ES1004 Econometrics by Example

Lecture 5: Heteroscedasticity

Dr. Hany Abdel-Latif

Swansea University, UK

Gujarati textbook, second edition

9th July 2016

CLRM Assumptions

A$_1$: model is linear in parameters

A$_2$: regressors are fixed non-stochastic

A$_3$: the expected value of the error term is zero $E(u_i|X) = 0$

A$_4$: homoscedastic or constant variance of errors $var(u_i|X) = \sigma^2$

A$_5$: no autocorrelation, $cov(u_i, u_j) = 0, i \neq j$

A$_6$: no multicollinearity; no perfect linear relationships among the Xs

A$_7$: no specification bias

Basic Idea

- CLRM assumes the variance of the error term u_i is constant A_4
- this may not be the case especially with cross-section data
  - the presence of outliers in the data
  - incorrect functional form of the regression model
  - incorrect transformation of data
  - mixing observations with different measures of scale
    - such as mixing high-income households with low-income households

Savings Model

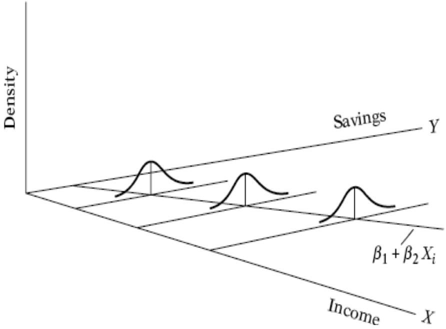
- consider the two-variable model

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

- Y and X represent savings and income, respectively
- as income increases, savings on the average also increase
- homoscedasticity $\rightarrow$ the variance of savings remains the same at all levels of income

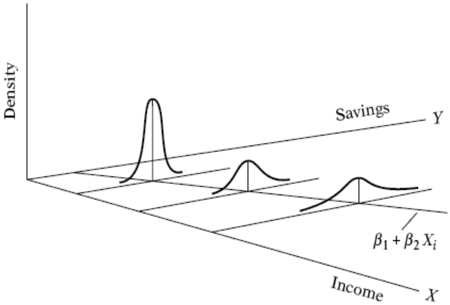
Homoscedastic Disturbances

- $var(u_i) = E(u_i^2|X_i) = \sigma^2$ i.e., constant variance
- equal (homo) spread (scedasticity) or equal variance



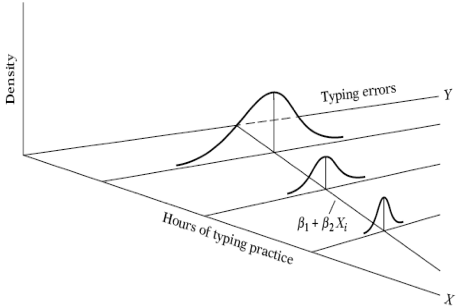
Heteroscedastic Disturbances

- $var(u_i) = E(u_i^2|X_i) = \sigma_i^2$
- the subscript of σ^2 refers to non-constant conditional variances of u_i



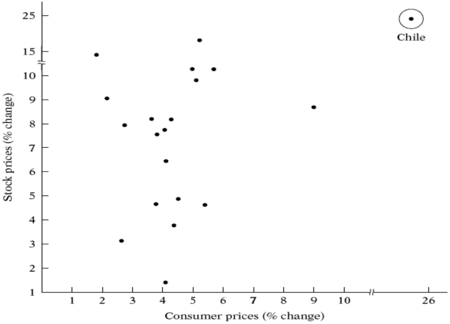
Example of Heteroscedasticity I

- error-learning models: as people learn, their errors of behaviour become smaller over time (σ_i^2 is expected to decrease)



Example of Heteroscedasticity II

- the presence of outliers
  - inclusion or exclusion of such an observation, especially if the sample size is small, can substantially alter the results of regression analysis



chile is an outlier because the given Y and X values are much larger than for the rest of the countries

Heteroscedasticity and OLS Estimation

- if heteroscedasticity exists, several consequences ensue
  - OLS estimators are still unbiased and consistent
  - yet they are less efficient
  - making statistical inference (estimated t values) less reliable
  - estimators are not BLUE
  - BLUE estimators are provided by the method of weighted least squares (WLS)

How to test for Heteroscedasticity

- graph histogram of squared residuals
- graph squared residuals against predicted Y
- Breusch-Pagan (BP) test
- White's test of heteroscedasticity
- other tests
  - Park, Glejser, Spearman's rank correlation and Goldfeld-Quandt tests of heteroscedasticity

Example: Abortion

- data `table5_1.xls` abortion rates in 50 US states and other characteristics
- religion, price, laws, funds, educ, income, and picket
- regress abortion rate on characteristics [variables] that explain abortion rate

Example: Abortion - Estimation

Dependent Variable: ABORTION
Method: Least Squares
Date: 04/01/14 Time: 13:09
Sample: 1 50
Included observations: 50

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	14.28396	15.07763	0.947361	0.3489
RELIGION	0.020071	0.086381	0.232355	0.8174
PRICE	-0.042363	0.022223	-1.906255	0.0635
LAWS	-0.873102	2.376566	-0.367380	0.7152
FUNDS	2.820003	2.783475	1.013123	0.3168
EDUC	-0.287255	0.199555	-1.439483	0.1574
INCOME	0.002401	0.000455	5.274041	0.0000
PICKET	-0.116871	0.042180	-2.770782	0.0083

R-squared	0.577426	Mean dependent var	20.57800
Adjusted R-squared	0.506997	S.D. dependent var	10.05863
S.E. of regression	7.062581	Akaike info criterion	6.893145
Sum squared resid	2094.962	Schwarz criterion	7.199069
Log likelihood	-164.3286	Hannan-Quinn criter.	7.009642
F-statistic	8.198706	Durbin-Watson stat	1.805932
Prob(F-statistic)	0.000003		

Example: Abortion - Graph

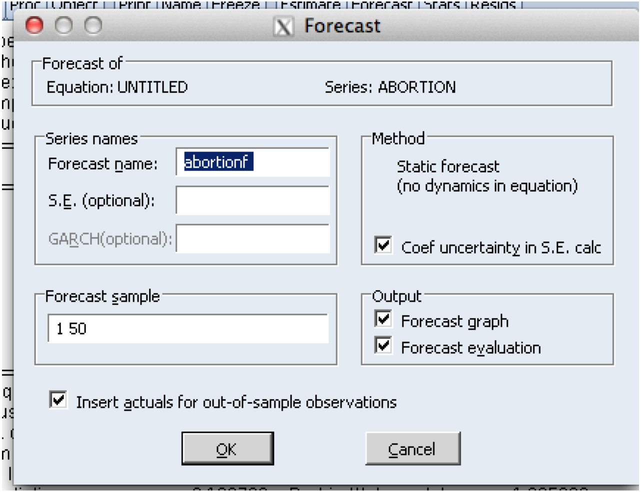
Equation: UNTITLED Workfile: TABLE5_1::Table5_1\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

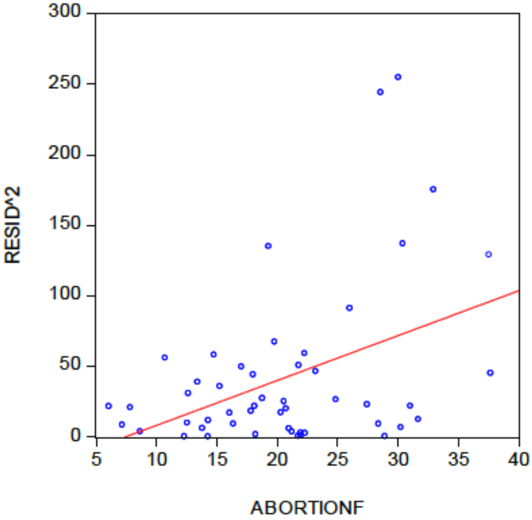
Dependent Variable: ABORTION
Method: Least Squares
Date: 07/09/16 Time: 14:18
Sample: 1 50
Included observations: 50

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	14.28396	15.07763	0.947361	0.3489
RELIGION	0.020071	0.086381	0.232355	0.8174
PRICE	-0.042363	0.022223	-1.906255	0.0635

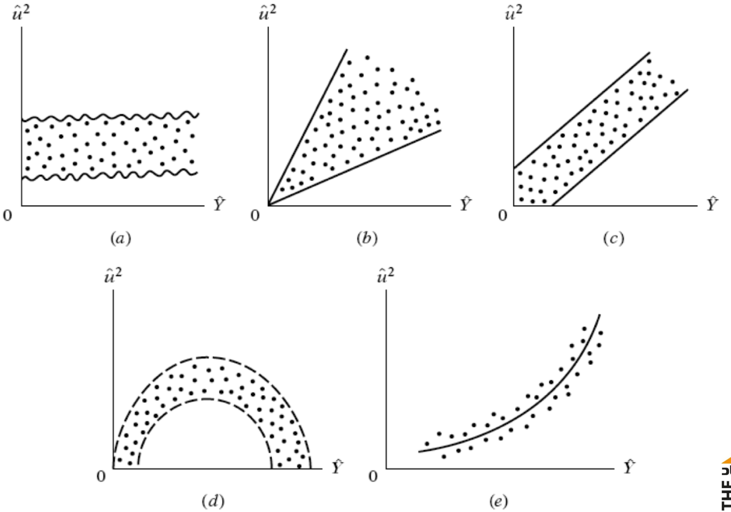
Example: Abortion - Graph



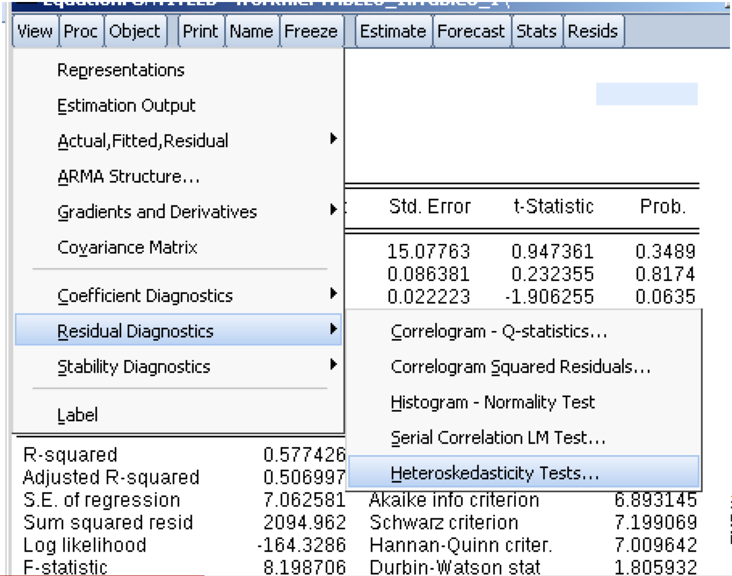
Example: Abortion - Graph



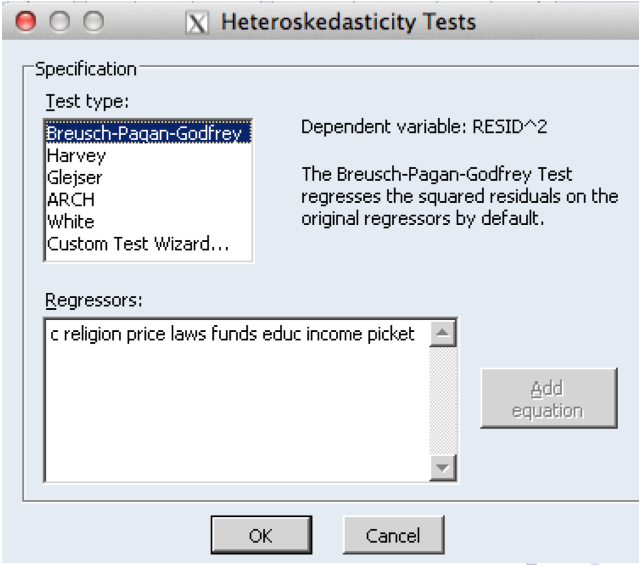
Graph



Formal Tests I



Formal Tests II



Breusch-Pagan Godfrey I

- estimate OLS regression, and obtain the squared OLS residuals from this regression
- regress the square residuals on k regressors in the model
- the null hypothesis: the error variance is homoscedastic
- use F statistic from this regression with (k-1) and (n-k) df to test this hypothesis
- if computed F-statistic is statistically significant $\rightarrow$ reject the hypothesis of homoscedasticity

Breusch-Pagan Godfrey II

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	2.823820	Prob. F(7,42)	0.0167
Obs*R-squared	16.00112	Prob. Chi-Square(7)	0.0251
Scaled explained SS	10.57563	Prob. Chi-Square(7)	0.1582

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 04/01/14 Time: 13:10
Sample: 1 50
Included observations: 50

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	16.68558	110.1532	0.151476	0.8803
RELIGION	-0.134865	0.631073	-0.213707	0.8318
PRICE	0.286153	0.162357	1.762492	0.0853
LAWS	-8.566472	17.36257	-0.493387	0.6243
FUNDS	24.30981	20.33533	1.195447	0.2386
EDUC	-1.590385	1.457893	-1.090879	0.2815
INCOME	0.004710	0.003325	1.416266	0.1641
PICKET	-0.576745	0.308155	-1.871606	0.0682

R-squared	0.320022	Mean dependent var	41.89925
Adjusted R-squared	0.206693	S.D. dependent var	57.93043
S.E. of regression	51.59736	Akaike info criterion	10.87046
Sum squared resid	111816.1	Schwarz criterion	11.17639
Log likelihood	-263.7616	Hannan-Quinn criter.	10.98696
F-statistic	2.823820	Durbin-Watson stat	2.258203
Prob(F-statistic)	0.016662		

White's Test I

- regress the squared residuals on the regressors, the squared terms of these regressors, and the pair-wise cross-product term of each regressor
- obtain the R^2 value from this regression and multiply it by the number of observations
- the null hypothesis: the error variance is homoscedastic
- follows Chi-square distribution with *df* equal the number of coefficients estimated

White's Test II

Heteroskedasticity Test: White

F-statistic	0.869647	Prob. F(33,16)	0.6455
Obs*R-squared	32.10224	Prob. Chi-Square(33)	0.5116
Scaled explained SS	21.21735	Prob. Chi-Square(33)	0.9437

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 04/01/14 Time: 13:32
Sample: 1 50
Included observations: 50
Collinear test regressors dropped from specification

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	406.3380	3370.544	0.120556	0.9055
RELIGION^2	-0.043923	0.142450	-0.308338	0.7618
RELIGION*PRICE	0.037087	0.057066	0.649904	0.5250
RELIGION*LAWS	0.123795	4.820159	0.025683	0.9798
RELIGION*FUNDS	5.197267	10.94371	0.474909	0.6413
RELIGION*EDUC	-0.052380	0.317205	-0.165131	0.8709
RELIGION*INCOME	0.000192	0.000692	0.277901	0.7846
RELIGION*PICKET	-0.013050	0.074687	-0.174728	0.8635
RELIGION	-7.095749	26.16119	-0.271232	0.7897
PRICE^2	0.003879	0.009101	0.426209	0.6756
PRICE*LAWS	-0.099400	0.868265	-0.114481	0.9103

How to Remedy for Heteroscedasticity

- use method of weighted least squares WLS
  - divide each observation by the heteroscedastic σ_i and estimate the transformed model by OLS [yet true variance is rarely known]
  - if the true error variance is proportional to the square of one of the regressors, we can divide both sides of the equation by that variable and run the transformed regression
- take natural log of dependent variable
- use White's heteroscedasticity-consistent standard errors or robust standard errors [valid in large samples]