ECONOMETRICS II

Topic 2.1: **HETEROSCEDASTICITY** AND AUTOCORRELATION

Luca Salvadori

UAB, BSE, TARC, IEB

Academic Year 22-23

1. Heteroscedasticity and autocorrelation

In topic 1, we reviewed the linear regression model $Y_t = X_t^{'}\beta + u_t$,

where:
- $t = 1, 2, ..., T$
- $X_t = (1, X_{2t}, ..., X_{kt})'$
- $\beta = (1, \beta_2, ..., \beta_k)'$

... and where the classic HPs hold:

1. $X_1, ... X_k$ and the Y are **iid.**
2. **there is not perfect multicollinearity**; $rg(X) = k$
3. X_t and u_t have nonzero finite fourth order moments.
4. $E(u_t/X) = 0 \Rightarrow E(u_t) = 0$, t = 1, 2, ..., T
5. $Var(u) = E(uu') = \sigma^2 I_T$.
   - $Var(u_t) = E(u^2) = \sigma^2$, $t = 1, 2, ..., T$ (**homoscedasticity**)
   - $Cov(u_t, u_s) = E(u_t u_s) = 0, \forall t \neq s$ (**u_t not serially correlated**)
6. **Normality:** $u \sim N(0, \sigma^2 I)$; $u_t \sim N(0, \sigma^2)$

1. Heteroscedasticity and autocorrelation

In this topic we will see the consequences of relaxing assumption 5 (and 1)

- **HETEROSCEDASTICITY:** ∃ *i,j* such that $V(u_i|X_i) \neq V(u_j|X_j)$,

- **AUTOCORRELATION:** ∃ *i,j* such that $Cov(u_i u_j|X) \neq 0$,

1.1 Heteroscedasticity

Specifically, let's start with the estimation of the regression model with heteroskedastic errors but without the presence of serial autocorrelation, that is to say, assumptions 2-4 are verified while 5 becomes:

$var(u_t) = E(u_t^2) = \sigma^2\omega_t, t = 1, 2, ..., T$ heteroscedasticity

$cov(u_t, u_s) = E(u_t u_s) = 0, \forall t \neq s$ u_i not serially correlated

$$Var(u) = E(uu') = \sigma^2 \Omega,$$ where

$$\Omega = \begin{bmatrix} \omega_1 & 0 & ... & 0 \\ 0 & \omega_2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \omega_T \end{bmatrix}$$

1.1 Heteroscedasticity

Estimation by OLS in the presence of heteroscedasticity:

1. The OLS estimator is unbiased: E($\widehat{\beta}$)= β

2. The variance is: $$Var(\widehat{\beta}|X) = \sigma^2(X'X)^{-1}\left(\sum_{t=1}^{T}\omega(X_t)X_tX_t'\right)(X'X)^{-1}$$

1.1 Heteroscedasticity

Estimation by OLS in the presence of heteroscedasticity:

$$\begin{aligned} Var(\widehat{\beta}|X) &= E\left[(\widehat{\beta}-E(\widehat{\beta}|X))(\widehat{\beta}-E(\widehat{\beta}|X))'|X\right] = E\left[(\widehat{\beta}-\beta)(\widehat{\beta}-\beta)'|X\right] \\ &= E\left[(X'X)^{-1}X'uu'X(X'X)^{-1}|X\right] = (X'X)^{-1}X'E(uu'|X)X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1} \end{aligned}$$

where

$$\begin{aligned} X'\Omega X &= \left[\, X_1,\ X_2,\ \ldots\ , X_T \,\right] \begin{bmatrix} \omega(X_1) & 0 & \cdots & 0 \\ 0 & \omega(X_2) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \omega(X_T) \end{bmatrix} \begin{bmatrix} X_1' \\ X_2' \\ \vdots \\ X_T' \end{bmatrix} \\ &= \left[\, \omega(X_1)X_1,\ \omega(X_2)X_2,\ \ldots\ , \omega(X_T)X_T \,\right] \begin{bmatrix} X_1' \\ X_2' \\ \vdots \\ X_T' \end{bmatrix} = \left(\sum_{t=1}^{T} \omega(X_t)X_tX_t' \right) \end{aligned}$$

1.1 Heteroscedasticity

Estimation by OLS in the presence of heteroscedasticity:

Given that $Var(\widehat{\beta}|X) \neq \sigma^2(X'X)^{-1}$

- Therefore, $\widehat{\sigma}^2(X'X)^{-1}$ is not an appropriate estimator for Var ($\widehat{\beta}_{OLS}$)

As a consequence of this result, the tests that we saw when the classical assumptions are met and that they were based on Var($\widehat{\beta}_{OLS}$) = $\widehat{\sigma}^2(X'X)^{-1}$ are not valid when the errors are heteroskedastic.

- Also, when the errors are heteroskedastic the **OLS** estimator **is not the best linear unbiased estimator (T. Gauss-Markov does not hold).**

1.1 Heteroscedasticity

Asymptotic properties:

- **Consistency:** if the observations are iid, and assumptions (a) and (c) hold, then the OLS estimator is consistent.

(a) $E(u_t|X_t) = 0 \Leftrightarrow E(Y_t|X_t) = X_t'\beta$

(c) $\Sigma = E(X_t X_t')$ is positive definite

Asymptotic normality: if the observations are iid, and the assumptions (a) and (c) hold, then

$$\sqrt{T}(\widehat{\beta} - \beta) \rightarrow_d N(0, \Sigma^{-1}\left(\sigma^2\Psi\right)\Sigma^{-1}) \qquad \Psi = E(\omega(X_t)X_t X_t')$$

The OLS estimator **is not asymptotically efficient**, since the assumption of homoscedasticity is not verified.

1.1 Heteroscedasticity

TYPE OF HETEROSCEDASTICITY: possible formulations and examples

$$Var(u_t) = \sigma^2 X_{jt}^{\alpha}$$
$$Var(u_t) = (\alpha' Z_t)^2$$
$$Var(u_t) = \exp(\alpha' Z_t)$$
$$Var(u_t) = (\alpha' Z_t)$$

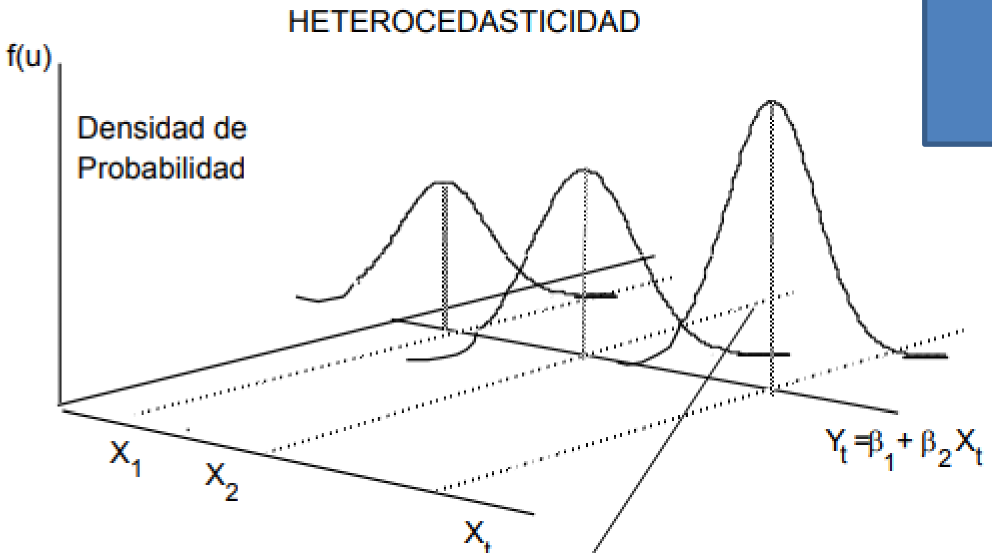
X_J is a model regressor
Z_t is a onecolumn vector of observed variables

$$\sigma_t^2 = \sigma^2 + \alpha_1 Y_t \quad \text{with} \quad \alpha_1 > 0$$
$$\sigma_t^2 = Z'_t \, \alpha_0 \quad \text{with} \quad \alpha_0 > 0$$
$$\sigma_t^2 = (\sigma + Z_t' \alpha)^2 \quad \text{with} \quad \alpha > 0$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2$$
$$\sigma_t^2 = \sum_{j=1}^{k} \alpha_j I_{jt} \quad \text{with} \quad \alpha_j > 0$$

1.1 Heteroscedasticity

EXAMPLE I:

Graphically

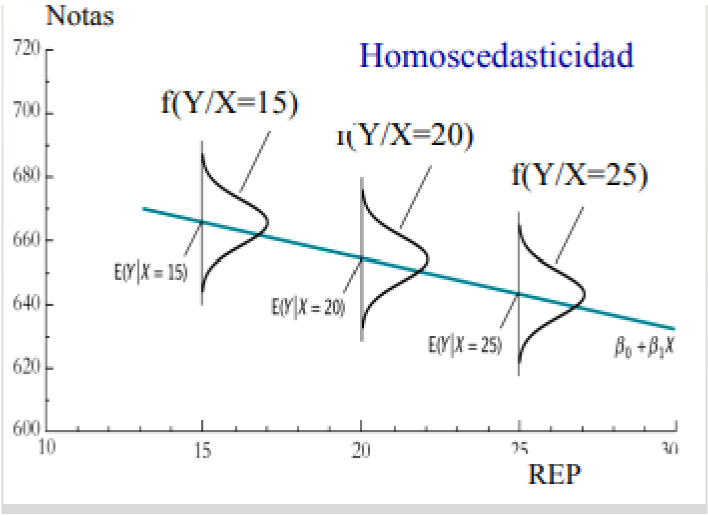


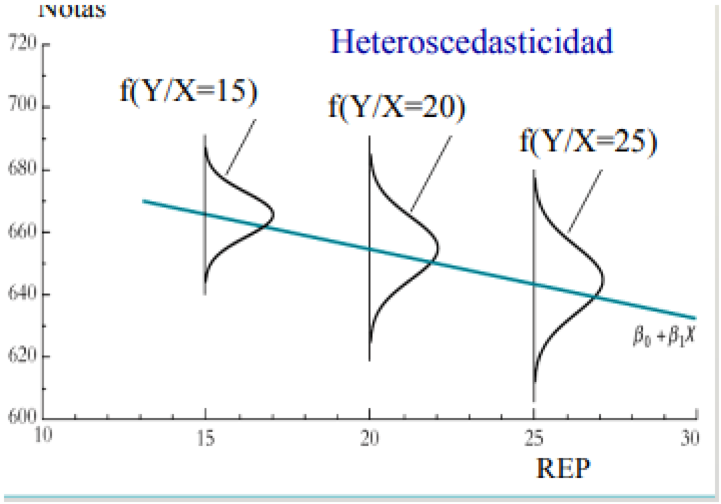
In this case, the variance decreases as the value of the independent variable increases

1.1 Heteroscedasticity

EXAMPLE II:

Graphically





1.1 Heteroscedasticity

EXAMPLE III: Given the regression

$$Y_{ij} = \beta_1 + \beta_2 X_{ij} + u_{ij} \quad i = 1, \ldots, N;\ j = 1, \ldots, n_i$$

where: $E(u_{ij}) = 0,\ E(u_{ij}^2) = \sigma_u^2\ y\ E(u_{ij}u_{kl}) = 0 \forall i \neq k, j \neq l,$

We can show that: $\bar{Y}_{i.} = \beta_1 + \beta_2 \bar{X}_{i.} + \bar{u}_{i.} \quad i = 1, \ldots, N$ where

$$\bar{Y}_{i.} = \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i} \quad \bar{X}_{i.} = \frac{\sum_{j=1}^{n_i} X_{ij}}{n_i} \quad \bar{u}_{i.} = \frac{\sum_{j=1}^{n_i} u_{ij}}{n_i}$$

presents heteroscedasticity. Let's calculate

$E(\bar{u}_{i.})$ $E(\bar{u}_{i.})^2$ $E(\bar{u}_{i.}\bar{u}_{k.})$

1.1 Heteroscedasticity

EXAMPLE III: Given the regression

$$Y_{ij} = \beta_1 + \beta_2 X_{ij} + u_{ij} \quad i = 1, \ldots, N;\ j = 1, \ldots, n_i$$

where: $E(u_{ij}) = 0,\ E(u_{ij}^2) = \sigma_u^2\ y\ E(u_{ij}u_{kl}) = 0\, \forall i \neq k, j \neq l,$

$$E(\bar{u}_{i.}) = E\left(\frac{\sum_{j=1}^{n_i} u_{ij}}{n_i}\right) = \frac{\sum_{j=1}^{n_i} E(u_{ij})}{n_i} = 0$$

$$E(\bar{u}_{i.})^2 = E\left(\frac{\sum_{j=1}^{n_i} u_{ij}}{n_i}\right)^2 = E\left(\frac{\sum_{j=1}^{n_i} u_{ij}^2 + \sum_{j\neq l}^{n_i} u_{ij}u_{il}}{n_i^2}\right) = \frac{\sum_{j=1}^{n_i} E(u_{ij}^2) + \sum_{j\neq l}^{n_i} E(u_{ij}u_{il})}{n_i^2} = \frac{\sigma_u^2}{n_i}$$

$$E(\bar{u}_{i.}\bar{u}_{k.}) = E\left(\frac{\sum_{j=1}^{n_i} u_{ij} \sum_{l=1}^{n_k} u_{kl}}{n_i^2}\right) = E\left(\frac{\sum_{j=1}^{n_i} \sum_{l=1}^{n_k} u_{ij}u_{kl}}{n_i^2}\right) = \frac{\sum_{j=1}^{n_i} \sum_{l=1}^{n_k} E(u_{ij}u_{kl})}{n_i^2} = 0$$

- We see that $E(\bar{u}_{i.})^2 = \sigma_u^2/n_i.$ is not constant.
- Variance-covariance matrix?

1.1 Heteroscedasticity

EXAMPLE IV:

Given the following model $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i, \quad u_i \sim N(0, \sigma_u^2)$

Suppose by mistake we specify the model $Y_i = \beta_1 + \beta_2 X_{2i} + e_i$

Therefore $e_i = \beta_3 X_{3i} + u_i$ presents heteroscedasticity since

$$E(e_i)^2 = \sigma_u^2 + \beta_3^2 X_{3i}^2$$

Variance-covariance matrix?

1.1 Heteroscedasticity

Example I:

Given the following model, consumption (Y_i) and income (X_i)

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad (i = 1, \ldots, n)$$

being

$$Var(ui) \equiv \sigma^2{}_i = \alpha_1 + \alpha_2 Xi$$

Variance-covariance matrix?

1.1 Heteroscedasticity

A. HETEROSCEDASTICITY DETECTION:

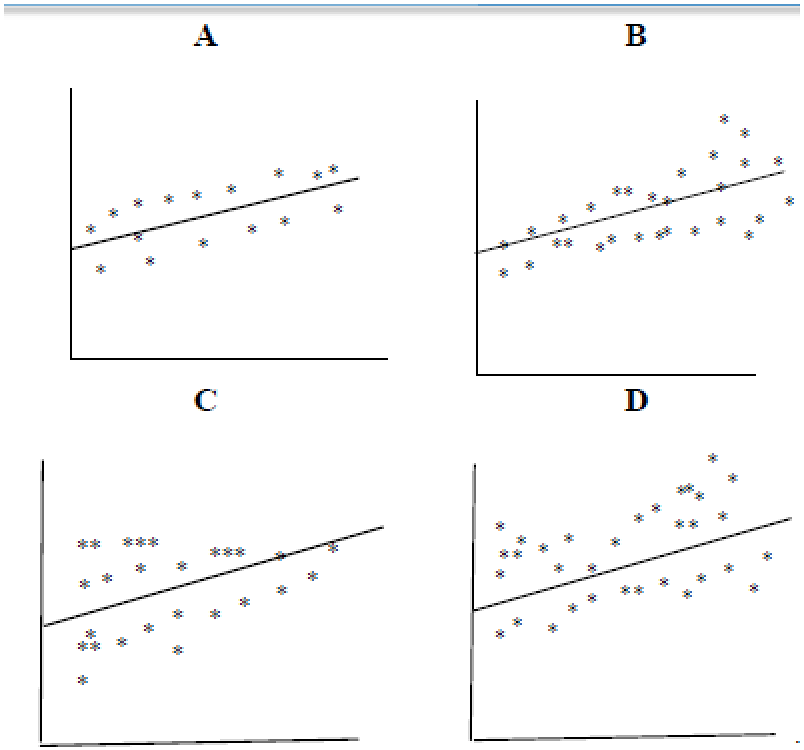
- Graphically:
  - ✓ Graph of the variables
  - ✓ Graph of the residuals
  - ✓ Graph of the squared residuals

- Hypothesis test
  - ✓ Golfeld-Quandt test. GQ
  - ✓ Breusch-Pagan test. BP
  - ✓ White test.

1.1 Heteroscedasticity

A. HETEROSCEDASTICITY DETECTION: Graphically:

- ✓ Graph of the variables



CASE A: homoscedastic

CASE B: increasing heteroscedasticity

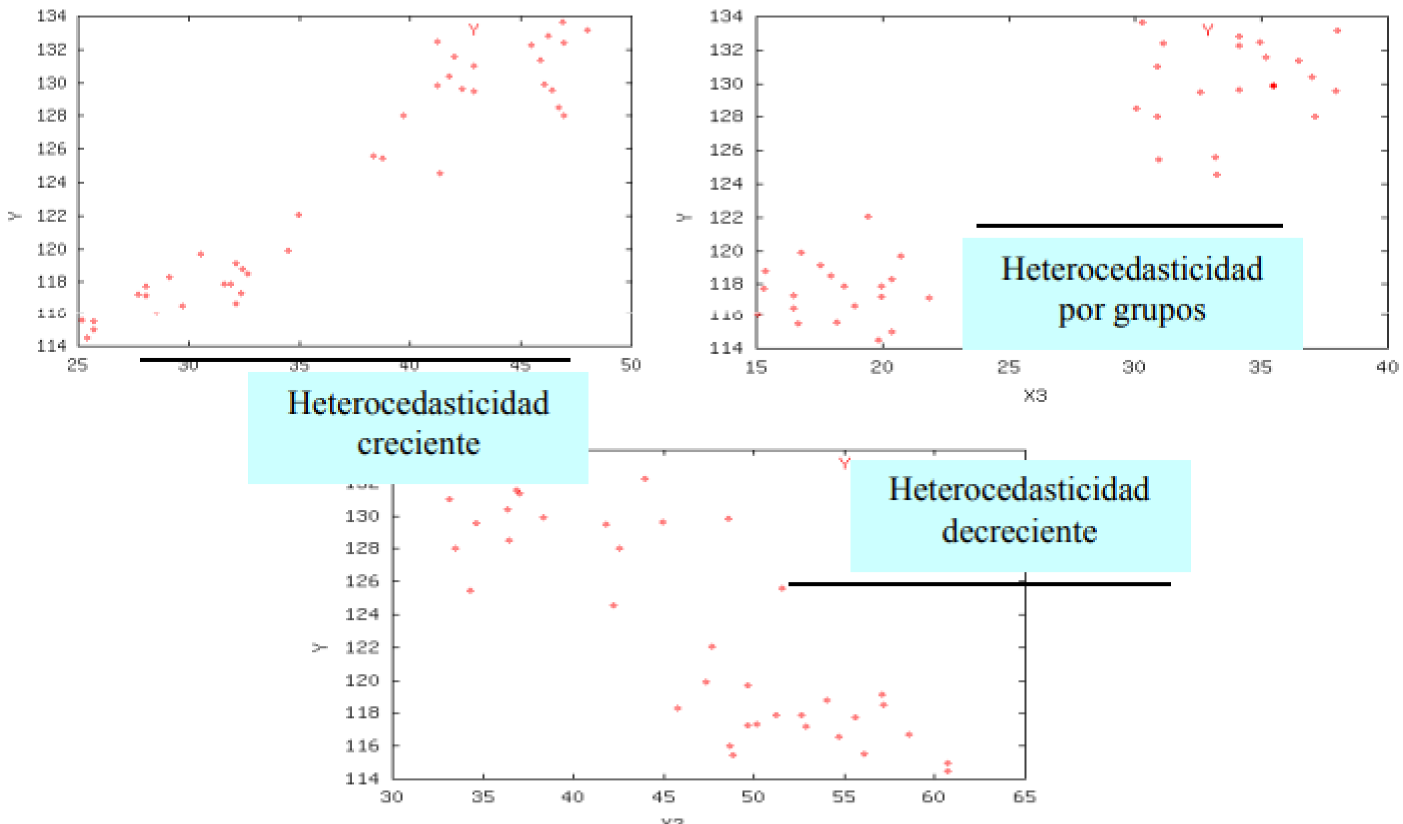
CASE C: decreasing heteroscedasticity

CASE D: squared heterocedasticity, it increases while stepping away from the mean on both sides.

1.1 Heteroscedasticity

A. HETEROSCEDASTICITY DETECTION: Graphically:

✓ With respect to independent variables (X_i):

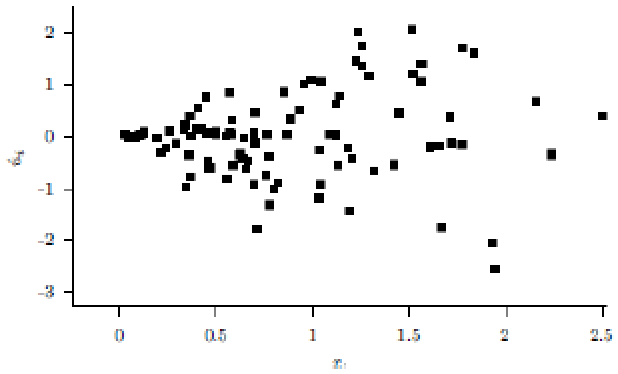


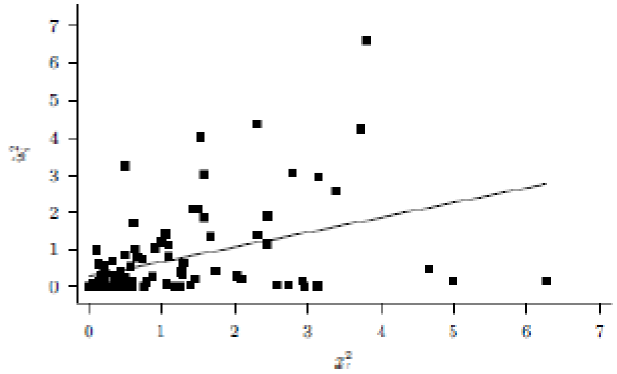
1.1 Heteroscedasticity

A. HETEROSCEDASTICITY DETECTION: Graphically:

✓ Residuals graph:

Scatter plot of $\hat{u}_i$ vs x_i and of $\hat{u}_i{}^2$ vs x_i^2

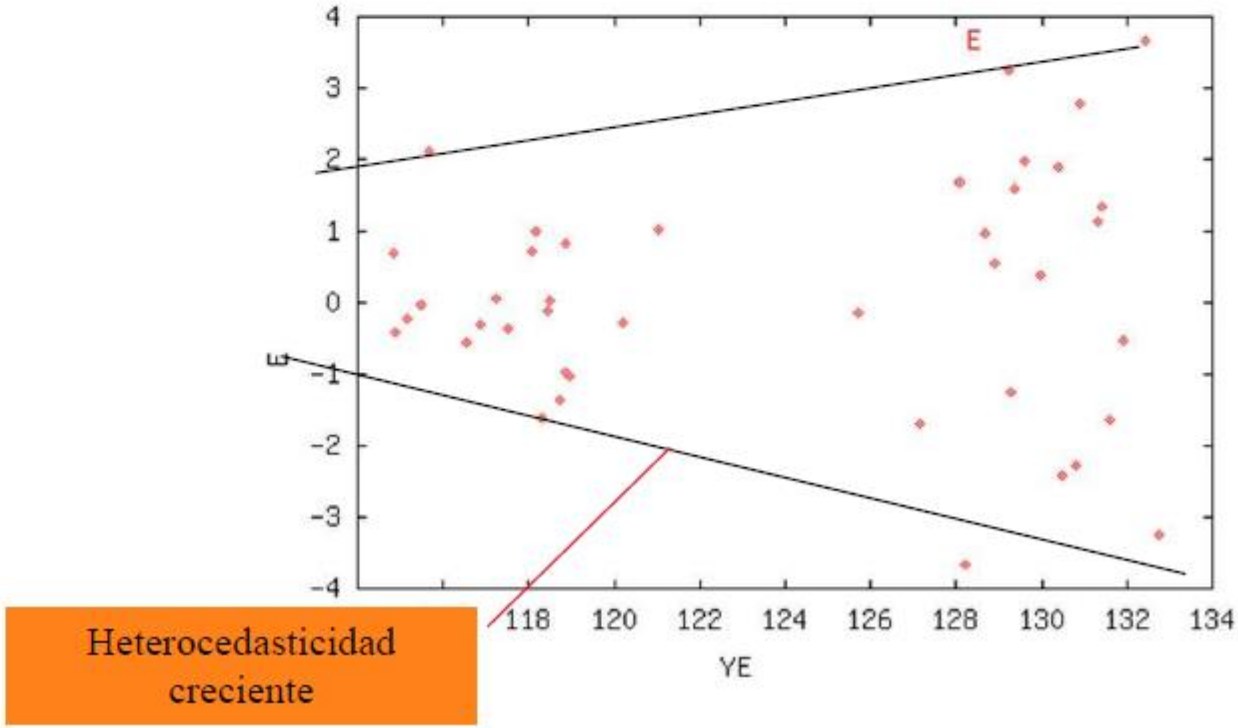




1.1 Heteroscedasticity

A. HETEROSCEDASTICITY DETECTION: Graphically:

✓ Graph of the residuals vs y_i

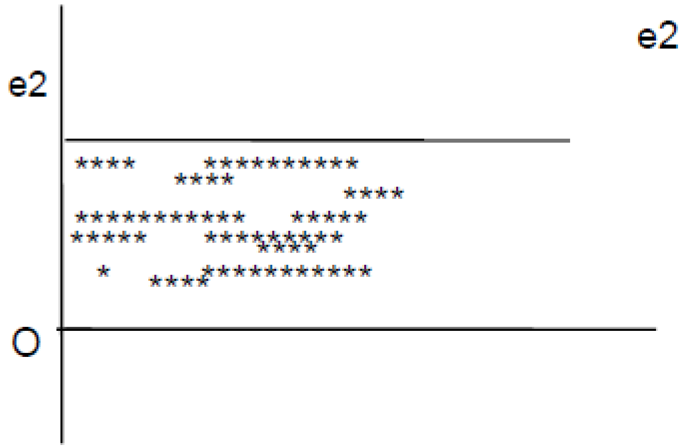


1.1 Heteroscedasticity

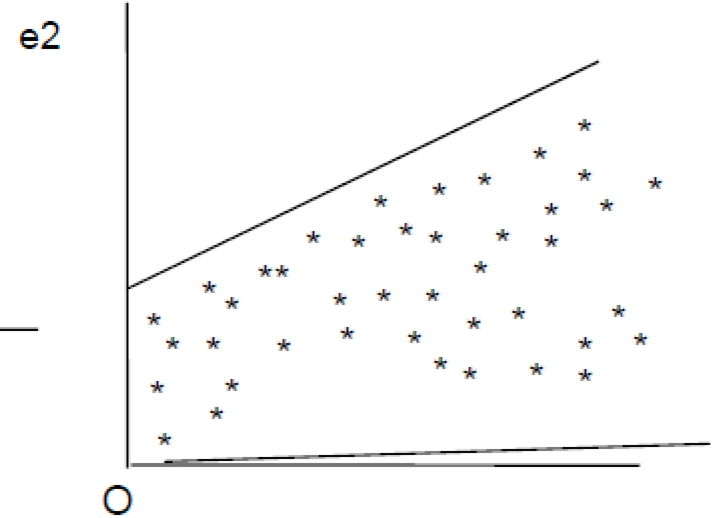
A. HETEROSCEDASTICITY DETECTION:Graphically:

- ✓ Graph of squared residuals

Hhomoscedasticity



Linear Heteroscedasticity



1.1 Heteroscedasticity

A. HETEROSCEDASTICITY DETECTION:

- Hypothesis testing
  - ✓ Golfeld-Quandt test. G-Q
  - ✓ Breusch-Pagan test. BP
  - ✓ White test.

Consider the model:

$$y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i, \quad i = 1, \ldots, n$$

1.1 Heteroscedasticity

A. HETEROSCEDASTICITY DETECTION: hypothesis testing

- ✓ Golfeld-Quandt test. G-Q

The test of Goldfeld-Quandt should be applied when we suspect that the variation of the error increases with the values of a known variable Z.

$$\sigma_i^2 = \sigma_u^2 Z_i$$

The test would be:

$$H_0 : \sigma_i^2 = \sigma_u^2$$

$$H_1 : \sigma_i^2 = \sigma_u^2 Z_i \qquad i = 2, \ldots, p$$

H_{or}:Homoscedasticity
H_1:Heteroscedasticity

1.1 Heteroscedasticity

A. HETEROSCEDASTICITY DETECTION: hypothesis testing

✓ **Golfeld-Quandt test. G-Q**

Steps of the test:

1. Identifying the variable causing heteroscedasticity, let's say Z.
2. Order the table of data by increasing values of Z
3. Split the table of data in subsamples. The central subsample with ***m*** observations, and the others two subsamples with ***(n – m)/2*** observations. Practically $m \approx n/3$
4. Skip the *m* central observations, and estimate by OLS the regression in each subsample.
5. Calculate the RSS_1 and RSS_2. We reject H0 if $RSS_2 > RSS_1$
6. Calculate the test statistics:

$$F = \frac{SCR_2}{SCR_1} \sim F_{\frac{n-m}{2}-k,\frac{n-m}{2}-k}$$

7. We reject H0 if $F > F_{\frac{n-m}{2}-k,\frac{n-m}{2}-k,\alpha}$

1.1 Heteroscedasticity

A. HETEROSCEDASTICITY DETECTION: hypothesis testing

✓ **Breush-Pagan test, BP**

Consider the following model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

With linear heteroscedasticity: $E(u^2) = \sigma^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \ldots + \delta_k x_k$

Which originates from the model $u^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \ldots + \delta_k x_k + v$

Where $E(v/X) = 0$

<u>Hypothesis to be tested:</u>

$H_0: \delta_1 = \delta_2 = \ldots = \delta_k = 0$ — H_{or}:Homoscedasticity

$H_1: \delta_i \neq 0$ *for any i=1,…,k* — H_1:Heteroscedasticity

1.1 Heteroscedasticity

A. HETEROSCEDASTICITY DETECTION: hypothesis testing

✓ **Breush-Pagan test, BP**

- We use the residuals to estimate the errors in the last equation:

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \ldots + \delta_k x_k + error$$

- let's make oan *F-test* of joint significance of the parameters in this model. The statistics in this case is:

$$F = \frac{R^2_{\hat{u}^2}/k}{(1 - R^2_{\hat{u}^2})/(n - k - 1)} \quad \sim \quad F_{k,\, n-k-1}$$

- We reject H_0 if $F > F_{k,\, n-k-1,\alpha}$
- Alternatively *LM* equivalent test (blackboard)

HETEROSCEDASTICITY DETECTION: hypothesis testing

✓ **Breush-Pagan test, BP**

Steps:

1.Estimate by OLS the equation

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

2.Get the residuals $\hat{u}_i$ and estimate by OLS the model

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \ldots + \delta_k x_k + error$$

3. Calculate the F-statistic

$$F = \frac{R^2_{\hat{u}^2}/k}{(1 - R^2_{\hat{u}^2})/(n - k - 1)} \sim F_{k,\, n-k-1}$$

4. We reject H_0 if $F > F_{k,\, n-k-1, a}$

1.1 Heteroscedasticity

HETEROSCEDASTICITY DETECTION: hypothesis contrast

- ✓ **White test**

Consider the following model: $y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i, \quad i = 1, \ldots, n$

Hypothesis to be tested $H_0 : E(u_i^2) = \sigma_u^2$

$H_1 : E(u_i^2) = \sigma_i^2$

H_{or}:Homoscedasticity

H_1:Heteroscedasticity

<u>Steps:</u>

1. Estimate by OLS and get the residuals $\hat{u}_i$

$$y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i, \quad i = 1, \ldots, n$$

2. Estimate by OLS, the following auxiliary regression and calculate R^2

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2_i} X_{3i} + e_i$$

1.1 Heteroscedasticity

HETEROSCEDASTICITY DETECTION: hypothesis testing

- ✓ **Test White**

3. Calculate test statistic:

$$LM = nR^2$$

distributed as a $X^2_{(p-1)}$, (i.e. auxiliary regression degrees of freedom)

4. We reject H_0 if $LM > X^2_{(p-1),\alpha}$

1.1 Heteroscedasticity

EXAMPLES: hypothesis testing

- ✓ **Test GQ**

Let B be the following model $y_i=\beta_1+\beta_2 V_i+u_i$ i=1,...,20

The estimate for OLS is:

Modelo 1: MCO, usando las observaciones 1-20
Variable dependiente: B

	Coeficiente	*Desv. Típica*	*Estadístico t*	*Valor p*	
const	10,2229	1,35823	7,5267	<0,00001	***
V	0,0638456	0,0112223	5,6892	0,00002	***

Media de la vble. dep.	17,64500	D.T. de la vble. dep.	2,751550
Suma de cuad. residuos	51,40902	D.T. de la regresión	1,689987

R-cuadrado	0,642619	R-cuadrado corregido	0,622765
F(1, 18)	32,36647	Valor p (de F)	0,000021
Log-verosimilitud	-37,81959	Criterio de Akaike	79,63917
Criterio de Schwarz	81,63063	Crit. de Hannan-Quinn	80,02792

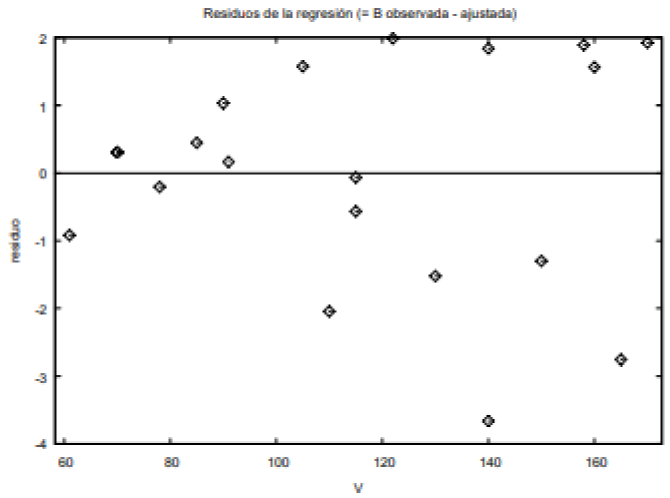
1.1 Heteroscedasticity

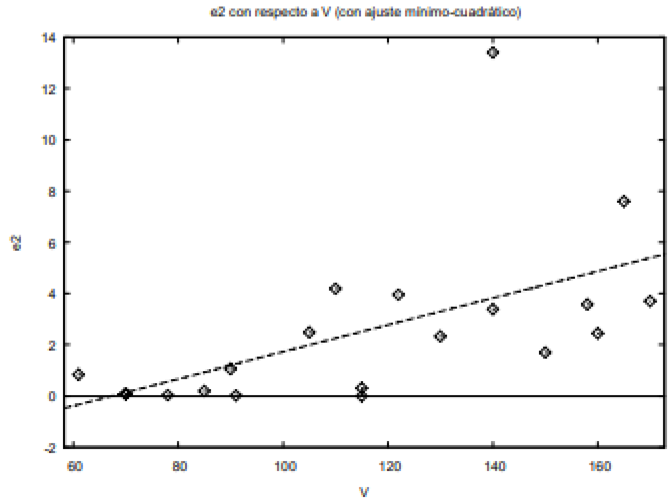
EXAMPLES: hypothesis testing

✓ **Test GQ**

Heteroscedasticity detection:

- Graphically:





We note that if sales increase, the dispersion of the residuals also increases, indicating the presence of heteroscedasticity

1.1 Heteroscedasticity

EXAMPLES: hypothesis testing

✓ **TestGQ**

Heteroscedasticity detection:

- GQ test:

$$H_0: \sigma_i^2 = \sigma_u^2$$

$$H_1: \sigma_i^2 = \sigma_u^2 Z_i$$

H_0:Homoscedasticity

H_1:Heteroscedasticity

1. We sort the sample in ascending order by the variable *Vi*
2. We divide the sample into three parts, 20/3≅ 6
3. We estimate by OLS the 1st and 3rd sample
4. We calculate the test statistic

   GQ= SQR2/SQR1~F_α(n-q)

   If α =5%; F(5.5)=5.05033

   Therefore 29.043>5.05033 We reject H_0

$$GQ = \frac{SCR2}{SCR1} = \frac{32{,}935}{1{,}134} = 29{,}043$$

1.1 Heteroscedasticity

EXAMPLES: hypothesis testing

✓ **TestGQ**

1st subsample

Modelo 1: MCO, usando las observaciones 1-7
Variable dependiente: B

	Coeficiente	*Desv. Típica*	*Estadístico t*	*Valor p*	
const	7,5343	1,34079	5,6193	0,00247	***
V	0,100477	0,0170653	5,8878	0,00201	***

Media de la vble. dep.	15,35714	D.T. de la vble. dep.	1,224550
Suma de cuad. residuos	1,134109	D.T. de la regresión	0,476258
R-cuadrado	0,873948	R-cuadrado corregido	0,848738
F(1, 5)	34,66614	Valor p (de F)	0,002009
Log-verosimilitud	-3,562349	Criterio de Akaike	11,12470
Criterio de Schwarz	11,01652	Crit. de Hannan-Quinn	9,787616

1.1 Heteroscedasticity

EXAMPLES: hypothesis testing

✓ **TestGQ**

3rd subsample

Modelo 2: MCO, usando las observaciones 14-20 (n = 7)
Variable dependiente: B

	Coeficiente	*Desv. Típica*	*Estadístico t*	*Valor p*
const	1,15735	13,7885	0,0839	0,93636
V	0,121975	0,0889012	1,3720	0,22841

Media de la vble. dep.	20,02857	D.T. de la vble. dep.	2,748766
Suma de cuad. residuos	32,93469	D.T. de la regresión	2,566503
R-cuadrado	0,273515	R-cuadrado corregido	0,128218
F(1, 5)	1,882452	Valor p (de F)	0,228412
Log-verosimilitud	-15,35273	Criterio de Akaike	34,70545
Criterio de Schwarz	34,59727	Crit. de Hannan-Quinn	33,36837

1.1 Heteroscedasticity

EXAMPLES: hypothesis testing

- ✓ **White test**

Consider a model where you want to study the relationship between employment and GDP. GDP=$\beta_1 + \beta_2 empl_i + u_i$

Test:

H_0:homecedasticity

H_1:heteroscedasticity

Test statistic: $\chi^2_{\text{exp}} = nR^2 \rightarrow \chi^2_{p-1}$

Where R2 is the coefficient of determination obtained from an auxiliary regression of the squared residuals on the explanatory variables, their squares and their cross products.

1.1 Heteroscedasticity

EXAMPLES: hypothesis testing

✓ **White test**

We get:

White Heteroskedasticity Test:

F-statistic	3.698591	Probability	0.049458
Obs*R-squared	**5.944912**	Probability	**0.051177**

Test Equation:
LS // Dependent Variable is RESID^2
Sample: 1 18
Included observations: 18

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.03E+11	2.39E+11	-1.266662	0.2246
EMPLEO	1.44E+09	6.56E+08	2.195207	0.0443
EMPLEO^2	-512657.3	302565.4	-1.694368	0.1109

R-squared	0.330273	Mean dependent var	2.52E+11
Adjusted R-squared	0.240976	S.D. dependent var	5.29E+11
S.E. of regression	4.61E+11	Akaike info criter	53.86446
Sum squared resid	3.19E+24	Schwarz criterion	54.01286
Log likelihood	-507.3210	F-statistic	3.698591
Durbin-Watson stat	1.230263	Prob(F-statistic)	0.049458

EXAMPLES: hypothesis testing

- ✓ **White test**

Test statistic: $NR^2=18x0.330273=5.944$

So:

$No^2 > X^2$ (p-1)=0.011778. We reject H_0

B. HETEROSCEDASTICITY TREATMENT

SOLUTIONS:

- Estimate by OLS and estimate an appropriate matrix of variance-covariance
- Estimate by Generalized Least Squares: GLS
- Estimate by Feasible Generalized Least Squares: FGLS

1.1 Heteroscedasticity

B. HETEROSCEDASTICITY TREATMENT

- **A first solution to the problem:** use OLS and estimate a variance-covariance matrix robust to HET

- When the errors are heteroscedastics, the matrix $\widehat{var_w}(\hat{\beta}_{OLS})$ is an appropriate estimator for the variance/covariance matrix of $\hat{\beta}_{OLS}$. It is called White's estimator, and allows us to make valid inference once we have estimated the model by OLS without the need to specify the type of heteroscedasticity.

$$\widehat{Var_W(\hat{\beta}_{MCO})} = (X'X)^{-1} \; \widehat{\Sigma_V} \; (X'X)^{-1}$$

Where $\widehat{\Sigma_V} = \frac{1}{n}\sum X_i X_i' \hat{u}_i^2$ is the estimator of $\Sigma_V \equiv Cov(X_i u_i) = E(X_i X_i' u_i^2)$

- This estimator makes sense,because,with the assumptions 2-4, OLS is consistent to β, thus $\widehat{u_i}$ will converge in probability to u_i.

1.1 Heteroscedasticity

B. HETEROSCEDASTICITY TREATMENT

ALTERNATIVE ESTIMATION METHODS:

1. **Generalized Least Squares: GLS**
2. **Feasible Generalized Least Squares: FGLS**

Given that: $Y = X\beta + \upsilon$ where υ is distributed as $N(0, \Omega)$

$$\Omega = diag\,(\sigma_1^2 \ldots\ldots \sigma_T^2) = \begin{pmatrix} \sigma_1^2 & \ldots & 0 \\ : & : & : \\ 0 & \ldots & \sigma_T^2 \end{pmatrix}$$

1.1 Heteroscedasticity

B. HETEROSCEDASTICITY TREATMENT

1. MQG:

If the matrix Ω is known, the transformation to be applied is:

$$P = \Omega^{-1/2} = \begin{pmatrix} 1/\sigma_1 & 0.... & 0 \\ .... & ... & ... \\ 0 & ..... & 1/\sigma_T \end{pmatrix}$$

where P′P= Ω^{-1}

And the transformation is:

$$\begin{cases} y_t^* = y_t / \sigma_t & t = 1...T \\ x_{jt}^* = x_{jt} / \sigma_t & t = 1...T \quad j = 0...j \end{cases}$$

being $\widehat{\beta}_{MCG} = (X^* X^{*\prime})^{-1} X^* Y^{\prime *}$ or $\hat{\beta}_{MCG} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$

1.1 Heteroscedasticity

GLS, a case specific of known heteroscedasticity: <u>multiplicative constant</u>

Given $Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i, \quad i = 1, \ldots, n$

where $E(u_i) = 0, \; E(u_i^2) = \sigma_u^2 \omega_i \text{ y } E(u_i u_j) = 0 \forall i \neq j,$

Heteroscedastic model, we estimate the model by GLS, $\hat{\beta}_{MCG} = (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{y}$

being $V(\hat{\beta}_{MCG}) = \sigma_u^2 (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}$ $\quad u_i^* = \dfrac{u_i}{\sqrt{\omega_i}}$

Then:

1. $$E(u_i^*) = E\left(\frac{u_i}{\sqrt{\omega_i}}\right) = \frac{E(u_i)}{\sqrt{\omega_i}} = 0$$

2. $$E({u_i^*}^2) = E(\frac{u_i^2}{\omega_i}) = \frac{E(u_i^2)}{\omega_i} = \frac{\sigma_u^2 \omega_i}{\omega_i} = \sigma_u^2$$

3. $$E(u_i^* u_j^*) = E\left(\frac{u_i u_j}{\sqrt{\omega_i}\sqrt{\omega_j}}\right) = \frac{E(u_i u_j)}{\omega_i \omega_j} = 0 \quad \forall i \neq j$$

u_i^* meets the basic assumptions

1.1 Heteroscedasticity

GLS, a case specific of known heteroscedasticity: <u>multiplicative constant</u>

Given $Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i, \quad i = 1, \ldots, n$

the transformed model is:

$$\frac{Y_i}{\sqrt{\omega_i}} = \beta_1 \frac{1}{\sqrt{\omega_i}} + \beta_2 \frac{X_{2i}}{\sqrt{\omega_i}} + \cdots + \beta_k \frac{X_{ki}}{\sqrt{\omega_i}} + \frac{u_i}{\sqrt{\omega_i}}, \quad i = 1, \ldots, n$$

$$Y_i^* = \beta_1 X_{1i}^* + \beta_2 X_{2i}^* + \cdots + \beta_k X_{ki}^* + u_i^*, \quad i = 1, \ldots, n$$

If $\Omega^{-1} = \mathbf{P}'\mathbf{P} \longrightarrow \hat{\beta}_{MCG} = (\mathbf{X}'\mathbf{P}'\mathbf{P}\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}'\mathbf{P}\mathbf{y}$

1.1 Heteroscedasticity

GLS, a case specific of known heteroscedasticity: <u>multiplicative constant</u>

where

$$\mathbf{P} = \begin{pmatrix} 1/\sqrt{\omega_1} & 0 & \dots & 0 \\ 0 & 1/\sqrt{\omega_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \dots & 1/\sqrt{\omega_n} \end{pmatrix}$$

and the transformed data is:

$$\mathbf{Py} = \begin{pmatrix} y_1/\sqrt{\omega_1} \\ y_2/\sqrt{\omega_2} \\ \vdots \\ y_n/\sqrt{\omega_n} \end{pmatrix} \quad \text{y} \quad \mathbf{PX} = \begin{pmatrix} 1/\sqrt{\omega_1} & X_{21}/\sqrt{\omega_1} & \dots & X_{k1}/\sqrt{\omega_1} \\ 1/\sqrt{\omega_2} & X_{22}/\sqrt{\omega_2} & \dots & X_{k2}/\sqrt{\omega_2} \\ \vdots & \vdots & \ddots & \\ 1/\sqrt{\omega_n} & X_{2n}/\sqrt{\omega_n} & \dots & X_{kn}/\sqrt{\omega_n} \end{pmatrix}$$

ECONOMETRICS II

1.1 Heteroscedasticity

EXAMPLE:

1. Consider the simple regression with heteroscedasticity $Y_i = \beta_1 + \beta_2 X_i + u_i$
where $E(u_i) = 0,\ E(u_i^2) = \sigma_u^2 X_i^2\ \ y\ E(u_i u_j) = 0 \forall i \neq j.$

Write the transformed model.

2. Consider the simple regression with heteroscedasticity $Y_i = \beta_1 + \beta_2 X_i + u_i$
where $E(u_i) = 0,\ E(u_i^2) = \sigma_u^2 e^{X_i}\ \ y\ E(u_i u_j) = 0 \forall i \neq j.$

Write the transformed model.

1.1 Heteroscedasticity

HETEROSCEDASTICITY TREATMENT

1. FGLS:

If the matrix Ω is unknown, the estimation process needs an additional step. We want to estimate

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i, \quad i = 1, \ldots, n$$

And we don't know the form of heteroscedasticity. If we suspect that:

$$\sigma_i^2 = \alpha_1 + \alpha_2 Z_{2i} + \cdots + \alpha_p Z_{pi} \qquad (1)$$

Since we don't know the $\sigma_i{}^2$ we need to estimate the following regression by OLS:

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 Z_{2i} + \cdots + \alpha_p Z_{pi} + e_i \qquad (2)$$

We need to get the $\widehat{\alpha}_i$ and substitute them in (1), and we obtain $\widehat{\sigma}_i{}^2$

$$\hat{\sigma}_i^2 = \hat{\alpha}_1 + \hat{\alpha}_2 Z_{2i} + \cdots + \hat{\alpha}_p Z_{pi}$$

HETEROSCEDASTICITY TREATMENT

1. MQGF:

Once we get $\widehat{\sigma}_i{}^2$, the original model can be transformed in this way

$$\frac{Y_i}{\hat{\sigma}_i} = \beta_1 \frac{1}{\hat{\sigma}_i} + \beta_2 \frac{X_{2i}}{\hat{\sigma}_i} + \cdots + \beta_k \frac{X_{ki}}{\hat{\sigma}_i} + \frac{u_i}{\hat{\sigma}_i}, \quad i = 1, \ldots, n$$

We estimate by OLS the transformed model and obtain:

$$\widehat{\beta}_{MQGF} = (X'\widehat{\Omega}^{-1}X)^{-1}X'\widehat{\Omega}^{-1}y$$