

ASIL SHAAR

FINN3302

النمذجة المالية

**CHAPTER 5: Classical linear
regression model assumptions
and diagnostics**

Chapter 5

Classical linear regression model assumption and diagnostic.

Classical linear regression model assumption

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$$

(1) $E(u_t) = 0$

(2) Homoskedasticity assumption $\text{Var}(u_t) \neq f(x)$

(3) $\text{Cov}(u_i, u_j) = 0$ (No serial autocorrelation) (error term are independent and one another)

(4) $\text{Cov}(u_t, x_t) = 0$ (The X matrix is non-stochastic or fixed in repeated samples)

(5) $u_t \sim N$

In general we could encounter any combination of 3 problem:

- the coefficient estimates are wrong.
- the associated standard errors are wrong.
- the distribution that we assumed for the test statistic will be inappropriate.

Solutions:

- the assumptions are no longer violated
- we work around the problem so that we use alternative techniques which are still valid.

* Assumption 1: $E(u_t) = 0$

Violated. $E(u_t) \neq 0$ \Rightarrow $E(u_t) = \text{constant}$ $\neq 0$ \Rightarrow systematic error \Rightarrow statistical test \Rightarrow to keep an intercept if you Regression equation to hold for systematic error

* Assumption 2: $\text{Var}(u_t) = \sigma^2 < \infty$

Violated $\leftarrow A_2$ all error terms \leftarrow

graph \Rightarrow (1)
 Statistical test \Rightarrow (2)

Homoskedasticity assumption \rightarrow The error terms have a constant variance $= \sigma^2$
 $\text{Var}(u_t) \neq f(x)$
 systematic error

To test if this assumption holds or not we can use one of the following methods:

- (a) graph.
- (b) Statistical test. (white test)

white test \Rightarrow
 \Rightarrow F test

white test:

1) estimate the regression equations

e.g: $Y_t = B_1 + B_2 X_{2t} + B_3 X_{3t} + u_t$

2) Collect the residuals

3) run an auxiliary regression:

$$\hat{u}_t^2 = \alpha_1 + \alpha_2 X_{2t} + \alpha_3 X_{3t} + \alpha_4 X_{2t}^2 + \alpha_5 X_{3t}^2 + \alpha_6 X_{2t} X_{3t} + v_t$$

regress \hat{u}_t^2 on all X 's, their squared terms and their cross product.

test if $\text{Var}(u_t) = \sigma^2$
 using \hat{u}_t

$$S^2 = \frac{RSS}{T-K}$$

\downarrow
 Variance of errors
 $S^2 = \frac{\sum \hat{u}_t^2}{T-K}$

4) Construct our hypotheses which are:

$H_0: \alpha_2 = 0$ and $\alpha_3 = 0$ and $\alpha_4 = 0$ and $\alpha_5 = 0$ and $\alpha_6 = 0$

$H_1: \alpha_2 \neq 0$ or $\alpha_3 \neq 0$ or $\alpha_4 \neq 0$ or $\alpha_5 \neq 0$ or $\alpha_6 \neq 0$

(joint test)

\Rightarrow

In words :

H_0 : The error terms are homoskedastic, they have a constant variance σ^2

H_1 : The error terms are heteroscedastic, they don't have a constant variance σ^2

(5) perform the test :

- ① F-version of the test.
- ② Chi squared - version of the test.

Chi squared - version of the test.

Chi stat $\leftarrow R^2 \cdot T$ \rightarrow number of observation.

from the auxiliary regression

$df = m$
 \rightarrow Restriction

Chi stat \sim Chi squared distribution.

get the critical value from the Chi squared distribution table.

rejection rule: If Chi stat $>$ Chi critical then Reject H_0 .

* How to deal with heteroscedasticity :-

- ① use another estimator (GLS)
- ② transform variables into natural logarithmic
- ③ use white standard errors.

* Assumption 3:

No serial autocorrelation $\rightarrow \text{Cov}(u_i, u_j) = 0 \quad i \neq j$

من لازم ان لا توجد الارتباط

To detect if this assumption is violated, we can use:

- ① graphs
- ② statistical test

\rightarrow

Statistical tests used to detect autocorrelation:

- ① Durbin ~~Watson~~ ^{watson} test
- ② B-G test

* Durbin ~~watson~~ ^{watson} test

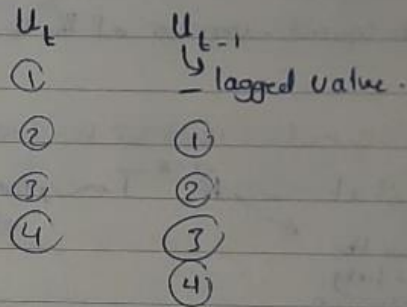
ليس يوجد إلا في علاقة بين error و error قبلها

Relationship is between an error and the previous one.

$$U_t = \rho U_{t-1} + U_t$$

\swarrow
 correlation coefficient

$H_0: \rho = 0$
 $H_1: \rho \neq 0$



approximation formula.

$$DW \text{ stat} = \frac{\sum_{t=2}^T (\hat{U}_t - \hat{U}_{t-1})^2}{\sum_{t=2}^T \hat{U}_t^2} \rightarrow \sum_{t=1}^T \hat{U}_t^2$$

$$DW \text{ stat} \approx 2(1 - \hat{\rho})$$

\swarrow
 Correlation coefficient

$\hat{\rho}$	-1	0	+1	
DW stat	4	2	0	$0 \leq DW \leq 4$
	↓	↓	↓	
	negative autocorrelation	no autocorrelation	positive autocorrelation	

get the critical values (2 critical value) from DW table.

D_U → upper

D_L → lower.



Reject H_0 :

positive autocorrelation

inconclusive

Do not Reject

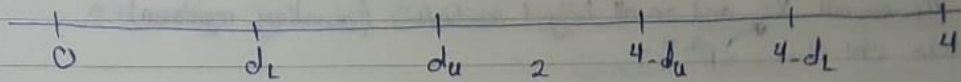
H_0 :

No evidence of autocorrelation

inconclusive

Reject H_0 :

negative autocorrelation



Weaknesses of Dw test:

- ① test for first order autocorrelation.
- ② Inconclusive regions.

Conditions which must be fulfilled for Dw to be a valid test:

1. Constant term in regression
2. Regressors are non-stochastic.
3. No lags of dependent variable.

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 Y_{t-1} + u_t$$

dynamic model.

K → number of parameters to be estimated in the regression excluding the constant term.

* BG test (Breusch-Godfrey Test)

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \dots + \rho_r u_{t-r} + u_t$$

$$H_0: \rho_1 = 0 \text{ and } \rho_2 = 0 \text{ and } \rho_3 = 0 \dots \text{ and } \rho_r = 0$$

$$H_1: \rho_1 \neq 0 \text{ or } \rho_2 \neq 0 \text{ or } \rho_3 \neq 0 \dots \text{ or } \rho_r \neq 0$$

→ H_0 : There is no serial autocorrelation, the error term are independent of one another.

H_1 : There is serial autocorrelation, the error term are not independent of one another.

- ① Estimate the regression equation.
- ② Collect the residuals.
- ③ regress \hat{u}_t on all X_t 's and on all lagged residuals. (auxiliary regression)
- ④ Calculate chi stat $= R^2 * T$
- ⑤ get chi critical
- ⑥ perform the test \rightarrow If chi stat $>$ chi critical then Reject H_0 .

* Assumption 4:

The X matrix is non-stochastic of fixed in repeated samples

$$\text{Cov}(u_t, X_t) = 0$$

Multicollinearity problem يكون عندي مشكلة اولى

This problem occurs when the explanatory variables are very highly correlated with each other.

هناي المشكلة ممكن اتوصف بـ two types :-

① the severe one \leftarrow يكون في عندي perfect multicollinearity \rightarrow correlation $>$ 100% يعني

② the moderate one \leftarrow يكون في عندي Near multicollinearity \rightarrow correlation $>$ 100% اقل من

Cannot estimate all the coefficient ①

ex: suppose $X_3 = 2X_2$ \rightarrow خطا

② يعني يكون في Relationship بين explanatory variables \rightarrow two or more \rightarrow بين عندي perfect

- R^2 will be high but the individual coefficient will have high standard error \leftarrow نتوقع بـ high

- The Regression becomes very sensitive to small changes in the specification.

- Thus Confidence intervals for the parameters will be very wide, and significance tests might therefore give inappropriate conclusions.

Measuring Multicollinearity

2 Variables \rightarrow $|r|$ Correlation coefficient Statistical test \rightarrow ما لي
Correlation matrix اذا كانا أكثر نحسب

* Solutions of the problem of Multicollinearity

The easiest ways to "cure" the problems are-

- drop one of the collinear variables (إزالة واحدة)
- transform the highly correlated variables into a ratio
- go out and collect more data
 - e.g. - a longer run of data
 - switch to a higher frequency

* Assumption 5: Normality التوزيع الطبيعي العادي و mean & Variance

H_0 : The error terms are normally distributed with skewness = 0 and excess Kurtosis = 0

H_1 : The error terms are not normally distributed with skewness $\neq 0$ or excess Kurtosis $\neq 0$

$$W = T \left[\frac{b_1^2}{8} + \frac{(b_2 - 3)^2}{24} \right] \quad \text{منه نطلبون نحسبها مالدستبان}$$

$W > \chi^2_{critical}$ then Reject H_0

* Testing for Departures from Normality:

- The Bera-Jarque normality test.
- A normal distribution is not skewed and is defined to have a coefficient of kurtosis of 3.
- The kurtosis of the normal distribution is 3 so its excess kurtosis ($b_2 - 3$) is zero.
- Skewness and kurtosis are the (standardised) third and fourth moments of a distribution.

$$\text{Kurtosis} = 3 \quad \text{excess Kurtosis} = \text{Kurtosis} - 3$$
$$3 - 3 = 0$$

* We estimate b_1 and b_2 using the residuals from the OLS regression.

What do we do if we find evidence of non-normality?

ايتم علاج تكون الكتل؟

- * you have to use different approach to conduct the analysis
- * to remove outliers of extreme residuals.
- * to use dummy variables.