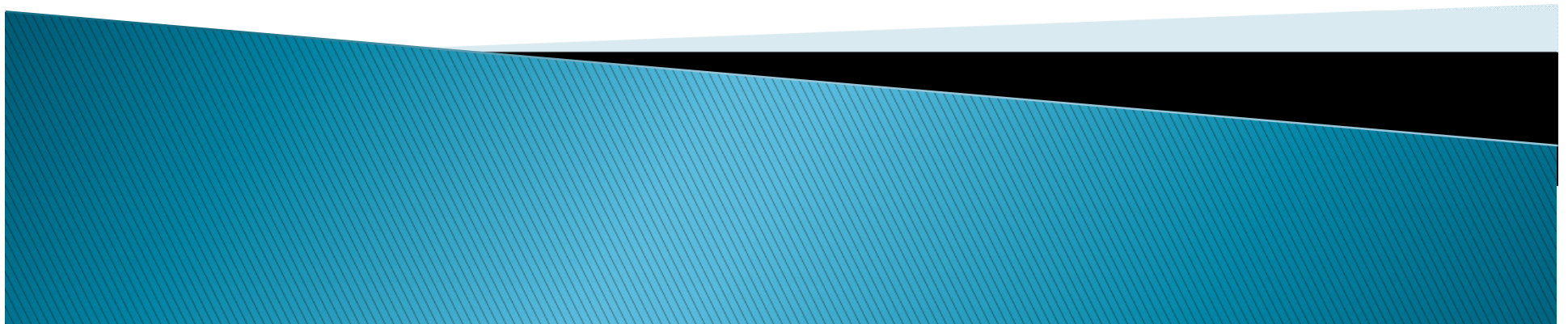


What happens if the error terms are correlated?

Autocorrelation



1. What is the nature of autocorrelation?
2. What are the theoretical and practical consequences of autocorrelation?
3. How does one detect it?
4. How does one remedy the problem of autocorrelation?

In cross sectional studies, data are often collected on the basis of a random sample of cross sectional units, such as households or firms so that there is no prior reason to believe that the error term pertaining to one household or firm is correlated with the error term of another household or firm.

If by chance such a correlation is observed in cross sectional units, it is called **spatial autocorrelation, that is correlation in space rather than time.**

If we are dealing with time series data, for the observations in such data follow a natural ordering over time so that successive observations are likely to exhibit intercorrelations, especially if the time interval between successive observations is short, such as a day, a week, or a month rather than a year.

The Nature of the Problem



The Nature of the Problem

Autocorrelation may be defined as “correlation between members of series of observations ordered in time (as in time series data) or space (as in cross-sectional data)”

CLRM assumes that such correlation does not exist in the disturbance u_i

$$\text{COV}(u_i, u_j | x_i, x_j) = E(u_i u_j) = 0 \quad i \neq j$$

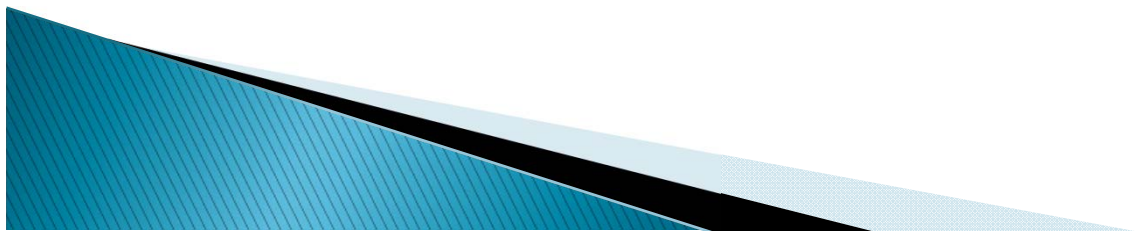
The disturbance term relating to any observation is not influenced by the disturbance term relating to any other observation.

Example

- ▶ If we are dealing with quarterly time series data involving the regression of output on labor and capital inputs and if, say there is a labor strike affecting output in one quarter, there is no reason to believe that this disruption will be carried over to the next quarter
- ▶ If output is lower this quarter, there is no reason to expect it to be lower next quarter



- ▶ If we are dealing with cross-sectional data involving the regression of family consumption expenditure on family income, the effect of an increase of one family's income on its consumption expenditure is not affect the consumption expenditure of another family



However, if there is such a dependence, we have autocorrelation

$$E(u_i u_j) \neq 0 \quad i \neq j$$

E.g. 😊

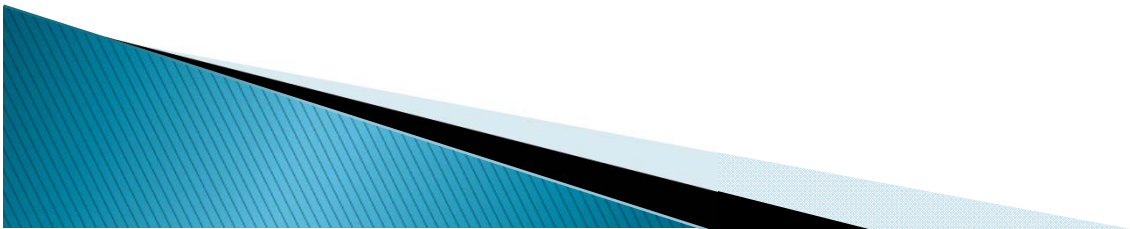
The disruption caused by a strike this quarter may very well affect output next quarter

The increases in the consumption expenditure of one family may very well prompt another family

to increase its consumption
expenditure if wants to keep
up with another family.

Autocorrelation defines as Lag correlation of a given series with itself, lagged by a number of time units

The correlation between two time series such as $u_1, u_2, \dots, u_{10}$ and $u_2, u_3, \dots, u_{11}$, where the former is the latter series lagged by one time period



The correlation between time series such as

$u_1, u_2, \dots, u_{10}$ and $v_2, v_3, \dots, v_{11}$, where u and v are two different time series, is called

Serial correlation

Although the distinction between the two terms may be useful, in this class we shall treat them synonymously.



No Autocorrelation

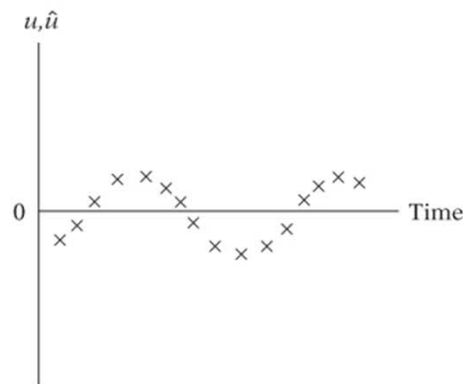
$$E(u_i u_j) = 0, \quad i \neq j$$

$$E(uu') = \begin{bmatrix} E(u_1^2) & E(u_1 u_2) & \dots & E(u_1 u_n) \\ E(u_2 u_1) & E(u_2^2) & \dots & E(u_2 u_n) \\ \dots & \dots & \dots & \dots \\ E(u_n u_1) & E(u_n u_2) & \dots & E(u_n^2) \end{bmatrix}$$

$$E(uu') = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 1 \end{bmatrix} = \sigma^2 I$$

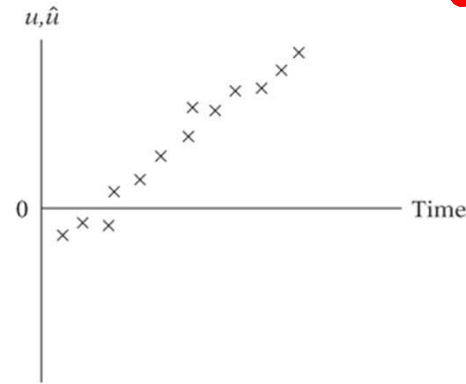


cyclical pattern →



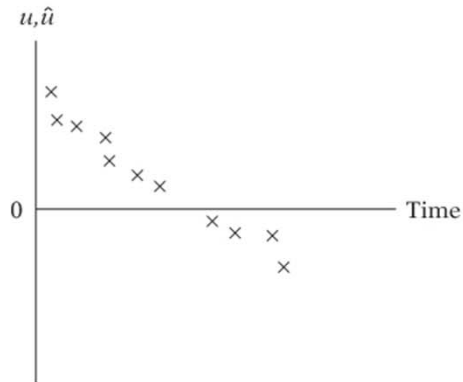
(a)

upward linear trend in the disturbances



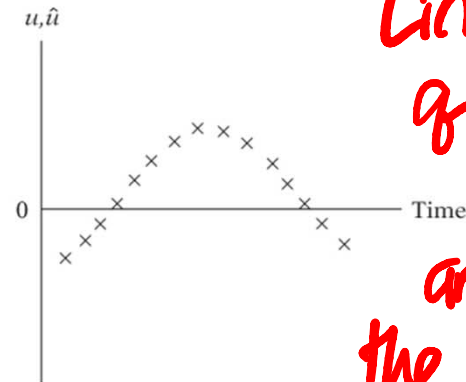
(b)

Downward linear trend in the disturbances ↘



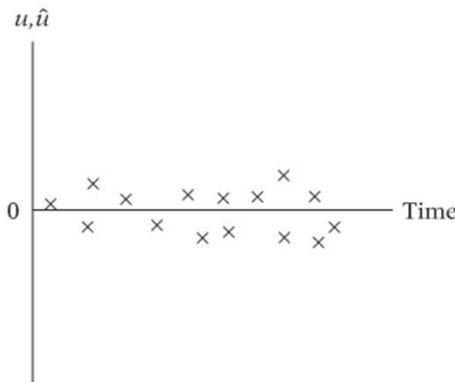
(c)

Linear and quadratic trend terms are present in the disturbances



(d)

Support the non autocorrelation assumption of CLRM →



(e)

No systematic pattern



Why does serial correlation occur?

Cobweb phenomenon

Lags

Data transformation

Inertia

Specification Bias

Excluded variables

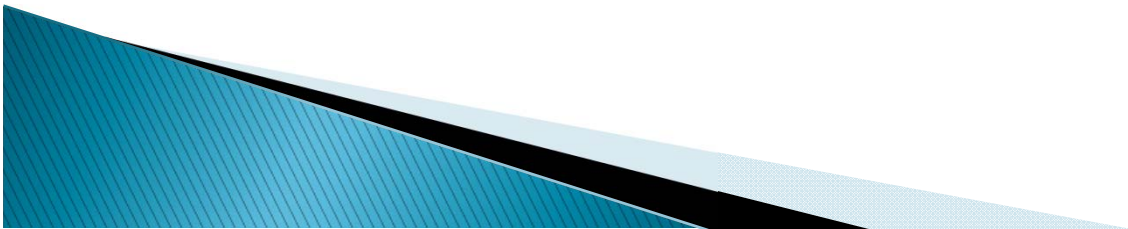
Incorrect functional form

Inertia

A salient feature of most economic time series is sluggishness

Example

GNP, price indexes, production, employment, and unemployment exhibit business cycles



Specification Bias: **excluded variables case**

correct model → $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t$

①

Y_t = quantity of beef demand

X_2 = price of beef

X_3 = consumer income

X_4 = price of pork

t = time

②

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + v_t$$

$$v_t = \beta_4 X_{4t} + u_t$$

reflect a systematic pattern, thus creating (false) autocorrelation.

The price of pork affects the consumption of beef, the error or disturbance term V will reflect a systematic pattern, thus creating (false) autocorrelation



Specification Bias: Incorrect Functional Form

Suppose the true or correct model in a cost-output study is as follows:

$$MARGINAL \text{ } COST_i = \beta_1 + \beta_2 output_i + \beta_3 output_i^2 + u_i \quad (1)$$

But we fit the following model:

$$MARGINAL \text{ } COST_i = \alpha_1 + \alpha_2 output_i + v_i \quad (2)$$

$$v_i = \beta_3 output_i^2 + u_i$$

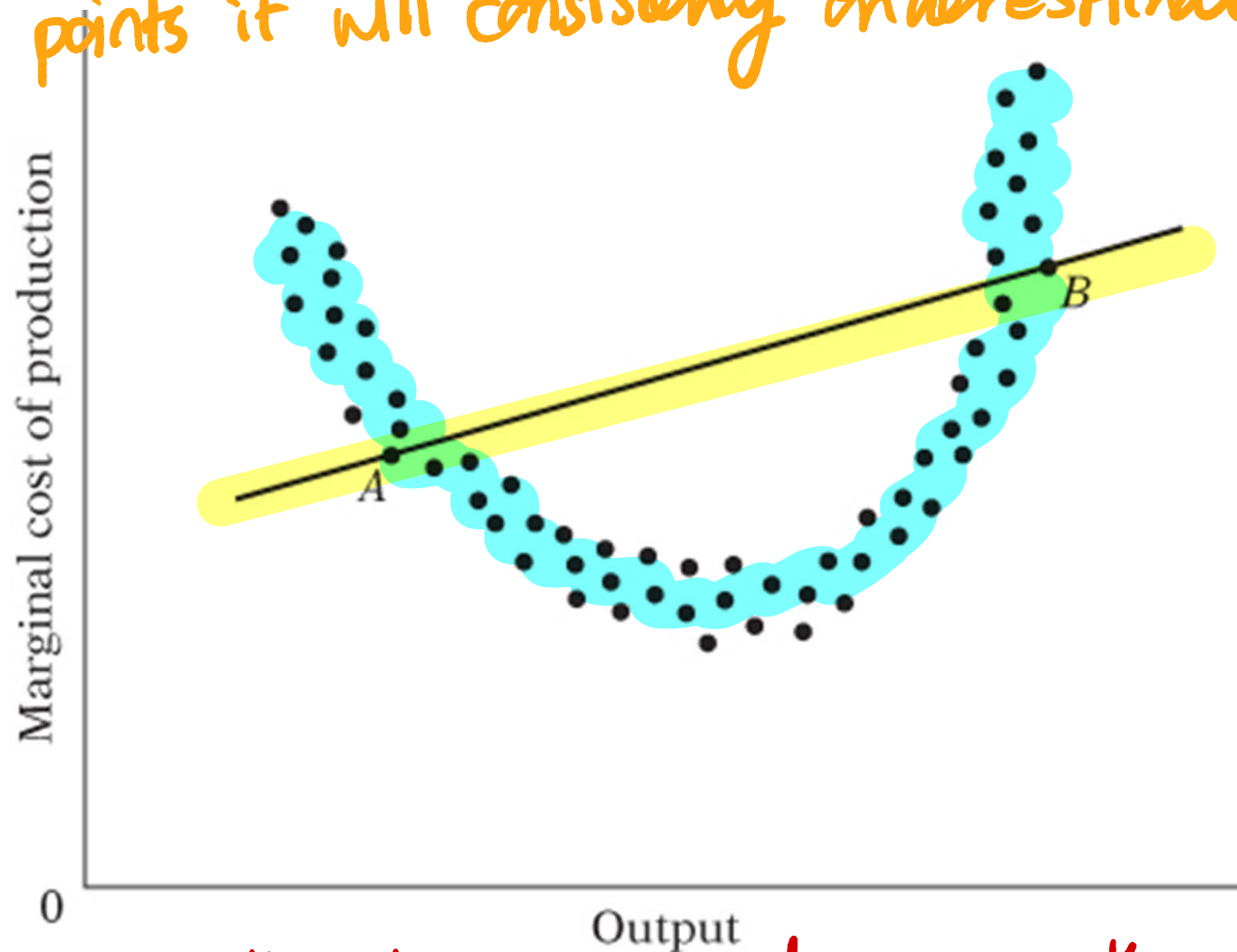
will catch the systematic effect of the output² term

The marginal cost curve corresponding to the true model is shown in the next figure along with the incorrect linear cost curve

v_i will reflect autocorrelation because the use of an incorrect functional form

marginal cost.

Between point A and B the linear marginal cost curve will consistently overestimate the true marginal cost, whereas beyond these points it will consistently underestimate the true MC.



This result is to be expected, because the disturbance term v_i is, in fact, equal to $\text{output}^2 + u_i$, and hence will catch the systematic effect of the output^2 term on marginal cost.

Cobweb Phenomenon

The supply of many agricultural commodities reflects the so-called cobweb phenomenon, where supply reacts to price with a lag of one time period because supply decisions take time to implement.

$$Supply_t = \beta_1 + \beta_2 P_{t-1} + u_t$$

At the beginning of this year's planting of crops, farmers are influenced by the price prevailing last year

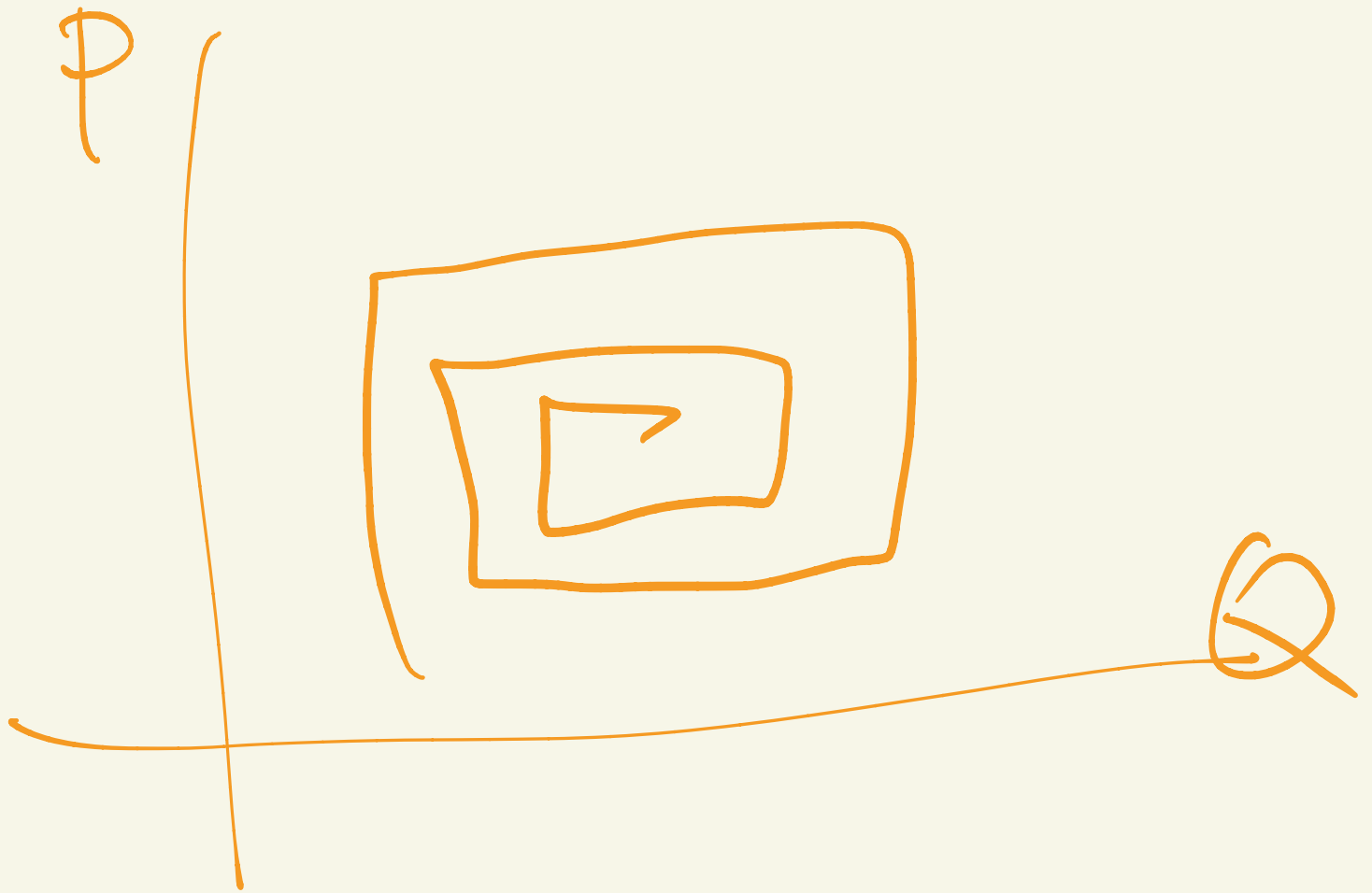


Suppose at the end of period t , price P_t turns out to be lower than P_{t-1} .

Therefore, in period $t+1$ farmers may very well decide to produce less than they did in period t .

The disturbances u_i are not expected to be random because if the farmers over-produce in year t , they are likely to

reduce their production in $t+1$,
and so on, leading to a
cobweb pattern.



Lags

$$\text{Consumption}_t = \beta_1 + \beta_2 \text{income} + \beta_3 \text{consumption}_{t-1} + u_t$$

Consumers do not change their habits readily for psychological, technological, or institutional reasons. Now if we neglect the lagged term, the resulting error term will reflect a systematic pattern due to the influence of lagged consumption on current consumption.

Data Transformation

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

①

$Y = \text{Consumption Expenditure}$

$X = \text{income}$

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + u_{t-1} \leftarrow \text{Level Form}$$

②

$$\Delta Y_t = \beta_2 \Delta X_t + \Delta u_t$$

③

$$\Delta Y_t = \beta_2 \Delta X_t + v_t$$

$$(X_t - X_{t-1})$$

$$v_t = \Delta u_t = (u_t - u_{t-1}) \leftarrow \text{Difference Form}$$

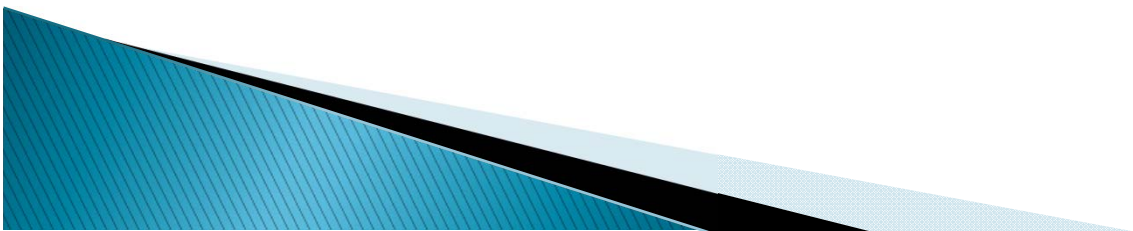
$$v_t = u_t - u_{t-1}$$

$$E(u) = 0$$

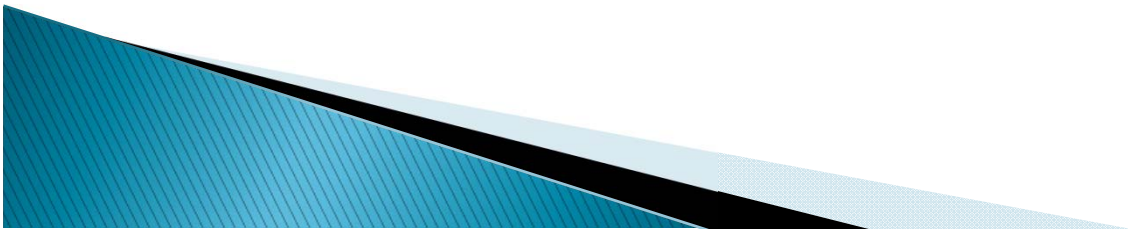
$$E(v_t) = E(u_t - u_{t-1}) = E(u_t) - E(u_{t-1}) = 0$$

$$\text{var}(v_t) = \text{var}(u_t - u_{t-1}) = \text{var}(u_t) - \text{var}(u_{t-1}) = 2\sigma^2$$

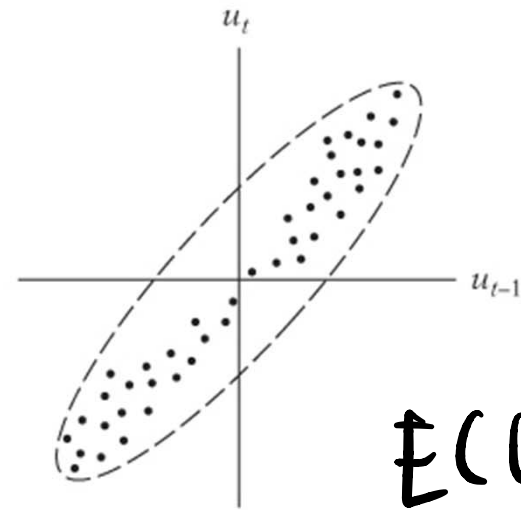
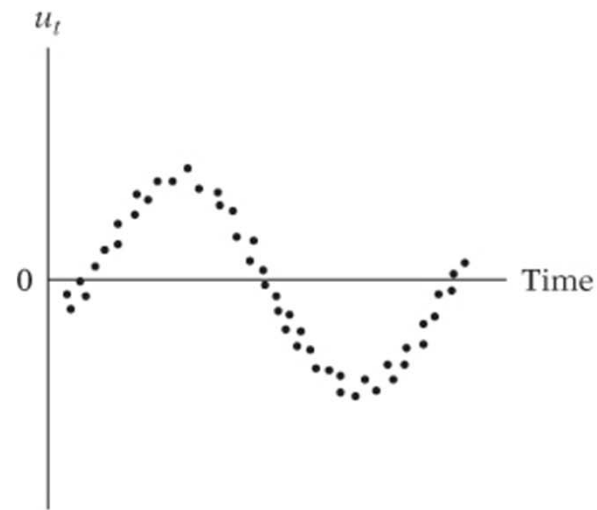
$$\text{cov}(v_t, v_{t-1}) = E(v_t v_{t-1}) = E[(u_t - u_{t-1})(u_{t-1} - u_{t-2})] = -\sigma^2$$



OLS Estimation in the Presence of Autocorrelation



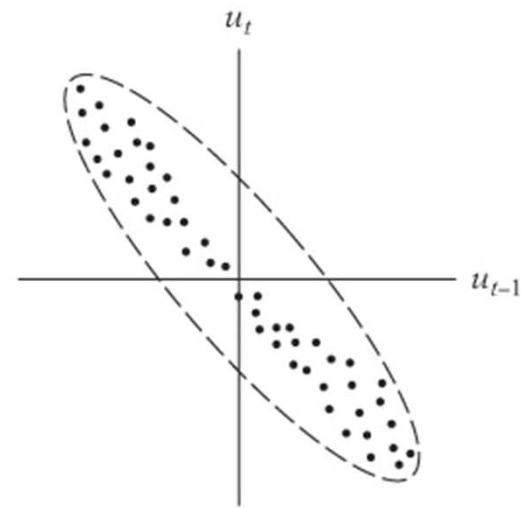
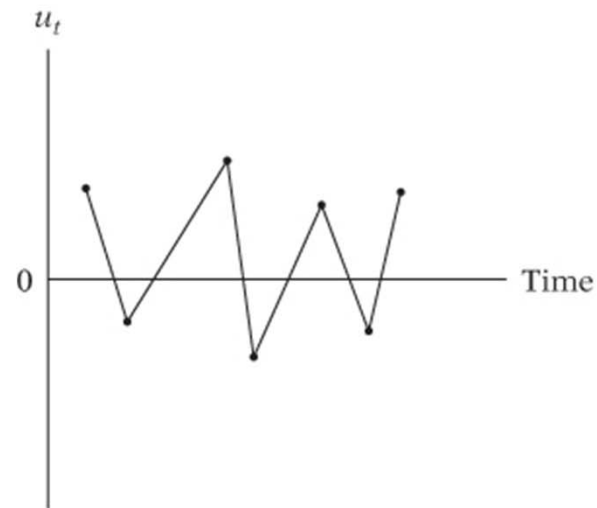
positive



$$E(u_i u_j) \neq 0$$

(a)

Negative

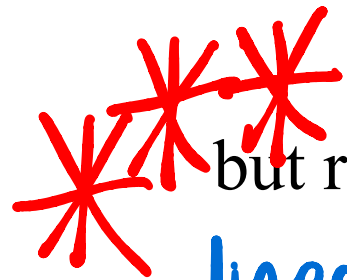
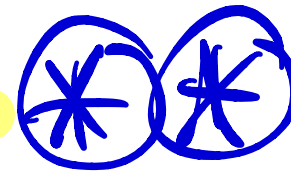


(b)



- What happens to the OLS estimators and their variances if we introduce autocorrelation in the disturbances by assuming that

$$E(u_t u_{t+s}) \neq 0 \quad (s \neq 0)$$



but retain all the other assumptions of classical model?

- linear

- unbiased

- not efficient

- no longer minimum variance



$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

①

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

②

ρ (coefficient of autocovariance)

$$E(\varepsilon_t) = 0$$

$$\text{var}(\varepsilon_t) = \sigma_\varepsilon^2$$

$$\text{cov}(\varepsilon_t, \varepsilon_{t+s}) = 0 \quad s \neq 0$$

ε (white noise error term)

where ε_t is the stochastic disturbance term such that it satisfies the standard OLS assumptions.



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

This equation is known as a Markov first order autoregressive scheme or first-order autoregressive scheme AR(1)

The regression of u_t on itself lagged one period.

E.g. AR(2) $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$



Under AR(1)

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

Therefore,

$$E(u_t) = \rho E(u_{t-1}) + E(\varepsilon_t) = 0 \quad (2)$$

$$\text{So, } \text{var}(u_t) = \rho^2 \text{var}(u_{t-1}) + \text{var}(\varepsilon_t) \quad (3)$$

because the u 's and ε 's are uncorrelated.

Since $\text{var}(u_t) = \text{var}(u_{t-1}) = \sigma^2$ and $\text{var}(\varepsilon_t) = \sigma_\varepsilon^2$, we get

$$\text{var}(u_t) = \frac{\sigma_\varepsilon^2}{1-\rho^2} \quad (4)$$

Now multiply eq(1) by u_{t-1} and take expectations on both sides to obtain

$$\begin{aligned}
 \text{Cov}(u_t, u_{t-1}) &= E(u_t u_{t-1}) \\
 &= E[\rho u_{t-1}^2 + u_{t-1} \varepsilon_t] \\
 &= \rho E(u_{t-1}^2)
 \end{aligned}$$

Noting that the covariance between u_{t-1} and ε_t is zero and that

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

$$\text{Cov}(u_t, u_{t-1}) = \rho \frac{\sigma_\varepsilon^2}{(1-\rho^2)} \quad (5)$$

The correlation coefficient is the ratio of covariance to variance.

$$\text{Hence, } \text{Corr}(u_t, u_{t-1}) = \rho$$

First-order coefficient of autocorrelation

AR (1)

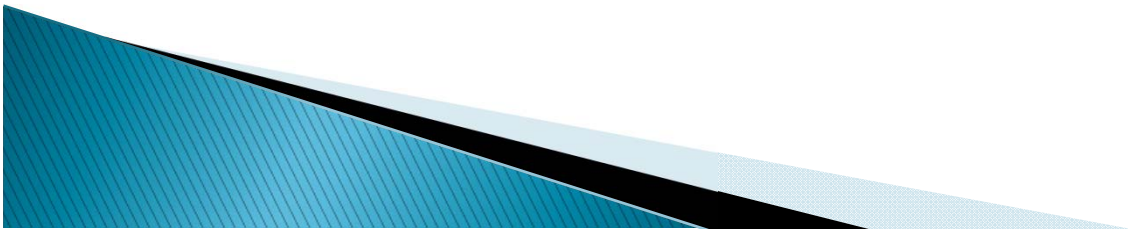
$$E(u_t) = \rho E(u_{t-1}) + E(\varepsilon) = 0$$

$$\text{var}(u_t) = E(u_t^2)$$

$$= \rho^2 \text{var}(u_{t-1}) + \text{var}(\varepsilon) = \frac{\sigma_\varepsilon^2}{1 - \rho^2}$$

$$\text{var}(u_t) = \text{var}(u_{t-1}) = \sigma^2 \quad \text{"still homoscedastic"}$$

$$\text{var}(\varepsilon_t) = \sigma_\varepsilon^2$$

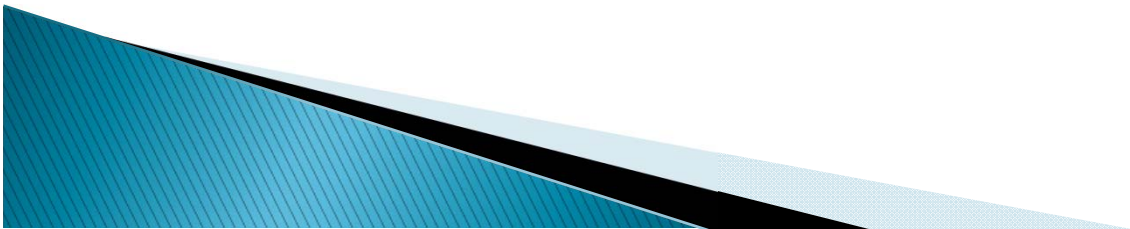


$$\text{cov}(u_t, u_{t+s}) = E(u_t, u_{t+s}) = \rho^s \frac{\sigma_\varepsilon^2}{1 - \rho^2}$$

Covariance between error terms s periods apart

The symmetry property of covariances

$$\text{cov}(u_t, u_{t+s}) = \text{cov}(u_t, u_{t-s})$$

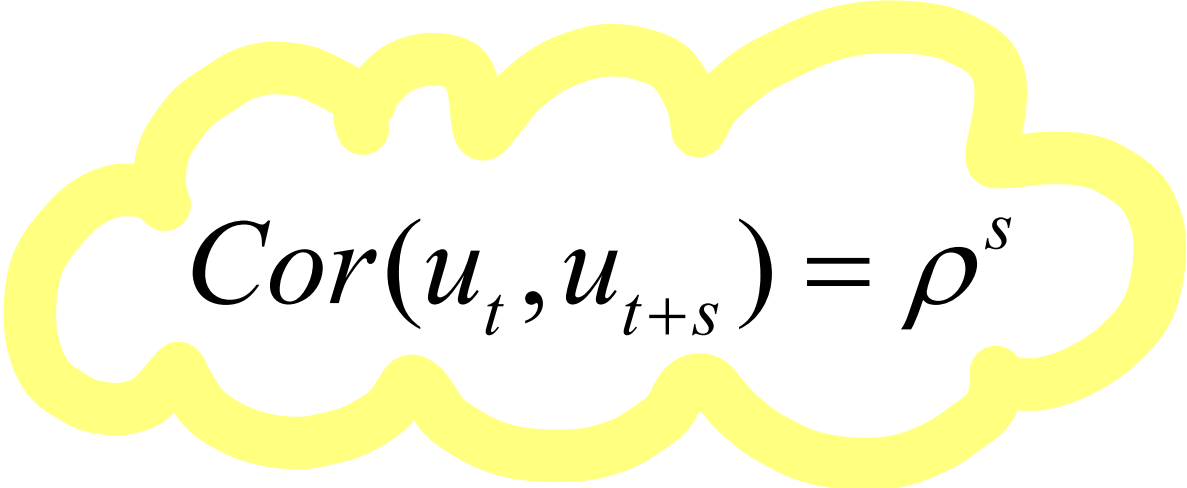


- ▶ By definition, the (population) coefficient of correlation between u_t and u_{t-1} is

$$\rho = \frac{E\{[u_t - E(u_t)][u_{t-1} - E(u_{t-1})]\}}{\sqrt{\text{var}(u_t)}\sqrt{\text{var}(u_{t-1})}} = \frac{E(u_t u_{t-1})}{\text{var}(u_{t-1})}$$

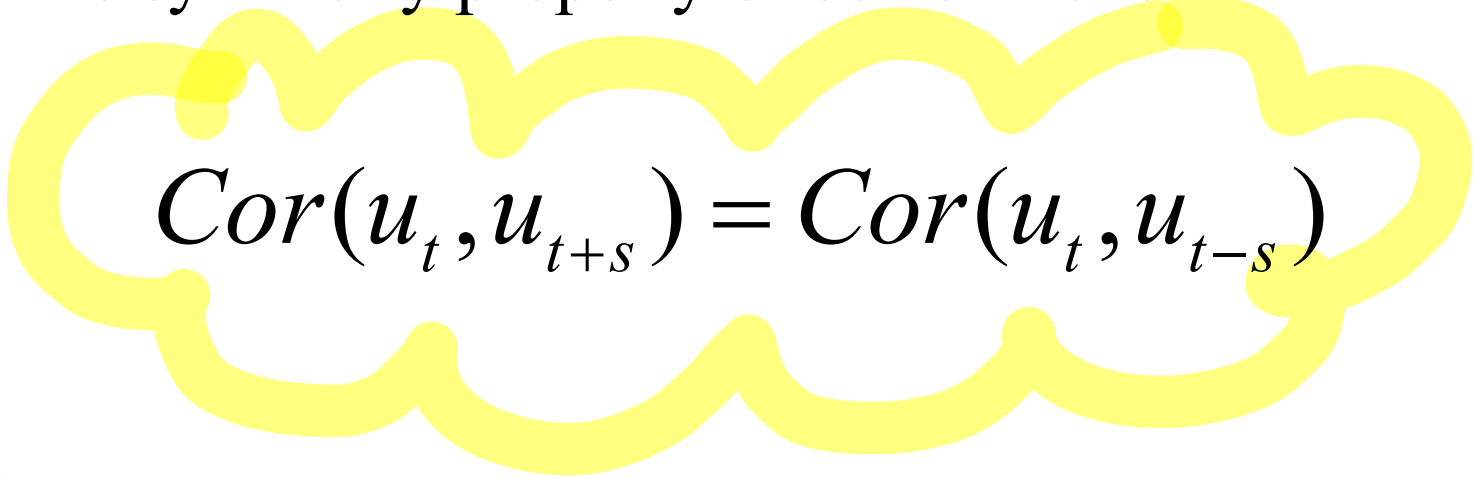
Since $E(u_t) = 0$ for each t and $\text{var}(u_t) = \text{var}(u_{t-1})$ because we are retaining the assumption of homoscedasticity.

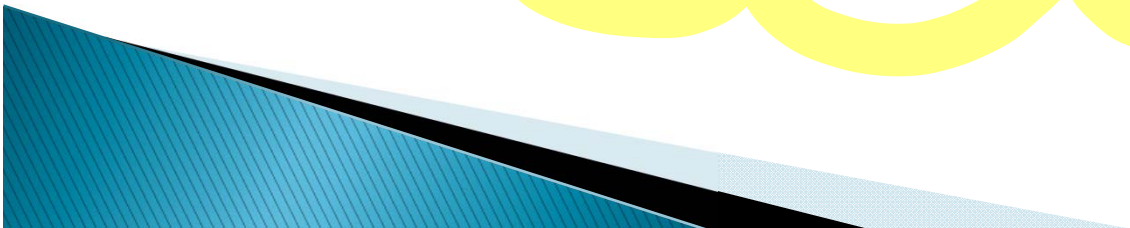



$$Cor(u_t, u_{t+s}) = \rho^s$$

Correlation between error terms s period apart

The symmetry property of correlations


$$Cor(u_t, u_{t+s}) = Cor(u_t, u_{t-s})$$



Since ρ is a constant between -1 and +1, , variance equation shows that under the AR(1) scheme, the variance of u_t is still homoscedastic, but u_t is correlated not only with its immediate past value but its values several periods in the past.



Two variable regression model

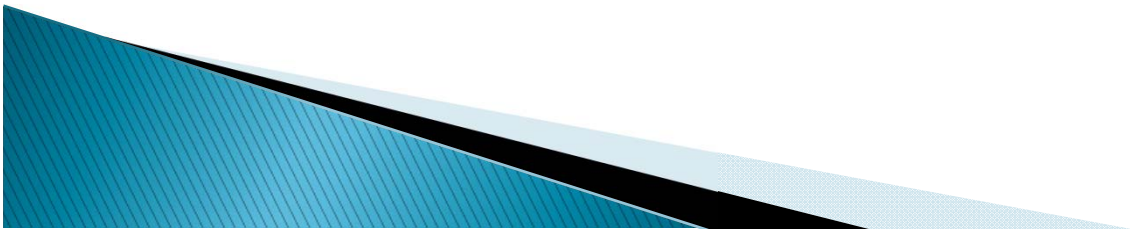
$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

OLS estimator of the slope coefficient is

$$\hat{\beta}_2 = \frac{\sum (X_t - \bar{X})(Y_t - \bar{Y})}{\sum (X_t - \bar{X})^2} = \frac{\sum x_t y_t}{\sum x_t^2}$$

and its variance is given by

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum (X_t - \bar{X})^2} = \frac{\sigma^2}{\sum x_t^2}$$



Now under the AR(1) scheme, it can be shown that the variance of this estimator is

$$\text{var}(\hat{\beta}_2)_{AR1} = \frac{\sigma^2}{\sum x_t^2} \left[1 + 2\rho \frac{\sum x_t x_{t-1}}{\sum x_t^2} + 2\rho^2 \frac{\sum x_t x_{t-2}}{\sum x_t^2} + \dots + 2\rho^{n-1} \frac{\sum x_1 x_n}{\sum x_t^2} \right]$$

where

$$x_t = (X_t - \bar{X})$$

$$x_{t-1} = (X_{t-1} - \bar{X})$$

$$\vdots$$

$$x_n = (X_n - \bar{X})$$



The difference between the $\text{var}(\hat{\beta}_2)$ and $\text{var}(\hat{\beta}_2)_{AR(1)}$, assume that the regressor X also follows the first order autoregressive scheme with a coefficient of autocorrelation of r

$$\text{var}(\hat{\beta}_2)_{AR(1)} = \frac{\sigma^2}{\sum x_t^2} \left(\frac{1 + r\rho}{1 - r\rho} \right) = \text{var}(\hat{\beta}_2)_{OLS} \left(\frac{1 + r\rho}{1 - r\rho} \right)$$



Example

$r = 0.6$ and $\rho = 0.8$

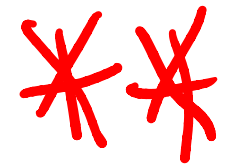
$$\text{Var}(\hat{\beta}_2)_{AR1} = 2.8461 \text{ Var}(\hat{\beta}_2)_{OLS}$$

$$\text{Var}(\hat{\beta}_2)_{OLS} = \frac{1}{2.8461} \text{Var}(\hat{\beta}_2)_{AR1}$$

$$= 0.3513 \text{ Var}(\hat{\beta}_2)_{AR1}$$

The usual OLS formula will underestimate the variance of $(\hat{\beta}_2)_{AR1}$ by about 65 percent.

The usual OLS formulas to compute the variances and standard errors of the OLS estimators could give seriously misleading results.



As in the case of **heteroscedasticity** in the presence of **autocorrelation** the OLS estimators are still linear unbiased as well as consistent and asymptotically normally distributed, but they are **no longer efficient (i.e., minimum variance)**

They may not be BLUE (Best Linear Unbiased Estimators)
As a result, the usual t , F and χ^2 may not be valid.



The blue estimator in the presence of autocorrelation



9

Continuing with the two-variable model and assuming the AR(1) process, we can show that the BLUE estimator of β_2 is given by the following expression

$$\widehat{\beta_2}_{GLS} = \frac{\sum_{t=2}^n (x_t - \rho x_{t-1})(y_t - \rho y_{t-1})}{\sum_{t=2}^n (x_t - \rho x_{t-1})^2} + C$$

Where C is a correction factor that may be disregarded in practice. Note that the subscript t now runs from t=2 to t=n.



$$\text{Var} \hat{\beta}_{2 \text{ OLS}} = \frac{\hat{\sigma}^2}{\sum (x_t - \bar{x})^2}$$

And its variance is given by

$$\text{Var} \hat{\beta}_2^{\text{GLS}} = \frac{\sigma^2}{\sum_{t=2}^n (x_t - \rho x_{t-1})^2} + D$$

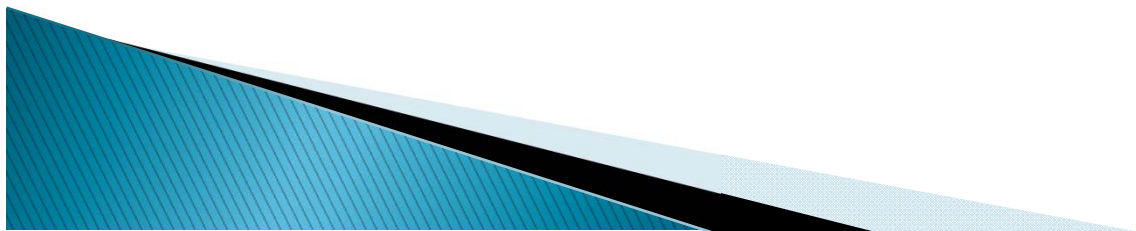
Where D too is a correction factor that may also be disregarded in practice



- ▶ In GLS we incorporate any additional information we have directly into the estimating procedure by transforming the variables, whereas in OLS such side information is not directly taken into consideration
- ▶ Under autocorrelation, it is the GLS estimator that is BLUE



Consequences of Using OLS in the Presence of Autocorrelation

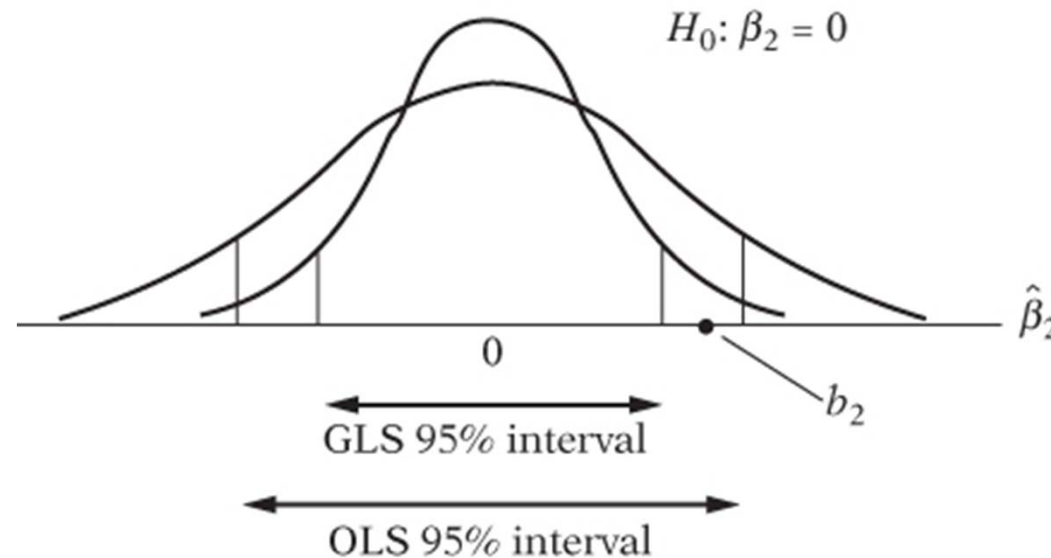


OLS estimation allowing for autocorrelation

$$\hat{Y}_t = \hat{\beta}_1 + \hat{\beta}_2 X_t + \hat{u}_t$$
$$Y_t = \hat{Y}_t + \bar{u}_t$$

$\hat{\beta}_2$ is not BLUE and even if we use $\text{var}(\hat{\beta}_2)_{AR1}$, the confidence intervals derived from there are likely to be **wider** than those based on the GLS procedure

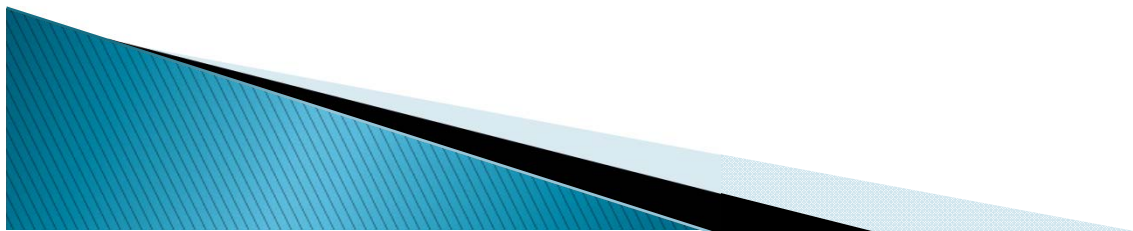
$$\hat{\beta}_i \pm t_{\alpha/2} \text{se}(\hat{\beta}_i)$$



OLS estimation disregarding autocorrelation

The situation is potentially very serious if we not only use $\hat{\beta}_2$ but also continue to use

$\text{var}(\hat{\beta}_2) = \sigma^2 / \sum (X_t - \bar{X})^2$, which completely disregards the problem of autocorrelation, that is, we mistakenly believe that the usual assumptions of the classical model hold true. Errors will arise for the following reasons



$$t = \frac{\hat{\beta}_i - \beta_i}{\text{se}(\hat{\beta}_i)}$$

$$\text{se}(\hat{\beta}_i) = \sqrt{\text{var}(\hat{\beta}_i)}$$

$$R^2 = \frac{\text{ESS}}{\text{TSS}}$$

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}}$$

1. The residual variance $\hat{\sigma}^2 = \frac{\sum \hat{u}_t^2}{(n-2)}$ is likely to **underestimate** the true σ^2
2. As a result, we are likely to **overestimate** R^2
3. Even if σ^2 is not underestimated, $\text{var}(\hat{\beta}_2)$ may be underestimated $\text{var}(\hat{\beta}_2)_{AR1}$ its variance under AR(1), even though the latter is **inefficient** compared to $\text{var}(\hat{\beta}_2)^{GLS}$
4. Therefore, the **usual t and F tests of significance are no longer valid**

$\text{var}(\hat{\beta}_2)_{AR1}$



$$TSS = ESS + RSS$$

$$R^2 = \frac{ESS}{TSS}$$

or

$$R^2 = 1 - \frac{RSS}{TSS}$$



Diagram illustrating the components of the Residual Sum of Squares (RSS):

The RSS is represented by the sum of squared residuals, $\sum \hat{u}_i^2$, divided by the degrees of freedom, $n-2$.

$\frac{\sum \hat{u}_i^2}{n-2}$ (circled) is connected to the circled label **RSS**.

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{(n-2)}$$

$$E(\hat{\sigma}^2) = \sigma^2$$

$$R^2 = 1 - \frac{RSS}{TSS}$$

$$= \frac{ESS}{TSS}$$

But if there is autocorrelation, given by AR(1), it can be shown that

$$E(\hat{\sigma}^2) = \frac{\sigma^2 \{ n - [2/(1-r)] - 2pr \}}{n-2}$$

where $r = \frac{\sum_{t=1}^{n-1} x_t x_{t-1}}{\sum_{t=1}^n x_t^2}$, which can be interpreted as the (sample) correlation coefficient between successive values of the x 's

If q and r are both positive,
 $E(\hat{\sigma}^2) < \sigma^2$; that is the usual
residual variance formula, on average,
will underestimate the true σ^2 .

In other words, $\hat{\sigma}^2$ will be biased
downward.

But even if σ^2 is not underestimated,
 $\text{var}(\hat{\beta}_2)$ is a biased estimator of
 $\text{var}(\hat{\beta}_2)_{\text{ARI}}$

If ρ is positive and the X 's are positively correlated, then

$$\text{var}(\hat{\beta}_2) < \text{var}(\hat{\beta}_2)_{\text{AR}_1}$$

that is, the usual OLS variance of $\hat{\beta}_2$ underestimates its variance under $\text{AR}(1)$. Therefore, if we use $\text{var}(\hat{\beta}_2)$, we shall inflate the precision or accuracy (i.e., underestimate the standard error) of the estimator $\hat{\beta}_2$.

As a result, in computing the t ratio

$$t = \frac{\hat{\beta}_2 - \beta_2}{\text{se}(\hat{\beta}_2)} \quad (\text{under } H_0: \beta_2 = 0),$$

we shall be overestimating the t-value and hence the statistical significance of the estimated β_2 .

The situation is likely to get worse if additionally σ^2 is underestimated.

PRF - Population Regression Function
SRF - Sample Regression Function

Example

To see how OLS_h is likely to underestimate σ^2 and the variance of $\hat{\beta}_2$, let us conduct the following **Monte Carlo experiment**

Suppose in the two-variable model

PRF $\Rightarrow$

$$Y_t = 1.0 + 0.8X_t + u_t$$

$$E(Y_t | X_t) = 1.0 + 0.8X_t$$

AR(1)

$$u_t = 0.7u_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0,1)$$



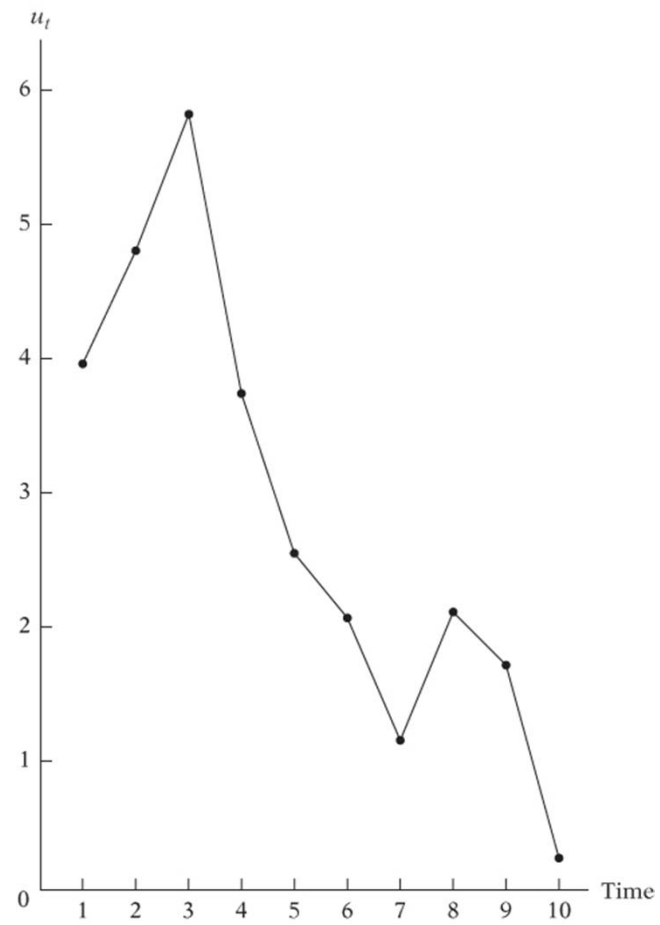
TABLE 12.1

**A Hypothetical
Example of Positively
Autocorrelated Error
Terms**

	ε_t	$u_t = 0.7u_{t-1} + \varepsilon_t$
0	0	$u_0 = 5$ (assumed)
1	0.464	$u_1 = 0.7(5) + 0.464 = 3.964$
2	2.026	$u_2 = 0.7(3.964) + 2.0262 = 4.8008$
3	2.455	$u_3 = 0.7(4.8010) + 2.455 = 5.8157$
4	-0.323	$u_4 = 0.7(5.8157) - 0.323 = 3.7480$
5	-0.068	$u_5 = 0.7(3.7480) - 0.068 = 2.5556$
6	0.296	$u_6 = 0.7(2.5556) + 0.296 = 2.0849$
7	-0.288	$u_7 = 0.7(2.0849) - 0.288 = 1.1714$
8	1.298	$u_8 = 0.7(1.1714) + 1.298 = 2.1180$
9	0.241	$u_9 = 0.7(2.1180) + 0.241 = 1.7236$
10	-0.957	$u_{10} = 0.7(1.7236) - 0.957 = 0.2495$

Note: ε_t data obtained from *A Million Random Digits and One Hundred Thousand Deviates*, Rand Corporation, Santa Monica, Calif., 1950.

$$u_{t-1} = 0$$
$$u_0 = 5$$



Now suppose the values of X are fixed at 1,2,3,...,10.

SRF

$$\hat{Y}_t = 6.5452 + 0.3051X_t$$

(0.6153) (0.0992)

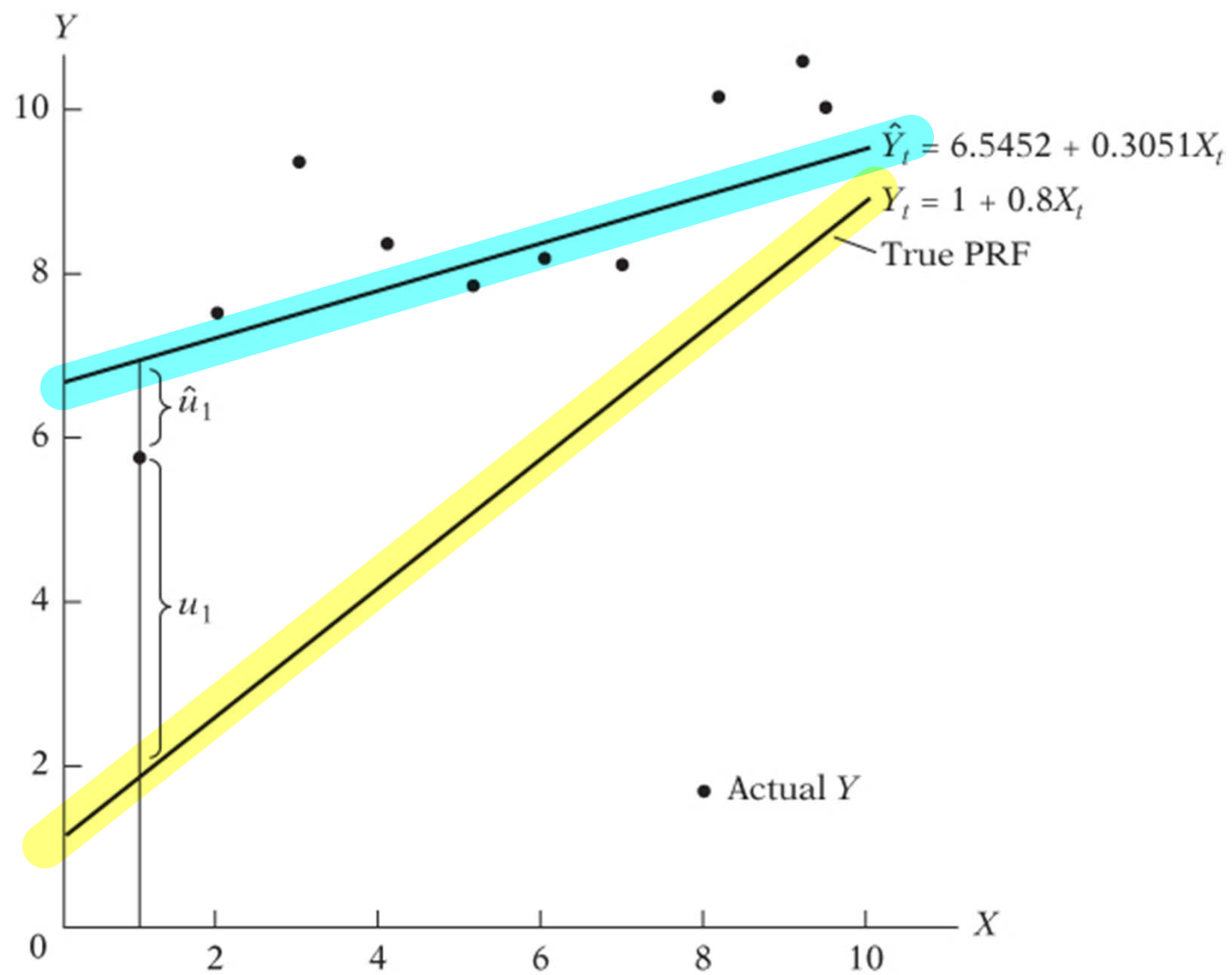
$$t = (10.6366) (3.0763)$$

$$r^2 = 0.5419$$

$$\hat{\sigma}^2 = 0.8114$$

Underestimates the true slope coefficient but overestimates the true intercept





No Autocorrelation

$$\rho = 0$$

TABLE 12.2

Generation of Y
Sample Values

X_t	u_t	$Y_t = 1.0 + 0.8X_t + u_t$
1	3.9640	$Y_1 = 1.0 + 0.8(1) + 3.9640 = 5.7640$
2	4.8010	$Y_2 = 1.0 + 0.8(2) + 4.8008 = 7.4008$
3	5.8157	$Y_3 = 1.0 + 0.8(3) + 5.8157 = 9.2157$
4	3.7480	$Y_4 = 1.0 + 0.8(4) + 3.7480 = 7.9480$
5	2.5556	$Y_5 = 1.0 + 0.8(5) + 2.5556 = 7.5556$
6	2.0849	$Y_6 = 1.0 + 0.8(6) + 2.0849 = 7.8849$
7	1.1714	$Y_7 = 1.0 + 0.8(7) + 1.1714 = 7.7714$
8	2.1180	$Y_8 = 1.0 + 0.8(8) + 2.1180 = 9.5180$
9	1.7236	$Y_9 = 1.0 + 0.8(9) + 1.7236 = 9.9236$
10	0.2495	$Y_{10} = 1.0 + 0.8(10) + 0.2495 = 9.2495$

Note: u_t data obtained from Table 12.1.



Keeping the X_t and ε_t given in Table 2.1 and Table 2.2, let us assume $\rho = 0$, that is, **no autocorrelation**

TABLE 12.3
Sample of Y Values
with Zero Serial
Correlation

X_t	$\varepsilon_t = u_t$	$Y_t = 1.0 + 0.8X_t + \varepsilon_t$
1	0.464	2.264
2	2.026	4.626
3	2.455	5.855
4	-0.323	3.877
5	-0.068	4.932
6	0.296	6.096
7	-0.288	6.312
8	1.298	8.698
9	0.241	8.441
10	-0.957	8.043

Note: Since there is no autocorrelation, the u_t and ε_t are identical. The ε_t are from Table 12.1.



The regression based on Table 12.3 is as follows:

SRF

$$\hat{Y}_t = 2.5345 + 0.6145X_t$$

$$(0.6796) \quad (0.1087)$$

$$t = (3.7910) \quad (5.6541)$$

$$r^2 = 0.7997$$

$$\hat{\sigma}^2 = 0.9752$$





Detecting Autocorrelation

- ▶ Durbin-Watson d Test



Durbin-Watson d Test

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

The ratio of the sum of squared differences in successive residuals to the RSS. Note that in the numerator of the d statistic the number of observations is n-1 because one observation is lost in taking successive differences



Assumptions underlying the d statistic

- ▶ The regression model includes the intercept term
- ▶ The explanatory variables, the X's are nonstochastic, or fixed in repeated sampling
- ▶ The disturbances are generated by the AR(1)

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

Therefore, it cannot be used to detect higher-order AR schemes

- ▶ The error term u_t is assumed to be normally distributed

$$\varepsilon_t \sim N(0, 1)$$


$$Y_t = \beta_1 + \beta_2 X_t + \cancel{\beta_t Y_{t-1}}$$

- ▶ The regression model does not include the lagged value(s) of the dependent variable as one of the explanatory variables
- ▶ There are no missing observations in the data

Obs	X_1	X_2	X_3	X_4
1	10	8		
2	5	10		
3	•	5		
4	7	•		

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

$$d = \frac{\sum \hat{u}_t^2 + \sum \hat{u}_{t-1}^2 - 2 \sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2}$$

$$d \approx 2 \left(1 - \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2} \right)$$

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

Since $\sum \hat{u}_t^2$ and $\sum \hat{u}_{t-1}^2$ differ in only one observation, they are approximately equal. Therefore, setting $\sum \hat{u}_{t-1}^2 \approx \sum \hat{u}_t^2$



$$\hat{\rho} = \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2}$$

$$d \approx 2(1 - \hat{\rho})$$

But since $-1 \leq \rho \leq 1$, $0 \leq d \leq 4$



$$\hat{u}_t = Y_t - \hat{Y}_t$$

true y
Approximate y

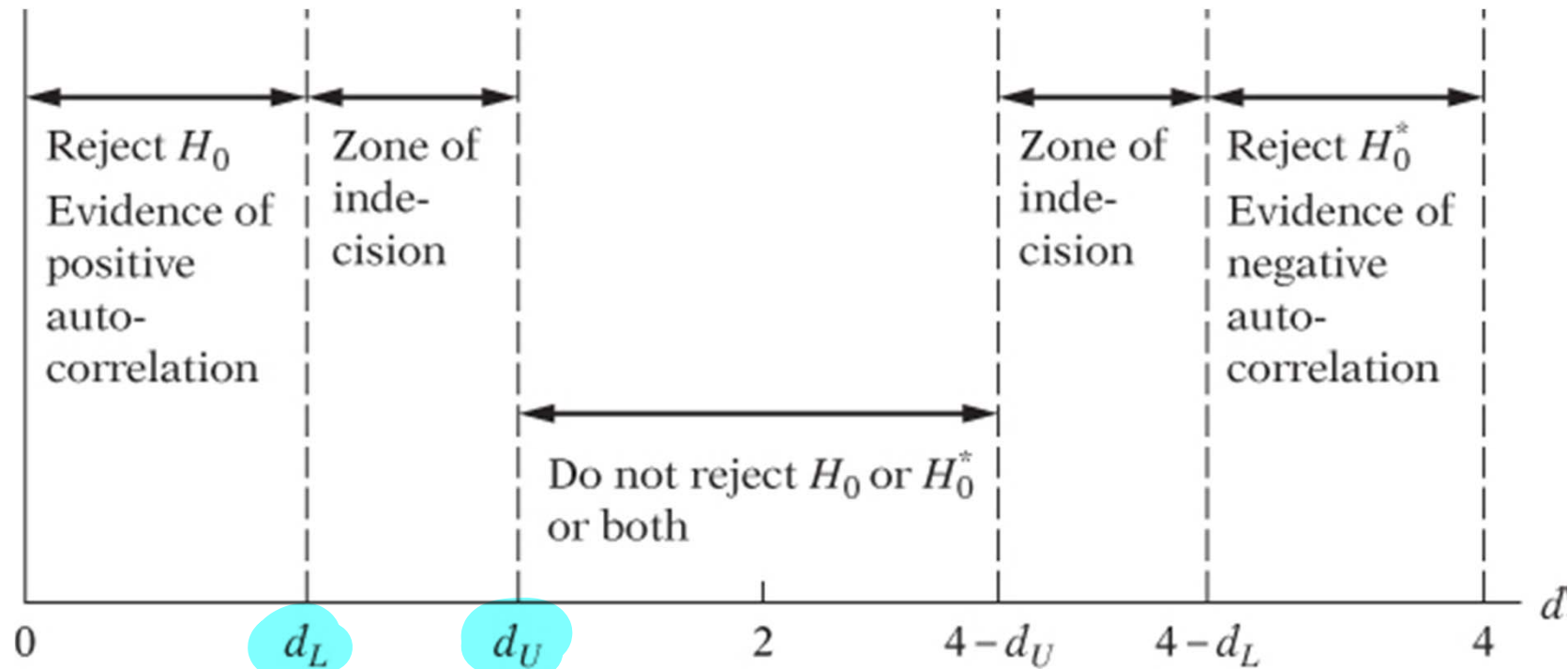
$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_2 + \dots$$

The mechanics of the Durbin-Watson test are $\hat{u}_t$

1. Run the OLS regression and obtain the residuals
2. Compute d
3. For the given sample size and given number of explanatory variables, find out the critical d_L and d_U values
4. Now follow the decision rules given in Table 12.6



Durbin-Watson d statistic



Legend

H_0 : No positive autocorrelation

H_0^* : No negative autocorrelation

H_1 : otherwise

TABLE 12.6**Durbin–Watson d
Test: Decision Rules**

Null Hypothesis	Decision	If
No positive autocorrelation	Reject	$0 < d < d_L$
No positive autocorrelation	No decision	$d_L \leq d \leq d_U$
No negative correlation	Reject	$4 - d_L < d < 4$
No negative correlation	No decision	$4 - d_U \leq d \leq 4 - d_L$
No autocorrelation, positive or negative	Do not reject	$d_U < d < 4 - d_U$



Example

- ▶ U.S. Consumption Expenditure for the period 1947-2000

$$\ln Consumption = \beta_1 + \beta_2 \ln income + \beta_3 \ln wealth + \beta_4 Interest$$



$$\ln Consumption = -0.4677 + 0.8049 \ln income + 0.2013 \ln wealth + 0.0027 Interest$$

Source	SS	df	MS
Model	16.1637474	3	5.3879158
Residual	.007120721	50	.000142414
Total	16.1708681	53	.305110719

Number of obs = 54
 F(3, 50) = 37832.66
 Prob > F = 0.0000
 R-squared = 0.9996
 Adj R-squared = 0.9995
 Root MSE = .01193

Inc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lnincome	.8048728	.0174978	46.00	0.000	.7697273	.8400182
lnwealth	.2012702	.0175926	11.44	0.000	.1659345	.236606
i	-.0026891	.0007619	-3.53	0.001	-.0042194	-.0011587
_cons	-.467712	.042778	-10.93	0.000	-.5536342	-.3817899



Hypothesis testing

Step 1

H_0 : *No positive autocorrelation*

H_0^* : *No negative autocorrelation*

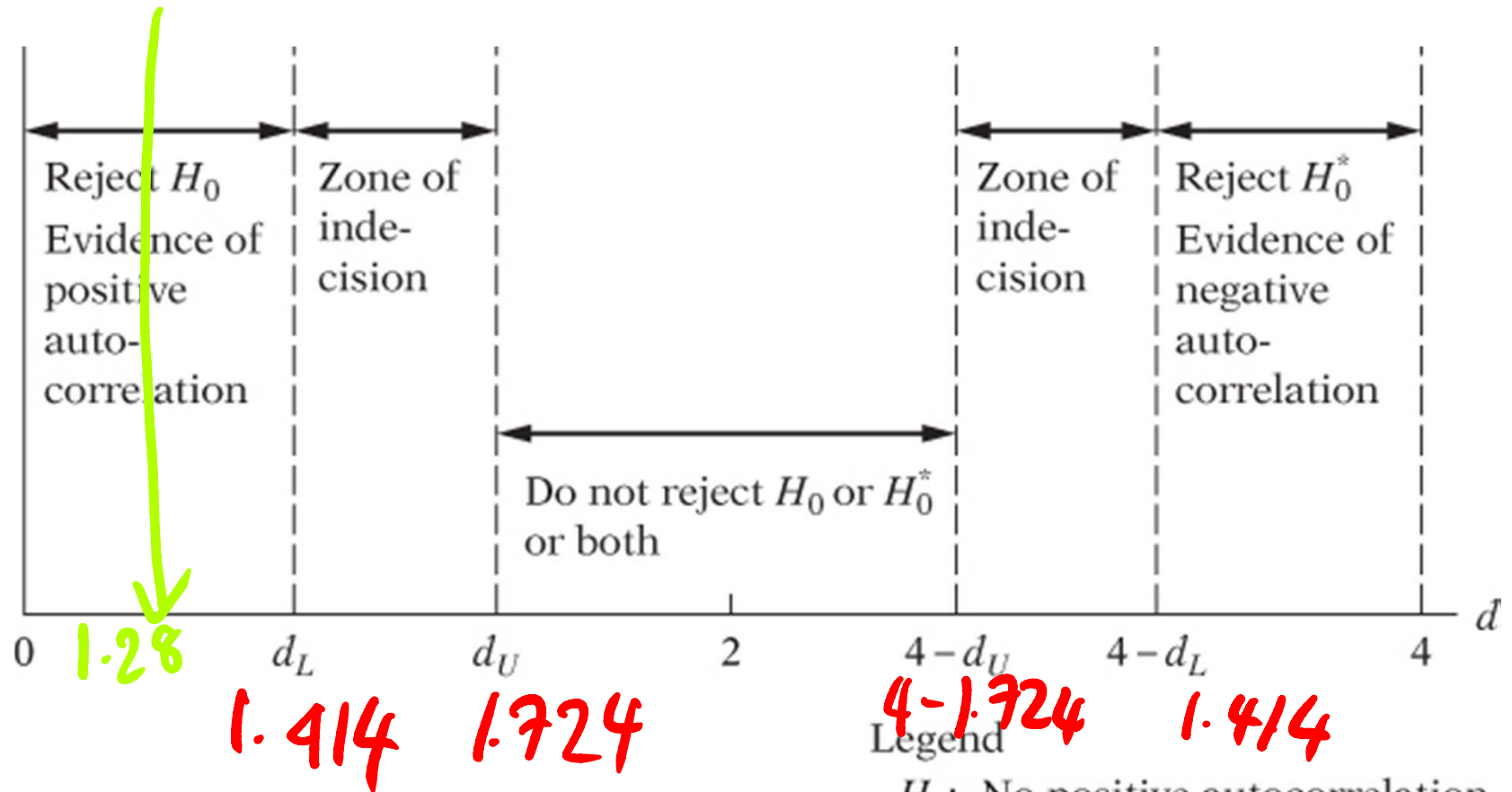
H_1 : *otherwise*

H_1^* : *otherwise*



Durbin-Watson d statistic

Step 3



Durbin-Watson test

```
. tsset year  
    time variable: year, 1947 to 2000  
      delta: 1 unit
```

```
. estat dwatson
```

Durbin-Watson d-statistic (4, 54) = 1.289232

$$d \approx 2(1 - \hat{\rho})$$

($n = 54, k = 4$) At 5% Significance level

$$d_L = 1.414$$

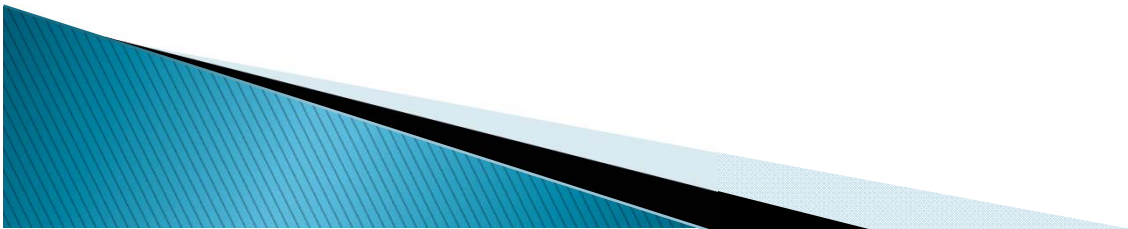
$$d_U = 1.724$$

From Durbin-Watson Table .

Reject null hypothesis

Evidence of positive autocorrelation

Correcting for (Pure) Autocorrelation



e

Correcting for (Pure) Autocorrelation

1. **The method of Generalized Least Squares (GLS) method**
2. The Newey West method – to obtain standard errors of OLS estimators that are corrected for autocorrelation

Robust command to correct std error.



The method of Generalized Least Squares (GLS)

The two-variable regression model:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \quad (0)$$

and assume that the error term follows the AR(1) scheme, namely,

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

Now we consider two cases:

- (1) ρ is known
- (2) ρ is not known but has to be estimated



When ρ is known

$$Y_t = \beta_1 + \beta_2 X_t + u_t \quad (1)$$

$$\rho Y_{t-1} = \rho\beta_1 + \rho\beta_2 X_{t-1} + \rho u_{t-1} \quad (2)$$

$$(Y_t - \rho Y_{t-1}) = \beta_1(1 - \rho) + \beta_2(X_t - \rho X_{t-1}) + (u_t - \rho u_{t-1}) \quad (3)$$

$$(Y_t - \rho Y_{t-1}) = \beta_1(1 - \rho) + \beta_2(X_t - \rho X_{t-1}) + \varepsilon_t$$

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

$$Y_t^* = \beta_1^* + \beta_2^* X_t^* + \varepsilon_t \quad (*)$$

Since the error term in

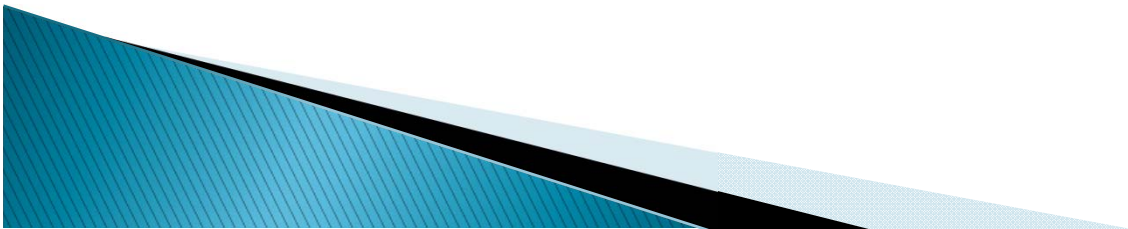
$$Y_t^* = \beta_1^* + \beta_2^* X_t^* + \varepsilon_t$$

satisfies the usual OLS assumptions, we can apply OLS to the transformed variables X^* and Y^* and obtain estimators with all the optimum properties, namely, BLUE.



When ρ is not known

- ▶ based on Durbin-Watson d Statistic



ρ based on Durbin–Watson d Statistic

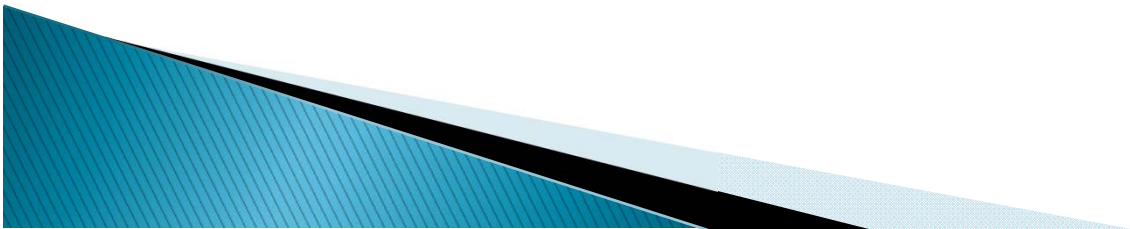
$$\hat{\rho} \approx 1 - \frac{d}{2}$$

In reasonably large sample one can obtain ρ and use it to transform the data as shown in the generalized difference equation

$$(Y_t - \rho Y_{t-1}) = \beta_1(1 - \rho) + \beta_2(X_t - \rho X_{t-1}) + \varepsilon_t$$

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

The Newey West method



The Newey West method of correcting the OLS standard errors

The corrected standard errors are known as **HAC (Heteroscedasticity and autocorrelation-consistent) standard errors** or simply **Newey West standard errors**



reg y x₁ x₂ , robust

The Newey West method

```
. newey Inc lnincome lnwealth i, lag(3)
```

Regression with Newey-West standard errors
maximum lag: 3

Number of obs = 54
F(3, 50) = 22341.20
Prob > F = 0.0000

Inc	Coef.	Newey-West Std. Err.	t	P> t	[95% Conf. Interval]	
lnincome	.8048728	.0171172	47.02	0.000	.7704919	.8392536
lnwealth	.2012702	.0154469	13.03	0.000	.1702441	.2322963
i	-.0026891	.0008798	-3.06	0.004	-.0044563	-.0009219
_cons	-.467712	.0439367	-10.65	0.000	-.5559616	-.3794625



Source

Gujarati, D.N. (2009) Basic Econometrics. 5th ed.
Singapore, McGraw-Hill.

