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

$$y_t = \beta_0 + B_1 X_{1t} + B_2 X_{2t} + B_3 X_{3t} + u_t$$

① $E(u_t) = 0$

② Homostidasticity assumption $var(u_t) \neq f(x)$

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

④ $Cov(u_t, x_t) = 0$ (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 coefficent estimates are wrong.
- the assocated standard errors are wrong.
- the distribution that we assumed for the test statistic will be inappropiate

Solutions:

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

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

* هاي ال Assumption ابداً Violated.

* ما في Statistical test لهاي الحالة وهي دائماً صفر. حتى لو طلعت Residuals ومش mean صفر بنعمل الحل

* الحل: to keep an intercept if you Regression equation to hold for systematic test error

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

① نشوف graph
② نستخدم Statistical test

عشان نعرف انه A_2 ← Violated أولاً.

Homoskedasticity assumption → The error terms have a constant variance $= \sigma^2$

$Var(u_t) \neq f(x)$ يعني ما في systimatic

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

ⓐ graph.
ⓑ Statistical test. (white test)

ح نستخدم white test
GQ test مش داخل.

white test:

1. estimate the regression equation:
   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 $Var(u_t) = \sigma^2$ using $\hat{u}_t$

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

↓

Variance of errors

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

4. Construct our hypotheses which are: (joint test)

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 $\alpha_6 \neq 0$

⇒

In words :

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

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

$\boxed{5}$ perform the test :

① F-version of the test.

→ ② Chi squared - version of the test.

Chi squared - Version of the test.

Chi stat = $R^2 * T$ → number of observation.

R^2: from the auxiliry regression

df = m
عدد Resstriction.

Chi stat ~ Chi squred distribution:

get the critical value from the Chi squred distribution table.

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

* How to deal with heteroscedasicity :-

① Use another estimator (GLS)

② transform variables into natural logarithmic

③ use white standard errors.

* Assumption 3 :

No serial autocorrelation → $Cov(u_i, u_j) = 0 \quad i \neq j$

متى لازم اتوقع الاشارة

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

① graphs

② statistical test

→

Statistical tests used to detect autocorrelation:

① Durbin watson test

② BG test

* Durbin watson test   بس بنحصر اذا في علاقة بين error وبين قبله

Relationship is between an error and the previous one.

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

ρ: Correlation coefficient

u_t	u_{t-1} (lagged value)
1	–
2	1
3	2
4	3
	4

$H_0: \rho = 0$

$H_1: \rho \neq 0$

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

approximation formula.

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

$\hat{\rho}$: Correlation coefficent

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

$0 \leq DW_{stat} \leq 4$

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

$D_U \rightarrow$ upper

$D_L \rightarrow$ lower.

$\Rightarrow$

Reject H_0: positive autocorrelation	inconclusive	Do not Reject H_0: No evidence of autocorrelation	inconclusive	Reject H_0: negative autocorrelation

$0 \quad d_L \quad d_u \quad 2 \quad 4-d_u \quad 4-d_L \quad 4$

Weaknesses of DW test:

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

Conditions which must be fullfilled for DW to be a valid test:

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

$$Y_t = B_1 + B_2 X_{2t} + B_3 Y_{t-1} + u_t$$

dynamic model.

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

* BG test (Breusch-Godfrey Test)

$$u_t = P_1 u_{t-1} + P_2 u_{t-2} + P_3 u_{t-3} + \dots + P_r u_{t-r} + u_t$$

$H_0: P_1 = 0$ and $P_2 = 0$ and $P_3 = 0$ ...... and $P_r = 0$

$H_1: P_1 \neq 0$ or $P_2 \neq 0$ or $P_3 \neq 0$ ...... or $P_r \neq 0$

→ 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's and on all lagged residuals. (auxiliary regression)
④ Calculate chi stat $= R^2 * T$
⑤ get chi critical
⑥ perform the test → If chi stat > chi critical then Reject H_0.

*** Assumption 4:**

The X matrix is non-stochastic of fixed in repeated samples
$Cov(u_t, X_t) = 0$

يكون عندي مشكلة اسمها Multicolinearity problem

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

هاي المشكلة ممكن اشوفها بصورتين :-

① الصوره الاولى ← يكون في عندي perfect multicolinearity يعني Correlation ٪100
② الصورة الثانية ← يكون في عندي Near multicolinearity اقل من Correlation ٪100

① Cannot estimate all the Coefficient
ex: suppose $X_3 = 2X_2$ وهيك

② يعني يكون في Relationship بين explanatory variables بين two or more بس مش perfect

شو بصير؟ ⇐
- R^2 will be high but the individual Coefficient will have high standard error
- 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 اذا شيئين Correlation coefficient منحسب Statistical test ماني
Correlation matrix اذا كانوا أكثر بحسب

* Solutions of the problem of Multicollinearity
  The easiest ways to "Cure" the problems are -
  - drop one of the Collinear Variables افضل طريقة
  - 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 — mean & variance النص اليمين زي الشكل ومنعرف

H_0: The error Terms are normaly distributed with skewness = 0 and excess Kurtosis = 0

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

$$W = T\left[\frac{b_1^2}{6} + \frac{(b_2-3)^2}{24}\right]$$

مش مطلوب نحسبها بالامتحان.

W > Chi critical then Reject Ho

* 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 it excess Kurtosis (b_2-3) is zero
  - Skewness and kurtosis are the (standardised) third and fourth moments of a distribution.

Kurtosis = 3 excess Kurtosis = 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 ?

ايش راح تكون الحلول؟

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