Chapter 5: The Generalized Linear Regression Model and Heteroscedasticity

Advanced Econometrics - HEC Lausanne

Christophe Hurlin

University of Orléans

December 15, 2013

Section 1

Introduction

1. Introduction

The outline of this chapter is the following:

Section 2. The generalized linear regression model

Section 3. Inefficiency of the Ordinary Least Squares

Section 4. Generalized Least Squares (GLS)

Section 5. Heteroscedasticity

Section 6. Testing for heteroscedasticity

1. Introduction

References

- Amemiya T. (1985), Advanced Econometrics. Harvard University Press.
- Greene W. (2007), Econometric Analysis, sixth edition, Pearson - Prentice Hil (**recommended**)
- Pelgrin, F. (2010), Lecture notes Advanced Econometrics, HEC Lausanne (**a special thank**)
- Ruud P., (2000) An introduction to Classical Econometric Theory, Oxford University Press.

1. Introduction

Notations: In this chapter, I will (try to...) follow some conventions of notation.

$f_Y(y)$	probability density or mass function
$F_Y(y)$	cumulative distribution function
$\Pr()$	probability
$\mathbf{y}$	vector
$\mathbf{Y}$	matrix

Be careful: in this chapter, I don't distinguish between a random vector (matrix) and a vector (matrix) of deterministic elements (except in section 2). For more appropriate notations, see:

Abadir and Magnus (2002), Notation in econometrics: a proposal for a standard, Econometrics Journal.

Section 2

The generalized linear regression model

2. The generalized linear regression model

Objectives

The objective of this section are the following:

1. Define the generalized linear regression model
2. Define the concept of heteroscedasticity
3. Define the concept of autocorrelation (or correlation) of disturbances

2. The generalized linear regression model

Consider the (population) **multiple linear regression model**:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon$$

where (cf. chapter 3):

- $\mathbf{y}$ is a $N \times 1$ vector of observations y_i for $i = 1, .., N$
- $\mathbf{X}$ is a $N \times K$ matrix of K explicative variables $\mathbf{x}_{ik}$ for $k = 1, ., K$ and $i = 1, .., N$
- $\boldsymbol{\varepsilon}$ is a $N \times 1$ vector of error terms ε_i.
- $\boldsymbol{\beta} = (\beta_1 .. \beta_K)^\top$ is a $K \times 1$ vector of parameters

2. The generalized linear regression model

In chapter 3 (linear regression model), we assume **spherical disturbances** (assumption A4):

$$\mathbb{V}\left(\left.\boldsymbol{\varepsilon}\right|\mathbf{X}\right)=\sigma^{2}\mathbf{I}_{N}$$

In this chapter, we will **relax** the assumption that the errors are independent and/or identically distributed and we will study:

1. Heteroscedasticity
2. Autocorrelation or correlation.

2. The generalized linear regression model

Definition (Generalized linear regression model)

The **generalized** linear regression model is defined as to be:

$$\mathbf{y} = \mathbf{X}\beta + \boldsymbol{\varepsilon}$$

where $\mathbf{X}$ is a matrix of fixed or random regressors, $\boldsymbol{\beta} \in \mathbb{R}^K$, and the error term $\boldsymbol{\varepsilon}$ satisfies:

$$\mathbb{E}\left(\left.\boldsymbol{\varepsilon}\right|\mathbf{X}\right) = \mathbf{0}_{N\times 1}$$

$$\mathbb{V}\left(\left.\boldsymbol{\varepsilon}\right|\mathbf{X}\right) = \boldsymbol{\Sigma} = \sigma^2\boldsymbol{\Omega}$$

where $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$ are symmetric positive definite matrices.

2. The generalized linear regression model

Reminder

$$\underbrace{\mathbb{V}\left(\varepsilon \middle| \mathbf{X}\right)}_{N\times N} = \underbrace{\mathbb{E}\left(\varepsilon\varepsilon^{\top} \middle| \mathbf{X}\right)}_{N\times N}$$

$$= \begin{pmatrix} \mathbb{V}\left(\varepsilon_1^2 \middle| \mathbf{X}\right) & \mathbb{C}ov\left(\varepsilon_1\varepsilon_2 \middle| \mathbf{X}\right) & .. & \mathbb{C}ov\left(\varepsilon_1\varepsilon_N \middle| \mathbf{X}\right) \\ \mathbb{E}\left(\varepsilon_2\varepsilon_1 \middle| \mathbf{X}\right) & \mathbb{V}\left(\varepsilon_2^2 \middle| \mathbf{X}\right) & .. & \mathbb{C}ov\left(\varepsilon_2\varepsilon_N \middle| \mathbf{X}\right) \\ .. & .. & .. & .. \\ \mathbb{C}ov\left(\varepsilon_N\varepsilon_1 \middle| \mathbf{X}\right) & .. & .. & \mathbb{V}\left(\varepsilon_N^2 \middle| \mathbf{X}\right) \end{pmatrix}$$

2. The generalized linear regression model

Remark

In the generalized linear regression model, we have

$$\mathbb{V}\left(\left.\boldsymbol{\varepsilon}\right|\mathbf{X}\right)=\boldsymbol{\Sigma}=\sigma^2\boldsymbol{\Omega}$$

with

$$\boldsymbol{\Sigma}=\begin{pmatrix}\sigma_1^2 & \sigma_{12} & .. & \sigma_{1N}\\ \sigma_{21} & \sigma_2^2 & .. & \sigma_{2N}\\ .. & .. & .. & ..\\ \sigma_{N1} & .. & .. & \sigma_N^2\end{pmatrix}=\sigma^2\begin{pmatrix}\omega_{11} & \omega_{12} & .. & \omega_{1N}\\ \omega_{21} & \omega_{22} & .. & \omega_{2N}\\ .. & .. & .. & ..\\ \omega_{N1} & .. & .. & \omega_{NN}\end{pmatrix}$$

and $\omega_{ij}=\sigma_{ij}/\sigma^2$.

2. The generalized linear regression model

Definition (Heteroscedasticity)

Disturbances are **heteroscedastic** when they have different (conditional) variances:

$$\mathbb{V}\left(\varepsilon_i \middle| \mathbf{X}\right) \neq \mathbb{V}\left(\varepsilon_j \middle| \mathbf{X}\right) \quad \text{for } i \neq j$$

2. The generalized linear regression model

Remarks

1. Heteroscedasticity often arises in volatile high-frequency time-series data such as daily observations in **financial markets.**

2. Heteroscedasticity often arises in **cross-section data** where the scale of the dependent variable and the explanatory power of the model tend to vary across observations. Microeconomic data such as expenditure surveys are typical

2. The generalized linear regression model

Example (Heteroscedasticity)

If the disturbances are **heteroscedastic** but they are still assumed to be uncorrelated across observations, so $\mathbf{\Omega}$ and $\mathbf{\Sigma}$ would be:

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 & .. & 0 \\ 0 & \sigma_2^2 & .. & 0 \\ .. & .. & .. & .. \\ 0 & .. & .. & \sigma_N^2 \end{pmatrix} = \sigma^2 \mathbf{\Omega} = \sigma^2 \begin{pmatrix} \omega_1 & 0 & .. & 0 \\ 0 & \omega_2 & .. & 0 \\ .. & .. & .. & .. \\ 0 & .. & .. & \omega_N \end{pmatrix}$$

with $\omega_i = \sigma_i^2 / \sigma^2$ for $i = 1, .., N$.

2. The generalized linear regression model

Definition (Autocorrelation)

Disturbances are **autocorrelated (or correlated)** when:

$$\mathbb{C}ov\left(\varepsilon_i, \varepsilon_j \middle| \mathbf{X}\right) \neq 0 \quad \text{for } i \neq j$$

2. The generalized linear regression model

Example (Autocorrelation)

For instance, **time-series data** are usually homoscedastic, but autocorrelated, so $\mathbf{\Omega}$ and $\mathbf{\Sigma}$ would be:

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma^2 & \sigma_{12} & .. & \sigma_{1N} \\ \sigma_{21} & \sigma^2 & .. & \sigma_{2N} \\ .. & .. & .. & .. \\ \sigma_{N1} & .. & .. & \sigma^2 \end{pmatrix} = \sigma^2 \mathbf{\Omega} = \sigma^2 \begin{pmatrix} 1 & \omega_{12} & .. & \omega_{1N} \\ \omega_{21} & 1 & .. & \omega_{2N} \\ .. & .. & .. & .. \\ \omega_{N1} & .. & .. & 1 \end{pmatrix}$$

with $\omega_{ij} = \sigma_{ij} / \sigma^2$ for $i = 1, .., N$ denotes the correlation (autocorrelation)

$$\omega_{ij} = \frac{\sigma_{ij}}{\sigma^2} = cor\left(\varepsilon_i, \varepsilon_j\right)$$

2. The generalized linear regression model

Key Concepts

1. The generalized linear regression model
2. Heteroscedasticity
3. Autocorrelation (or correlation) of disturbances

Section 3

Inefficiency of the Ordinary Least Squares

3. Inefficiency of the Ordinary Least Squares

Objectives

The objective of this section are the following:

1. Study the properties of the OLS estimator in the generalized linear regression model
2. Study the finite sample properties of the OLS
3. Study the asymptotic properties of the OLS
4. Introduce the concept of robust / non-robust inference

3. Inefficiency of the Ordinary Least Squares

Introduction

Assume that the data are generated by the **generalized linear regression model**:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\mathbb{E}\left(\boldsymbol{\varepsilon} \middle| \mathbf{X} \right) = \mathbf{0}_{N\times 1}$$

$$\mathbb{V}\left(\boldsymbol{\varepsilon} \middle| \mathbf{X} \right) = \sigma^2 \boldsymbol{\Omega} = \boldsymbol{\Sigma}$$

Now consider the OLS estimator, denoted $\widehat{\boldsymbol{\beta}}_{OLS}$, of the parameters $\boldsymbol{\beta}$:

$$\widehat{\boldsymbol{\beta}}_{OLS} = \left(\mathbf{X}^{\top}\mathbf{X} \right)^{-1} \mathbf{X}^{\top}\mathbf{y}$$

We will study its **finite sample** and **asymptotic properties**.

3. Inefficiency of the Ordinary Least Squares

Definition (Assumption 3: Strict exogeneity of the regressors)

The regressors are exogenous in the sense that:

$$\mathbb{E}\left(\left.\varepsilon\right|\mathbf{X}\right)=\mathbf{0}_{N\times 1}$$

Finite sample properties of the OLS estimator

3. Inefficiency of the Ordinary Least Squares

Definition (Bias)

In the generalized linear regression model, under the assumption A3 (exogeneity), the OLS estimator is **unbiased**:

$$\mathbb{E}\left(\widehat{\boldsymbol{\beta}}_{OLS}\right) = \boldsymbol{\beta}_0$$

where $\boldsymbol{\beta}_0$ denotes the true value of the parameters.

3. Inefficiency of the Ordinary Least Squares

Remark

Heteroscedasticity and/or autocorrelation **don't induce a bias** for the OLS estimator

Proof

$$\widehat{\beta}_{OLS} = \left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\left(\mathbf{X}^{\top}\mathbf{y}\right) = \beta_0 + \left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\left(\mathbf{X}^{\top}\varepsilon\right)$$

So we have:

$$\mathbb{E}\left(\widehat{\beta}_{OLS}\middle|\,\mathbf{X}\right) = \beta_0 + \left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\left(\mathbf{X}^{\top}\mathbb{E}\left(\varepsilon|\,\mathbf{X}\right)\right)$$

Under assumption A3 (exogeneity), $\mathbb{E}\left(\varepsilon|\,\mathbf{X}\right) = \mathbf{0}$. Then, we get:

$$\mathbb{E}\left(\widehat{\beta}_{OLS}\middle|\,\mathbf{X}\right) = \beta_0$$

3. Inefficiency of the Ordinary Least Squares

Proof (cont'd)

$$\mathbb{E}\left(\left.\widehat{\boldsymbol{\beta}}_{OLS}\right|\mathbf{X}\right)=\boldsymbol{\beta}_0$$

So, we have:

$$\mathbb{E}\left(\widehat{\boldsymbol{\beta}}_{OLS}\right)=\mathbb{E}_X\left(\mathbb{E}\left(\left.\widehat{\boldsymbol{\beta}}_{OLS}\right|\mathbf{X}\right)\right)=\mathbb{E}_X\left(\boldsymbol{\beta}_0\right)=\boldsymbol{\beta}_0$$

where $\mathbb{E}_X$ denotes the expectation with respect to the distribution of $\mathbf{X}$.

The OLS estimator is unbiased:

$$\mathbb{E}\left(\widehat{\boldsymbol{\beta}}_{OLS}\right)=\boldsymbol{\beta}_0 \quad \square$$

3. Inefficiency of the Ordinary Least Squares

Definition (Bias)

In the generalized linear regression model, under the assumption A3 (exogeneity), the OLS estimator has a conditional **variance covariance matrix** given by

$$\mathbb{V}\left(\widehat{\beta}_{OLS}\middle|\mathbf{X}\right)=\sigma_0^2\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\Omega\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}$$

and a variance covariance matrix given by:

$$\mathbb{V}\left(\widehat{\beta}_{OLS}\right)=\mathbb{E}_X\left(\mathbb{V}\left(\widehat{\beta}_{OLS}\middle|\mathbf{X}\right)\right)$$

3. Inefficiency of the Ordinary Least Squares

Proof

$$\widehat{\beta}_{OLS} = \left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\left(\mathbf{X}^{\top}\mathbf{y}\right) = \beta_0 + \left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\left(\mathbf{X}^{\top}\varepsilon\right)$$

So we have:

$$\begin{aligned}
\mathbb{V}\left(\widehat{\beta}_{OLS}\middle|\mathbf{X}\right) &= \mathbb{E}\left(\left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\mathbf{X}^{\top}\varepsilon\varepsilon^{\top}\mathbf{X}\left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\middle|\mathbf{X}\right) \\
&= \left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\mathbf{X}^{\top}\mathbb{E}\left(\varepsilon\varepsilon^{\top}\middle|\mathbf{X}\right)\mathbf{X}\left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1} \\
&= \sigma_0^2\left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\mathbf{X}^{\top}\Omega\mathbf{X}\left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1} \quad \square
\end{aligned}$$

3. Inefficiency of the Ordinary Least Squares

Definition (Variance estimator)

An **estimator** of the variance covariance matrix of the OLS estimator $\widehat{\beta}_{OLS}$ is given by

$$\widehat{\mathbb{V}}\left(\widehat{\beta}_{OLS}\right) = \widehat{\sigma}^2 \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \widehat{\mathbf{\Omega}} \mathbf{X} \left(\mathbf{X}^\top \mathbf{X}\right)^{-1}$$

where $\widehat{\sigma}^2 \widehat{\mathbf{\Omega}}$ is a consistent estimator of $\mathbf{\Sigma} = \sigma^2 \mathbf{\Omega}$. This estimator holds whether $\mathbf{X}$ is stochastic or non-stochastic.

3. Inefficiency of the Ordinary Least Squares

Definition (Normality assumption)

Under assumptions A3 (exogeneity) and A6 (normality), the OLS estimator obtained in the generalized linear regression model has an (exact) **normal conditional distribution:**

$$\widehat{\beta}_{OLS} \Big| \mathbf{X} \sim \mathcal{N}\left(\beta_0, \sigma^2 \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \Omega \mathbf{X} \left(\mathbf{X}^\top \mathbf{X}\right)^{-1}\right)$$

Asymptotic properties of the OLS estimator

3. Inefficiency of the Ordinary Least Squares

Assumptions

$$\text{plim}\,\frac{1}{N}\mathbf{X}^{\top}\mathbf{X} = \mathbf{Q}$$

$$\text{plim}\,\frac{1}{N}\mathbf{X}^{\top}\Omega\mathbf{X} = \mathbf{Q}^{*}$$

where:

1. $\mathbf{Q}^{*}$ is a $K \times K$ finite (non null) definite positive matrix
2. $\mathbf{Q}$ is a $K \times K$ finite (non null) definite positive matrix with

$$\text{rank}\,(\mathbf{Q}) = K$$

3. Inefficiency of the Ordinary Least Squares

Definition (Consistency of the OLS estimator)

If plim $N^{-1}\mathbf{X}^{\top}\mathbf{\Omega}\mathbf{X}$ and plim $N^{-1}\mathbf{X}^{\top}\mathbf{X}$ are both finite positive definite matrices, then $\widehat{\boldsymbol{\beta}}_{OLS}$ is a consistent estimator of $\boldsymbol{\beta}$:

$$\widehat{\boldsymbol{\beta}}_{OLS} \xrightarrow{p} \boldsymbol{\beta}_0$$

3. Inefficiency of the Ordinary Least Squares

Proof

$$\widehat{\beta}_{OLS} = \beta_0 + \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \left(\mathbf{X}^\top \varepsilon\right)$$

We know that under assumption A3 (exogeneity):

$$\text{plim} \frac{1}{N} \mathbf{X}^\top \varepsilon = \mathbf{0}_{K \times 1}$$

$$\text{plim} \frac{1}{N} \mathbf{X}^\top \mathbf{X} = \mathbf{Q}$$

So, we have

$$\text{plim } \widehat{\beta}_{OLS} = \beta_0$$

So, the estimator $\widehat{\beta}$ is consistent. $\square$

3. Inefficiency of the Ordinary Least Squares

Definition (Asymptotic distribution of the OLS)

If the regressors are sufficiently well behaved and the off-diagonal terms in diminish sufficiently rapidly, then the least squares estimator is asymptotically normally distributed with

$$\sqrt{N}\left(\widehat{\beta}_{OLS}-\beta_0\right)\xrightarrow{d}\mathcal{N}\left(\mathbf{0},\sigma^2\mathbf{Q}^{-1}\mathbf{Q}^{*}\mathbf{Q}^{-1}\right)$$

where

$$\mathbf{Q}=\text{plim}\frac{1}{N}\mathbf{X}^{\top}\mathbf{X}\qquad\mathbf{Q}^{*}=\text{plim}\frac{1}{N}\mathbf{X}^{\top}\Omega\mathbf{X}$$

3. Inefficiency of the Ordinary Least Squares

Remark

1. Regularity conditions include the exogeneity conditions, but also (i) the regressors are sufficiently well-behaved and (ii) the off-diagonal terms of the variance-covariance matrix diminish sufficiently rapidly (relative to the diagonal elements).

2. For a formal proof in a general case, see Amemiya (1985, p. 187).

Amemiya T. (1985), Advanced Econometrics. Harvard University Press.

3. Inefficiency of the Ordinary Least Squares

Definition (Asymptotic variance)

Under suitable regularity conditions, the asymptotic variance covariance matrix of the OLS estimator $\widehat{\beta}$ is given by:

$$\mathbb{V}_{asy}\left(\widehat{\beta}_{OLS}\right) = \frac{\sigma^2}{N}\mathbf{Q}^{-1}\mathbf{Q}^{*}\mathbf{Q}^{-1}$$

with

$$\mathbf{Q} = \text{plim}\frac{1}{N}\mathbf{X}^{\top}\mathbf{X} \qquad \mathbf{Q}^{*} = \text{plim}\frac{1}{N}\mathbf{X}^{\top}\boldsymbol{\Omega}\mathbf{X}$$

3. Inefficiency of the Ordinary Least Squares

Fact (Non-robust inference)

*Because the variance of the least squares estimator is not $\sigma^2 \left(\mathbf{X}^\top \mathbf{X} \right)^{-1}$ statistical inference (**non-robust inference**) based on $\widehat{\sigma}^2 \left(\mathbf{X}^\top \mathbf{X} \right)^{-1}$ may be **misleading**. For instance the t-test-statistic:*

$$t_{\beta_k} = \frac{\widehat{\beta}_k}{\widehat{\sigma} \sqrt{m_{kk}}}$$

where m_{kk} is k^{th} diagonal element of $\mathbf{X}^\top \mathbf{X}$ do not have a Student distribution.

Robust / Non-robust inference

- As a consequence, the **familiar inference** procedures based on the F and t distributions will no longer be appropriate.
- The question is to know how to **estimate** $\mathbb{V}\left(\widehat{\beta}_{OLS}\right)$ in the context of the linear generalized regression model in order to make **robust inference**.

3. Inefficiency of the Ordinary Least Squares

Definition (Estimator of the asymptotic variance covariance matrix)

If $\mathbf{\Sigma} = \sigma^2 \mathbf{\Omega}$ were known, the consistent **estimator** of the (asymptotic) variance covariance of $\widehat{\beta}_{OLS}$ would be:

$$\widehat{\mathbb{V}}_{asy}\left(\widehat{\beta}_{OLS}\right) = \frac{\sigma^2}{N}\left(\frac{1}{N}\mathbf{X}^\top\mathbf{X}\right)^{-1}\left(\frac{1}{N}\mathbf{X}^\top\mathbf{\Omega}\mathbf{X}\right)\left(\frac{1}{N}\mathbf{X}^\top\mathbf{X}\right)^{-1}$$

3. Inefficiency of the Ordinary Least Squares

Proof

By definition:

$$\mathbf{Q} = \text{plim} \frac{1}{N} \mathbf{X}^{\top} \mathbf{X}$$

$$\mathbf{Q}^{*} = \text{plim} \frac{1}{N} \mathbf{X}^{\top} \Omega \mathbf{X}$$

So,

$$\begin{aligned} \text{plim } \widehat{\mathbb{V}}_{asy}\left(\widehat{\beta}_{OLS}\right) &= \text{plim} \frac{\sigma^2}{N} \left(\frac{1}{N}\mathbf{X}^{\top}\mathbf{X}\right)^{-1} \left(\frac{1}{N}\mathbf{X}^{\top}\Omega\mathbf{X}\right) \left(\frac{1}{N}\mathbf{X}^{\top}\mathbf{X}\right)^{-1} \\ &= \frac{\sigma^2}{N} \mathbf{Q}^{-1} \mathbf{Q}^{*} \mathbf{Q}^{-1} \end{aligned}$$

Or equivalently

$$\widehat{\mathbb{V}}_{asy}\left(\widehat{\beta}_{OLS}\right) \xrightarrow{p} \mathbb{V}_{asy}\left(\widehat{\beta}_{OLS}\right) \quad \square$$

3. Inefficiency of the Ordinary Least Squares

Reminder

$$\mathbf{X}^{\top}\mathbf{X} = \sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i^{\top}$$

$$\mathbf{X}^{\top}\boldsymbol{\Omega}\mathbf{X} = \sum_{i=1}^{N} \sum_{j=1}^{N} \omega_{ij} \mathbf{x}_i \mathbf{x}_i^{\top}$$

$$\mathbf{X}^{\top}\boldsymbol{\Sigma}\mathbf{X} = \sum_{i=1}^{N} \sum_{j=1}^{N} \sigma_{ij} \mathbf{x}_i \mathbf{x}_i^{\top} = \sigma^2 \sum_{i=1}^{N} \sum_{j=1}^{N} \omega_{ij} \mathbf{x}_i \mathbf{x}_i^{\top}$$

Remark

The estimator

$$\widehat{\mathbb{V}}_{asy}\left(\widehat{\beta}_{OLS}\right)=\frac{\sigma^2}{N}\left(\frac{1}{N}\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\left(\frac{1}{N}\mathbf{X}^{\top}\Omega\mathbf{X}\right)\left(\frac{1}{N}\mathbf{X}^{\top}\mathbf{X}\right)^{-1}$$

can also be written as

$$\widehat{\mathbb{V}}_{asy}\left(\widehat{\beta}_{OLS}\right)=\frac{\sigma^2}{N}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^{\top}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\omega_{ij}\mathbf{x}_i\mathbf{x}_i^{\top}\right)\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^{\top}\right)^{-1}$$

3. Inefficiency of the Ordinary Least Squares

Remark

In the next section, we will give a **feasible** estimator $\widehat{\mathbb{V}}_{asy}\left(\widehat{\beta}_{OLS}\right)$ in the specific case of an heteroscedastic model.

3. Inefficiency of the Ordinary Least Squares

Summary

In the GLR model, under some regularity conditions:

1. The OLS estimator is **unbiased**

2. The OLS estimator is (weakly) **consistent**

3. The OLS estimator is **asymptotically normally distributed**

3. Inefficiency of the Ordinary Least Squares

Summary

But...

1. The inference based on the estimator $\sigma^2 \left(\mathbf{X}^\top \mathbf{X}\right)^{-1}$ is **misleading**.

2. The OLS is **inefficient**.

$$\mathbb{V}\left(\widehat{\boldsymbol{\beta}}_{OLS}\right) - I_N^{-1}\left(\boldsymbol{\beta}_0\right) \text{ is a positive definite matrix}$$

3. Inefficiency of the Ordinary Least Squares

Key Concepts

1. OLS estimator in the generalized regression model
2. Finite sample properties
3. Asymptotic variance covariance matrix of the OLS estimator

Section 4

Generalized Least Squares (GLS)

4. Generalized Least Squares (GLS)

Objectives

The objective of this section are the following:

1. Define the Generalized Least Squares (GLS)
2. Define the Feasible Generalized Least Squares (FGLS)
3. Study the statistical properties of the GLS and FGLS estimators

4. Generalized Least Squares (GLS)

Consider the generalized linear regression model with

$$\mathbb{V}\left(\boldsymbol{\varepsilon}\right|\mathbf{X}) = \boldsymbol{\Sigma} = \sigma^2\boldsymbol{\Omega}$$

We will distinguish two cases:

Case 1: the variance covariance matrix $\boldsymbol{\Sigma}$ is known (unrealistic case)

Case 2: the variance covariance matrix $\boldsymbol{\Sigma}$ is unknown

Case 1: Σ is known

The Generalized Least Squares (GLS) estimator

4. Generalized Least Squares (GLS)

Definition (Factorisation)

Since $\mathbf{\Omega}$ is a positive definite matrix, it can factored as follows:

$$\mathbf{\Omega} = \mathbf{C}\mathbf{\Lambda}\mathbf{C}^{\top}$$

where the columns of $\mathbf{C}$ are the characteristics vectors of $\mathbf{\Omega}$, the characteristic roots of $\mathbf{\Omega}$ are arrayed in the diagonal matrix $\mathbf{\Lambda}$, and

$$\mathbf{C}^{\top}\mathbf{C} = \mathbf{C}\mathbf{C}^{\top} = \mathbf{I}_N$$

where $\mathbf{I}$ denotes the identity matrix.

4. Generalized Least Squares (GLS)

Definition

We define the matrix $\mathbf{P}$ such that

$$\mathbf{P}^{\top} = \mathbf{C}\boldsymbol{\Lambda}^{-1/2}$$

so that

$$\boldsymbol{\Omega}^{-1} = \mathbf{P}^{\top}\mathbf{P}$$

4. Generalized Least Squares (GLS)

Proof

$$\mathbf{P}^{\top} = \mathbf{C}\mathbf{\Lambda}^{-1/2}$$

Since $\mathbf{\Lambda}$ is diagonal, $\mathbf{\Lambda}^{-1/2}\mathbf{\Lambda}^{-1/2} = \mathbf{\Lambda}^{-1}$, and we have:

$$\mathbf{P}^{\top}\mathbf{P} = \mathbf{C}\mathbf{\Lambda}^{-1/2}\mathbf{\Lambda}^{-1/2}\mathbf{C}^{\top} = \mathbf{C}\mathbf{\Lambda}^{-1}\mathbf{C}^{\top}$$

Consider the quantity $\mathbf{P}^{\top}\mathbf{P}\mathbf{\Omega}$:

$$\begin{aligned} \mathbf{P}^{\top}\mathbf{P}\mathbf{\Omega} &= \mathbf{C}\mathbf{\Lambda}^{-1}\mathbf{C}^{\top}\mathbf{C}\mathbf{\Lambda}\mathbf{C}^{\top} \\ &= \mathbf{C}\mathbf{\Lambda}^{-1}\mathbf{\Lambda}\mathbf{C}^{\top} \\ &= \mathbf{C}\mathbf{C}^{\top} \\ &= \mathbf{I}_N \end{aligned}$$

Since $\mathbf{C}$ satisfies $\mathbf{C}\mathbf{C}^{\top} = \mathbf{I}_N$. Then, $\mathbf{P}^{\top}\mathbf{P} = \mathbf{\Omega}^{-1}$ $\square$

4. Generalized Least Squares (GLS)

GLS estimator

Premultiply the generalized linear regression model by $\mathbf{P}$ to obtain

$$\mathbf{P}\mathbf{y} = \mathbf{P}\mathbf{X}\beta + \mathbf{P}\varepsilon$$

or equivalently

$$\mathbf{y}^{*} = \mathbf{X}^{*}\boldsymbol{\beta} + \boldsymbol{\varepsilon}^{*}$$

The conditional variance of $\boldsymbol{\varepsilon}^{*}$ is

$$\begin{aligned}
\mathbb{V}\left(\boldsymbol{\varepsilon}^{*}\middle|\,\mathbf{X}\right) &= \mathbb{E}\left(\boldsymbol{\varepsilon}^{*}\boldsymbol{\varepsilon}^{*\top}\middle|\,\mathbf{X}\right) \\
&= \mathbf{P}\mathbb{E}\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\top}\middle|\,\mathbf{X}\right)\mathbf{P}^{\top} \\
&= \sigma^{2}\mathbf{P}\boldsymbol{\Omega}\mathbf{P}^{\top} \\
&= \sigma^{2}\boldsymbol{\Lambda}^{-1/2}\mathbf{C}^{\top}\mathbf{C}\boldsymbol{\Lambda}\mathbf{C}^{\top}\mathbf{C}\boldsymbol{\Lambda}^{-1/2} \\
&= \sigma^{2}\mathbf{I}_{N}
\end{aligned}$$

4. Generalized Least Squares (GLS)

GLS estimator (cont'd)

$$\mathbf{y}^* = \mathbf{X}^* \beta + \varepsilon^*$$

$$\mathbb{V}\left(\varepsilon^* \middle| \mathbf{X}\right) = \sigma^2 \mathbf{I}_N$$

The classical regression model applies to this transformed model.

If Ω is assumed to be known, $\mathbf{y}^* = \mathbf{Py}$ and $\mathbf{X}^* = \mathbf{PX}$ are observed data.

So, we can apply the ordinary least squares to this transformed model:

$$\widehat{\beta} = \left(\mathbf{X}^{*\top}\mathbf{X}^*\right)^{-1}\left(\mathbf{X}^{*\top}\mathbf{y}^*\right)$$

4. Generalized Least Squares (GLS)

GLS estimator (cont'd)

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} &= \left(\mathbf{X}^{*\top}\mathbf{X}^{*}\right)^{-1}\left(\mathbf{X}^{*\top}\mathbf{y}^{*}\right) \\
&= \left(\mathbf{X}^{\top}\mathbf{P}^{\top}\mathbf{P}\mathbf{X}\right)^{-1}\left(\mathbf{X}^{\top}\mathbf{P}^{\top}\mathbf{P}\mathbf{y}\right) \\
&= \left(\mathbf{X}^{\top}\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}^{\top}\boldsymbol{\Omega}^{-1}\mathbf{y}\right)
\end{aligned}$$

This estimator is the generalized least squares (GLS) estimator of $\boldsymbol{\beta}$.

4. Generalized Least Squares (GLS)

Definition (GLS estimator)

The **Generalized Least Squares (GLS)** estimator of β is defined as to be:

$$\widehat{\beta}_{GLS} = \left(\mathbf{X}^{\top}\mathbf{\Omega}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}^{\top}\mathbf{\Omega}^{-1}\mathbf{y}\right)$$

4. Generalized Least Squares (GLS)

Definition (Bias)

Under the exogeneity assumption (A3), the estimator $\widehat{\boldsymbol{\beta}}_{GLS}$ is **unbiased**:

$$\mathbb{E}\left(\widehat{\boldsymbol{\beta}}_{GLS}\right) = \boldsymbol{\beta}_0$$

where $\boldsymbol{\beta}_0$ denotes the true value of the parameters.

Proof

We have:

$$\widehat{\beta}_{GLS} = \left(\mathbf{X}^{\top}\mathbf{\Omega}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}^{\top}\mathbf{\Omega}^{-1}\mathbf{y}\right) = \beta_0 + \left(\mathbf{X}^{\top}\mathbf{\Omega}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}^{\top}\mathbf{\Omega}^{-1}\varepsilon\right)$$

So,

$$\mathbb{E}\left(\widehat{\beta}_{GLS}\middle|\,\mathbf{X}\right) = \beta_0 + \left(\mathbf{X}^{\top}\mathbf{\Omega}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}^{\top}\mathbf{\Omega}^{-1}\mathbb{E}\left(\varepsilon|\,\mathbf{X}\right)\right)$$

Under the exogeneity assumption A3, $\mathbb{E}\left(\varepsilon|\,\mathbf{X}\right) = 0$, so we have

$$\mathbb{E}\left(\widehat{\beta}_{GLS}\middle|\,\mathbf{X}\right) = \beta_0$$

and

$$\mathbb{E}\left(\widehat{\beta}_{GLS}\right) = \mathbb{E}_X\left(\mathbb{E}\left(\widehat{\beta}_{GLS}\middle|\,\mathbf{X}\right)\right) = \mathbb{E}_X\left(\beta_0\right) = \beta_0 \quad \square$$

4. Generalized Least Squares (GLS)

Definition (Variance covariance matrix)

The conditional **variance covariance matrix** of the estimator $\widehat{\beta}_{GLS}$ is defined as to be:

$$\mathbb{V}\left(\widehat{\beta}_{GLS}\middle|\mathbf{X}\right) = \sigma^2\left(\mathbf{X}^{\top}\mathbf{\Omega}^{-1}\mathbf{X}\right)^{-1}$$

The variance covariance matrix is given by

$$\mathbb{V}\left(\widehat{\beta}_{GLS}\right) = \sigma^2\mathbb{E}_X\left(\left(\mathbf{X}^{\top}\mathbf{\Omega}^{-1}\mathbf{X}\right)^{-1}\right)$$

Proof

Consider the definition of $\widehat{\beta}_{GLS}$ in the transformed model:

$$\widehat{\beta}_{GLS} = \beta_0 + \left(\mathbf{X}^{*\top}\mathbf{X}^{*}\right)^{-1}\left(\mathbf{X}^{*\top}\varepsilon^{*}\right)$$

$$\mathbb{V}\left(\widehat{\beta}_{GLS}\middle|\,\mathbf{X}\right) = \left(\mathbf{X}^{*\top}\mathbf{X}^{*}\right)^{-1}\mathbf{X}^{*\top}\mathbb{E}\left(\varepsilon^{*}\varepsilon^{*\top}\middle|\,\mathbf{X}\right)\mathbf{X}^{*}\left(\mathbf{X}^{*\top}\mathbf{X}^{*}\right)^{-1}$$

Since $\mathbb{E}\left(\varepsilon^{*}\varepsilon^{*\top}\middle|\,\mathbf{X}\right) = \sigma^2\mathbf{I}_N$, we have

$$\begin{aligned}
\mathbb{V}\left(\widehat{\beta}_{GLS}\middle|\,\mathbf{X}\right) &= \sigma^2\left(\mathbf{X}^{*\top}\mathbf{X}^{*}\right)^{-1}\mathbf{X}^{*\top}\mathbf{X}^{*}\left(\mathbf{X}^{*\top}\mathbf{X}^{*}\right)^{-1} \\
&= \sigma^2\left(\mathbf{X}^{*\top}\mathbf{X}^{*}\right)^{-1} \\
&= \sigma^2\left(\mathbf{X}^{\top}\mathbf{P}^{\top}\mathbf{P}\mathbf{X}\right)^{-1} \\
&= \sigma^2\left(\mathbf{X}^{\top}\mathbf{\Omega}^{-1}\mathbf{X}\right)^{-1} \quad \square
\end{aligned}$$

4. Generalized Least Squares (GLS)

Definition (Consistency)

Under the exogeneity assumption A3, the GLS estimator $\widehat{\beta}_{GLS}$ is (weakly) **consistent**:

$$\widehat{\beta}_{GLS} \xrightarrow{p} \beta_0$$

as soon as

$$\text{plim} \frac{1}{N} \mathbf{X}^{*\top} \mathbf{X}^{*} = \mathbf{Q}^{*}$$

where $\mathbf{Q}^{*}$ is a finite positive definite matrix.

4. Generalized Least Squares (GLS)

Proof

$$\widehat{\beta}_{GLS} = \beta_0 + \left(\mathbf{X}^\top \Omega^{-1} \mathbf{X}\right)^{-1} \left(\mathbf{X}^\top \Omega^{-1} \varepsilon\right)$$

Under the assumption A3 (exogeneity):

$$\text{plim} \frac{1}{N} \mathbf{X}^\top \Omega^{-1} \varepsilon = \mathbf{0}_{K \times 1}$$

$$\text{plim} \frac{1}{N} \mathbf{X}^\top \Omega^{-1} \mathbf{X} = \mathbf{Q}^*$$

So, we have

$$\text{plim } \widehat{\beta}_{GLS} = \beta_0$$

The estimator $\widehat{\beta}_{GLS}$ is weakly consistent. □

4. Generalized Least Squares (GLS)

Definition (Asymptotic distribution)

Under some regularity conditions, the GLS estimator $\widehat{\beta}_{GLS}$ is **asymptotically normally distributed**:

$$\sqrt{N}\left(\widehat{\beta}_{GLS}-\beta_0\right)\xrightarrow{d}\mathcal{N}\left(\mathbf{0},\sigma^2\mathbf{Q}^{*-1}\right)$$

where

$$\mathbf{Q}^{*}=\text{plim}\frac{1}{N}\mathbf{X}^{*\top}\mathbf{X}^{*}=\text{plim}\frac{1}{N}\mathbf{X}^{\top}\boldsymbol{\Omega}^{-1}\mathbf{X}$$

4. Generalized Least Squares (GLS)

Definition (Asymptotic variance covariance matrix)

The asymptotic **variance covariance matrix** of the estimator $\widehat{\beta}_{GLS}$ is:

$$\mathbb{V}_{asy}\left(\widehat{\beta}_{GLS}\right) = \frac{\sigma^2}{N}\mathbf{Q}^{*-1}$$

If $\mathbf{\Sigma} = \sigma^2\mathbf{\Omega}$ is known, a consistent estimator is given by:

$$\widehat{\mathbb{V}}_{asy}\left(\widehat{\beta}_{GLS}\right) = \frac{\sigma^2}{N}\left(\mathbf{X}^{\top}\mathbf{\Omega}^{-1}\mathbf{X}\right)^{-1}$$

This estimator holds whether **X** is stochastic or non-stochastic.

4. Generalized Least Squares (GLS)

Theorem (BLUE estimator)

*The GLS estimator $\widehat{\beta}_{GLS}$ is the minimum variance linear unbiased estimator (***BLUE estimator***) in the* **semi-parametric** *generalized linear regression model. In particular, the matrix defined by:*

$$\mathbb{V}_{asy}\left(\widehat{\beta}_{OLS}\right) - \mathbb{V}_{asy}\left(\widehat{\beta}_{GLS}\right)$$

is a positive semi definite matrix.

4. Generalized Least Squares (GLS)

Theorem (Efficiency)

Under suitable regularity conditions, in a **parametric** *generalized linear regression model, the GLS estimator* $\widehat{\boldsymbol{\beta}}_{GLS}$ *is efficient*

$$\mathbb{V}\left(\widehat{\boldsymbol{\beta}}_{GLS}\right) = I_N^{-1}\left(\boldsymbol{\beta}_0\right)$$

where $I_N^{-1}\left(\boldsymbol{\beta}_0\right)$ *denotes the FDCR or Cramer-Rao bound.*

4. Generalized Least Squares (GLS)

Remark

In a **Gaussian** generalized linear regression model (under assumption A6), the likelihood of the sample is given by:

$$
\begin{aligned}
L_N\left(\theta; y \middle| x\right) &= \left(2\pi\sigma^2\right)^{-N/2} \left|\mathbf{\Omega}\right|^{-N/2} \\
&\quad \times \exp\left(-\frac{1}{2\sigma^2}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)^{\top}\mathbf{\Omega}^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right)
\end{aligned}
$$

The log-likelihood is defined as to be:

$$
\begin{aligned}
\ell_N\left(\theta; y \middle| x\right) &= -\frac{N}{2}\ln\left(2\pi\sigma^2\right) - \frac{N}{2}\log\left(\left|\mathbf{\Omega}\right|\right) \\
&\quad -\frac{1}{2\sigma^2}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)^{\top}\mathbf{\Omega}^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)
\end{aligned}
$$

Remark

For testing hypotheses, we can apply the full set of results in Chapter 4 to the **transformed model**. For instance, for testing the p linear constraints $\mathrm{H}_0 : \mathbf{R}\beta = \mathbf{q}$, the appropriate test-statistic is:

$$\mathrm{F} = \frac{1}{p}\left(\mathbf{R}\widehat{\beta}_{GLS} - \mathbf{q}\right)^{\top}\left(\sigma^2\mathbf{R}\left(\mathbf{X}^{\top}\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}\mathbf{R}^{\top}\right)^{-1}\left(\mathbf{R}\widehat{\beta}_{GLS} - \mathbf{q}\right)$$

4. Generalized Least Squares (GLS)

Fact

To summarize, all the results for the **classical model**, *including the usual inference procedures, apply to the* **transformed model.**

Case 2: Σ is unknown

The Feasible Generalized Least Squares (FGLS) estimator

4. Generalized Least Squares (GLS)

Introduction

1. If $\mathbf{\Sigma}$ contains unknown parameters that must be estimated, then generalized least squares is not feasible.
2. With an unrestricted matrix $\mathbf{\Sigma} = \sigma^2 \mathbf{\Omega}$, there are $N(N+1)/2$ additional parameters (since $\mathbf{\Sigma}$ is symmetric) to estimate
3. This number is far too many to estimate with N observations.
4. Obviously, some **structure** must be imposed on the model if we are to proceed.

4. Generalized Least Squares (GLS)

Definition (Structure of variance covariance matrix)

We assume that the conditional variance covariance matrix of the disturbances can be expressed as a function of a small set of parameters $\boldsymbol{\alpha}$:

$$\mathbb{V}\left(\boldsymbol{\varepsilon}\right|\mathbf{X}) = \sigma^2 \mathbf{\Omega}\left(\boldsymbol{\alpha}\right)$$

4. Generalized Least Squares (GLS)

Example (Time series)

For instance, a commonly used formula in time-series settings is

$$\mathbf{\Omega}(\rho) = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & .. & \rho^{N-1} \\ \rho & 1 & \rho & \rho^2 & .. & \rho^{N-2} \\ \rho^2 & \rho & 1 & \rho & .. & \rho^{N-3} \\ \rho^3 & \rho^2 & \rho & 1 & .. & .. \\ .. & .. & .. & .. & .. & .. \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & .. & .. & 1 \end{pmatrix}$$

4. Generalized Least Squares (GLS)

Example (Heteroscedascticity)

If we consider a heteroscedastic model, where the variance of ε_i depends on a variable z_i, with

$$\mathbb{V}\left(\varepsilon_i \middle| \mathbf{X}\right) = \sigma^2 z_i^{\theta}$$

we have

$$\boldsymbol{\Omega}\left(\theta\right) = \begin{pmatrix} z_1^{\theta} & 0 & 0 & .. & 0 \\ 0 & z_2^{\theta} & 0 & .. & 0 \\ 0 & 0 & z_3^{\theta} & .. & 0 \\ .. & .. & .. & .. & .. \\ 0 & 0 & 0 & .. & z_N^{\theta} \end{pmatrix}$$

4. Generalized Least Squares (GLS)

Definition (Feasible Generalized Least Squares (FGLS))

Consider a consistent estimator $\widehat{\boldsymbol{\alpha}}$ of $\boldsymbol{\alpha}$, then the Feasible Least Generalized Squares (FGLS) estimator of $\boldsymbol{\beta}$ is defined as to be:

$$\widehat{\boldsymbol{\beta}}_{FGLS} = \left(\mathbf{X}^{\top} \widehat{\boldsymbol{\Omega}}^{-1} \mathbf{X} \right)^{-1} \left(\mathbf{X}^{\top} \widehat{\boldsymbol{\Omega}}^{-1} \mathbf{y} \right)$$

where $\widehat{\boldsymbol{\Omega}} = \boldsymbol{\Omega} \left(\widehat{\boldsymbol{\alpha}} \right)$ is a consistent estimator of $\boldsymbol{\Omega} \left(\boldsymbol{\alpha} \right)$.

4. Generalized Least Squares (GLS)

Remark

If

$$\text{plim}\left(\left(\frac{1}{N}\mathbf{X}^\top\widehat{\mathbf{\Omega}}^{-1}\mathbf{X}\right)-\left(\frac{1}{N}\mathbf{X}^\top\mathbf{\Omega}^{-1}\mathbf{X}\right)\right)=0$$

$$\text{plim}\left(\left(\frac{1}{N}\mathbf{X}^\top\widehat{\mathbf{\Omega}}^{-1}\mathbf{y}\right)-\left(\frac{1}{N}\mathbf{X}^\top\mathbf{\Omega}^{-1}\mathbf{y}\right)\right)=0$$

Then the GLS and FGLS estimators are asymptotically equivalent

$$\widehat{\beta}_{FGLS}-\widehat{\beta}_{GLS}\xrightarrow{p}\mathbf{0}_{K\times 1}$$

4. Generalized Least Squares (GLS)

Theorem (Efficiency)

An asymptotically efficient FGLS estimator does not require that we have an efficient estimator of $\boldsymbol{\alpha}$*; only a consistent one is required to achieve full efficiency for the FGLS estimator.*

Remark

If the estimator $\widehat{\boldsymbol{\alpha}}$ is consistent

$$\widehat{\boldsymbol{\alpha}} \xrightarrow{p} \boldsymbol{\alpha}$$

then the FGLS estimator has the same asymptotic properties (consistency, efficiency, asymptotic distribution etc.) than the GLS estimator.

Key Concepts

1. Factorisation of the variance covariance matrix
2. Generalized Least Squares (GLS) estimator
3. Feasible Generalized Least Squares (FGLS) estimator

Section 5

Heteroscedasticity

Objectives

The objective of this section are the following:

1. To determine the properties of the OLS in presence of heteroscedasticity
2. To estimate the asymptotic variance covariance matrix of the OLS estimator in presence of heteroscedasticity
3. To introduce the concept of robust inference (to heteroscedasticity)

Introduction

In the rest of this chapter, we will focus on the case of heteroscedastic disturbances.

$$\mathbb{V}\left(\varepsilon_i \middle| \mathbf{X}\right) = \sigma_i^2 \quad \text{for } i = 1, .., N$$

Heteroscedasticity arises in numerous applications, in both cross-section and time-series data.

For example, even after accounting for firm sizes, we expect to observe greater variation in the profits of large firms than in those of small ones.

5. Heteroscedasticity

Assumption: We assume that the disturbances are **pairwise uncorrelated** and **heteroscedastic:**

$$\mathbb{V}\left(\left.\boldsymbol{\varepsilon}\right|\mathbf{X}\right) = \boldsymbol{\Sigma} = \sigma^2 \boldsymbol{\Omega}$$

with

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 & .. & 0 \\ 0 & \sigma_2^2 & .. & 0 \\ .. & .. & .. & .. \\ 0 & .. & .. & \sigma_N^2 \end{pmatrix} = \sigma^2 \boldsymbol{\Omega} = \sigma^2 \begin{pmatrix} \omega_1 & 0 & .. & 0 \\ 0 & \omega_2 & .. & 0 \\ .. & .. & .. & .. \\ 0 & .. & .. & \omega_N \end{pmatrix}$$

with $\omega_i = \sigma_i^2 / \sigma^2$ for $i = 1, .., N$.

5. Heteroscedasticity

Definition (Scaling)

The fact to scale the variances as

$$\sigma_i^2 = \sigma^2 \omega_i \quad \text{for } i = 1, .., N$$

allows us to use a normalisation on $\mathbf{\Omega}$

$$\text{trace}\left(\mathbf{\Omega}\right) = \sum_{i=1}^{N} \omega_i = N$$

Introduction (cont'd)

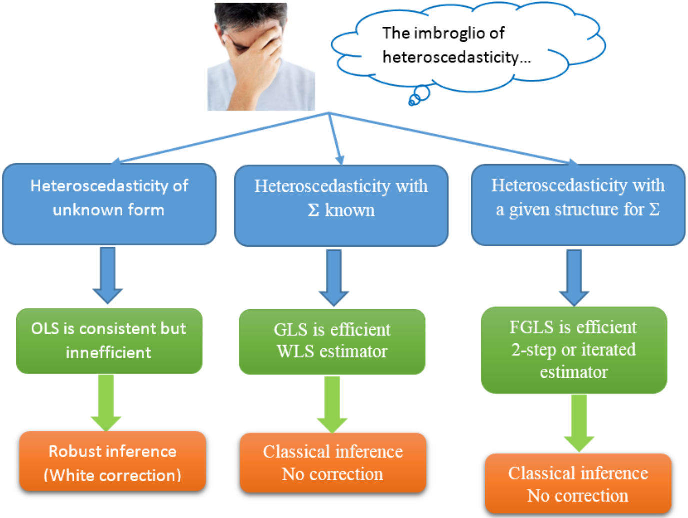
We will consider three cases:

Case 1: the heteroscedasticity form (structure) is unknown: OLS estimator and robust inference

Case 2: the variance covariance matrix $\mathbf{\Sigma}$ is known: GLS or Weighted Least Square (WLS)

Case 3: the variance covariance matrix $\mathbf{\Sigma}$ is unknown but its form (structure) is known: two-steps or iterated FGLS estimator

5. Heteroscedasticity



Case 1: Heteroscedasticity of unknown form

OLS and robust inference

5. Heteroscedasticity

Assumption: We assume that the variances σ_i^2 are unknown for $i = 1, ..N$ and no particular form (structure) is imposed on $\mathbf{\Omega}$ (or $\mathbf{\Sigma}$).

Introduction

1. The GLS cannot be implemented since $\mathbf{\Sigma}$ is unknown.
2. The FGLS estimator requires to estimate (in a first step) N parameters $\sigma_1^2, .., \sigma_N^2$. With N observations, the FGLS is not feasible.
3. The **only solution to estimate β** consists in using the OLS.
4. Under suitable regularity conditions, the OLS estimator is unbiased, consistent, asymptotically normally distributed but... **inefficient**.

5. Heteroscedasticity

Introduction (cont'd)

Consider the OLS estimator:

$$\widehat{\beta}_{OLS} = \left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\mathbf{X}^{\top}\mathbf{y}$$

We know that

$$\widehat{\beta}_{OLS} \overset{asy}{\approx} \mathcal{N}\left(\beta_0, \frac{\sigma^2}{N}\mathbf{Q}^{-1}\mathbf{Q}^{*}\mathbf{Q}^{-1}\right)$$

$$\mathbb{V}_{asy}\left(\widehat{\beta}_{OLS}\right) = \frac{\sigma^2}{N}\mathbf{Q}^{-1}\mathbf{Q}^{*}\mathbf{Q}^{-1}$$

with

$$\mathbf{Q} = \text{plim}\frac{1}{N}\mathbf{X}^{\top}\mathbf{X} \qquad \mathbf{Q}^{*} = \text{plim}\frac{1}{N}\mathbf{X}^{\top}\Omega\mathbf{X}$$

Problem (Robust inference with OLS)

The conventionally estimated covariance matrix for the least squares estimator $\sigma^2 \left(\mathbf{X}^\top \mathbf{X}\right)^{-1}$ *is inappropriate; the appropriate matrix is* $\sigma^2 \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \left(\mathbf{X}^\top \mathbf{\Omega} \mathbf{X}\right)^{-1} \left(\mathbf{X}^\top \mathbf{X}\right)^{-1}$. *It is unlikely that these two would coincide, so the usual estimators of the standard errors are likely to be erroneous. The inference (test-statistics) based* $\sigma^2 \left(\mathbf{X}^\top \mathbf{X}\right)^{-1}$ *is misleading.*

Question

How to estimate $\mathbb{V}_{asy}\left(\widehat{\beta}_{OLS}\right)$ and to make **robust inference**?

$$\mathbb{V}_{asy}\left(\widehat{\beta}_{OLS}\right) = \frac{\sigma^2}{N}\mathbf{Q}^{-1}\mathbf{Q}^{*}\mathbf{Q}^{-1}$$

$$\mathbf{Q} = \text{plim}\frac{1}{N}\mathbf{X}^{\top}\mathbf{X} \qquad \mathbf{Q}^{*} = \text{plim}\frac{1}{N}\mathbf{X}^{\top}\Omega\mathbf{X}$$

5. Heteroscedasticity

We seek an estimator for

$$\mathbf{Q}^{*} = \text{plim} \frac{1}{N} \mathbf{X}^{\top} \mathbf{\Omega} \mathbf{X} = \text{plim} \frac{1}{N} \sum_{i=1}^{N} \omega_i \mathbf{x}_i \mathbf{x}_i^{\top} = \mathbb{E}_X \left(\omega_i \mathbf{x}_i \mathbf{x}_i^{\top} \right)$$

or equivalently of

$$\mathbf{Q}^{**} = \text{plim} \frac{1}{N} \mathbf{X}^{\top} \mathbf{\Sigma} \mathbf{X} = \text{plim} \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 \mathbf{x}_i \mathbf{x}_i^{\top} = \mathbb{E}_X \left(\sigma_i \mathbf{x}_i \mathbf{x}_i^{\top} \right)$$

with

$$\mathbf{Q}^{**} = \sigma^2 \mathbf{Q}^{*}$$

5. Heteroscedasticity

$$\mathbf{Q}^{**} = \text{plim}\frac{1}{N}\mathbf{X}^{\top}\mathbf{\Sigma}\mathbf{X} = \text{plim}\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2\mathbf{x}_i\mathbf{x}_i^{\top}$$

White (1980) shows that under very general condition, the estimator

$$\mathbf{S}_0 = \frac{1}{N}\sum_{i=1}^{N}\widehat{\varepsilon}_i^2\mathbf{x}_i\mathbf{x}_i^{\top}$$

where $\widehat{\varepsilon}_i = y_i - \mathbf{x}_i^{\top}\widehat{\boldsymbol{\beta}}_{OLS}$, converges to $\mathbf{Q}^{**} = \sigma^2\mathbf{Q}^{*}$

$$\mathbf{S}_0 \xrightarrow{p} \mathbf{Q}^{**} = \sigma^2\mathbf{Q}^{*}$$

White, H. "A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity." Econometrica, 48, 1980, pp. 817–838.

5. Heteroscedasticity

$$\mathbb{V}_{asy}\left(\widehat{\beta}_{OLS}\right)=\frac{\sigma^2}{N}\mathbf{Q}^{-1}\mathbf{Q}^{*}\mathbf{Q}^{-1}$$

We know that:

$$\mathbf{S}_0=\frac{1}{N}\sum_{i=1}^{N}\widehat{\varepsilon}_i^2\mathbf{x}_i\mathbf{x}_i^{\top}\xrightarrow{p}\sigma^2\mathbf{Q}^{*}$$

$$\left(\frac{1}{N}\mathbf{X}^{\top}\mathbf{X}\right)^{-1}=\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^{\top}\right)^{-1}\xrightarrow{p}\mathbf{Q}^{-1}$$

So,

$$\frac{1}{N}\left(\frac{1}{N}\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\mathbf{S}_0\left(\frac{1}{N}\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\xrightarrow{p}\mathbb{V}_{asy}\left(\widehat{\beta}_{OLS}\right)$$

5. Heteroscedasticity

Definition (White heteroscedasticity consistent estimator)

The **White consistent** estimator of the **asymptotic variance-covariance** matrix of the ordinary least squares estimator $\widehat{\beta}_{OLS}$ in the generalized linear regression model is defined to be:

$$\widehat{\mathbb{V}}_{asy}\left(\widehat{\beta}_{OLS}\right) = N\left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\mathbf{S}_0\left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}$$

$$\widehat{\mathbb{V}}_{asy}\left(\widehat{\beta}_{OLS}\right) \xrightarrow{p} \mathbb{V}_{asy}\left(\widehat{\beta}_{OLS}\right)$$

with

$$\mathbf{S}_0 = \frac{1}{N}\sum_{i=1}^{N}\widehat{\varepsilon}_i^2\mathbf{x}_i\mathbf{x}_i^{\top}$$

5. Heteroscedasticity

Corollary (White heteroscedasticity consistent estimator)

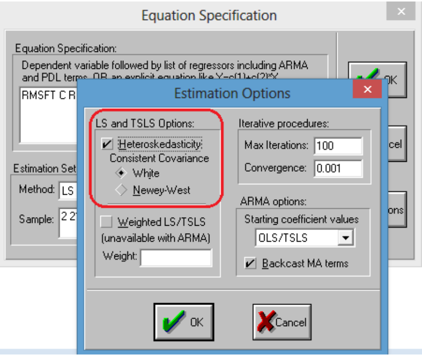
The **White consistent** *estimator can written as:*

$$\widehat{\mathbb{V}}_{asy}\left(\widehat{\beta}_{OLS}\right)=\frac{1}{N}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^{\top}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\widehat{\varepsilon}_i^{2}\mathbf{x}_i\mathbf{x}_i^{\top}\right)\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^{\top}\right)^{-1}$$

Remarks

1. This result is extremely important and useful. It implies that without actually specifying the type of heteroscedasticity, we can still make appropriate inferences based on the results of least squares.
2. This implication is especially useful if we are unsure of the precise nature of the heteroscedasticity (which is probably most of the time).

5. Heteroscedasticity



Dependent Variable: RMSFT
Method: Least Squares
Date: 12/14/13 Time: 16:12
Sample: 2 21
Included observations: 20
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.001189	0.001160	1.025585	0.3187
RSP500	1.989787	0.311130	6.395357	0.0000

R-squared	0.690203	Mean dependent var	-0.000180
Adjusted R-squared	0.672992	S.D. dependent var	0.009272
S.E. of regression	0.005302	Akaike info criterion	-7.546873
Sum squared resid	0.000506	Schwarz criterion	-7.447300
Log likelihood	77.46873	F-statistic	40.10263
Durbin-Watson stat	1.955366	Prob(F-statistic)	0.000006

Remark

Given the normalisation $\text{trace}(\mathbf{\Omega}) = N$, we have:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2$$

5. Heteroscedasticity

Definition (SSR)

The **least squares estimator** $\widehat{\sigma}^2$ defined by:

$$\widehat{\sigma}^2 = \frac{\widehat{\boldsymbol{\varepsilon}}^\top \widehat{\boldsymbol{\varepsilon}}}{N-K} = \frac{1}{N-K} \sum_{i=1}^{N} \widehat{\varepsilon}_i^2$$

converges to the probability limit of the average variance of the disturbances

$$\widehat{\sigma}^2 \xrightarrow{p} \lim_{N \to \infty} \sigma^2 = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2$$

5. Heteroscedasticity

Example (White robust estimator. Source: Greene (2012))

Consider the generalized linear regression model:

$$\text{AVGEXP}_i = \beta_1 + \beta_2 \text{AGE}_i + \beta_3 \text{Ownrent}_i + \beta_4 \text{Income}_i + \beta_5 \text{Income}_i^2 + \varepsilon_i$$

where AVGEXP denotes the Avg. monthly credit card expenditure, Ownrent denotes a binary variable (individual owns (1) or rents (0) home), Age denotes the age in years, Income denotes the income divided by 10,000. The data are available in file `Chapter5_data.xls`. Question: write a Matlab code to (1) estimate the parameters by OLS, (2) compute the standard errors and the robust standard errors and (3) compare your results with Eviews.

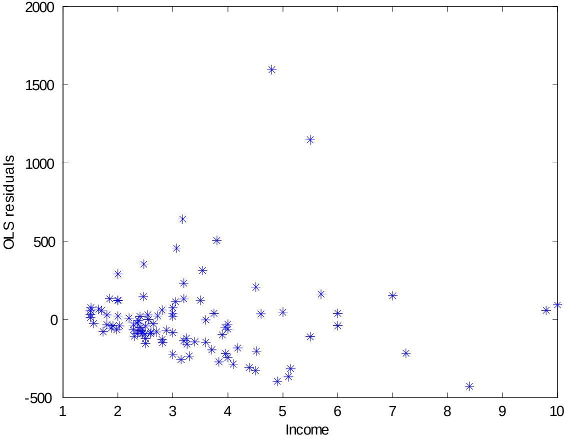
5. Heteroscedasticity

```
clear all; clc ; close all
data=xlsread('Chapter5_data.xls');
Age=data(:,3);
Income=data(:,4);
Avgexp=data(:,5);
Ownrent=data(:,6);

y=Avgexp;                                      % Dependent variable
N=length(y);                                   % Sample size
X=[ones(N,1) Age Ownrent Income Income.^2];   % Matrix X
K=size(X,2);                                   % Number of explicative variables
beta=X\y;                                      % OLS estimates
res=y-X*beta;                                  % Residuals
v=sum(res.^2)/(N-K);                           % Variance of the disturbances
V=v*inv(X'*X);                                 % Estimated asymptotic variance
std=sqrt(diag(V));                             % Standard errors
S0=zeros(K,K);
for i=1:N
        S0=S0+(res(i)^2)*X(i,:)'*X(i,:);
end
S0=S0/N;
V_robust=N*inv(X'*X)*S0*inv(X'*X);            % White estimator
std_robust=sqrt(diag(V_robust));               % Robust standard errors

disp(' ') , disp(' Beta        std       Robust std')
disp([beta std std_robust])
```

5. Heteroscedasticity



This graph is the sign of **heteroscedasticity**.. the variance of the residuals seems to depend on the income.

5. Heteroscedasticity

Beta	std	Robust std
-115.9914	157.8311	148.1444
-3.6537	3.7522	2.3843
60.8815	61.9485	66.1458
156.4672	63.9536	71.2170
-9.0760	6.2024	5.9867

Dependent Variable: AVGEXP
Method: Least Squares
Date: 12/14/13 Time: 17:10
Sample: 1 100
Included observations: 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-115.9914	157.8311	-0.734909	0.4642
AGE	-3.653724	3.752179	-0.973760	0.3326
OWNRENT	60.88148	61.94852	0.982775	0.3282
INCOME	156.4672	63.95355	2.446575	0.0163
INCOME2	-9.075987	6.202363	-1.463311	0.1467

R-squared	0.178845	Mean dependent var	189.0231
Adjusted R-squared	0.144270	S.D. dependent var	294.2446
S.E. of regression	272.1930	Akaike info criterion	14.09961
Sum squared resid	7038457.	Schwarz criterion	14.22986
Log likelihood	-699.9803	F-statistic	5.172665
Durbin-Watson stat	1.785912	Prob(F-statistic)	0.000818

The values are the same.. perfect

5. Heteroscedasticity

Beta	std	Robust std
-115.9914	157.8311	148.1444
-3.6537	3.7522	2.3843
60.8815	61.9485	66.1458
156.4672	63.9536	71.2170
-9.0760	6.2024	5.9867

Dependent Variable: AVGEXP
Method: Least Squares
Date: 12/14/13 Time: 16:56
Sample: 1 100
Included observations: 100
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-115.9914	151.9929	-0.763137	0.4473
AGE	-3.653724	2.446277	-1.493585	0.1386
OWNRENT	60.88148	67.86418	0.897108	0.3719
INCOME	156.4672	73.06713	2.141417	0.0348
INCOME2	-9.075987	6.142213	-1.477641	0.1428

R-squared	0.178845	Mean dependent var	189.0231
Adjusted R-squared	0.144270	S.D. dependent var	294.2446
S.E. of regression	272.1930	Akaike info criterion	14.09961
Sum squared resid	7038457.	Schwarz criterion	14.22986
Log likelihood	-699.9803	F-statistic	5.172665
Durbin-Watson stat	1.785912	Prob(F-statistic)	0.000818

The values are different... Why?

Remark

This difference is due to the fact that Eviews uses a **finite sample correction** for $\mathbf{S}_0$ (Davidson and MacKinnon, 1993)

$$\mathbf{S}_0 = \frac{1}{N-K}\sum_{i=1}^{N}\widehat{\varepsilon}_i^2 \mathbf{x}_i \mathbf{x}_i^\top$$

Davidson, R. and J. MacKinnon. Estimation and Inference in Econometrics. New York: Oxford University Press, 1993.

5. Heteroscedasticity

```
clear all; clc ; close all
data=xlsread('Chapter5_data.xls');
Age=data(:,3);
Income=data(:,4);
Avgexp=data(:,5);
Ownrent=data(:,6);

y=Avgexp;                                     % Dependent variable
N=length(y);                                  % Sample size
X=[ones(N,1) Age Ownrent Income Income.^2];   % Matrix X
K=size(X,2);                                  % Number of explicative variables
beta=X\y;                                     % OLS estimates
res=y-X*beta;                                 % Residuals
v=sum(res.^2)/(N-K);                          % Variance of the disturbances
V=v*inv(X'*X);                                % Estimated asymptotic variance
std=sqrt(diag(V));                            % Standard errors
S0=zeros(K,K);
for i=1:N
        S0=S0+(res(i)^2)*X(i,:)'*X(i,:);
end
S0=S0/(N-K);
V_robust=N*inv(X'*X)*S0*inv(X'*X);            % White estimator
std_robust=sqrt(diag(V_robust));              % Robust standard errors

disp(' ') , disp(' Beta        std       Robust std')
disp([beta std std_robust])
```

5. Heteroscedasticity

Beta	std	Robust std
-115.9914	157.8311	151.9929
-3.6537	3.7522	2.4463
60.8815	61.9485	67.8642
156.4672	63.9536	73.0671
-9.0760	6.2024	6.1422

Dependent Variable: AVGEXP
Method: Least Squares
Date: 12/14/13 Time: 16:56
Sample: 1 100
Included observations: 100
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-115.9914	151.9929	-0.763137	0.4473
AGE	-3.653724	2.446277	-1.493585	0.1386
OWNRENT	60.88148	67.86418	0.897108	0.3719
INCOME	156.4672	73.06713	2.141417	0.0348
INCOME2	-9.075987	6.142213	-1.477641	0.1428

R-squared	0.178845	Mean dependent var	189.0231
Adjusted R-squared	0.144270	S.D. dependent var	294.2446
S.E. of regression	272.1930	Akaike info criterion	14.09961
Sum squared resid	7038457.	Schwarz criterion	14.22986
Log likelihood	-699.9803	F-statistic	5.172665
Durbin-Watson stat	1.785912	Prob(F-statistic)	0.000818

The values are now identical.

Case 2: Heteroscedasticity with known Σ

GLS and Weighted Least Squares

5. Heteroscedasticity

Assumption: We assume that the disturbances are heteroscedastic with

$$\mathbb{V}\left(\boldsymbol{\varepsilon} \middle| \mathbf{X}\right) = \boldsymbol{\Sigma} = \sigma^2 \boldsymbol{\Omega}$$

with

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 & .. & 0 \\ 0 & \sigma_2^2 & .. & 0 \\ .. & .. & .. & .. \\ 0 & .. & .. & \sigma_N^2 \end{pmatrix} = \sigma^2 \boldsymbol{\Omega} = \sigma^2 \begin{pmatrix} \omega_1 & 0 & .. & 0 \\ 0 & \omega_2 & .. & 0 \\ .. & .. & .. & .. \\ 0 & .. & .. & \omega_N \end{pmatrix}$$

where the parameters σ_i^2 and ω_i are known for $i = 1, ..N$.

5. Heteroscedasticity

Definition (GLS estimator)

In presence of heteroscedasticity, the Generalized Least Squares (GLS) estimator of $\boldsymbol{\beta}$ is defined as to:

$$\widehat{\boldsymbol{\beta}}_{GLS} = \left(\sum_{i=1}^{N} \frac{\mathbf{x}_i \mathbf{x}_i^\top}{\omega_i} \right)^{-1} \left(\sum_{i=1}^{N} \frac{\mathbf{x}_i y_i}{\omega_i} \right)$$

or equivalently by

$$\widehat{\boldsymbol{\beta}}_{GLS} = \left(\sum_{i=1}^{N} \frac{\mathbf{x}_i \mathbf{x}_i^\top}{\sigma_i^2} \right)^{-1} \left(\sum_{i=1}^{N} \frac{\mathbf{x}_i y_i}{\sigma_i^2} \right)$$

Proof

In general, whatever the form of $\mathbf{\Sigma} = \sigma^2 \mathbf{\Omega}$, we have:

$$\widehat{\beta}_{GLS} = \left(\mathbf{X}^\top \mathbf{\Omega}^{-1} \mathbf{X} \right)^{-1} \left(\mathbf{X}^\top \mathbf{\Omega}^{-1} \mathbf{y} \right)$$

Since $\mathbf{\Omega}$ is diagonal:

$$\mathbf{X}^\top \mathbf{\Omega}^{-1} \mathbf{X} = \sum_{i=1}^{N} \frac{\mathbf{x}_i \mathbf{x}_i^\top}{\omega_i}$$

$$\mathbf{X}^\top \mathbf{\Omega}^{-1} \mathbf{y} = \sum_{i=1}^{N} \frac{\mathbf{x}_i y_i}{\omega_i}$$

As a consequence:

$$\widehat{\beta}_{GLS} = \left(\sum_{i=1}^{N} \frac{\mathbf{x}_i \mathbf{x}_i^\top}{\omega_i} \right)^{-1} \left(\sum_{i=1}^{N} \frac{\mathbf{x}_i y_i}{\omega_i} \right) \quad \square$$

Remark

$$\widehat{\beta}_{GLS} = \left(\sum_{i=1}^{N} \frac{\mathbf{x}_i \mathbf{x}_i^\top}{\omega_i} \right)^{-1} \left(\sum_{i=1}^{N} \frac{\mathbf{x}_i y_i}{\omega_i} \right)$$

This formula is similar to that obtained for a Weighted Least Squares (WLS).

$$\widehat{\beta}_{WLS} = \left(\sum_{i=1}^{N} \delta_i \mathbf{x}_i \mathbf{x}_i^\top \right)^{-1} \left(\sum_{i=1}^{N} \delta_i \mathbf{x}_i y_i \right)$$

5. Heteroscedasticity

Fact (GLS and WLS)

In presence of heteroscedasticity, the GLS estimator is a particular case of the **Weighted Least Squares (WLS)** *estimator.*

$$\widehat{\beta}_{WLS} = \left(\sum_{i=1}^{N} \delta_i \mathbf{x}_i \mathbf{x}_i^{\top} \right)^{-1} \left(\sum_{i=1}^{N} \delta_i \mathbf{x}_i y_i \right)$$

where δ_i is an arbitrary weight. For $\delta_i = 1/\omega_i$, we have $\widehat{\beta}_{WLS} = \widehat{\beta}_{GLS}$.

Remark

1. The WLS estimator is consistent regardless of the weights used, as long as the weights are uncorrelated with the disturbances.

2. In general, we consider a weight which is proportional to one explicative variable (the income in the last example):

$$\sigma_i^2 = \sigma^2 x_{ik}^2 \Longleftrightarrow \delta_i = \frac{1}{x_{ik}^2}$$

Case 3: Heteroscedasticity for a given structure

FGLS and two-step or iterated estimators

5. Heteroscedasticity

Assumption: We assume that the disturbances are heteroscedastic with

$$\mathbb{V}\left(\left.\boldsymbol{\varepsilon}\right|\mathbf{X}\right)=\boldsymbol{\Sigma}\left(\boldsymbol{\alpha}\right)=\sigma^{2}\boldsymbol{\Omega}\left(\boldsymbol{\alpha}\right)$$

where $\boldsymbol{\alpha}$ denotes a set of parameters.

5. Heteroscedasticity

Example (Restriction)

We assume that

$$\mathbb{V}\left(\varepsilon_i \middle| \mathbf{X}\right) = \sigma_i^2\left(\boldsymbol{\alpha}\right) = \sigma^2\left(\mathbf{z}_i^\top \boldsymbol{\alpha}\right)^2$$

where $\boldsymbol{\alpha} = \left(\alpha_1 : .. : \alpha_H\right)^\top$ is a $H \times 1$ vector of parameters and $\mathbf{z}_i$ is $H \times 1$ of explicative variables (not necessarily the same as in $\mathbf{x}_i$).

5. Heteroscedasticity

Example (Harvey's (1976) restriction)

Harvey (1976) considers a restriction of the form:

$$\mathbb{V}\left(\varepsilon_i \middle| \mathbf{X}\right) = \sigma_i^2\left(\boldsymbol{\alpha}\right) = \exp\left(\mathbf{x}_i^\top \boldsymbol{\alpha}\right)$$

where $\boldsymbol{\alpha} = \left(\alpha_1 : .. : \alpha_H\right)^\top$ is a $H \times 1$ vector of parameters and $\mathbf{z}_i$ is $H \times 1$ of explicative variables (not necessarily the same as in $\mathbf{x}_i$).

5. Heteroscedasticity

We know that the GLS estimator is defined by:

$$\widehat{\beta}_{GLS} = \left(\sum_{i=1}^{N} \frac{\mathbf{x}_i \mathbf{x}_i^{\top}}{\sigma_i^2 \left(\boldsymbol{\alpha} \right)} \right)^{-1} \left(\sum_{i=1}^{N} \frac{\mathbf{x}_i y_i}{\sigma_i^2 \left(\boldsymbol{\alpha} \right)} \right)$$

S, the feasible GLS (FGLS) estimator is:

$$\widehat{\beta}_{FGLS} = \left(\sum_{i=1}^{N} \frac{\mathbf{x}_i \mathbf{x}_i^{\top}}{\sigma_i^2 \left(\widehat{\boldsymbol{\alpha}} \right)} \right)^{-1} \left(\sum_{i=1}^{N} \frac{\mathbf{x}_i y_i}{\sigma_i^2 \left(\widehat{\boldsymbol{\alpha}} \right)} \right)$$

5. Heteroscedasticity

If we assume for instance that

$$\mathbb{V}\left(\varepsilon_i \middle| \mathbf{X}\right) = \sigma_i^2\left(\boldsymbol{\alpha}\right) = \exp\left(\mathbf{z}_i^\top \boldsymbol{\alpha}\right)$$

where $\mathbf{z}_i$ is a vector of H variables, a way to estimate $\boldsymbol{\alpha}$ consists in considering the model:

$$\ln\left(\widehat{\varepsilon}_i^2\right) = \mathbf{z}_i^\top \boldsymbol{\alpha} + v_i$$

and to estimate $\boldsymbol{\alpha}$ by OLS. The OLS is consistent even it is inefficient (due to the heteroscedasticity). Given $\widehat{\boldsymbol{\alpha}}$, we have a consistent estimator for σ_i^2 :

$$\widehat{\sigma}_i^2 = \exp\left(\mathbf{z}_i^\top \widehat{\boldsymbol{\alpha}}\right) \xrightarrow{p} \sigma_i^2\left(\boldsymbol{\alpha}\right)$$

5. Heteroscedasticity

Problem

In order to estimate $\boldsymbol{\beta}$ by the GLS, we need $\widehat{\boldsymbol{\alpha}}$, and to estimate $\boldsymbol{\alpha}$, we need the residuals $\widehat{\varepsilon}_i = y_i - \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}_{GLS} \ldots$

5. Heteroscedasticity

Two solutions

1. A two steps FGLS estimator
2. An iterative FGLS estimator

5. Heteroscedasticity

Definition (Two-steps FGLS estimator)

First step: estimate the parameters $\boldsymbol{\beta}$ by OLS. Compute the residuals $\widehat{\varepsilon}_i = y_i - \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}_{OLS}$ and estimate the parameters $\boldsymbol{\alpha}$ according to the appropriate model. **Second step:** compute the estimated variances $\sigma_i^2 \left(\widehat{\boldsymbol{\alpha}} \right)$ and compute the FGLS estimator:

$$\widehat{\boldsymbol{\beta}}_{FGLS} = \left(\sum_{i=1}^{N} \frac{\mathbf{x}_i \mathbf{x}_i^\top}{\sigma_i^2 \left(\widehat{\boldsymbol{\alpha}} \right)} \right)^{-1} \left(\sum_{i=1}^{N} \frac{\mathbf{x}_i y_i}{\sigma_i^2 \left(\widehat{\boldsymbol{\alpha}} \right)} \right)$$

5. Heteroscedasticity

Definition (Iterated FGLS estimator)

Estimate the parameters $\boldsymbol{\beta}$ by OLS. Compute the residuals $\widehat{\varepsilon}_i = y_i - \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}_{OLS}$ and estimate the parameters $\boldsymbol{\alpha}$ according to the appropriate model. Compute the estimated variances $\sigma_i^2\left(\widehat{\boldsymbol{\alpha}}\right)$ and compute the FGLS estimator:

$$\widehat{\boldsymbol{\beta}}_{FGLS}^{(1)} = \left(\sum_{i=1}^{N} \frac{\mathbf{x}_i \mathbf{x}_i^\top}{\sigma_i^2\left(\widehat{\boldsymbol{\alpha}}\right)} \right)^{-1} \left(\sum_{i=1}^{N} \frac{\mathbf{x}_i y_i}{\sigma_i^2\left(\widehat{\boldsymbol{\alpha}}\right)} \right)$$

Compute the residuals $\widehat{\varepsilon}_i = y_i - \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}_{FGLS}^{(1)}$ and estimate the parameters $\boldsymbol{\alpha}$ according to the appropriate model. Compute the FGLS $\widehat{\boldsymbol{\beta}}_{FGLS}^{(2)}$ and so on...The procedure stop when

$$\sup_{j=1,..,K} \left| \widehat{\boldsymbol{\beta}}_{j,FGLS}^{(i)} - \widehat{\boldsymbol{\beta}}_{j,FGLS}^{(i-1)} \right| < \text{threshold (ex: 0.001)}$$

Example (Harvey's (1976) multiplicative model of heteroscedasticity)

Consider the generalized linear regression model:

$$\text{AVGEXP}_i = \beta_1 + \beta_2 \text{AGE}_i + \beta_3 \text{Ownrent}_i + \beta_4 \text{Income}_i + \beta_5 \text{Income}_i^2 + \varepsilon_i$$

where the heteroscedasticity satisfies the Harvey's (1976) specification

$$\mathbb{V}\left(\varepsilon_i \middle| \mathbf{X}\right) = \sigma_i^2 = \exp\left(\alpha_1 + \alpha_2 \text{Income}_i\right)$$

The data are available in file `Chapter5_data.xls`. Question: write a Matlab code to estimate the parameters by FGLS by using a two-step and an iterative estimator.

Remark

A way to get the estimates of the parameters α_1 and α_2 is to consider the regression:

$$\ln\left(\widehat{\varepsilon}_i^2\right) = \alpha_1 + \alpha_2 \text{Income}_i + v_i$$

and to estimate the parameters by OLS.

5. Heteroscedasticity

```
clear all; clc ; close all
data=xlsread('Chapter5_data.xls');
Age=data(:,3);
Income=data(:,4);
Avgexp=data(:,5);
Ownrent=data(:,6);

y=Avgexp;                                    % Dependent variable
N=length(y);                                 % Sample size
X=[ones(N,1) Age Ownrent Income Income.^2]; % Matrix X
K=size(X,2);                                 % Number of explicative variables

% First step
beta=X\y;                                    % OLS estimates
res=y-X*beta;                                % Residuals
W=[ones(N,1) Income];                        % Matrix W
alpha=W\log(res.^2);

% Second step
Sigma=diag(exp(W*alpha));                        % Matrix Sigma
beta_FGLS2=inv(X'*inv(Sigma)*X)*X'*inv(Sigma)*y; % FGLS

disp(' ') , disp('OLS       FGLS (two-steps)')
disp([beta beta_FGLS2])
```

5. Heteroscedasticity

OLS	FGLS (two-steps)
-115.9914	-35.1646
-3.6537	-3.7218
60.8815	45.5433
156.4672	110.8203
-9.0760	-3.0666

5. Heteroscedasticity

```
clear all; clc ; close all
data=xlsread('Chapter5_data.xls');
Age=data(:,3);
Income=data(:,4);
Avgexp=data(:,5);
Ownrent=data(:,6);

y=Avgexp;                                        % Dependent variable
N=length(y);                                     % Sample size
X=[ones(N,1) Age Ownrent Income Income.^2];      % Matrix X
K=size(X,2);                                     % Number of explicative variables
beta=X\y;                                        % OLS estimates
dif=ones(K,1);

while max(dif)>0.001

    res=y-X*beta;                                        % Residuals
    W=[ones(N,1) Income];                                % Matrix W
    alpha=W\log(res.^2);                                 % Estimated parameters alpha
    Sigma=diag(exp(W*alpha));                            % Matrix Sigma
    beta_FGLS=inv(X'*inv(Sigma)*X)*X'*inv(Sigma)*y;      % FGLS
    dif=beta_FGLS-beta;
    disp([beta beta_FGLS dif])
    beta=beta_FGLS;
end
```

5. Heteroscedasticity

OLS	FGLS (iterated)
-115.9914	8.8438
-3.6537	-3.6947
60.8815	44.0512
156.4672	79.8858
-9.0760	1.6777

Key Concepts

1. OLS and robust inference
2. White heteroscedasticity consistent estimator
3. GLS and Weighted Least Squares (WLS)
4. FGLS: two-steps and iterated estimators

Section 6

Testing for Heteroscedasticity

6. Testing for heteroscedasticity

Objectives

The objective of this section are to introduce the following tests for heteroscedasticity:

1. White general test
2. The Breusch-Pagan / Godfrey LM test

6. Testing for heteroscedasticity

Definition (White test for heteroscedasticity)

The **White test for heteroscedasticity** is based on:

$$\mathrm{H}_0 : \sigma_i^2 = \sigma^2 \quad \text{for } i = 1, .., N$$

$$\mathrm{H}_1 : \sigma_i^2 \neq \sigma_j^2 \quad \text{for at least one pair } \; (i, j)$$

6. Testing for heteroscedasticity

The intuition of the test is based on the following idea:

1. If there is no heteroscedasticity (under the null H_0):

$$\mathbb{V}_{asy}\left(\widehat{\beta}_{OLS}\right) = \sigma^2 \mathbf{Q}^{-1}$$

$$\widehat{\mathbb{V}}_{asy}\left(\widehat{\beta}_{OLS}\right) = \sigma^2 \left(\mathbf{X}^\top \mathbf{X}\right)^{-1}$$

2. Under the alternative (heteroscedasticity):

$$\mathbb{V}_{asy}\left(\widehat{\beta}_{OLS}\right) = \sigma^2 \mathbf{Q}^{-1} \mathbf{Q}^{*} \mathbf{Q}^{-1}$$

$$\widehat{\mathbb{V}}_{asy}\left(\widehat{\beta}_{OLS}\right) = \sigma^2 \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \Omega \mathbf{X} \left(\mathbf{X}^\top \mathbf{X}\right)^{-1}$$

6. Testing for heteroscedasticity

White (1980) proposes the following procedure and test-statistic:

Step 1: Estimation of the model using the OLS estimator of $\boldsymbol{\beta}$.

Step 2: Determine the residuals $\widehat{\varepsilon}_i = y_i - \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}_{OLS}$.

Step 3: Regress $\widehat{\varepsilon}_i^2$ on a constant and all unique columns vectors contained in $\mathbf{X}$ and all the squares and cross-products of the column vectors in $\mathbf{X}$.

Step 4: Determine the coefficient of determination, R^2, of the previous regression.

6. Testing for heteroscedasticity

Definition (White test for heteroscedasticity)

Under the null, the **White test-statistic** $N \times \mathrm{R}^2$ converges:

$$N \times \mathrm{R}^2 \underset{\mathrm{H}_0}{\overset{d}{\rightarrow}} \chi^2 \left(m - 1 \right)$$

where m is the number of explanatory variables in the regression of $\widehat{\varepsilon}_i^2$. The critical region of size α is

$$\mathrm{W} = \left\{ y : N \times \mathrm{R}^2 > \chi^2_{1-\alpha} \right\}$$

where $\chi^2_{1-\alpha}$ denotes the 1-α critical value of the $\chi^2 \left(m - 1 \right)$ distribution.

6. Testing for heteroscedasticity

Example (White's (1980) test for heteroscedasticity)

Consider the generalized linear regression model:

$$\text{AVGEXP}_i = \beta_1 + \beta_2 \text{AGE}_i + \beta_3 \text{Ownrent}_i + \beta_4 \text{Income}_i + \beta_5 \text{Income}_i^2 + \varepsilon_i$$

The data are available in file `Chapter5_data.xls`. Question: write a Matlab code to compute the White test-statistic for heteroscedasticity and its p-value. What is you conclusion for a significance level of 5%? Compare your results with Eviews.

6. Testing for heteroscedasticity

```
clear all; clc ; close all
data=xlsread('Chapter5_data.xls');
Age=data(:,3);
Income=data(:,4);
Avgexp=data(:,5);
Ownrent=data(:,6);

y=Avgexp;                                    % Dependent variable
N=length(y);                                 % Sample size
X=[ones(N,1) Age Ownrent Income Income.^2]; % Matrix X
K=size(X,2);                                 % Number of explicative variables
beta=X\y;                                    % OLS estimates
res=y-X*beta;                                % Residuals
W=[ones(N,1) Age Age.^2 Age.*Ownrent Age.*Income ...
    Age.*Income.^2 Ownrent Ownrent.*Income ...
    Ownrent.*Income.^2 Income Income.^2 Income.*Income.^2 Income.^4];
gam=W\(res.^2);                              % Estimate of the regression of eps^2
res2=res.^2-W*gam;
R2=1-var(res2)/var(res.^2);                  % R2
White=R2*N;                                  % White statistic
pvalue=1-chi2cdf(White,size(W,2)-1);         % pvalue
```

6. Testing for heteroscedasticity

White Heteroskedasticity Test:

F-statistic	1.244819	Probability	0.266541
Obs*R-squared	14.65386	Probability	0.260914

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 12/14/13 Time: 21:00
Sample: 1 100
Included observations: 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	876511.9	913863.8	0.959128	0.3402
AGE	28775.90	31660.00	0.908904	0.3659
AGE^2	-644.2271	425.9743	-1.512361	0.1341
AGE*OWNRENT	5681.491	8776.134	0.647380	0.5191
AGE*INCOME	6853.915	11227.53	0.610456	0.5432
AGE*INCOME2	-647.8628	1274.148	-0.508467	0.6124
OWNRENT	195763.1	474111.1	0.412905	0.6807
OWNRENT*INCOME	-177650.5	199416.6	-0.890851	0.3755
OWNRENT*INCOME2	11325.35	21530.66	0.526010	0.6002
INCOME	-1509045.	778264.9	-1.938986	0.0557
INCOME^2	498964.2	253154.3	1.970989	0.0519
INCOME*INCOME2	-63934.08	34454.00	-1.855636	0.0669
INCOME2^2	2820.726	1630.189	1.730306	0.0871

R-squared	0.146539	Mean dependent var	70384.57
Adjusted R-squared	0.028820	S.D. dependent var	287729.4
S.E. of regression	283552.9	Akaike info criterion	28.06892
Sum squared resid	6.99E+12	Schwarz criterion	28.40759
Log likelihood	-1390.446	F-statistic	1.244819
Durbin-Watson stat	1.745177	Prob(F-statistic)	0.266541

```
White =

   14.6539

pvalue =

    0.2609
```

6. Testing for heteroscedasticity

Definition (Breusch and Pagan test)

Breusch and Pagan (1979) have devised a **Lagrange multiplier test** of the hypothesis that

$$\sigma_i^2 = \sigma^2 f\left(\alpha_0 + \mathbf{z}_i^\top \boldsymbol{\alpha}\right)$$

where $\mathbf{z}_i = (z_{i1}..z_{ip})^\top$ is a $p \times 1$ vector of independent variables. The test is:

$$\mathrm{H}_0 : \boldsymbol{\alpha} = \mathbf{0}_{p\times 1} \quad \text{(homoscedasticity)}$$

$$\mathrm{H}_1 : \boldsymbol{\alpha} \neq \mathbf{0}_{p\times 1} \quad \text{(heteroscedasticity)}$$

6. Testing for heteroscedasticity

The test can be carried out with a simple regression of

$$g_i = N \frac{\widehat{\varepsilon}_i^2}{\widehat{\boldsymbol{\varepsilon}}^\top \widehat{\boldsymbol{\varepsilon}}} - 1 = N \frac{\widehat{\varepsilon}_i^2}{\sum_{i=1}^N \widehat{\varepsilon}_i^2} - 1$$

on the variables z_{ik} for $k = 1, ., N$ and a constant term.

$$g_i = \alpha_0 + \alpha_1 z_{i1} + ... + \alpha_p z_{ip} + v_i$$

6. Testing for heteroscedasticity

Definition (Breusch and Pagan test-statistic)

Define $\mathbf{Z}$ the $N \times (p+1)$ matrix of observations on $(1, \mathbf{z}_i)$ and let $\mathbf{g}$ be the $N \times 1$ vector of observations

$$g_i = N \frac{\widehat{\varepsilon}_i^2}{\widehat{\boldsymbol{\varepsilon}}^\top \widehat{\boldsymbol{\varepsilon}}} - 1$$

Then, the **Breusch and Pagan's test-statistic** is defined by:

$$\mathrm{LM} = \frac{1}{2} \mathbf{g}^\top \mathbf{Z} \left(\mathbf{Z}^\top \mathbf{Z} \right)^{-1} \mathbf{Z}^\top \mathbf{g}$$

Under the null, we have:

$$\mathrm{LM} \underset{\mathrm{H}_0}{\overset{d}{\to}} \chi^2 (p)$$

6. Testing for heteroscedasticity

Example (Breusch and Pagan's (1979) test for heteroscedasticity)

Consider the generalized linear regression model:

$$\text{AVGEXP}_i = \beta_1 + \beta_2 \text{AGE}_i + \beta_3 \text{Ownrent}_i + \beta_4 \text{Income}_i + \beta_5 \text{Income}_i^2 + \varepsilon_i$$

The data are available in file `Chapter5_data.xls`. Question: write a Matlab code to compute the Breusch and Pagan test-statistic for heteroscedasticity with $\mathbf{z}_i = \mathbf{x}_i$ and its p-value. What is you conclusion for a significance level of 5%?

6. Testing for heteroscedasticity

```
clear all; clc ; close all
data=xlsread('Chapter5_data.xls');
Age=data(:,3);
Income=data(:,4);
Avgexp=data(:,5);
Ownrent=data(:,6);

y=Avgexp;                                      % Dependent variable
N=length(y);                                   % Sample size
X=[ones(N,1) Age Ownrent Income Income.^2];    % Matrix X
K=size(X,2);                                   % Number of explicative variables
beta=X\y;                                      % OLS estimates
res=y-X*beta;                                  % Residuals

g=N*res.^2/sum(res.^2)-1;                      % G vector
Z=X;                                           % We use z=x
LM=0.5*g'*Z*inv(Z'*Z)*Z'*g;                    % LM test-statistic
pvalue=1-chi2cdf(LM,size(Z,2)-1);              % The constant is not considered in the DF
```

6. Testing for heteroscedasticity

```
LM =

   59.7983


pvalue =

  3.1982e-012
```

6. Testing for heteroscedasticity

Key Concepts

1. White test for heteroscedasticity
2. Breusch and Pagan test for heteroscedasticity

End of Chapter 5

Christophe Hurlin (University of Orléans)