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

- Pelgrin, F. (2010), Lecture notes Advanced Econometrics, HEC Lausanne (a special thank)
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

\[ y \]  vector

\[ Y \]  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:

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:

\[ y = X\beta + \varepsilon \]

where (cf. chapter 3):

- \( y \) is a \( N \times 1 \) vector of observations \( y_i \) for \( i = 1, \ldots, N \)
- \( X \) is a \( N \times K \) matrix of \( K \) explicative variables \( x_{ik} \) for \( k = 1, \ldots, K \) and \( i = 1, \ldots, N \)
- \( \varepsilon \) is a \( N \times 1 \) vector of error terms \( \varepsilon_i \).
- \( \beta = (\beta_1 \ldots \beta_K)^\top \) 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):

\[ \mathbb{V}(\varepsilon | X) = \sigma^2 I_N \]

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:

\[ y = X\beta + \varepsilon \]

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

\[
\begin{align*}
\mathbb{E}(\varepsilon|X) &= 0_{N \times 1} \\
\mathbb{V}(\varepsilon|X) &= \Sigma = \sigma^2 \Omega
\end{align*}
\]

where \( \Omega \) and \( \Sigma \) are symmetric positive definite matrices.
2. The generalized linear regression model

Reminder

$$\mathbb{V} (\varepsilon | X) = \mathbb{E} (\varepsilon \varepsilon^T | X)$$

$$\mathbb{E} (\varepsilon_1\varepsilon_2 | X) \ldots \mathbb{Cov} (\varepsilon_1\varepsilon_N | X)$$

$$\mathbb{V} (\varepsilon_1 | X) \ldots \mathbb{Cov} (\varepsilon_1\varepsilon_2 | X)$$

$$\mathbb{E} (\varepsilon_2\varepsilon_1 | X) \ldots \mathbb{V} (\varepsilon_2^2 | X)$$

$$\ldots \ldots \ldots \ldots$$

$$\mathbb{Cov} (\varepsilon_N\varepsilon_1 | X) \ldots \mathbb{V} (\varepsilon_N^2 | X)$$
2. The generalized linear regression model

**Remark**

In the generalized linear regression model, we have

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

with

$$\Sigma = \begin{pmatrix} \sigma^2_1 & \sigma_{12} & \ldots & \sigma_{1N} \\ \sigma_{21} & \sigma^2_2 & \ldots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \ldots & \ldots & \sigma^2_N \end{pmatrix} = \sigma^2 \begin{pmatrix} \omega_{11} & \omega_{12} & \ldots & \omega_{1N} \\ \omega_{21} & \omega_{22} & \ldots & \omega_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{N1} & \ldots & \ldots & \omega_{NN} \end{pmatrix}$$

and \( \omega_{ij} = \sigma_{ij} / \sigma^2 \).
2. The generalized linear regression model

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

\[
V(\epsilon_i | X) \neq V(\epsilon_j | X) \quad \text{for} \ i \neq j
\]
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 \( \Omega \) and \( \Sigma \) would be:

\[
\Sigma = \begin{pmatrix}
\sigma_1^2 & 0 & \ldots & 0 \\
0 & \sigma_2^2 & \ldots & 0 \\
\ldots & \ldots & \ldots & \ldots \\
0 & \ldots & \ldots & \sigma_N^2
\end{pmatrix} = \sigma^2 \Omega = \sigma^2 \begin{pmatrix}
\omega_1 & 0 & \ldots & 0 \\
0 & \omega_2 & \ldots & 0 \\
\ldots & \ldots & \ldots & \ldots \\
0 & \ldots & \ldots & \omega_N
\end{pmatrix}
\]

with \( \omega_i = \frac{\sigma_i^2}{\sigma^2} \) for \( i = 1, \ldots, N \).
Definition (Autocorrelation)

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

\[ \text{Cov}(\varepsilon_i, \varepsilon_j | X) \neq 0 \quad \text{for} \quad i \neq j \]
2. The generalized linear regression model

Example (Autocorrelation)

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

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

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

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

\[ y = X\beta + \varepsilon \]

\[ \mathbb{E}(\varepsilon|X) = 0_{N \times 1} \]

\[ \mathbb{V}(\varepsilon|X) = \sigma^2 \Omega = \Sigma \]

Now consider the OLS estimator, denoted \( \hat{\beta}_{OLS} \), of the parameters \( \beta \):

\[ \hat{\beta}_{OLS} = \left( X^\top X \right)^{-1} X^\top y \]

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:

\[ \mathbb{E} (\varepsilon \mid X) = 0_{N \times 1} \]
3. Inefficiency of the Ordinary Least Squares

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**:

\[
E \left( \hat{\beta}_{OLS} \right) = \beta_0
\]

where \( \beta_0 \) 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.
3. Inefficiency of the Ordinary Least Squares

Proof

\[ \hat{\beta}_{OLS} = (X^\top X)^{-1} (X^\top y) = \beta_0 + (X^\top X)^{-1} (X^\top \varepsilon) \]

So we have:

\[ \mathbb{E} \left( \hat{\beta}_{OLS} \mid X \right) = \beta_0 + (X^\top X)^{-1} (X^\top \mathbb{E} (\varepsilon \mid X)) \]

Under assumption A3 (exogeneity), \( \mathbb{E} (\varepsilon \mid X) = 0 \). Then, we get:

\[ \mathbb{E} \left( \hat{\beta}_{OLS} \mid X \right) = \beta_0 \]
Proof (cont’d)

\[ \mathbb{E} \left( \hat{\beta}_{OLS} \mid X \right) = \beta_0 \]

So, we have:

\[ \mathbb{E} \left( \hat{\beta}_{OLS} \right) = \mathbb{E}_X \left( \mathbb{E} \left( \hat{\beta}_{OLS} \mid X \right) \right) = \mathbb{E}_X (\beta_0) = \beta_0 \]

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

The OLS estimator is unbiased:

\[ \mathbb{E} \left( \hat{\beta}_{OLS} \right) = \beta_0 \] \( \Box \)
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

\[
V \left( \hat{\beta}_{OLS} \mid \mathbf{x} \right) = \sigma_0^2 \left( \mathbf{x}^\top \mathbf{x} \right)^{-1} \mathbf{x}^\top \Omega \mathbf{x} \left( \mathbf{x}^\top \mathbf{x} \right)^{-1}
\]

and a variance covariance matrix given by:

\[
V \left( \hat{\beta}_{OLS} \right) = \mathbb{E}_X \left( V \left( \hat{\beta}_{OLS} \mid \mathbf{x} \right) \right)
\]
3. Inefficiency of the Ordinary Least Squares

Proof

\[ \hat{\beta}_{OLS} = (X^\top X)^{-1} (X^\top y) = \beta_0 + (X^\top X)^{-1} (X^\top \varepsilon) \]

So we have:

\[ \text{Var} \left( \hat{\beta}_{OLS} \mid X \right) = \mathbb{E} \left( \left( (X^\top X)^{-1} X^\top \varepsilon \varepsilon^\top X \right) \left( (X^\top X)^{-1} \right) \right) \]
\[ = (X^\top X)^{-1} X^\top \mathbb{E} \left( \varepsilon \varepsilon^\top \mid X \right) X \left( (X^\top X)^{-1} \right) \]
\[ = \sigma_0^2 (X^\top X)^{-1} X^\top \Omega X \left( (X^\top X)^{-1} \right) \]
3. Inefficiency of the Ordinary Least Squares

Definition (Variance estimator)

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

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

where $\hat{\sigma}^2 \hat{\Omega}$ is a consistent estimator of $\Sigma = \sigma^2 \Omega$. 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:

\[ \hat{\beta}_{OLS} \mid X \sim \mathcal{N} \left( \beta_0, \sigma^2 \left( X^\top X \right)^{-1} X^\top \Omega X \left( X^\top X \right)^{-1} \right) \]
3. Inefficiency of the Ordinary Least Squares

Asymptotic properties of the OLS estimator
3. Inefficiency of the Ordinary Least Squares

Assumptions

\[
\text{plim} \frac{1}{N} X^\top X = Q
\]
\[
\text{plim} \frac{1}{N} X^\top \Omega X = Q^*
\]

where:

1. \(Q^*\) is a \(K \times K\) finite (non null) definite positive matrix

2. \(Q\) is a \(K \times K\) finite (non null) definite positive matrix with

\[
\text{rank} (Q) = K
\]
3. Inefficiency of the Ordinary Least Squares

Definition (Consistency of the OLS estimator)

If \( \lim N^{-1}X^\top \Omega X \) and \( \lim N^{-1}X^\top X \) are both finite positive definite matrices, then \( \hat{\beta}_{OLS} \) is a consistent estimator of \( \beta \):

\[
\hat{\beta}_{OLS} \xrightarrow{p} \beta_0
\]
3. Inefficiency of the Ordinary Least Squares

Proof

\[ \hat{\beta}_{OLS} = \beta_0 + \left( X^\top X \right)^{-1} \left( X^\top \epsilon \right) \]

We know that under assumption A3 (exogeneity):

\[ \operatorname{plim} \frac{1}{N} X^\top \epsilon = 0_{K \times 1} \]

\[ \operatorname{plim} \frac{1}{N} X^\top X = Q \]

So, we have

\[ \operatorname{plim} \hat{\beta}_{OLS} = \beta_0 \]

So, the estimator \( \hat{\beta} \) 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

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

where

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

3. Inefficiency of the Ordinary Least Squares

**Definition (Asymptotic variance)**

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

\[
V_{asy} \left( \hat{\beta}_{OLS} \right) = \frac{\sigma^2}{N} Q^{-1} Q^* Q^{-1}
\]

with

\[
Q = \text{plim} \frac{1}{N} X^\top X \\
Q^* = \text{plim} \frac{1}{N} X^\top \Omega X
\]
## 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

\[
\hat{\sigma}^2 \left( \mathbf{X}^\top \mathbf{X} \right)^{-1}
\]

may be misleading. For instance the t-test-statistic:

\[
t_{\beta_k} = \frac{\hat{\beta}_k}{\hat{\sigma} \sqrt{m_{kk}}}
\]

where \( m_{kk} \) is \( k^{th} \) diagonal element of \( \mathbf{X}^\top \mathbf{X} \) do not have a Student distribution.
3. Inefficiency of the Ordinary Least Squares

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

$$\hat{\Sigma}_{asy} \left( \hat{\beta}_{OLS} \right) = \frac{\sigma^2}{N} \left( \frac{1}{N} x^\top x \right)^{-1} \left( \frac{1}{N} x^\top \Omega x \right) \left( \frac{1}{N} x^\top x \right)^{-1}$$
3. Inefficiency of the Ordinary Least Squares

Proof

By definition:

\[ Q = \text{plim} \frac{1}{N} X^\top X \]

\[ Q^* = \text{plim} \frac{1}{N} X^\top \Omega X \]

So,

\[ \text{plim} \widehat{V}_{asy} \left( \hat{\beta}_{OLS} \right) = \text{plim} \frac{\sigma^2}{N} \left( \frac{1}{N} X^\top X \right)^{-1} \left( \frac{1}{N} X^\top \Omega X \right) \left( \frac{1}{N} X^\top X \right)^{-1} \]

\[ = \frac{\sigma^2}{N} Q^{-1} Q^* Q^{-1} \]

Or equivalently

\[ \widehat{V}_{asy} \left( \hat{\beta}_{OLS} \right) \xrightarrow{p} \mathbb{V}_{asy} \left( \hat{\beta}_{OLS} \right) \]
3. Inefficiency of the Ordinary Least Squares

Reminder

\[
X^T X = \sum_{i=1}^{N} x_i x_i^T
\]

\[
X^T \Omega X = \sum_{i=1}^{N} \sum_{j=1}^{N} \omega_{ij} x_i x_i^T
\]

\[
X^T \Sigma X = \sum_{i=1}^{N} \sum_{j=1}^{N} \sigma_{ij} x_i x_i^T = \sigma^2 \sum_{i=1}^{N} \sum_{j=1}^{N} \omega_{ij} x_i x_i^T
\]
3. Inefficiency of the Ordinary Least Squares

**Remark**

The estimator

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

can also be written as

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

In the next section, we will give a feasible estimator $\hat{V}_{asy}(\hat{\beta}_{OLS})$ 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 \( \sigma^2 \left( X^\top X \right)^{-1} \) is misplaced.

2. The OLS is inefficient.

\[
\nabla \left( \hat{\beta}_{OLS} \right) - I_N^{-1}(\beta_0) \text{ is a positive definite matrix}
\]
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

\[ \mathbb{V}(\varepsilon|X) = \Sigma = \sigma^2 \Omega \]

We will distinguish two cases:

**Case 1:** the variance covariance matrix \( \Sigma \) is known (unrealistic case)

**Case 2:** the variance covariance matrix \( \Sigma \) is unknown
4. Generalized Least Squares (GLS)

Case 1: $\Sigma$ is known

The Generalized Least Squares (GLS) estimator
4. Generalized Least Squares (GLS)

**Definition (Factorisation)**

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

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

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

$$\mathbf{C}^\top \mathbf{C} = \mathbf{C} \mathbf{C}^\top = I_N$$

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

We define the matrix \( P \) such that

\[
P^\top = C \Lambda^{-1/2}
\]

so that

\[
\Omega^{-1} = P^\top P
\]
4. Generalized Least Squares (GLS)

Proof

\[ P^T = C \Lambda^{-1/2} \]

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

\[ P^T P = C \Lambda^{-1/2} \Lambda^{-1/2} C^T = C \Lambda^{-1} C^T \]

Consider the quantity \( P^T P \Omega \):

\[
\begin{align*}
P^T P \Omega &= C \Lambda^{-1} C^T C \Lambda C^T \\
&= C \Lambda^{-1} \Lambda C^T \\
&= C C^T \\
&= I_N
\end{align*}
\]

Since \( C \) satisfies \( CC^T = I_N \). Then, \( P^T P = \Omega^{-1} \) \( \square \)
4. Generalized Least Squares (GLS)

**GLS estimator**

Premultiply the generalized linear regression model by $P$ to obtain

$$Py = PX\beta + P\varepsilon$$

or equivalently

$$y^* = X^* \beta + \varepsilon^*$$

The conditional variance of $\varepsilon^*$ is

$$V(\varepsilon^* | X) = \mathbb{E} \left( \varepsilon^* \varepsilon^{*\top} \middle| X \right)$$

$$= PE \left( \varepsilon \varepsilon^{\top} \middle| X \right) P^{\top}$$

$$= \sigma^2 P \Omega P^{\top}$$

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

$$= \sigma^2 I_N$$
4. Generalized Least Squares (GLS)

GLS estimator (cont’d)

\[ y^* = X^* \beta + \varepsilon^* \]
\[ \mathbb{V} (\varepsilon^* | X) = \sigma^2 I_N \]

The classical regression model applies to this transformed model.

If \( \Omega \) is assumed to be known, \( y^* = Py \) and \( X^* = PX \) are observed data. So, we can apply the ordinary least squares to this transformed model:

\[ \hat{\beta} = \left( X^{*\top} X^* \right)^{-1} \left( X^{*\top} y^* \right) \]
4. Generalized Least Squares (GLS)

**GLS estimator (cont’d)**

\[
\hat{\beta} = (X^* \mathbf{X}^*)^{-1} (X^* \mathbf{y}^*) \\
= (X^T \mathbf{P}^T \mathbf{PX})^{-1} (X^T \mathbf{P}^T \mathbf{Py}) \\
= (X^T \mathbf{\Omega}^{-1} \mathbf{X})^{-1} (X^T \mathbf{\Omega}^{-1} \mathbf{y})
\]

This estimator is the generalized least squares (GLS) estimator of \( \beta \).
Definition (GLS estimator)

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

$$\hat{\beta}_{GLS} = \left(X^\top \Omega^{-1} X\right)^{-1} \left(X^\top \Omega^{-1} y\right)$$
4. Generalized Least Squares (GLS)

**Definition (Bias)**

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

$$\mathbb{E} \left( \hat{\beta}_{GLS} \right) = \beta_0$$

where $\beta_0$ denotes the true value of the parameters.
4. Generalized Least Squares (GLS)

Proof

We have:

\[ \hat{\beta}_{GLS} = \left( X^\top \Omega^{-1} X \right)^{-1} \left( X^\top \Omega^{-1} y \right) = \beta_0 + \left( X^\top \Omega^{-1} X \right)^{-1} \left( X^\top \Omega^{-1} \varepsilon \right) \]

So,

\[ \mathbb{E} \left( \hat{\beta}_{GLS} \bigg| X \right) = \beta_0 + \left( X^\top \Omega^{-1} X \right)^{-1} \left( X^\top \Omega^{-1} \mathbb{E} \left( \varepsilon \big| X \right) \right) \]

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

\[ \mathbb{E} \left( \hat{\beta}_{GLS} \bigg| X \right) = \beta_0 \]

and

\[ \mathbb{E} \left( \hat{\beta}_{GLS} \right) = \mathbb{E}_X \left( \mathbb{E} \left( \hat{\beta}_{GLS} \bigg| X \right) \right) = \mathbb{E}_X \left( \beta_0 \right) = \beta_0 \]
4. Generalized Least Squares (GLS)

**Definition (Variance covariance matrix)**

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

$$\text{Var} \left( \hat{\beta}_{GLS} \left| X \right. \right) = \sigma^2 \left( X^\top \Omega^{-1} X \right)^{-1}$$

The variance covariance matrix is given by

$$\text{Var} \left( \hat{\beta}_{GLS} \right) = \sigma^2 \mathbb{E}_X \left( \left( X^\top \Omega^{-1} X \right)^{-1} \right)$$
4. Generalized Least Squares (GLS)

Proof

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

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

$$V \left( \hat{\beta}_{GLS} \Big| X \right) = \left( X^*^\top X^* \right)^{-1} X^*^\top E \left( \varepsilon^* \varepsilon^{*\top} \Big| X \right) X^* \left( X^*^\top X^* \right)^{-1}$$

Since $E \left( \varepsilon^* \varepsilon^{*\top} \Big| X \right) = \sigma^2 I_N$, we have

$$V \left( \hat{\beta}_{GLS} \Big| X \right) = \sigma^2 \left( X^*^\top X^* \right)^{-1} X^*^\top X^* \left( X^*^\top X^* \right)^{-1}$$

$$= \sigma^2 \left( X^*^\top X^* \right)^{-1}$$

$$= \sigma^2 \left( X^\top P^\top PX \right)^{-1}$$

$$= \sigma^2 \left( X^\top \Omega^{-1} X \right)^{-1} \quad \Box$$
4. Generalized Least Squares (GLS)

**Definition (Consistency)**

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

$$\hat{\beta}_{GLS} \overset{p}{\rightarrow} \beta_0$$

as soon as

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

where $Q^*$ is a finite positive definite matrix.
4. Generalized Least Squares (GLS)

**Proof**

\[
\hat{\beta}_{GLS} = \beta_0 + \left( X^\top \Omega^{-1} X \right)^{-1} \left( X^\top \Omega^{-1} \varepsilon \right)
\]

Under the assumption A3 (exogeneity):

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

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

So, we have

\[
\text{plim} \hat{\beta}_{GLS} = \beta_0
\]

The estimator \( \hat{\beta}_{GLS} \) is weakly consistent. \( \square \)
4. Generalized Least Squares (GLS)

**Definition (Asymptotic distribution)**

Under some regularity conditions, the GLS estimator \( \hat{\beta}_{GLS} \) is asymptotically normally distributed:

\[
\sqrt{N} \left( \hat{\beta}_{GLS} - \beta_0 \right) \overset{d}{\rightarrow} \mathcal{N} \left( 0, \sigma^2 Q^{*-1} \right)
\]

where

\[
Q^* = \text{plim} \frac{1}{N} \mathbf{X}^* \mathbf{X}^* = \text{plim} \frac{1}{N} \mathbf{X}^\top \Omega^{-1} \mathbf{X}
\]
4. Generalized Least Squares (GLS)

Definition (Asymptotic variance covariance matrix)

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

$$V_{asy} \left( \hat{\beta}_{GLS} \right) = \frac{\sigma^2}{N} Q^{-1}$$

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

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

This estimator holds whether $X$ is stochastic or non-stochastic.
4. Generalized Least Squares (GLS)

Theorem (BLUE estimator)

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

$$V_{asy} (\hat{\beta}_{OLS}) - V_{asy} (\hat{\beta}_{GLS})$$

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 $\hat{\beta}_{GLS}$ is efficient

$$\nabla \left( \hat{\beta}_{GLS} \right) = I_{N}^{-1}(\beta_0)$$

where $I_{N}^{-1}(\beta_0)$ 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:

\[
L_N (\theta; y \mid x) = \left(2\pi\sigma^2\right)^{-N/2} \left| \Omega \right|^{-N/2} \times \exp \left(-\frac{1}{2\sigma^2} (y - X\beta)^\top \Omega^{-1} (y - X\beta) \right)
\]

The log-likelihood is defined as to be:

\[
\ell_N (\theta; y \mid x) = -\frac{N}{2} \ln (2\pi\sigma^2) - \frac{N}{2} \log (|\Omega|) - \frac{1}{2\sigma^2} (y - X\beta)^\top \Omega^{-1} (y - X\beta)
\]
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 $H_0 : R\beta = q$, the appropriate test-statistic is:

$$F = \frac{1}{p} \left( \hat{R}\beta_{GLS} - q \right)^\top \left( \sigma^2 R \left( X^\top \Omega^{-1} X \right)^{-1} R^\top \right)^{-1} \left( \hat{R}\beta_{GLS} - q \right)$$
Fact

To summarize, all the results for the **classical model**, including the usual inference procedures, apply to the **transformed model**.
Case 2: $\Sigma$ is unknown

The Feasible Generalized Least Squares (FGLS) estimator
4. Generalized Least Squares (GLS)

Introduction

1. If $\Sigma$ contains unknown parameters that must be estimated, then generalized least squares is not feasible.

2. With an unrestricted matrix $\Sigma = \sigma^2 \Omega$, there are $N(N + 1)/2$ additional parameters (since $\Sigma$ 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 \( \alpha \):

\[
\mathbb{V} (\varepsilon | X) = \sigma^2 \Omega (\alpha)
\]
4. Generalized Least Squares (GLS)

Example (Time series)

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

\[
\Omega(\rho) = \begin{pmatrix}
1 & \rho & \rho^2 & \rho^3 & \ldots & \rho^{N-1} \\
\rho & 1 & \rho & \rho^2 & \ldots & \rho^{N-2} \\
\rho^2 & \rho & 1 & \rho & \ldots & \rho^{N-3} \\
\rho^3 & \rho^2 & \rho & 1 & \ldots & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots & \ddots \\
\rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \ldots & \ldots & 1
\end{pmatrix}
\]
Example (Heteroscedasticity)

If we consider a heteroscedastic model, where the variance of \( \varepsilon_i \) depends on a variable \( z_i \), with

\[
\mathbb{V} \left( \varepsilon_i \mid X \right) = \sigma^2 z_i^{\theta}
\]

we have

\[
\Omega \left( \theta \right) = \begin{pmatrix}
z_1^{\theta} & 0 & 0 & \ldots & 0 \\
0 & z_2^{\theta} & 0 & \ldots & 0 \\
0 & 0 & z_3^{\theta} & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & z_N^{\theta}
\end{pmatrix}
\]
4. Generalized Least Squares (GLS)

**Definition (Feasible Generalized Least Squares (FGLS))**

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

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

where $\hat{\Omega} = \Omega(\hat{\alpha})$ is a consistent estimator of $\Omega(\alpha)$. 
4. Generalized Least Squares (GLS)

**Remark**

If

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

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

Then the GLS and FGLS estimators are asymptotically equivalent

$$\hat{\beta}_{FGLS} - \hat{\beta}_{GLS} \xrightarrow{p} 0_{K \times 1}$$
4. Generalized Least Squares (GLS)

Theorem (Efficiency)

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

If the estimator $\hat{\alpha}$ is consistent

$$\hat{\alpha} \xrightarrow{p} \alpha$$

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

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
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)
5. Heteroscedasticity

**Introduction**

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

\[ \nabla (\epsilon_i \mid \mathbf{X}) = \sigma_i^2 \quad \text{for } i = 1, \ldots, N \]

**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:

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

with

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

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

Definition (Scaling)

The fact to scale the variances as

\[ \sigma_i^2 = \sigma^2 \omega_i \quad \text{for } i = 1, \ldots, N \]

allows us to use a normalisation on \( \Omega \)

\[ \text{trace} (\Omega) = \sum_{i=1}^{N} \omega_i = N \]
5. Heteroscedasticity

**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 $\Sigma$ is known: GLS or Weighted Least Square (WLS)

**Case 3:** the variance covariance matrix $\Sigma$ is unknown but its form (structure) is known: two-steps or iterated FGLS estimator
5. Heteroscedasticity

The imbroglio of heteroscedasticity...

- Heteroscedasticity of unknown form
  - OLS is consistent but inefficient
    - Robust inference (White correction)

- Heteroscedasticity with $\Sigma$ known
  - GLS is efficient WLS estimator
    - Classical inference No correction

- Heteroscedasticity with a given structure for $\Sigma$
  - FGLS is efficient 2-step or iterated estimator
    - Classical inference No correction
5. Heteroscedasticity

Case 1: Heteroscedasticity of unknown form

OLS and robust inference
5. Heteroscedasticity

**Assumption:** We assume that the variances $\sigma_i^2$ are unknown for $i = 1, \ldots, N$ and no particular form (structure) is imposed on $\Omega$ (or $\Sigma$).
5. Heteroscedasticity

Introduction

1. The GLS cannot be implemented since $\Sigma$ is unknown.

2. The FGLS estimator requires to estimate (in a first step) $N$ parameters $\sigma_1^2, \ldots, \sigma_N^2$. With $N$ observations, the FGLS is not feasible.

3. The only solution to estimate $\beta$ 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:

\[ \hat{\beta}_{OLS} = \left( X^\top X \right)^{-1} X^\top y \]

We know that

\[ \hat{\beta}_{OLS} \overset{asy}{\sim} \mathcal{N} \left( \beta_0, \frac{\sigma^2}{N} Q^{-1} Q^* Q^{-1} \right) \]

\[ \mathbb{V}_{asy} \left( \hat{\beta}_{OLS} \right) = \frac{\sigma^2}{N} Q^{-1} Q^* Q^{-1} \]

with

\[ Q = \text{plim} \frac{1}{N} X^\top X \quad Q^* = \text{plim} \frac{1}{N} X^\top \Omega X \]
5. Heteroscedasticity

Problem (Robust inference with OLS)

The conventionally estimated covariance matrix for the least squares estimator \( \sigma^2 (X^\top X)^{-1} \) is inappropriate; the appropriate matrix is \( \sigma^2 (X^\top X)^{-1} (X^\top \Omega X)^{-1} (X^\top X)^{-1} \). 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 \( \sigma^2 (X^\top X)^{-1} \) is misleading.
5. Heteroscedasticity

**Question**

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

\[
\mathbb{V}_{asy} \left( \hat{\beta}_{OLS} \right) = \frac{\sigma^2}{N} Q^{-1} Q^* Q^{-1}
\]

\[
Q = \text{plim} \frac{1}{N} X^\top X \quad Q^* = \text{plim} \frac{1}{N} X^\top \Omega X
\]
5. Heteroscedasticity

We seek an estimator for

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

or equivalently of

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

with

$$Q^{**} = \sigma^2 Q^*$$
5. Heteroscedasticity

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

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

\[ S_0 = \frac{1}{N} \sum_{i=1}^{N} \hat{\varepsilon}_i^2 \mathbf{x}_i \mathbf{x}_i^\top \]

where \( \hat{\varepsilon}_i = y_i - \mathbf{x}_i^\top \hat{\beta}_{OLS} \), converges to \( Q^{**} = \sigma^2 Q^* \)

\[ S_0 \xrightarrow{p} Q^{**} = \sigma^2 Q^* \]

5. Heteroscedasticity

\[ \text{\textbf{V}}_{\text{asy}} \left( \hat{\beta}_{\text{OLS}} \right) = \frac{\sigma^2}{N} Q^{-1} Q^* Q^{-1} \]

We know that:

\[ S_0 = \frac{1}{N} \sum_{i=1}^{N} \varepsilon_i^2 x_i x_i^\top \overset{p}{\to} \sigma^2 Q^* \]

\[ \left( \frac{1}{N} X^\top X \right)^{-1} = \left( \frac{1}{N} \sum_{i=1}^{N} x_i x_i^\top \right)^{-1} \overset{p}{\to} Q^{-1} \]

So,

\[ \frac{1}{N} \left( \frac{1}{N} X^\top X \right)^{-1} S_0 \left( \frac{1}{N} X^\top X \right)^{-1} \overset{p}{\to} \text{\textbf{V}}_{\text{asy}} \left( \hat{\beta}_{\text{OLS}} \right) \]
5. Heteroscedasticity

**Definition (White heteroscedasticity consistent estimator)**

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

$$\hat{V}_{asy} \left( \hat{\beta}_{OLS} \right) = N \left( X^\top X \right)^{-1} S_0 \left( X^\top X \right)^{-1}$$

$$\hat{V}_{asy} \left( \hat{\beta}_{OLS} \right) \xrightarrow{p} V_{asy} \left( \hat{\beta}_{OLS} \right)$$

with

$$S_0 = \frac{1}{N} \sum_{i=1}^{N} \hat{\varepsilon}_i^2 x_i^\top x_i^\top$$
5. Heteroscedasticity

**Corollary (White heteroscedasticity consistent estimator)**

The White consistent estimator can written as:

\[
\hat{V}_{asy} (\hat{\beta}_{OLS}) = \frac{1}{N} \left( \frac{1}{N} \sum_{i=1}^{N} x_i x_i^\top \right)^{-1} \left( \frac{1}{N} \sum_{i=1}^{N} \hat{\varepsilon}_i^2 x_i x_i^\top \right) \left( \frac{1}{N} \sum_{i=1}^{N} x_i x_i^\top \right)^{-1}
\]
5. Heteroscedasticity

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

White Heteroskedasticity-Consistent Standard Errors & Covariance

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.001189</td>
<td>0.001160</td>
<td>1.025585</td>
<td>0.3187</td>
</tr>
<tr>
<td>RSP500</td>
<td>1.989787</td>
<td>0.311130</td>
<td>6.395357</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

R-squared 0.690203, Adjusted R-squared 0.672992, 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
5. Heteroscedasticity

Remark

Given the normalisation trace($\Omega$) = $N$, we have:

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

**Definition (SSR)**

The **least squares estimator** \( \hat{\sigma}^2 \) defined by:

\[
\hat{\sigma}^2 = \frac{\hat{\mathbf{e}}^\top \hat{\mathbf{e}}}{N-K} = \frac{1}{N-K} \sum_{i=1}^{N} \hat{\varepsilon}_i^2
\]

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

\[
\hat{\sigma}^2 \xrightarrow{p} \lim_{N \to \infty} \sigma^2 = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2
\]
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;
N=length(y);
X=[ones(N,1) Age Ownrent Income Income.^2];
X=size(X,2);
beta=X\y;
res=y-X*beta;
v=sum(res.^2)/(N-K);
V=v*inv(X'\X);
std=sqrt(diag(V));
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);
std_robust=sqrt(diag(V_robust));

disp('') , disp(' Beta std Robust std')
disp([beta std std_robust])
This graph is the sign of heteroscedasticity. The variance of the residuals seems to depend on the income.
5. Heteroscedasticity

The values are the same.. perfect
5. Heteroscedasticity

The values are different... Why?
5. Heteroscedasticity

**Remark**

This difference is due to the fact that Eviews uses a **finite sample correction** for $S_0$ (Davidson and MacKinnon, 1993)

$$S_0 = \frac{1}{N-K} \sum_{i=1}^{N} \hat{\epsilon}_i^2 x_i x_i^T$$

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;
N=length(y);
X=[ones(N,1) Age Ownrent Income Income.^2];
K=size(X,2);
beta=X\y;
res=y-X'*beta;
% Dependent variable
% Sample size
% Matrix X
% Number of explicative variables
% OLS estimates
% Residuals
% Variance of the disturbances
% Estimated asymptotic variance
% Standard errors

for i=1:N
    S0=S0+(res(i)^2)*X(i,:)'*X(i,:);
end
S0=S0/(N-K);

V_robust=N*X'*X*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

The values are now identical.
5. Heteroscedasticity

Case 2: Heteroscedasticity with known $\Sigma$

GLS and Weighted Least Squares
5. Heteroscedasticity

**Assumption:** We assume that the disturbances are heteroscedastic with

$$ \mathbb{V}(\epsilon|\mathbf{X}) = \Sigma = \sigma^2 \Omega $$

with

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

where the parameters $\sigma_i^2$ and $\omega_i$ are known for $i = 1, \ldots, N$. 
5. Heteroscedasticity

**Definition (GLS estimator)**

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

$$\hat{\beta}_{GLS} = \left( \sum_{i=1}^{N} \frac{x_i x_i^\top}{\omega_i} \right)^{-1} \left( \sum_{i=1}^{N} \frac{x_i y_i}{\omega_i} \right)$$

or equivalently by

$$\hat{\beta}_{GLS} = \left( \sum_{i=1}^{N} \frac{x_i x_i^\top}{\sigma_i^2} \right)^{-1} \left( \sum_{i=1}^{N} \frac{x_i y_i}{\sigma_i^2} \right)$$
5. Heteroscedasticity

Proof

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

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

Since $\Omega$ is diagonal:

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

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

As a consequence:

$$\hat{\beta}_{GLS} = \left( \sum_{i=1}^{N} \frac{x_i x_i^\top}{\omega_i} \right)^{-1} \left( \sum_{i=1}^{N} \frac{x_i y_i}{\omega_i} \right) \square$$
5. Heteroscedasticity

Remark

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

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

\[ \hat{\beta}_{WLS} = \left( \sum_{i=1}^{N} \delta_i \mathbf{x}_i \mathbf{x}_i^\top \right)^{-1} \left( \sum_{i=1}^{N} \delta_i \mathbf{x}_i y_i \right) \]
5. Heteroscedasticity

**Fact (GLS and WLS)**

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

\[
\hat{\beta}_{WLS} = \left( \sum_{i=1}^{N} \delta_i x_i x_i^\top \right)^{-1} \left( \sum_{i=1}^{N} \delta_i x_i y_i \right)
\]

where \( \delta_i \) is an arbitrary weight. For \( \delta_i = 1/\omega_i \), we have \( \hat{\beta}_{WLS} = \hat{\beta}_{GLS} \).
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):

$$\sigma_i^2 = \sigma^2 x_{ik}^2 \iff \delta_i = \frac{1}{x_{ik}^2}$$
5. Heteroscedasticity

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

\[ \nabla (\varepsilon | X) = \Sigma (\alpha) = \sigma^2 \Omega (\alpha) \]

where \( \alpha \) denotes a set of parameters.
Example (Restriction)

We assume that

$$\nabla (\varepsilon_i | X) = \sigma_i^2 (\alpha) = \sigma^2 (z_i^\top \alpha)^2$$

where $\alpha = (\alpha_1 : \ldots : \alpha_H)^\top$ is a $H \times 1$ vector of parameters and $z_i$ is $H \times 1$ of explicative variables (not necessarily the same as in $x_i$).
Example (Harvey’s (1976) restriction)

Harvey (1976) considers a restriction of the form:

$$\nabla (\varepsilon_i \mid X) = \sigma_i^2 (\alpha) = \exp \left( x_i^\top \alpha \right)$$

where $\alpha = (\alpha_1 : \ldots : \alpha_H)^\top$ is a $H \times 1$ vector of parameters and $z_i$ is $H \times 1$ of explicative variables (not necessarily the same as in $x_i$).
5. Heteroscedasticity

We know that the GLS estimator is defined by:

$$\hat{\beta}_{GLS} = \left( \sum_{i=1}^{N} \frac{x_i x_i^\top}{\sigma_i^2(\alpha)} \right)^{-1} \left( \sum_{i=1}^{N} \frac{x_i y_i}{\sigma_i^2(\alpha)} \right)$$

S, the feasible GLS (FGLS) estimator is:

$$\hat{\beta}_{FGLS} = \left( \sum_{i=1}^{N} \frac{x_i x_i^\top}{\sigma_i^2(\hat{\alpha})} \right)^{-1} \left( \sum_{i=1}^{N} \frac{x_i y_i}{\sigma_i^2(\hat{\alpha})} \right)$$
5. Heteroscedasticity

If we assume for instance that

$$
V (\varepsilon_i | X) = \sigma_i^2 (\alpha) = \exp \left( z_i^\top \alpha \right)
$$

where $z_i$ is a vector of $H$ variables, a way to estimate $\alpha$ consists in considering the model:

$$
\ln \left( \hat{\varepsilon}_i^2 \right) = z_i^\top \alpha + v_i
$$

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

$$
\hat{\sigma}_i^2 = \exp \left( z_i^\top \hat{\alpha} \right) \xrightarrow{p} \sigma_i^2 (\alpha)
$$
5. Heteroscedasticity

Problem

*In order to estimate $\beta$ by the GLS, we need $\hat{\alpha}$, and to estimate $\alpha$, we need the residuals $\hat{\varepsilon}_i = y_i - x_i^T \hat{\beta}_{GLS}$...*
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 $\beta$ by OLS. Compute the residuals $\hat{\varepsilon}_i = y_i - x_i^\top \hat{\beta}_{OLS}$ and estimate the parameters $\alpha$ according to the appropriate model. **Second step:** compute the estimated variances $\sigma_i^2 (\hat{\alpha})$ and compute the FGLS estimator:

$$
\hat{\beta}_{FGLS} = \left( \sum_{i=1}^{N} \frac{x_i x_i^\top}{\sigma_i^2 (\hat{\alpha})} \right)^{-1} \left( \sum_{i=1}^{N} \frac{x_i y_i}{\sigma_i^2 (\hat{\alpha})} \right)
$$
5. Heteroscedasticity

**Definition (Iterated FGLS estimator)**

Estimate the parameters $\beta$ by OLS. Compute the residuals
\[ \hat{\varepsilon}_i = y_i - x_i^\top \hat{\beta}_{OLS} \]
and estimate the parameters $\alpha$ according to the appropriate model. Compute the estimated variances $\sigma_i^2(\hat{\alpha})$ and compute the FGLS estimator:

\[
\widehat{\beta}_{FGLS}^{(1)} = \left( \sum_{i=1}^{N} \frac{x_i x_i^\top}{\sigma_i^2(\hat{\alpha})} \right)^{-1} \left( \sum_{i=1}^{N} \frac{x_i y_i}{\sigma_i^2(\hat{\alpha})} \right)
\]

Compute the residuals $\hat{\varepsilon}_i = y_i - x_i^\top \hat{\beta}_{FGLS}^{(1)}$ and estimate the parameters $\alpha$ according to the appropriate model. Compute the FGLS $\widehat{\beta}_{FGLS}^{(2)}$ and so on... The procedure stop when
\[
\sup_{j=1,\ldots,K} \left| \widehat{\beta}_{j,FGLS}^{(i)} - \widehat{\beta}_{j,FGLS}^{(i-1)} \right| < \text{threshold (ex: 0.001)}
\]
Example (Harvey’s (1976) multiplicative model of heteroscedasticity)

Consider the generalized linear regression model:

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

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

$$ \nabla (\epsilon_i | X) = \sigma_i^2 = \exp (\alpha_1 + \alpha_2 \text{Income}_i) $$

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.


5. Heteroscedasticity

**Remark**

A way to get the estimates of the parameters $\alpha_1$ and $\alpha_2$ is to consider the regression:

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

and to estimate the parameters by OLS.
5. Heteroscedasticity

```matlab
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

\[
\begin{array}{cc}
\text{OLS} & \text{FGLS (two-steps)} \\
-115.9914 & -35.1646 \\
-3.6537 & -3.7218 \\
60.8815 & 45.5433 \\
156.4672 & 110.8203 \\
-9.0760 & -3.0666
\end{array}
\]
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

<table>
<thead>
<tr>
<th>OLS</th>
<th>FGLS (iterated)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-115.9914</td>
<td>8.8438</td>
</tr>
<tr>
<td>-3.6537</td>
<td>-3.6947</td>
</tr>
<tr>
<td>60.8815</td>
<td>44.0512</td>
</tr>
<tr>
<td>156.4672</td>
<td>79.8858</td>
</tr>
<tr>
<td>-9.0760</td>
<td>1.6777</td>
</tr>
</tbody>
</table>
5. Heteroscedasticity

**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
Definition (White test for heteroscedasticity)

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

\[ H_0 : \sigma_i^2 = \sigma^2 \quad \text{for } i = 1, \ldots, N \]

\[ H_1 : \sigma_i^2 \neq \sigma_j^2 \quad \text{for at least one pair } (i, j) \]
6. Testing for heteroscedasticity

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

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

$$V_{asy} \left( \hat{\beta}_{OLS} \right) = \sigma^2 Q^{-1}$$

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

2. Under the alternative (heteroscedasticity):

$$V_{asy} \left( \hat{\beta}_{OLS} \right) = \sigma^2 Q^{-1} Q^* Q^{-1}$$

$$\hat{V}_{asy} \left( \hat{\beta}_{OLS} \right) = \sigma^2 \left( X^\top X \right)^{-1} X^\top \Omega X \left( X^\top X \right)^{-1}$$
6. Testing for heteroscedasticity

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

**Step 1:** Estimation of the model using the OLS estimator of $\beta$.

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

**Step 3:** Regress $\hat{\varepsilon}_i^2$ 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** $N \times R^2$ converges:

$$N \times R^2 \xrightarrow{d}_{H_0} \chi^2 (m - 1)$$

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

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

where $\chi^2_{1-\alpha}$ denotes the $1-\alpha$ critical value of the $\chi^2 (m - 1)$ distribution.
6. Testing for heteroscedasticity

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

Consider the generalized linear regression model:

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

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 your conclusion for a significance level of 5%? Compare your results with Eviews.
6. Testing for heteroscedasticity

```matlab
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:

<table>
<thead>
<tr>
<th>F-statistic</th>
<th>1.244819</th>
<th>Probability</th>
<th>0.266541</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs*R-squared</td>
<td>14.65386</td>
<td>Probability</td>
<td>0.260914</td>
</tr>
</tbody>
</table>

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

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

White = 14.6539

p-value = 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

\[ \sigma_i^2 = \sigma^2 f \left( \alpha_0 + z_i^\top \alpha \right) \]

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

\[
H_0 : \alpha = 0_{p \times 1} \quad \text{(homoscedasticity)} \\
H_1 : \alpha \neq 0_{p \times 1} \quad \text{(heteroscedasticity)}
\]
6. Testing for heteroscedasticity

The test can be carried out with a simple regression of

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

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

$$g_i = \alpha_0 + \alpha_1 z_{i1} + \ldots + \alpha_p z_{ip} + v_i$$
6. Testing for heteroscedasticity

**Definition (Breusch and Pagan test-statistic)**

Define \( Z \) the \( N \times (p + 1) \) matrix of observations on \((1, z_i)\) and let \( g \) be the \( N \times 1 \) vector of observations

\[
g_i = N \frac{\hat{\varepsilon}_i^2}{\hat{\varepsilon}^\top \hat{\varepsilon}} - 1
\]

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

\[
LM = \frac{1}{2} g^\top Z \left( Z^\top Z \right)^{-1} Z^\top g
\]

Under the null, we have:

\[
LM \overset{d}{\rightarrow} \chi^2 (p)
\]
6. Testing for heteroscedasticity

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

Consider the generalized linear regression model:

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

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 $z_i = x_i$ and its p-value. What is your 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;
N=length(y);
X=[ones(N,1) Age Ownrent Income Income.^2];
K=size(X,2);
beta=X\y;
res=y-X*beta;

% Dependent variable
% Sample size
% Matrix X
% Number of explicative variables
% OLS estimates
% Residuals

g=N*res.^2/sum(res.^2)-1;
Z=X;
LM=0.5*g'*Z*inv(Z'*Z)*Z'*g;
pvalue=1-chi2cdf(LM,size(Z,2)-1);
% G vector
% We use z=x
% LM test-statistic
% 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)