FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

BACHELOR THESIS

Filip Dávidík

Parametric variance modelling within a feasible weighted least squares estimator

Department of Probability and Mathematical Statisticst

Supervisor of the bachelor thesis: RNDr. Šárka Hudecová, Ph.D.
Study programme: Financial Mathematics
Study branch: Faculty of Mathematics and Physics

Prague 2024

I declare that I carried out this bachelor thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In .... date ....

Author's signature

I want to dedicate this thesis to my thesis supervisor RNDr. Šárka Hudecová, Ph.D., whose support and guidance have been invaluable. Their insightful advice, diligent reviews, and excellent planning have significantly contributed to this work. I am deeply grateful for their assistance and encouragement.

Title: Parametric variance modelling within a feasible weighted least squares estimator

Author: Filip Dávidík

Department: Department of Probability and Mathematical Statisticst

Supervisor: RNDr. Šárka Hudecová, Ph.D., Department of Probability and Mathematical Statisticst

Abstract: This thesis explores the implications of heteroscedasticity in regression models, where the variance of errors is not constant across observations. Traditional estimators such as Ordinary Least Squares (OLS) rely on the assumption of homoscedasticity, but real-world data often deviate from this ideal. In response, Weighted Least Squares (WLS) estimation is introduced to address known forms of heteroscedasticity, alongside the Feasible Weighted Least Squares (FWLS) estimation method, which only requires partial knowledge of heteroscedasticity's form. The theoretical contribution establishes the efficiency of the WLS over the OLS under known heteroscedasticity, and the introduction of the FWLS as a viable alternative. Simulation studies further illustrate the nuanced behavior of the FWLS estimators, offering a comprehensive comparison of the various candidate FWLS estimators under varying model specifications (including misspecified variance models) and insights into their performance relative to the OLS estimator. Recommendations are provided to guide method selection based on specific model characteristics, highlighting the importance of accounting for heteroscedasticity in empirical research.

Keywords: Heteroscedasticity Regression Weighted least squares Ordinary least squares Feasible weighted least squares

Contents

Introduction **2**

1 Simple linear regression model **3**
1.1 Method of moments estimation . . . 3
1.1.1 Ordinary least squares estimation . . . 5

2 Multiple linear regression model **7**
2.1 MLR model . . . 7
2.2 Ordinary least squares method (OLS) . . . 8
2.3 Properties of the OLS estimate . . . 11

3 Heteroscedasticity data **15**
3.1 Heteroscedasticity . . . 15
3.2 Properties of the OLS estimate . . . 16
3.3 Weighted least squares . . . 18
3.4 Feasible weighted least squares . . . 23
3.4.1 Two-step estimation . . . 23
3.4.2 Iterative estimation . . . 26

4 Simulation studies **27**
4.1 Study 1 . . . 28
4.2 Study 2 . . . 31
4.3 Study 3 . . . 33
4.4 Study 4 . . . 34
4.5 Study 5 . . . 36
4.6 Study 6 . . . 38
4.7 Conclusion to simulation studies . . . 41

Conclusion **42**

A Attachments **47**
A.1 First Attachment . . . 47

Introduction

In regression models, the assumption of a constant variance of errors across observations (homoscedasticity) is crucial for reliable predictions and meaningful conclusions. Traditional estimators like ordinary least squares (OLS) rely on this assumption. However, real-world data often exhibit heteroscedasticity, where the variance of errors is a non-constant function of regressors. In such a case, statistical inference based on the OLS may be incorrect, and the corresponding conclusions misleading.

To tackle such situations, we use weighted least squares to address known heteroscedasticity forms or a method called Feasible Weighted Least Squares, which requires only partial knowledge of the heteroscedasticity's form.

In Chapter 1 we present a simple linear regression model and present methods to estimate unknown parameters within the model. In Chapter 2, we extend the simple linear regression model by introducing the multiple linear regression model and state assumptions that we follow for the entirety of the thesis. Finally, we then define a generalized version of the ordinary least squares estimation and discuss its properties. In Chapter 3 we introduce the presence of heteroscedasticity within the model and explore its implications on the ordinary least squares estimate. Additionally, to address heteroscedasticity, we define the weighted least square estimate and later feasible weighted least square estimate.

The author's contribution to the theoretical part of the thesis involves a more detailed elaboration of certain proofs and rearranging the content from professional literature to align with the topic of the thesis. The main contribution of the author resides in the practical section, where through simulation studies, they compare the OLS, WLS, and FWLS estimators in terms of their efficiency and bias across various scenarios, to validate the material described in the theory. Furthermore, the author provides recommendations regarding the selection between OLS and FWLS methods based on specific model characteristics.

1. Simple linear regression model

This chapter is based on Greene, (2003) and Wooldridge, (2013). It aims to introduce a two-variable model called the simple linear regression model and means to estimate unknown parameters within the model. The author's contribution lies in a detailed description of a connection between the moment method and the ordinary least squares method.

As an introduction to regression analysis, we consider a case with two random variables Y and X, to which we assume, that the linear relationship between Y and X has a form of

$$Y = \beta_0 + \beta_1 X + \epsilon, \tag{1.1}$$

where $\beta_0 \in \mathbb{R}$ and $\beta_1 \in \mathbb{R}$ are both unknown constants and ϵ is a random variable, such that $\mathsf{E}[\epsilon \mid X] = 0$. Consequently, this gives us $\mathsf{E}[Y \mid X] = \beta_0 + \beta_1 X$, implying that the conditional expected value of Y given X is linear in both the random variable X and constants β_0 and β_1.

Let us contemplate a set of n observations, where for each observation $i = 1, 2, \ldots, n$, such that (Y_i, X_i) are independent and identically distributed (iid) copies of (Y, X) from (1.1). We have

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad i = 1, 2, \ldots, n, \tag{1.2}$$

where ϵ_i are iid random variables, satisfying $\mathsf{E}[\epsilon_i \mid X_i] = 0$. We designate Y_i as the dependent variables, and X_i as the independent variables, also referred to as regressors.

Following up on (1.2), let $\mathbf{Y}$ be a $n \times 1$ random vector $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)^\top$ and $\mathbf{X}$ be a $n \times 1$ random vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)^\top$, so that $(Y_i, X_i), i = 1, \ldots, n$ represent all n observations from (1.2). We proceed by defining a two-variable model called the simple linear regression model (SLR)

$$\mathbf{Y} = \beta_0 + \beta_1 \mathbf{X} + \boldsymbol{\epsilon}, \tag{1.3}$$

where β_0 is an unknown constant called intercept, β_1 is an unknown constant called slope parameter, and $\boldsymbol{\epsilon}$ is an $n \times 1$ vector $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)^\top$ of iid random variables ϵ_i called the error terms or disturbances. $\boldsymbol{\epsilon}$ represents all unobserved factors affecting $\mathbf{Y}$.

1.1 Method of moments estimation

To estimate unknown values β_0 and β_1 we recall that

$$\mathsf{E}[\epsilon_i \mid X_i] = 0,$$

for $i = 1, \ldots, n$. This gives us

$$\mathsf{E}[\epsilon_i] = \mathsf{E}\left[\mathsf{E}[\epsilon_i \mid X_i]\right] = \mathsf{E}[0] = 0,$$

then

$$\mathsf{E}[X_i\epsilon_i] = \mathsf{E}\left[X_i\mathsf{E}\left[\epsilon_i \mid X_i\right]\right] = 0.$$

Finally, we get

$$\mathsf{Cov}[X_i, \epsilon_i] = \mathsf{E}[X_i\epsilon_i] - \mathsf{E}[\epsilon_i]\mathsf{E}[X_i] = 0.$$

Given these assumptions, we can assert that

$$\mathsf{E}[Y_i - \beta_0 - \beta_1 X_i] = 0, \tag{1.4}$$

additionally

$$\mathsf{E}[X_i(Y_i - \beta_0 - \beta_1 X_i)] = 0, \tag{1.5}$$

for $i = 1, \ldots, n$.

We can now construct estimate $\widehat{\beta}_0$ of β_0 and estimate $\widehat{\beta}_1$ of β_1 using a method of moments. Hence, by estimating expected values (1.4) and (1.5) by the corresponding sample averages, we get

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i) = 0 \tag{1.6}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}X_i(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i) = 0. \tag{1.7}$$

From (1.6), we deduce that

$$\widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1\overline{X}, \tag{1.8}$$

where

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n}Y_i.$$

and

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n}X_i. \tag{1.9}$$

We can now rewrite (1.7) as

$$\sum_{i=1}^{n}X_i(Y_i - (\overline{Y} - \widehat{\beta}_1\overline{X}) - \widehat{\beta}_1 X_i) = 0.$$

By rearranging it so that $\widehat{\beta}_1$ is on the right

$$\sum_{i=1}^{n}X_i(Y_i - \overline{Y}) = \widehat{\beta}_1\sum_{i=1}^{n}X_i(X_i - \overline{X}),$$

we express $\widehat{\beta}_1$ as

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}X_i(Y_i - \overline{Y})}{\sum_{i=1}^{n}X_i(X_i - \overline{X})}, \tag{1.10}$$

where we assume that $\sum_{i=1}^{n}(X_i - \overline{X})^2 > 0$, which means that there exist at least two distinct values of X_i.

Using the fact that

$$\begin{aligned}\sum_{i=1}^{n}(X_i-\overline{X})^2 &= \sum_{i=1}^{n}(X_i^2-2X_i\overline{X}+\overline{X}^2)\\ &= \sum_{i=1}^{n}X_i^2-2n\overline{X}^2+n\overline{X}^2\\ &= \sum_{i=1}^{n}X_i^2-n\overline{X}^2=\sum_{i=1}^{n}X_i^2-\sum_{i=1}^{n}\overline{X}^2\\ &= \sum_{i=1}^{n}(X_i^2-\overline{X}^2)=\sum_{i=1}^{n}X_i(X_i-\overline{X}),\end{aligned}$$

and

$$\begin{aligned}\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y}) &= \sum_{i=1}^{n}(X_iY_i-Y_i\overline{X}-X_i\overline{Y}+\overline{XY})\\ &= \sum_{i=1}^{n}X_iY_i-n\overline{XY}-n\overline{YX}+n\overline{XY}\\ &= \sum_{i=1}^{n}X_iY_i-n\overline{XY}=\sum_{i=1}^{n}X_iY_i-\sum_{i=1}^{n}\overline{XY}\\ &= \sum_{i=1}^{n}(X_iY_i-\overline{XY})=\sum_{i=1}^{n}X_i(Y_i-\overline{Y})\end{aligned}$$

we rewrite (1.10), as

$$\widehat{\beta}_1=\frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sum_{i=1}^{n}(X_i-\overline{X})^2}.$$

1.1.1 Ordinary least squares estimation

We will now introduce a different method to estimate β_0 and β_1, known as ordinary least squares (denoted as OLS), and compare it to the method of moments.

The fundamental concept behind the OLS method is to determine $\widehat{\beta}_0$ and $\widehat{\beta}_1$ that minimizes the sum of squared deviations across all n data points. First, we define a function g as

$$g(\beta_0,\beta_1)=\sum_{i=1}^{n}(Y_i-\beta_0-\beta_1X_i)^2. \tag{1.11}$$

Given our objective is minimizing the sum of squares, we aim to identify

$$(\hat{\beta}_0,\hat{\beta}_1)^\top=\underset{(\beta_0,\beta_1)^\top\in\mathbb{R}^2}{\arg\min}\; g(\beta_0,\beta_1), \tag{1.12}$$

which we seek as a stationary point. Hence, we proceed by calculating partial derivatives

$$\begin{aligned}\frac{\partial g}{\partial\beta_0} &= -2\sum_{i=1}^{n}(Y_i-\beta_0-\beta_1X_i)\\ \frac{\partial g}{\partial\beta_1} &= -2\sum_{i=1}^{n}X_i(Y_i-\beta_0-\beta_1X_i).\end{aligned}$$

and solve

$$\frac{\partial g}{\partial \widehat{\beta}_0} = 0$$
$$\frac{\partial g}{\partial \widehat{\beta}_1} = 0.$$

The uniqueness of the solution to (1.12) can be demonstrated by obtaining the determinant of the Hessian matrix

$$H(g) = \begin{vmatrix} \frac{\partial^2 g}{\partial \beta_0^2} & \frac{\partial^2 g}{\partial \beta_0 \partial \beta_1} \\ \frac{\partial^2 g}{\partial \beta_1 \partial \beta_0} & \frac{\partial^2 g}{\partial \beta_1^2} \end{vmatrix} = \begin{vmatrix} 2n & 2n\overline{X} \\ 2n\overline{X} & 2(n\overline{X})^2 \end{vmatrix} = 4n(n\overline{X})^2 - 4(n\overline{X})^2 = 4(n-1)(n\overline{X})^2.$$

For determinant of $H(g)$, we have $4(n-1)(n\overline{X})^2 > 0$ for $n > 1$, while for its minor, $2n > 0$ for $n > 0$, implies that the Hessian matrix $H(g)$ is positive definite, and g is a convex function. Consequently, this yields the same problem as (1.6) and (1.7). Therefore, we obtain equivalent estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$ as those in the method of moments.

Figure 1.1 visualizes the idea behind the function g (1.11) for OLS estimation. Blue points on the graph are singular observations (Y_i, X_i), $i = 1, 2, \ldots, n$, while the line is described by $\widehat{\beta}_0 + \widehat{\beta}_1 X$. We denote the difference between Y_i and $\widehat{\beta}_0 + \widehat{\beta}_1 X_i$ as $\widehat{\epsilon}_i$.

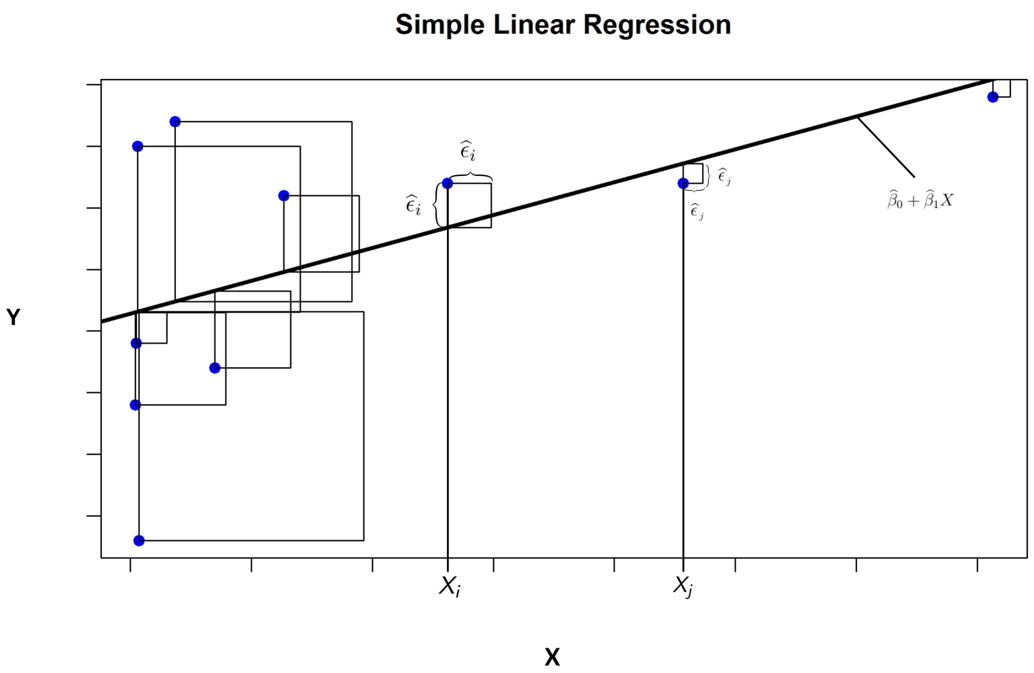


Figure 1.1: Illustration of the idea of the ordinary least squares estimation method.

2. Multiple linear regression model

This chapter is based on Greene, (2003) and Wooldridge, (2013). It aims to introduce a multiple linear regression model and discuss the generalized form of the OLS method from the first chapter. Finally, we show the properties of the OLS estimate. The author's contribution is in providing a more detailed breakdown and commentary on certain proofs and derivations.

Multiple linear regression (denoted as MLR) is a statistical technique used to analyze the relationship between a dependent variable and two or more independent variables. It is an extension of simple linear regression, which is used to model the relationship between two variables.

2.1 MLR model

Following up on the first chapter, let us have n observations, where for each observation $i = 0, 1, \dots, n$, we measure a random variable Y_i and k random variables $X_{ij},\ j = 1, \dots, k$. We define a $(k+1) \times 1$ random vector $\mathbf{X}_i = (1, X_{i1}, \dots, X_{ik})^\top$, assuming that the linear relationship between Y_i and $\mathbf{X}_i$, follows the form

$$Y_i = \mathbf{X}_i^\top \boldsymbol{\beta} + \epsilon_i, \tag{2.1}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)^\top$ is a $(k+1) \times 1$ vector of unknown real constants and $\epsilon_i,\ i = 0, 1, \dots, n$, are random variables satisfying

$$\mathsf{E}[\epsilon_i \mid \mathbf{X}_i] = 0, \quad i = 0, 1, \dots, n.$$

We refer to β_0 the intercept and $\beta_j,\ j = 1, \dots, k$ as slope parameters.

As in the previous chapter, Y_i is referred to as the dependent variable, while $X_{ij},\ j = 0, 1, \dots, k$ are referred to as independent variables or regressors.

Before defining the MLR model, let us introduce an $n \times (k+1)$ matrix $\mathbf{X}$ defined as

$$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & \dots & X_{1k} \\ 1 & X_{21} & \dots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \dots & X_{nk} \end{bmatrix} = \begin{bmatrix} X_{10} & X_{11} & \dots & X_{1k} \\ X_{20} & X_{21} & \dots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n0} & X_{n1} & \dots & X_{nk} \end{bmatrix},$$

where row i of matrix $\mathbf{X}$ represents a $(k+1) \times 1$ vector $\mathbf{X}_i = (X_{i0}, X_{i1}, \dots, X_{ik})^\top$, $i = 1, 2, \dots, n$, with the assumption that $X_{i0} = 1,\ i = 0, 1, \dots, n$. Intuitively, $\mathbf{X}$ represents a specific dataset used in the MLR model.

Additionally, we define $n \times 1$ random vector $\mathbf{Y}$ as

$$\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top.$$

For the remainder of the thesis, we assume the following:

Assumption of full rank. The $n \times (k+1)$ matrix $\mathbf{X}$ has a full rank $(k+1)$ with probability one, indicating that all columns of $\mathbf{X}$ are linearly independent.

Assumption of independent and identically distributed random sampling. The sample data $\{(Y_i, X_{i0}, X_{i1}, X_{i2}, \ldots, X_{ik}) : i = 1, 2, \ldots, n\}$ are independent and identically distributed.

MLR model. Using previously defined $\mathbf{Y}$ and $\mathbf{X}$ we write down the MLR model as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{2.2}$$

such that $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)^\top$ is a $n \times 1$ vector of random variables called the error terms or disturbances. Vector $\boldsymbol{\epsilon}$ represents all unobserved factors affecting the dependent variables.

Under the MLR model, we state the following assumption for $\boldsymbol{\epsilon}$.

A1: Assumption of zero conditional mean. Assume

$$\mathsf{E}[\boldsymbol{\epsilon} \mid \mathbf{X}] = \begin{pmatrix} \mathsf{E}[\epsilon_1 \mid \mathbf{X}] \\ \mathsf{E}[\epsilon_2 \mid \mathbf{X}] \\ \vdots \\ \mathsf{E}[\epsilon_n \mid \mathbf{X}] \end{pmatrix} = \mathbf{0}.$$

Assumption **A1** implies that

$$\mathsf{E}[\boldsymbol{\epsilon}] = \mathsf{E}\left[\mathsf{E}\left[\boldsymbol{\epsilon} \mid \mathbf{X}\right]\right] = \mathsf{E}[\mathbf{0}] = \mathbf{0},$$

and

$$\mathsf{E}[\mathbf{X}^\top \boldsymbol{\epsilon}] = \mathsf{E}[\mathbf{X}_1 \epsilon_1] + \ldots + \mathsf{E}[\mathbf{X}_n \epsilon_n] = \mathsf{E}\left[\mathbf{X}_1^\top \mathsf{E}\left[\epsilon_1 \mid \mathbf{X}\right]\right] + \ldots + \mathsf{E}\left[\mathbf{X}_n^\top \mathsf{E}\left[\epsilon_n \mid \mathbf{X}\right]\right] = \mathbf{0}.$$

Finally, we have

$$\mathsf{Cov}[\mathbf{X}_i, \epsilon_i] = \mathsf{E}[\mathbf{X}_i \epsilon_i] - \mathsf{E}[\epsilon_i]\mathsf{E}[\mathbf{X}_i] = \mathbf{0}. \tag{2.3}$$

In the subsequent section, we delve into the generalized form of the OLS method discussed in the first chapter.

2.2 Ordinary least squares method (OLS)

The main principle of OLS is to minimize the sum of the squared deviations (the least squares). This means finding the values of the intercept and the slope parameters that minimize the sum of squared deviations across all n data points.

Based on the Section 1.1.1, we continue with a general case of the OLS method. We aim to obtain $(k+1) \times 1$ vector estimate $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, by defining a function g as

$$g(\boldsymbol{\beta}) = \sum_{i=1}^{n} (Y_i - \mathbf{X}_i^\top \boldsymbol{\beta})^2 = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}).$$

We write the estimate $\widehat{\boldsymbol{\beta}}$ as

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^{k+1}}{\arg\min}\ g(\boldsymbol{\beta}).$$

By multiplying the expression

$$(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}),$$

we obtain

$$\begin{aligned}\mathbf{Y}^\top\mathbf{Y} - \mathbf{Y}^\top\mathbf{X}\boldsymbol{\beta} - \mathbf{Y}\boldsymbol{\beta}^\top\mathbf{X}^\top + \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} \\ = \mathbf{Y}^\top\mathbf{Y} - 2\mathbf{Y}^\top\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta}.\end{aligned}$$

Thus, we get

$$g(\boldsymbol{\beta}) = \mathbf{Y}^\top\mathbf{Y} - 2\mathbf{Y}^\top\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta}.$$

For $\widehat{\boldsymbol{\beta}}$ to be minimal a first-order condition must hold

$$\frac{\partial g(\widehat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^\top\mathbf{Y} + 2\mathbf{X}^\top\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{0}. \tag{2.4}$$

Finally, by rearranging (2.4), we obtain

$$\mathbf{X}^\top\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}^\top\mathbf{Y}.$$

Theorem 1. *If matrix $\mathbf{X}^\top\mathbf{X}$ has full rank, then*

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y} \tag{2.5}$$

minimizes function g.

Proof. Since under the assumption of full rank, $\mathbf{X}^\top\mathbf{X}$ is a regular matrix with rank $(k+1)$ we can multiply the equation by its inverse and obtain

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}.$$

Lastly, $\widehat{\boldsymbol{\beta}}$ minimizes function g when the second partial derivation of g is a positive definite matrix. We have

$$\frac{\partial^2 g(\widehat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = 2\mathbf{X}^\top\mathbf{X}. \tag{2.6}$$

Let us have $\mathbf{c} \in \mathbb{R}^{k+1}$, and $\mathbf{p} = \mathbf{X}\mathbf{c}$. The matrix $\mathbf{X}^\top\mathbf{X}$ is positive definite when

$$\mathbf{c}^\top\mathbf{X}^\top\mathbf{X}\mathbf{c} = \mathbf{p}^\top\mathbf{p} = \sum_{i=1}^{n} p_i^2 > 0,$$

for all $\mathbf{c} \in \mathbb{R}^{k+1}$, such that $\mathbf{c} \neq \mathbf{0}$. Unless each element of $\mathbf{p}$ is zero, $\mathbf{c}^\top\mathbf{X}^\top\mathbf{X}\mathbf{c}$ is positive. If vector $\mathbf{p} = \mathbf{0}$ then $\mathbf{X}\mathbf{c} = \mathbf{0}$. This would imply that there is a linear combination of the $(k+1)$ columns of $\mathbf{X}$ that is equal to zero, which contradicts the assumption of $\mathbf{X}$ having full rank. $\square$

With Theorem 1, we can now establish that $\widehat{\boldsymbol{\beta}}$ is the only solution to the OLS.

Next, we define a fitted value for each of our n observations. The fitted value of observation i is

$$\widehat{Y_i} = \mathbf{X}_i^\top \widehat{\boldsymbol{\beta}}.$$

Hence, for $\widehat{\mathbf{Y}} = (\widehat{Y_1}, \widehat{Y_2}, \ldots, \widehat{Y_n})^\top$, we get

$$\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}.$$

We also define $\widehat{\epsilon_i}$ as a **residual**, where

$$\widehat{\epsilon_i} = Y_i - \widehat{Y_i}.$$

Again, for $\widehat{\boldsymbol{\epsilon}} = (\widehat{\epsilon}_1, \widehat{\epsilon}_2, \ldots, \widehat{\epsilon}_n)^\top$

$$\widehat{\boldsymbol{\epsilon}} = \mathbf{Y} - \widehat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}.$$

Example 1. As an example of the OLS estimation to obtain $\boldsymbol{\beta}$, under the assumption **A1**, we use a SLR model (1.3) to analyze a dataset of $n = 100$ random samples relating video game user ratings Y_i to their global sales X_i. The dataset was obtained from Shukla, (2019).

$$Y_i = \beta_0 + \beta_1 X_i + \boldsymbol{\epsilon}$$

Using OLS estimates of β_0 and β_1 we obtain fitted line

$$\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_i,$$

where we estimated the intercept β_0 as 7.040 and the slope parameter β_1 as 0.251. This yields

$$\widehat{Y}_i = 7.040 + 0.251 X_i.$$

Figure 2.1 showcases a line obtained by the OLS method, estimating the relation between user ratings Y_i and the global sales data X_i. Blue points on the graph are singular observations (Y_i, X_i), $i = 1, 2, \ldots, n$, while the line is described by $\widehat{\beta}_0 + \widehat{\beta}_1 X$.

Finally, we aim to estimate the covariance matrix of $\widehat{\boldsymbol{\beta}}$. For this purpose, we state the following assumption

A2: Assumption of homoscedasticity. Assume

$$\mathsf{Var}[\boldsymbol{\epsilon}|\mathbf{X}] = \mathsf{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top|\mathbf{X}] = \sigma^2 \mathbf{I}_n,$$

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

For independent and identically distributed (iid) data, assumption **A2** is equivalent to

$$\mathsf{Var}[\epsilon_i \mid \mathbf{X}_i] = \mathsf{E}[\epsilon_i \epsilon_i^\top \mid \mathbf{X}] = \sigma^2.$$

Under the assumption **A2**, we define the OLS estimate $\widehat{\sigma}^2$ of σ^2, as

$$\widehat{\sigma}^2 = \frac{\sum_{i=1}^n \widehat{\epsilon}_i^2}{n-k} = \frac{\widehat{\boldsymbol{\epsilon}}^\top \widehat{\boldsymbol{\epsilon}}}{n-k}. \tag{2.7}$$

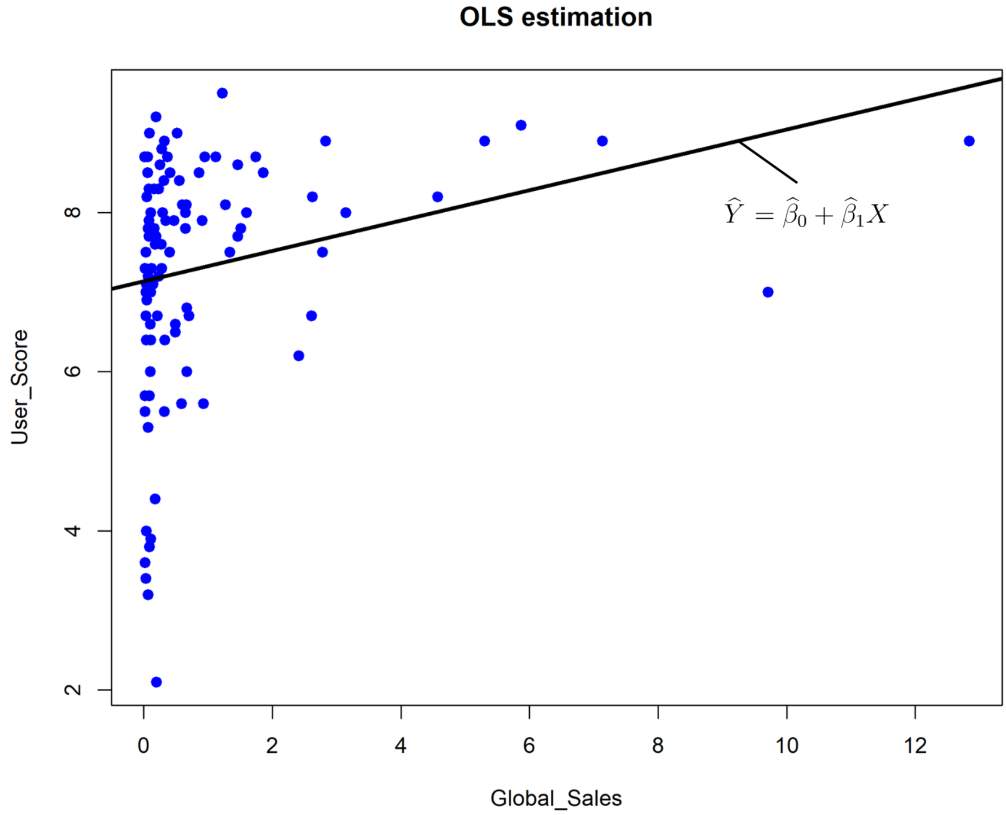


Figure 2.1: Estimating user ratings Y_i from video game sales X_i with the OLS estimate $\widehat{\boldsymbol{\beta}}$.

2.3 Properties of the OLS estimate

Utilizing the assumption of zero conditional mean **A1** and the assumption of homoscedasticity **A2** we will now derive multiple properties of OLS.

Theorem 2 (Unbiasedness of OLS)**.** *Under the assumption* **A1***, the expected value of the OLS estimate* $\widehat{\boldsymbol{\beta}}$ *is given by*

$$E[\widehat{\boldsymbol{\beta}}] = \boldsymbol{\beta},$$

therefore $\widehat{\boldsymbol{\beta}}$ *is an unbiased estimator of* $\boldsymbol{\beta}$*.*

Proof. For OLS we defined estimate $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ as

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}.$$

Since, $\mathbf{Y}$ satisfies $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, we obtain

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\epsilon}.$$

The conditional expected value of $\widehat{\boldsymbol{\beta}}$ given $\mathbf{X}$ is

$$\mathsf{E}[\widehat{\boldsymbol{\beta}} \mid \mathbf{X}] = \mathsf{E}[\boldsymbol{\beta} \mid \mathbf{X}] + \mathsf{E}[(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\epsilon} \mid \mathbf{X}] = \boldsymbol{\beta} + (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathsf{E}[\boldsymbol{\epsilon} \mid \mathbf{X}].$$

Because of the assumption **A1**, the second term is **0**, which yields

$$\mathsf{E}[\widehat{\boldsymbol{\beta}} \mid \mathbf{X}] = \boldsymbol{\beta}.$$

Finally, we can state that

$$\mathsf{E}[\widehat{\boldsymbol{\beta}}] = \mathsf{E}[\mathsf{E}[\widehat{\boldsymbol{\beta}} \mid \mathbf{X}]] = \mathsf{E}[\boldsymbol{\beta}] = \boldsymbol{\beta}.$$

□

Theorem 3. *Under the assumptions* **A1** *and* **A2**, *the OLS estimate* $\widehat{\sigma}^2$ *defined in* (2.7) *is an unbiased estimate of* σ^2, *that is*

$$\mathsf{E}[\widehat{\sigma}^2] = \sigma^2.$$

Proof. We refer to Greene, (2003) 4.6 page 76, for the proof of this claim. □

Theorem 4 (Variance of the OLS estimate)**.** *Under the assumptions* **A1** *and* **A2**, *the conditional covariance matrix of the OLS estimator given* $\mathbf{X}$ *is*

$$\mathsf{Var}[\widehat{\boldsymbol{\beta}} \mid \mathbf{X}] = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}.$$

Proof. Once more, we utilize the fact that

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$$

and

$$\begin{aligned}\mathsf{Var}[\widehat{\boldsymbol{\beta}} \mid \mathbf{X}] &= \mathsf{E}[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \mid \mathbf{X}] \\ &= \mathsf{E}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} \mid \mathbf{X}] \\ &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathsf{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top \mid \mathbf{X}]\ \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}.\end{aligned}$$

By the assumption of homoscedasticity $\mathsf{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top \mid \mathbf{X}] = \sigma^2\mathbf{I}_n$, therefore the covariance matrix of the OLS estimator can be written as

$$\mathsf{Var}[\widehat{\boldsymbol{\beta}} \mid \mathbf{X}] = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}.$$

□

We can further obtain unconditioned covariance matrix of $\widehat{\boldsymbol{\beta}}$ by employing a decomposition of variance

$$\mathsf{Var}[\widehat{\boldsymbol{\beta}}] = \mathsf{E}[\mathsf{Var}[\widehat{\boldsymbol{\beta}} \mid \mathbf{X}]] + \mathsf{Var}[\mathsf{E}[\widehat{\boldsymbol{\beta}} \mid \mathbf{X}]], \tag{2.8}$$

where $\mathsf{Var}[\mathsf{E}[\widehat{\boldsymbol{\beta}} \mid \mathbf{X}]] = \mathbf{0}$, due to $\mathsf{E}[\widehat{\boldsymbol{\beta}} \mid \mathbf{X}] = \boldsymbol{\beta}$ being a vector of constants. Therefore,

$$\mathsf{Var}[\widehat{\boldsymbol{\beta}}] = \mathsf{E}[\mathsf{Var}[\widehat{\boldsymbol{\beta}} \mid \mathbf{X}]] = \mathsf{E}[\sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}] = \sigma^2\mathsf{E}[(\mathbf{X}^\top\mathbf{X})^{-1}]. \tag{2.9}$$

In the following text, we understand the term the best linear unbiased estimator (BLUE), as a linear unbiased estimator, with the minimum covariance matrix. By that, we mean that the covariance matrix of the BLUE estimator differs from a covariance matrix of any linear unbiased estimator by a positive semidefinite matrix. In what follows, we use this notation. Let $\mathbf{A}$ and $\mathbf{B}$ be $p \times p$ square matrices. By $\mathbf{A} \geq \mathbf{B}$, we understand that the difference $\mathbf{A} - \mathbf{B} = \mathbf{C}$ is a positive-definite matrix.

We aim to determine whether the OLS estimate $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is the best linear unbiased estimator. To address this, we formulate the Gauss-Markov Theorem.

Theorem 5 (Gauss-Markov Theorem)**.** *Under the assumptions* **A1** *and* **A2***, it holds that:*

1. *The OLS estimate $\widehat{\boldsymbol{\beta}}$ is the best linear unbiased estimator of $\boldsymbol{\beta}$.*
2. *Let $\mathbf{c}$ be a vector of constant values, then $\mathbf{c}^\top \widehat{\boldsymbol{\beta}}$ is the best linear unbiased estimator of $\mathbf{c}^\top \boldsymbol{\beta}$.*

Proof. Let $\tilde{\boldsymbol{\beta}}$ be a linear unbiased estimator of $\boldsymbol{\beta}$ different from $\widehat{\boldsymbol{\beta}}$ so that

$$\tilde{\boldsymbol{\beta}} = \mathbf{Z}\mathbf{Y},$$

where $(k+1) \times n$ matrix $\mathbf{Z}$ is a function of $\mathbf{X}$. The expected value of $\tilde{\boldsymbol{\beta}}$ can be expressed as

$$\mathsf{E}[\tilde{\boldsymbol{\beta}}] = \mathsf{E}[\mathbf{Z}\mathbf{Y} \mid \mathbf{X}] = \mathsf{E}[\mathbf{Z}\mathbf{X}\boldsymbol{\beta} \mid \mathbf{X}] + \mathsf{E}[\mathbf{Z}\boldsymbol{\epsilon} \mid \mathbf{X}] = \boldsymbol{\beta}.$$

Under the assumption **A1**, it holds that $\mathsf{E}[\mathbf{Z}\boldsymbol{\epsilon} \mid \mathbf{X}] = 0$, we get $\mathbf{Z}\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}$, given any $\boldsymbol{\beta}$. Following that, we make an observation that $\mathbf{Z}\mathbf{X} = \mathbf{I_{k+1}}$. The covariance matrix of $\tilde{\boldsymbol{\beta}}$ given $\mathbf{X}$ is

$$\mathsf{Var}[\tilde{\boldsymbol{\beta}} \mid \mathbf{X}] = \sigma^2 \mathbf{Z}\mathbf{Z}^\top.$$

Now let us define a matrix $\mathbf{D}$ as $\mathbf{D} = \mathbf{Z} - (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$. Rewriting the previous equation in the following way yields

$$\begin{aligned}
\mathsf{Var}[\tilde{\boldsymbol{\beta}} \mid \mathbf{X}] &= \sigma^2(\mathbf{D} + (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top)(\mathbf{D} + (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top)^\top \\
&= \sigma^2(\mathbf{D} + (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top)(\mathbf{D}^\top + \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}) \\
&= \sigma^2(\mathbf{D}\mathbf{D}^\top + (\mathbf{X}^\top\mathbf{X})^{-1} + \mathbf{D}\mathbf{X}(\mathbf{X}\mathbf{X}^\top)^{-1} + (\mathbf{X}^\top\mathbf{X})^{-1}(\mathbf{D}\mathbf{X})^\top).
\end{aligned}$$

Combining the definition of $\mathbf{D}$ and the fact that $\mathbf{Z}\mathbf{X} = \mathbf{I}_{k+1}$

$$\mathbf{D}\mathbf{X} + (\mathbf{X}^\top\mathbf{X})^{-1}(\mathbf{X}^\top\mathbf{X}) = \mathbf{Z}\mathbf{X} = \mathbf{I}_{k+1},$$

we can state that $\mathbf{D}\mathbf{X} = \mathbf{0}$, which gives us

$$\mathsf{Var}[\tilde{\boldsymbol{\beta}} \mid \mathbf{X}] = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1} + \sigma^2(\mathbf{D}\mathbf{D}^\top) = \mathsf{Var}[\widehat{\boldsymbol{\beta}} \mid \mathbf{X}] + \sigma^2(\mathbf{D}\mathbf{D}^\top)$$

where the matrix $\mathbf{D}\mathbf{D}^\top$ is positive-semidefinite.

This means that the covariance matrix of $\tilde{\boldsymbol{\beta}}$ given $\mathbf{X}$, differs from the covariance matrix of $\widehat{\boldsymbol{\beta}}$ given $\mathbf{X}$, by a positive-semidefinite matrix $\mathbf{D}\mathbf{D}^\top$. Hence,

$$\mathsf{Var}[\tilde{\boldsymbol{\beta}} \mid \mathbf{X}] \geq \mathsf{Var}[\widehat{\boldsymbol{\beta}} \mid \mathbf{X}]. \tag{2.10}$$

Given that, for any $(k+1) \times 1$ column vector $\mathbf{p}$ and $(k+1) \times (k+1)$ random matrices $\mathbf{V}_1$, $\mathbf{V}_2$

$$\mathbf{p}^\top (\mathbf{V}_1 - \mathbf{V}_2)\mathbf{p} \geq 0,$$

implies

$$\mathbf{p}^\top \mathsf{E}[\mathbf{V}_1 - \mathbf{V}_2]\mathbf{p} \geq 0.$$

We can then write

$$\mathsf{E}[\mathsf{Var}[\tilde{\boldsymbol{\beta}} \mid \mathbf{X}]] \geq \mathsf{E}[\mathsf{Var}[\widehat{\boldsymbol{\beta}} \mid \mathbf{X}]],$$

which, according to (2.8) is an equivalent to

$$\mathsf{Var}[\tilde{\boldsymbol{\beta}}] \geq \mathsf{Var}[\widehat{\boldsymbol{\beta}}].$$

Therefore, $\widehat{\boldsymbol{\beta}}$ is a linear unbiased estimator of $\boldsymbol{\beta}$ with the minimum covariance matrix. $\square$

3. Heteroscedasticity data

This chapter is based on Greene, (2003), Wooldridge, (2013), Heij et al., (2004), Harvey, (1976), Romano and Wolf, (2016). It aims to explore how the presence of heteroscedasticity impacts the ordinary least squares (OLS) estimate. Additionally, in Chapter 3.3 we define the weighted least square estimate, and later feasible weighted least square estimate.

The author's contribution lies in providing a comprehensive breakdown and analysis of specific proofs and derivations. Additionally, they provide a discussion on the corollaries of such proofs.

In the previous chapter, we have shown that the OLS estimator is the best linear unbiased estimator under the assumptions **A1** and **A2**. In this chapter, we examine the change in this behavior when the assumption of homoscedasticity **A2** is omitted, in other words when heteroscedasticity is present.

3.1 Heteroscedasticity

In the previous chapter, we assumed homoscedasticity as

$$\mathsf{Var}[\epsilon_i \mid \mathbf{X}_i] = \mathsf{E}[\epsilon_i \epsilon_i^\top \mid \mathbf{X}] = \sigma^2,$$

where $i = 1, 2, \dots, n$ and σ^2 is a positive constant.

We say heteroscedasticity is present when the conditional variance of the unobserved errors ϵ_i is not constant across observations. Hence,

$$\mathsf{Var}[\epsilon_i \mid \mathbf{X}_i] = \sigma_i^2 = h(\mathbf{X}_i), \tag{3.1}$$

where $h : \mathbb{R}^{k+1} \to \mathbb{R}$ is a non-constant function.

Let us denote $\mathbf{H}$ as a $n \times n$ diagonal matrix, so that

$$\mathsf{Var}[\boldsymbol{\epsilon} \mid \mathbf{X}] = \mathbf{H} = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix}. \tag{3.2}$$

In some special cases, we can express (3.2) as

$$\mathsf{Var}[\boldsymbol{\epsilon} \mid \mathbf{X}] = \sigma^2 \boldsymbol{\Omega} = \sigma^2 \begin{bmatrix} \omega_1 & 0 & \dots & 0 \\ 0 & \omega_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \omega_n \end{bmatrix}, \tag{3.3}$$

where σ^2 depicts a scale and $\boldsymbol{\Omega}$ is a $n \times n$ diagonal matrix with ω_i as i-th diagonal element. $\boldsymbol{\Omega}$ represents the form of heteroscedasticity for the given model.

Example 2. Taking a look back at the data sample from Example 1, Figure 3.1 illustrates a greater variance for lower values of X, suggesting the presence of heteroscedasticity.

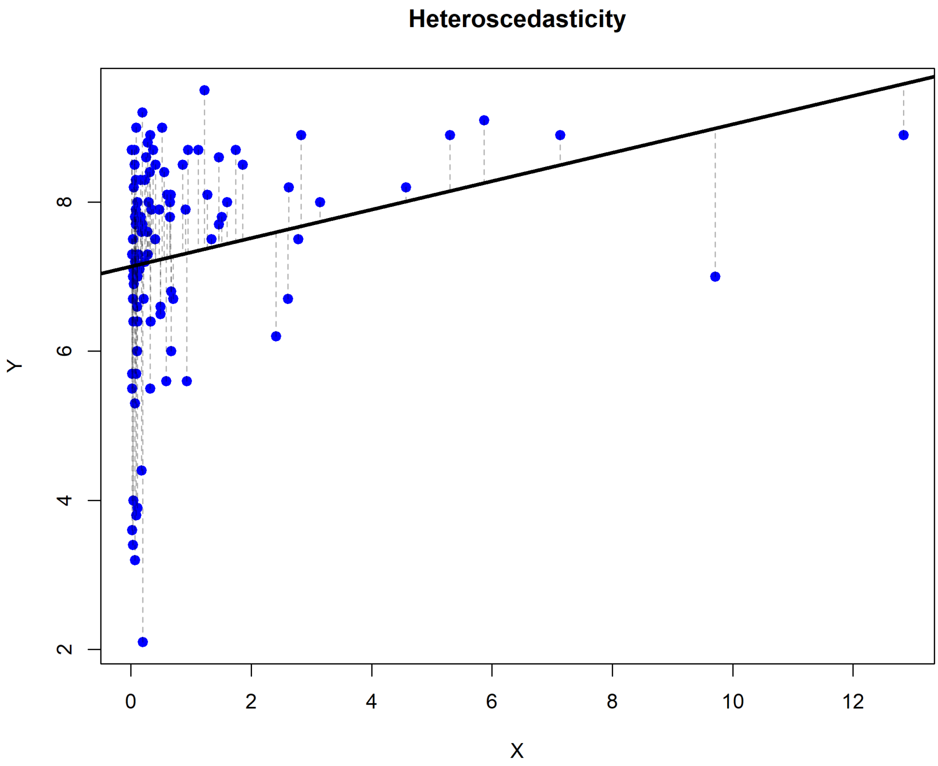


Figure 3.1: Data showcasing heteroscedasticity

Next, we will discuss the properties of the OLS estimate.

3.2 Properties of the OLS estimate

We can show that in the presence of heteroscedasticity, the least squares estimator $\widehat{\boldsymbol{\beta}}$ maintains its properties of being unbiased, consistent, and asymptotically normally distributed.

In Theorem 2, we proved that under the assumption **A1**, $\widehat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}$. Therefore, the presence of heteroscedasticity does not alter the unbiasedness of the OLS estimate.

Theorem 6. *Under the assumption* **A1**, *let*

$$\mathbf{Q} = \mathsf{E}[\mathbf{X}_i \mathbf{X}_i^\top] \tag{3.4}$$

be a finite positive definite matrix, then $\widehat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}$.

Proof. We have

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} = \boldsymbol{\beta} + (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\epsilon}, \tag{3.5}$$

To examine asymptotic properties of $\widehat{\boldsymbol{\beta}}$ we rearrange the equation (3.5) in a following way

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \left(\frac{\mathbf{X}^\top\mathbf{X}}{n}\right)^{-1}\left(\frac{\mathbf{X}^\top\boldsymbol{\epsilon}}{n}\right) \tag{3.6}$$

then

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\frac{\mathbf{X}^\top\mathbf{X}}{n}\right)^{-1}\left(\frac{\mathbf{X}^\top\boldsymbol{\epsilon}}{n}\right). \tag{3.7}$$

We can further dissolve the right side of the equation into two parts $\left(\frac{\mathbf{X}^\top\mathbf{X}}{n}\right)^{-1}$ and $\left(\frac{\mathbf{X}^\top\boldsymbol{\epsilon}}{n}\right)$, where for $\left(\frac{\mathbf{X}^\top\mathbf{X}}{n}\right)^{-1}$, by using the law of large numbers, the first part can be expressed as

$$\left(\frac{\mathbf{X}^\top\mathbf{X}}{n}\right) = \frac{1}{n}\left[\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i^\top\right] \xrightarrow{P} \mathsf{E}[\mathbf{X}_i\mathbf{X}_i^\top] = \mathbf{Q},$$

which is a finite positive definite matrix, by the assumption made in (3.4).

For the second part $\left(\frac{\mathbf{X}^\top\boldsymbol{\epsilon}}{n}\right)$, under the assumption **A1**, we apply the law of large numbers and utilize the fact that $\mathsf{Cov}[\mathbf{X}_i, \epsilon_i] = \mathbf{0}$ from (2.3), to derive

$$\left(\frac{\mathbf{X}^\top\boldsymbol{\epsilon}}{n}\right) = \frac{1}{n}\sum_{i=1}^{n}\epsilon_i\mathbf{X}_i \xrightarrow{P} \mathsf{E}[\epsilon_i\mathbf{X}_i] = \mathsf{Cov}[\mathbf{X}_i, \epsilon_i] = \mathbf{0}.$$

Finally, for equation (3.6) we obtain

$$\widehat{\boldsymbol{\beta}} \xrightarrow{P} \boldsymbol{\beta} + \mathbf{Q}^{-1}\cdot\mathbf{0} = \boldsymbol{\beta}.$$

Hence, $\widehat{\boldsymbol{\beta}}$ is a consistent estimate of $\boldsymbol{\beta}$. □

We showed that in the presence of heteroscedasticity, $\widehat{\boldsymbol{\beta}}$ remains both unbiased and consistent. The covariance matrix of OLS estimator given $\mathbf{X}$ is defined as

$$\begin{aligned}\mathsf{Var}[\widehat{\boldsymbol{\beta}}|\mathbf{X}] &= \mathsf{E}[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top|\mathbf{X}] \\ &= \mathsf{E}[(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top(\mathbf{X}^\top\mathbf{X})^{-1}|\mathbf{X}].\end{aligned}$$

Utilizing (3.5) and (3.2), we obtain

$$\mathsf{Var}[\widehat{\boldsymbol{\beta}}|\mathbf{X}] = (\mathbf{X}^\top\mathbf{X})^{-1}(\mathbf{X}^\top\mathbf{H}\mathbf{X})(\mathbf{X}^\top\mathbf{X})^{-1}. \tag{3.8}$$

Theorem 7. *With prerequisites from Theorem 6 and function $h(\mathbf{X}_i)$ from* (3.1)*, let*

$$\mathbf{W} = \mathsf{E}\left[h(\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^\top\right],$$

be a finite $(k+1)\times(k+1)$ matrix. Then

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{D} \mathsf{N}(\mathbf{0}, \mathbf{Q}^{-1}\mathbf{W}\mathbf{Q}^{-1}),$$

where $\mathbf{Q}$ is defined in (3.4).

Proof. To describe asymptotic distribution of $\widehat{\boldsymbol{\beta}}$, we first multiply (3.7) by $\sqrt{n}$

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)=\left(\frac{\mathbf{X}^\top\mathbf{X}}{n}\right)^{-1}\left(\frac{\sum_{i=1}^{n}(\epsilon_i\mathbf{X}_i)}{\sqrt{n}}\right)=\left(\frac{1}{n}\left[\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i^\top\right]\right)^{-1}\left(\frac{\sum_{i=1}^{n}(\epsilon_i\mathbf{X}_i)}{\sqrt{n}}\right).$$

We denote $\mathbf{U}_i$ as $\mathbf{U}_i=\epsilon_i\mathbf{X}_i$. Under the assumption **A1**

$$\mathsf{E}[\mathbf{U}_i]=\mathsf{E}[\epsilon_i\mathbf{X}_i]=\mathsf{E}\left[\mathsf{E}\left[\epsilon_i\mathbf{X}_i\mid\mathbf{X}\right]\right]=\mathbf{0}$$

and

$$\mathsf{Var}\left[\mathbf{U}_i\right]=\mathsf{E}\left[\mathbf{U}_i\mathbf{U}_i^\top\right]=\mathsf{E}\left[\epsilon_i^2\mathbf{X}_i\mathbf{X}_i^\top\right]=\mathsf{E}\left[\mathsf{E}\left[\epsilon_i^2\mathbf{X}_i\mathbf{X}_i^\top\mid\mathbf{X}\right]\right],$$

whereby applying (3.1), we obtain

$$\mathsf{Var}[\mathbf{U}_i]=\mathsf{E}\left[h(\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^\top\right]=\mathbf{W}.$$

Since random vectors $\mathbf{U}_i$ are independent and identically distributed with a finite covariance matrix, we apply the central limit theorem and gain

$$\frac{\sum_{i=1}^{n}\mathbf{U}_i}{\sqrt{n}}\xrightarrow{D}\mathsf{N}(\mathbf{0},\mathbf{W}).$$

Finally, using the fact that $\left(\frac{\mathbf{X}^\top\mathbf{X}}{n}\right)^{-1}\xrightarrow{P}\mathbf{Q}^{-1}$, we apply Slutsky's theorem to obtain

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\xrightarrow{D}\mathsf{N}(\mathbf{0},\mathbf{Q}^{-1}\mathbf{W}\mathbf{Q}^{-1}).$$

□

In this section, we have proven that removing the assumption of homoscedasticity doesn't affect the OLS estimate unbiasedness. However, it will no longer remain the most efficient estimate. In the following section, we will explore a new estimation approach that generally outperforms OLS in terms of efficiency.

3.3 Weighted least squares

This section discusses techniques used to estimate $\boldsymbol{\beta}$, when $\mathbf{H}$ (3.2) can be expressed as (3.3), so that $\boldsymbol{\Omega}$ is a known function of $\mathbf{X}$ and σ^2 is an unknown parameter.

Under the assumption **A1**, consider a multiple linear regression (MLR) model (2.2)

$$\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\epsilon},\tag{3.9}$$

where we assume that $\boldsymbol{\Omega}$ is a known positive definite $n\times n$ diagonal matrix

$$\boldsymbol{\Omega}=\begin{bmatrix}\omega_1&0&\dots&0\\0&\omega_2&\dots&0\\\vdots&\vdots&\ddots&\vdots\\0&0&\dots&\omega_n\end{bmatrix}.$$

Knowing the form of heteroscedasticity will allow us to transform our model so that the transformed model satisfies the assumption **A2**.

First, given that $\mathbf{\Omega}$ is a positive definite matrix, we can define a matrix $\mathbf{\Omega}^{-\frac{1}{2}}$ such that

$$\mathbf{\Omega}^{-\frac{1}{2}}\mathbf{\Omega}^{-\frac{1}{2}} = \mathbf{\Omega}^{-1}. \tag{3.10}$$

This allows us to transform the MLR model by multiplying it with $\mathbf{\Omega}^{-\frac{1}{2}}$ from the left, yielding

$$\mathbf{\Omega}^{-\frac{1}{2}}\mathbf{Y} = \mathbf{\Omega}^{-\frac{1}{2}}\mathbf{X}\boldsymbol{\beta} + \mathbf{\Omega}^{-\frac{1}{2}}\boldsymbol{\epsilon}. \tag{3.11}$$

We reffer to (3.11), as

$$\mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\epsilon}^*, \tag{3.12}$$

where $\mathbf{Y}^* = \mathbf{\Omega}^{-\frac{1}{2}}\mathbf{Y}, \mathbf{X}^* = \mathbf{\Omega}^{-\frac{1}{2}}\mathbf{X}$ and $\boldsymbol{\epsilon}^* = \mathbf{\Omega}^{-\frac{1}{2}}\boldsymbol{\epsilon}$.

Secondly, the covariance matrix of the vector of unobserved errors $\boldsymbol{\epsilon}$ can then be written as

$$\mathsf{Var}[\boldsymbol{\epsilon}^* \mid \mathbf{X}] = \mathsf{Var}\left[\mathbf{\Omega}^{-\frac{1}{2}}\boldsymbol{\epsilon} \mid \mathbf{X}\right] = \sigma^2\mathbf{\Omega}^{-\frac{1}{2}}\mathbf{\Omega}\mathbf{\Omega}^{-\frac{1}{2}} = \sigma^2\mathbf{I}_n,$$

where $\mathbf{I}_n$ is a $n \times n$ unit matrix.

Furthermore, for the transformed model (3.12), we find that

$$\mathsf{E}[\mathbf{\Omega}^{-\frac{1}{2}}\boldsymbol{\epsilon} \mid \mathbf{X}] = \begin{pmatrix} \mathsf{E}[\omega_1^{-\frac{1}{2}}\epsilon_1 \mid \mathbf{X}] \\ \mathsf{E}[\omega_2^{-\frac{1}{2}}\epsilon_2 \mid \mathbf{X}] \\ \vdots \\ \mathsf{E}[\omega_n^{-\frac{1}{2}}\epsilon_n \mid \mathbf{X}] \end{pmatrix} = \mathbf{0}.$$

Hence, the transformed model (3.12) satisfies both assumptions **A1** and **A2**. Consequently, we obtain the OLS estimate $\widehat{\boldsymbol{\beta}}_{WLS}$ of $\boldsymbol{\beta}$ in terms of the transformed model (3.12). We have

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_{WLS} &= (\mathbf{X}^{*\top}\mathbf{X}^*)^{-1}\mathbf{X}^{*\top}\mathbf{Y}^* \\ &= ((\mathbf{\Omega}^{-\frac{1}{2}}\mathbf{X})^\top\mathbf{\Omega}^{-\frac{1}{2}}\mathbf{X})^{-1}(\mathbf{\Omega}^{-\frac{1}{2}}\mathbf{X})^\top\mathbf{\Omega}^{-\frac{1}{2}}\mathbf{Y} \\ &= (\mathbf{X}^\top\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{\Omega}^{-1}\mathbf{Y}. \end{aligned} \tag{3.13}$$

The weighted least square estimator $\widehat{\boldsymbol{\beta}}_{WLS}$ is by the Gauss-Markov theorem 5, the best linear unbiased estimator (BLUE).

Weighted least squares estimate. We derived the estimate $\widehat{\boldsymbol{\beta}}_{WLS}$ as the OLS estimate in the transformed model (3.12), given by

$$\widehat{\boldsymbol{\beta}}_{WLS} = \underset{\boldsymbol{\beta}\in\mathbb{R}^{k+1}}{\arg\min}(\mathbf{Y}^* - \mathbf{X}^*\boldsymbol{\beta})^\top(\mathbf{Y}^* - \mathbf{X}^*\boldsymbol{\beta}),$$

minimizes the sum of squares in the transformed model (3.12). By expressing the objective function we aim to minimize in terms of the original model (3.9), we obtain

$$\begin{aligned} (\mathbf{Y}^* - \mathbf{X}^*\boldsymbol{\beta})^\top(\mathbf{Y}^* - \mathbf{X}^*\boldsymbol{\beta}) &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top\mathbf{\Omega}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \sum_{i=1}^{n}\frac{1}{\omega_i}(Y_i - \mathbf{X}_i^\top\boldsymbol{\beta})^2. \end{aligned}$$

Hence, $\widehat{\boldsymbol{\beta}}_{WLS}$ can be rewritten as

$$\widehat{\boldsymbol{\beta}}_{WLS} = \underset{\beta \in \mathbb{R}^{k+1}}{\arg\min} \sum_{i=1}^{n} \frac{1}{\omega_i}(Y_i - \mathbf{X}_i^\top \boldsymbol{\beta})^2.$$

Due to that, $\widehat{\boldsymbol{\beta}}_{WLS}$ is termed a weighted least squares (denoted as WLS) estimate. The WLS estimate of the original model (3.9) is equivalent to the OLS estimate of the transformed model (3.12).

Finally, utilizing (3.13), we can rewrite $\widehat{\boldsymbol{\beta}}_{WLS}$ as

$$\widehat{\boldsymbol{\beta}}_{WLS} = \left[\sum_{i=1}^{n} w_i \mathbf{X}_i \mathbf{X}_i^\top\right]^{-1} \left[\sum_{i=1}^{n} w_i \mathbf{X}_i Y_i\right], \tag{3.14}$$

where $w_i = \frac{1}{\omega_i}$ are called weigths.

Lemma 8. *Let us have weigths $w_i > 0$, for $i = 1, 2, \ldots, n$ and a constant $\gamma \in \mathbb{R}_+ \setminus \{0\}$. If $\widehat{\boldsymbol{\beta}}_{WLS}$ is a solution to* (3.14) *with weigths w_i and $\tilde{\boldsymbol{\beta}}_{WLS}$ is a solution to* (3.14) *with weigths $\gamma \cdot w_i$, then $\widehat{\boldsymbol{\beta}}_{WLS} = \tilde{\boldsymbol{\beta}}_{WLS}$.*

Proof. From (3.14) we have

$$\begin{aligned}
\tilde{\boldsymbol{\beta}}_{WLS} &= \left[\sum_{i=1}^{n} \gamma \cdot w_i \mathbf{X}_i \mathbf{X}_i^\top\right]^{-1} \left[\sum_{i=1}^{n} \gamma \cdot w_i \mathbf{X}_i Y_i\right] \\
&= \left[\sum_{i=1}^{n} w_i \mathbf{X}_i \mathbf{X}_i^\top\right]^{-1} \left[\sum_{i=1}^{n} w_i \mathbf{X}_i Y_i\right] = \widehat{\boldsymbol{\beta}}_{WLS}.
\end{aligned}$$

□

In the following examples, we will discuss two forms of $\boldsymbol{\Omega}$. Before we do that, let us have a model (2.2)

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

and $\boldsymbol{\Omega}^{-\frac{1}{2}}$ (3.10)

$$\boldsymbol{\Omega}^{-\frac{1}{2}} = \begin{bmatrix} \sqrt{\frac{1}{\omega_1}} & 0 & \cdots & 0 \\ 0 & \sqrt{\frac{1}{\omega_2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\frac{1}{\omega_n}} \end{bmatrix}.$$

By rewriting the form of transformations defined in (3.12), we get

$$\mathbf{Y}^* = \boldsymbol{\Omega}^{-\frac{1}{2}}\mathbf{Y} = \begin{bmatrix} \frac{Y_1}{\sqrt{\omega_1}} \\ \frac{Y_2}{\sqrt{\omega_2}} \\ \vdots \\ \frac{Y_n}{\sqrt{\omega_n}} \end{bmatrix}, \mathbf{X}^* = \boldsymbol{\Omega}^{-\frac{1}{2}}\mathbf{X} = \begin{bmatrix} \frac{\mathbf{X_1}^\top}{\sqrt{\omega_1}} \\ \frac{\mathbf{X_2}^\top}{\sqrt{\omega_2}} \\ \vdots \\ \frac{\mathbf{X_n}^\top}{\sqrt{\omega_n}} \end{bmatrix}, \boldsymbol{\epsilon}^* = \boldsymbol{\Omega}^{-\frac{1}{2}}\boldsymbol{\epsilon} = \begin{bmatrix} \frac{\boldsymbol{\epsilon_1}}{\sqrt{\omega_1}} \\ \frac{\boldsymbol{\epsilon_2}}{\sqrt{\omega_2}} \\ \vdots \\ \frac{\boldsymbol{\epsilon_n}}{\sqrt{\omega_n}} \end{bmatrix}.$$

Example 3. **Variance proportional to a regressor**

For some fixed $j = 1, \dots, k$, assume that

$$\mathsf{Var}[\epsilon_i|\mathbf{X}_i] = \sigma_i^2 = \sigma^2 X_{ij}.$$

The form of the transformed model (3.12) is then

$$\frac{Y_i}{\sqrt{X_{ij}}} = \beta_j\sqrt{X_{ij}} + \frac{\beta_0}{\sqrt{X_{ij}}} + \frac{\beta_1 X_{i1}}{\sqrt{X_{ij}}} + \ldots + \frac{\beta_k X_{ik}}{\sqrt{X_{ij}}} + \frac{\epsilon_i}{\sqrt{X_{ij}}}. \tag{3.15}$$

The WLS estimate (3.14) can be expressed as

$$\widehat{\boldsymbol{\beta}}_{WLS} = \left[\sum_{i=1}^{n} \frac{1}{X_{ij}}\mathbf{X}_i\mathbf{X}_i^\top\right]^{-1}\left[\sum_{i=1}^{n} \frac{1}{X_{ij}}\mathbf{X}_i Y_i\right].$$

Now let us compare the WLS estimate with the OLS estimate, using a simple model

$$Y_i = \beta X_i + \epsilon_i,$$

and

$$\mathsf{Var}[\epsilon_i|\mathbf{X}] = \sigma_i^2 = \sigma^2 X_i,$$

where the matrix $\mathbf{X} = (X_1, \dots, X_n)^\top$.

From (3.15) we get

$$\frac{Y_i}{\sqrt{X_i}} = \beta\sqrt{X_i} + \frac{\epsilon_i}{\sqrt{X_i}}.$$

The WLS estimate (3.14) can be written as

$$\widehat{\beta}_{WLS} = \left[\sum_{i=1}^{n} \frac{1}{X_i}X_i^2\right]^{-1}\left[\sum_{i=1}^{n} \frac{1}{X_i}X_i Y_i\right] = \left[\sum_{i=1}^{n} X_i\right]^{-1}\left[\sum_{i=1}^{n} Y_i\right],$$

while the OLS estimate as

$$\widehat{\beta} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}\mathbf{Y} = \left[\sum_{i=1}^{n} X_i^2\right]^{-1}\left[\sum_{i=1}^{n} X_i Y_i\right].$$

It is clear, that we obtained two different unbiased estimates, for which the general theory states that the WLS estimate $\widehat{\beta}_{WLS}$ has lesser variance than the OLS estimate $\widehat{\beta}$.

Example 4. **Variance proportional to squared regressor**

Following the previous example, let us now assume that

$$\mathsf{Var}[\epsilon_i|\mathbf{X}_i] = \sigma_i^2 = \sigma^2 X_{ij}^2.$$

Then (3.12) takes form

$$\frac{Y_i}{X_{ij}} = \beta_j + \frac{\beta_0}{X_{ij}} + \frac{\beta_1 X_{i1}}{X_{ij}} + \ldots + \frac{\beta_k X_{ik}}{X_{ij}} + \frac{\epsilon_i}{X_{ij}}.$$

The WLS estimate (3.14) is

$$\widehat{\boldsymbol{\beta}}_{WLS} = \left[\sum_{i=1}^{n} \frac{1}{X_{ij}^2} \mathbf{X}_i \mathbf{X}_i^\top\right]^{-1} \left[\sum_{i=1}^{n} \frac{1}{X_{ij}^2} \mathbf{X}_i Y_i\right].$$

We will again compare the WLS estimate with the OLS estimate using a model

$$Y_i = \beta X_i + \epsilon_i,$$

and

$$\mathsf{Var}[\epsilon_i|\mathbf{X}] = \sigma_i^2 = \sigma^2 X_i^2.$$

where $\mathbf{X} = (X_1, \ldots, X_n)^\top$.

Transforming the model according to (3.12) gives us

$$\frac{Y_i}{X_i} = \beta + \frac{\epsilon_i}{X_i}.$$

The WLS estimate (3.14) can be written as

$$\widehat{\beta}_{WLS} = \left[\sum_{i=1}^{n} \frac{1}{X_i^2} X_i^2\right]^{-1} \left[\sum_{i=1}^{n} \frac{1}{X_i} Y_i\right] = \left[\sum_{i=1}^{n} \frac{Y_i}{X_i}\right],$$

While the OLS estimate as

$$\widehat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X} \mathbf{Y} = \left[\sum_{i=1}^{n} X_i^2\right]^{-1} \left[\sum_{i=1}^{n} X_i Y_i\right].$$

Similarly to the previous example, we obtained two different unbiased estimates, so that the WLS estimate $\widehat{\beta}_{WLS}$ has lesser variance than the OLS estimate $\widehat{\beta}$.

Corollary 1. Under the assumption **A1**, if

$$\mathbf{Q} = \mathsf{E}[w_i \mathbf{X}_i \mathbf{X}_i^\top] \tag{3.16}$$

is a finite positive definite matrix, then $\widehat{\boldsymbol{\beta}}_{WLS}$ is a consistent estimator of $\boldsymbol{\beta}$.

Proof. The proof follows the same steps as in Theorem 6, applied on the transformed model (3.12). □

According to the Corollary 1, the WLS estimate (3.13) is consistent, under the assumption **A1**, for all forms of $\boldsymbol{\Omega}$ (given that $\mathbf{Q}$ 3.16 is finite). However, in the case of more complex MLR models, identifying the exact form of $\boldsymbol{\Omega}$ is often impossible, and choosing a wrong form of $\boldsymbol{\Omega}$ may result in an inefficient estimate. In the following section, we will explore estimation techniques applicable in scenarios where the form of $\boldsymbol{\Omega}$ is partially unknown.

3.4 Feasible weighted least squares

Previously we assumed

$$\mathsf{Var}[\epsilon_i \mid \mathbf{X}_i] = h(\mathbf{X}_i),$$

where h is a non-constant function. In the preceding section, we addressed the case where

$$h(x) = \sigma^2 \omega(x),$$

such that ω is a known function. Now, we broaden this scope to situations where $\mathbf{X}$ depends on some finite-dimensional parameter $\boldsymbol{\alpha}$. Consequently, the matrix

$$\boldsymbol{\Omega} = \boldsymbol{\Omega}(\boldsymbol{\alpha})$$

becomes dependent on $\boldsymbol{\alpha}$, making the straightforward application of the WLS estimate unfeasible. We proceed by first estimating $\boldsymbol{\alpha}$, thereby obtaining $\widehat{\boldsymbol{\Omega}} = \Omega(\widehat{\boldsymbol{\alpha}})$, which is used to obtain the feasible weighted least squares (denoted as FWLS) estimate $\widehat{\boldsymbol{\beta}}_{FWLS}$, defined as

$$\widehat{\boldsymbol{\beta}}_{FWLS} = (\mathbf{X}^\top \widehat{\boldsymbol{\Omega}}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \widehat{\boldsymbol{\Omega}}^{-1} \mathbf{Y}. \tag{3.17}$$

Example 5. In Example 4 we had $\mathsf{Var}[\epsilon_i \mid \mathbf{X}_i] = \sigma^2 X_i^2$ for a model

$$Y_i = \beta X_i + \epsilon_i.$$

Assuming $X_i > 0$ we can generalize the power exponent in the Example 4 so that

$$\mathsf{Var}[\epsilon_i \mid \mathbf{X}_i] = \sigma^2 X_i^\alpha,$$

where $\alpha \in \mathbb{R}$ is an unknown parameter. Consequently, our problem shifts towards finding a consistent estimate $\widehat{\alpha}$ of α.

Based on the previous example, we formulate a parametric model for the conditional variance of unobserved errors as follows

$$\mathsf{Var}[\epsilon_i \mid \mathbf{X}_i] = h(\mathbf{X}_i, \boldsymbol{\alpha}), \tag{3.18}$$

where h is known and $\boldsymbol{\alpha} \in \mathbb{R}^d$ is a vector of unknown parameters.

3.4.1 Two-step estimation

We can summarize two-step estimation into the following steps:

1. Construct an estimate $\widehat{\boldsymbol{\alpha}}$ of $\boldsymbol{\alpha}$. This is usually done by firstly obtaining the OLS estimate $\widehat{\boldsymbol{\beta}}$ and using it to calculate the residuals $\widehat{\boldsymbol{\epsilon}}$. Secondly, we estimate $\widehat{\boldsymbol{\alpha}}$ from a auxiliary model using transformed residuals $\widehat{\boldsymbol{\epsilon}}$.

2. Use $\widehat{\boldsymbol{\alpha}}$ to calculate the FWLS estimate $\widehat{\boldsymbol{\beta}}_{FWLS}$ from (3.17).

We will demonstrate the first step of the two-step estimation method for several selected common models.

Multiplicative model. In this particular model, we assume the form of heteroscedasticity to be

$$\mathsf{Var}[\epsilon_i \mid \mathbf{X}_i] = h(\mathbf{X}_i, \boldsymbol{\alpha}) = \exp[\boldsymbol{\alpha}^\top \mathbf{X}_i],$$

where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_k)^\top$ is a $(k+1) \times 1$ vector of unknown parameters.

We define a random variable V_i, so that

$$V_i = \frac{\epsilon_i^2}{h(\mathbf{X}_i, \boldsymbol{\alpha})}.$$

Therefore, $\mathsf{E}[V_i] = 1$. We can now write

$$\epsilon_i^2 = \exp[\boldsymbol{\alpha}^\top \mathbf{X}_i] V_i. \tag{3.19}$$

By applying the log transform to (3.19), we get

$$\log(\epsilon_i^2) = \boldsymbol{\alpha}^\top \mathbf{X}_i + \log(V_i), \tag{3.20}$$

Since the unobserved errors ϵ_i are unknown, we replace them with the residuals $\widehat{\epsilon_i}$ obtained by the OLS method. Let $\widehat{\boldsymbol{\beta}}$ be the OLS estimate, and

$$\widehat{\epsilon_i} = Y_i - \mathbf{X}_i^\top \widehat{\boldsymbol{\beta}}, \quad \epsilon_i = Y_i - \mathbf{X}_i^\top \boldsymbol{\beta}.$$

Then

$$\widehat{\epsilon_i} = \epsilon_i - \mathbf{X}_i^\top (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \epsilon_i - \gamma_i.$$

However, since $\widehat{\boldsymbol{\beta}} \xrightarrow{P} \boldsymbol{\beta}$, which was shown in Theorem 6, γ_i will asymptotically become negligible.
Finally, we rewrite the equation (3.20) to get a multiplicative model

$$\log(\widehat{\epsilon_i}^2) = \boldsymbol{\alpha}^\top \mathbf{X}_i + \log(V_i) - \log(\epsilon_i^2) + \log(\widehat{\epsilon_i}^2) = \boldsymbol{\alpha}^\top \mathbf{X}_i + e_i, \tag{3.21}$$

where $e_i = \log(V_i) - \log(\epsilon_i^2) + \log(\widehat{\epsilon_i}^2)$ is a random variable.

Using the OLS method on the model, we then obtain an estimate $\widehat{\boldsymbol{\alpha}} = (\widehat{\alpha}_0, \widehat{\alpha}_1, \dots, \widehat{\alpha}_k)^\top$ of $\boldsymbol{\alpha}$, where $\widehat{\alpha}_0$ is not a consistent estimate of α_0 while $\widehat{\alpha}_j, j = 1, 2, \dots, k$, are consistent estimates of α_j Harvey, (1976) page 463. We define vectors $\widehat{\boldsymbol{\alpha}}^* = (\widehat{\alpha}_1, \dots, \widehat{\alpha}_k)^\top$ and $\mathbf{X}_i^* = (X_{i1}, X_{i2}, \dots, X_{ik})^\top$, that is the original $\mathbf{X}_i$, from which we remove the first element, i.e., the absolute term. We can now form following equation

$$\exp[\widehat{\boldsymbol{\alpha}}^\top \mathbf{X}_i] = \exp[\widehat{\alpha}_0] \cdot \exp[\widehat{\boldsymbol{\alpha}}^{*\top} \mathbf{X}_i^*]. \tag{3.22}$$

Applying Lemma 8 to equation (3.22) implies that the inconsistency of $\widehat{\alpha}_0$ does not affect the WLS estimate, we aim to obtain.

Finally, utilizing $\widehat{\boldsymbol{\alpha}}$, we can derive the estimate $h(\mathbf{X}_i, \widehat{\boldsymbol{\alpha}})$ of $h(\mathbf{X}_i, \boldsymbol{\alpha})$ as

$$h(\mathbf{X}_i, \widehat{\boldsymbol{\alpha}}) = \exp[\widehat{\boldsymbol{\alpha}}^\top \mathbf{X}_i].$$

This enables us to employ the WLS method (3.14) with weights $\frac{1}{h(\mathbf{X}_i, \widehat{\boldsymbol{\alpha}})}$ to compute $\widehat{\boldsymbol{\beta}}_{FWLS}$ (3.17).

Additive model. We assume the form of heteroscedasticity to be

$$\mathsf{Var}[\epsilon_i \mid \mathbf{X}_i] = h(\mathbf{X}_i, \boldsymbol{\alpha}) = \boldsymbol{\alpha}^\top \mathbf{X}_i,$$

so that both $X_{ij} > 0$ and $\alpha_j > 0$, for $j = 0, 1, \ldots, k$.

We define a random variable V_i, so that

$$V_i = \epsilon_i^2 - h(\mathbf{X}_i, \boldsymbol{\alpha}),$$

resulting in $\mathsf{E}[V_i] = 0$.

Thus, we express

$$\epsilon_i^2 = \boldsymbol{\alpha}^\top \mathbf{X}_i + V_i. \tag{3.23}$$

Using the same reasoning as for the multiplicative model, we estimate alpha from the auxiliary model

$$\widehat{\epsilon}_i^{\,2} = \boldsymbol{\alpha}^\top \mathbf{X}_i + e_i. \tag{3.24}$$

Power model. We assume the form of heteroscedasticity to be

$$\mathsf{Var}[\epsilon_i \mid \mathbf{X}_i] = h(\mathbf{X}_i, \alpha) = \left(\mathbf{X}_i^\top \boldsymbol{\beta}\right)^\alpha,$$

where $\alpha \in \mathbb{R}$ and $Y_i > 0$.

Utilizing V_i from the multiplicative model, we write

$$\epsilon_i^2 = \left(\mathbf{X}_i^\top \boldsymbol{\beta}\right)^\alpha V_i. \tag{3.25}$$

By applying the log transform to (3.25), we get

$$\log(\epsilon_i^2) = \alpha \log\left(\mathbf{X}_i^\top \boldsymbol{\beta}\right) + \log(V_i).$$

As for the previous models, we replace ϵ_i with $\widehat{\epsilon}_i$. Additionally, using the fact that $\widehat{\boldsymbol{\beta}} \xrightarrow{P} \boldsymbol{\beta}$, we replace $\boldsymbol{\beta}$ with $\widehat{\boldsymbol{\beta}}$. Thus, we obtain an auxiliary model

$$\log(\widehat{\epsilon}_i^2) = \alpha \log\left(\mathbf{X}_i^\top \widehat{\boldsymbol{\beta}}\right) + e_i = \alpha \log(\widehat{Y}_i) + e_i. \tag{3.26}$$

Power in X_i model. The final model we present applies to the SLR model. We assume the form of heteroscedasticity to be

$$\mathsf{Var}[\epsilon_i \mid X_i] = h(X_i, \alpha) = X_i^\alpha,$$

where $\alpha \in \mathbb{R}$ and $Y_i > 0$.

Utilizing V_i from the multiplicative model, we write

$$\epsilon_i^2 = X_i^\alpha V_i. \tag{3.27}$$

By applying the log transform to (3.27), we get

$$\log(\epsilon_i^2) = \alpha \log\left(X_i\right) + \log(V_i).$$

From (3.27), we get

$$\log(\hat{\epsilon_i}^2) = \alpha \log(X_i) + e_i. \tag{3.28}$$

FWLS estimate $\widehat{\boldsymbol{\beta}}_{FWLS}$ obtained by two-step estimation is no longer unbiased. However, it is consistent and should be asymptotically more efficient than the OLS estimate $\widehat{\boldsymbol{\beta}}$, given Corollary 1 Heij et al., (2004) page 336. Additionally, some authors suggest that iterating the two-step method may provide better asymptotic properties Greene, (2003) page 228.

Next, we introduce an algorithm to iterate the two-step method.

3.4.2 Iterative estimation

We can describe the iterative estimation method in the following steps:

1. Obtain an OLS estimate $\widehat{\boldsymbol{\beta}}$, and calculate residuals $\widehat{\boldsymbol{\epsilon}}$.
2. Use $\widehat{\boldsymbol{\epsilon}}$ in a suitable model to obtain $\widehat{\boldsymbol{\alpha}}$.
3. Obtain $\widehat{\boldsymbol{\beta}}_{FWLS}$ and use it to calcualte residuals $\widehat{\boldsymbol{\epsilon}}$.
4. Iterate step 2 and 3 until $||\widehat{\boldsymbol{\beta}}^i - \widehat{\boldsymbol{\beta}}^{(i+1)}|| < \theta$, where $\theta > 0$ and $\widehat{\boldsymbol{\beta}}^i_{FWLS}$ is obtained by i-th iteration.

In the next chapter, we will run simulation studies to examine and compare the asymptotic properties of the models using both the two-step estimation and iterative two-step estimation.

4. Simulation studies

In this chapter, we compare the ordinary least squares (OLS), weighted least squares (WLS or oracle), and feasible weighted least squares (FWLS) estimators in terms of their performance (efficiency) across specific studies. The metric of interest is the standard deviation (denoted as *sd*) of the estimator's deviation from the true slope coefficients and the intercept, as it allows us to compare the efficiency of each estimate. Secondly, we measure the bias of each estimate and discuss its asymptotic behavior.

The regressors for each study are generated using a uniform distribution $\mathcal{U}(0, 10)$. Error terms are generated using a conditioned normal distribution $\mathcal{N}(0, h(\mathbf{X}_i))$ with $h(\mathbf{X}_i)$ defined in (3.1) being specified in each study. Simulations are executed across varying sample sizes $n \in \{30, 50, 100, 300, 500\}$, with each configuration repeated 1000 times. We then calculate *sd*, and bias for each n. Tables with simulation results can be found in the Attachments A.1 section.

We estimate $\boldsymbol{\beta}$ by 10 candidate estimators:

- OLS = Ordinary least squares (2.5),
- WLS = Weighted least squares (3.13),
- FWLS_mult = FWLS, using two-step estimation with a multiplicative model (3.21),
- FWLS_add = FWLS, using two-step estimation with an additive model (3.24),
- FWLS_power = FWLS, using two-step estimation with power model (3.26),
- FWLS_powerinX = FWLS, using two-step estimation with Power in X_i model (3.28),
- imm_FWLS_mult = FWLS, using iterative (3.4.2) two-step estimation with a multiplicative model (3.21),
- iam_FWLS_add = FWLS, using iterative (3.4.2) two-step estimation with an additive model (3.24),
- ipm_FWLS_power = FWLS, using iterative (3.4.2) two-step estimation with power model (3.26),
- ipx_FWLS_powerinX = FWLS, using iterative (3.4.2) two-step estimation with power in X_i model (3.28).

Power in X_i models are only applicable to the SLR model. Consequently, we don't measure them in studies under the MLR model.

Let us have estimate $\widehat{\beta}_{ji,est}$ of $\beta_j, j = 0, 1, \ldots, k$, where $i = 0, 1, \ldots, N$, for $N = 1000$, denotes the iteration in which we obtained the estimate, and est is one of the candidate estimates. We calculate the metrics sd and bias in the following way:

- $sd = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\widehat{\beta}_{ji,est} - \beta_j)^2}$
- $bias = \frac{1}{N}\sum_{i=1}^{N}\widehat{\beta}_{ji,est} - \beta_j$

4.1 Study 1

In the first study, we consider a SLR model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i = 100 + 20X_i + \epsilon_i,$$

where we assume

$$\epsilon_i | X_i \sim \mathcal{N}(0, \sigma^2 X_i^{\alpha}) = \mathcal{N}(0, 10X_i^2).$$

The following graph 4.1 shows us an example of a data sample of 500 observations generated by the Study 1 configuration.

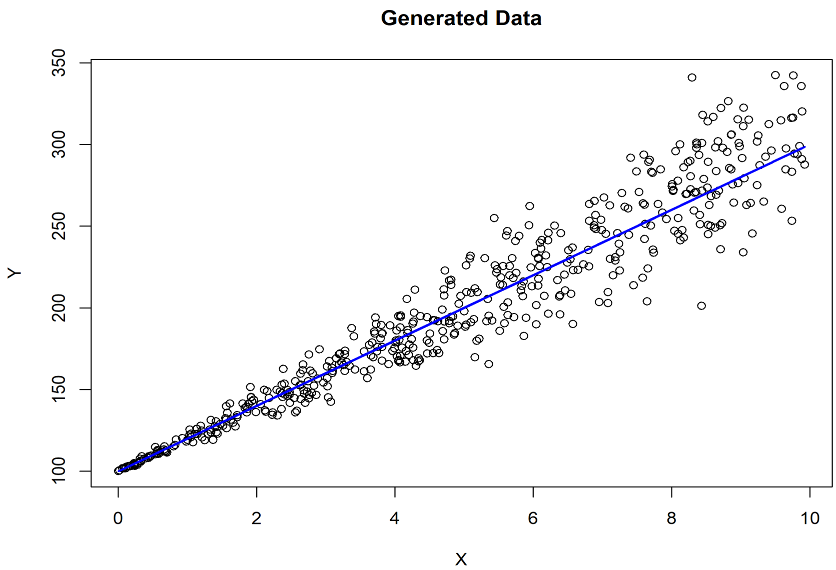
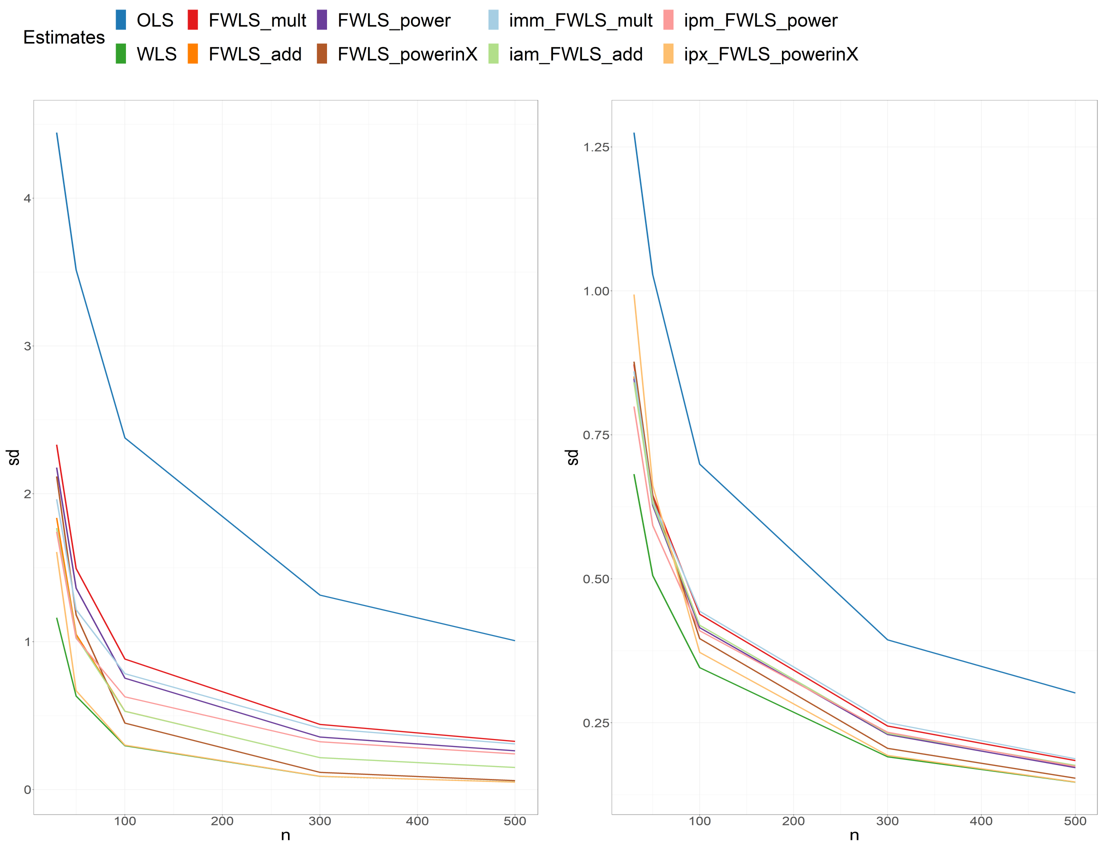


Figure 4.1: Study 1: Generated data around $100 + 20X_i$ line (blue solid line).

We will now run the simulations as described in the previous section. On the graphs in Figure 4.2 we showcase the dependence of the sd on the sample size n for all 10 candidate estimators, both for the intercept β_0 in Figure (a) and the slope parameter β_1 in Figure (b). Additionally, the simulation results are listed in Table 4.1 for β_0 and the Table 4.2 for β_1.



(a) Intercept β_0.

(b) Slope parameter β_1.

Figure 4.2: Study 1: Dependence of the *sd* on the sample size n for the candidate estimators.

n	OLS	WLS	mult	add	pwr	PinX	imm	iam	ipm	ipx
30	4.443	1.163	2.332	1.838	2.178	2.118	1.963	1.769	1.743	1.607
50	3.514	0.632	1.495	1.050	1.362	1.181	1.216	1.034	1.024	0.669
100	2.378	0.297	0.883	0.530	0.754	0.450	0.784	0.529	0.627	0.301
300	1.315	0.090	0.441	0.216	0.356	0.117	0.415	0.216	0.323	0.091
500	1.007	0.051	0.326	0.150	0.262	0.060	0.309	0.150	0.241	0.052

Table 4.1: Study 1: Dependence of the *sd* on the sample size n for the candidate estimators, for parameter β_0

n	OLS	WLS	mult	add	pwr	PinX	imm	iam	ipm	ipx
30	1.275	0.682	0.872	0.852	0.848	0.877	0.860	0.841	0.799	0.994
50	1.028	0.506	0.646	0.633	0.628	0.644	0.634	0.631	0.593	0.662
100	0.699	0.346	0.439	0.420	0.415	0.396	0.444	0.419	0.410	0.372
300	0.394	0.191	0.245	0.232	0.230	0.206	0.250	0.232	0.234	0.193
500	0.302	0.147	0.184	0.176	0.172	0.154	0.188	0.176	0.174	0.147

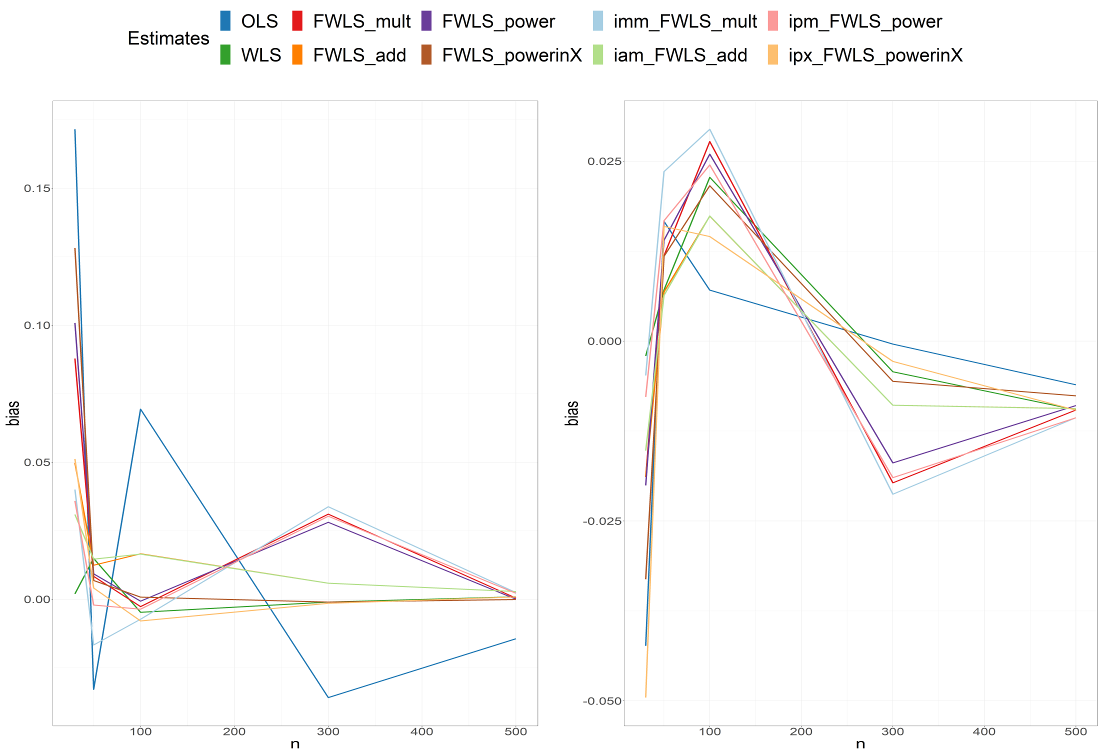
Table 4.2: Study 1: Dependence of the *sd* on the sample size n for the candidate estimators, for parameter β_1

In line with the theory, the WLS estimate outperformed other estimators, a trend we expect to continue across our studies. This fact will not be mentioned again unless something unexpected occurs.

We can observe that the OLS estimate had the worst performance, almost doubling the other candidate estimators in terms of sd for all n.

As for the FWLS models, the iterative and one-step (non-iterative) power in X_i models performed the best. This is likely the case because they assume $\mathsf{Var}[\epsilon_i|X_i] = X_i^\alpha$ and are not influenced by changes in σ^2, as shown in Lemma 8. The simulations also indicated that the power models outperform additive models in β_1 estimation, but the opposite is true for β_0 estimation. However, the difference is insignificant, hence the power models are comparable to the additive models, respectively. Out of the FWLS models, the multiplicative model was the least efficient but still performed much better than the OLS model.

The results of simulations measuring bias dependence on the sample size n for the candidate estimators are shown in Figure 4.3, both for the intercept β_0 in Figure (a) and the slope parameter $\beta 1$ in Figure (b). Additionally, the simulation results are listed in Table 4.3 for β_0 and Table 4.4 for β_1.



(a) Intercept β_0.

(b) Slope parameter β_1.

Figure 4.3: Study 1: Dependence of the bias on the sample size n for the candidate estimators.

n	OLS	WLS	mult	add	pwr	PinX	imm	iam	ipm	ipx
30	0.17148	0.00195	0.08786	0.04977	0.10083	0.12816	0.04003	0.03093	0.03590	0.05119
50	-0.03289	0.01490	0.00835	0.01240	0.00928	0.00684	-0.01669	0.01463	-0.00212	0.00412
100	0.06936	-0.00476	-0.00269	0.01658	-0.00070	0.00076	-0.00729	0.01645	-0.00362	-0.00792
300	-0.03588	-0.00103	0.03105	0.00585	0.02808	-0.00103	0.03374	0.00585	0.03028	-0.00148
500	-0.01441	0.00089	0.00039	0.00280	0.00001	-0.00010	0.00241	0.00281	0.00217	0.00083

Table 4.3: Study 1: Dependence of the bias on the sample size n for the candidate estimators, for parameter β_0

n	OLS	WLS	mult	add	pwr	PinX	imm	iam	ipm	ipx
30	-0.04234	-0.00206	-0.01888	-0.01888	-0.02005	-0.03310	-0.00478	-0.01522	-0.00775	-0.04954
50	0.01664	0.00711	0.01186	0.00683	0.01401	0.01178	0.02355	0.00636	0.01673	0.01594
100	0.00708	0.02276	0.02774	0.01737	0.02598	0.02160	0.02944	0.01739	0.02448	0.01453
300	-0.00041	-0.00426	-0.01969	-0.00891	-0.01693	-0.00559	-0.02128	-0.00891	-0.01897	-0.00282
500	-0.00607	-0.00955	-0.00957	-0.00939	-0.00896	-0.00761	-0.01064	-0.00939	-0.01065	-0.00960

Table 4.4: Study 1: Dependence of the bias on the sample size n for the candidate estimators, for parameter β_1

From these results, we can observe that all estimates are asymptotically unbiased. Consequently, since the bias values for each estimate are all in the proximity of 0, we won't be discussing bias for other studies.

Another important factor that we can observe from the simulations is that all the estimates are in fact consistent as $n \to \infty$. We have proven that in Theorem 6 and Corrolarly 1.

The results of Study 1, demonstrated that the FWLS estimators, which assume the correct partial form of heteroscedasticity (in this case power in X_i models), outperformed FWLS estimators that misspecify it. Furthermore, it may be advisable to prefer FWLS estimators over the OLS model when accommodating a multiplicative form of heteroscedasticity.

4.2 Study 2

In the second case, the data are generated from

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i = 2000 + 20X_i + \epsilon_i,$$

where

$$\epsilon_i | X_i \sim \mathcal{N}(0, \exp[3 + 1 \cdot X_i]).$$

The following graph 4.4 shows us an example of a data sample of 500 observations generated by the Study 2 configuration.

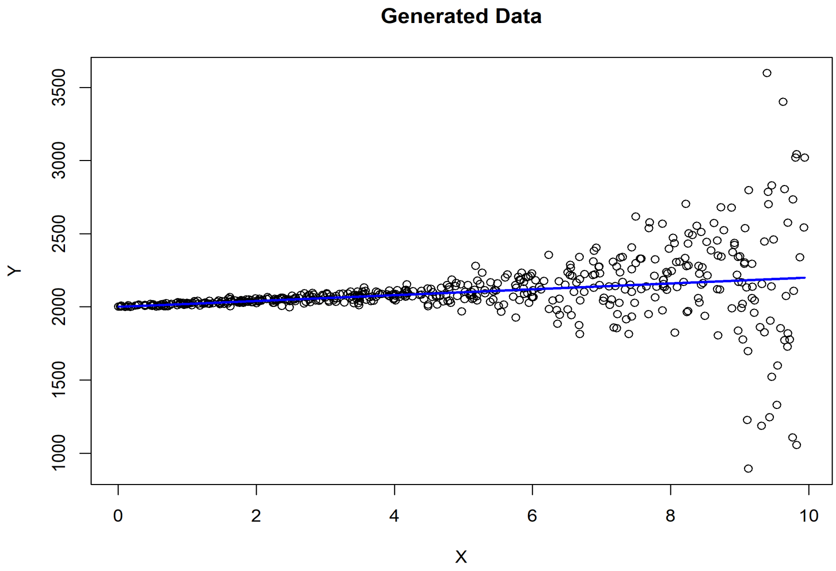
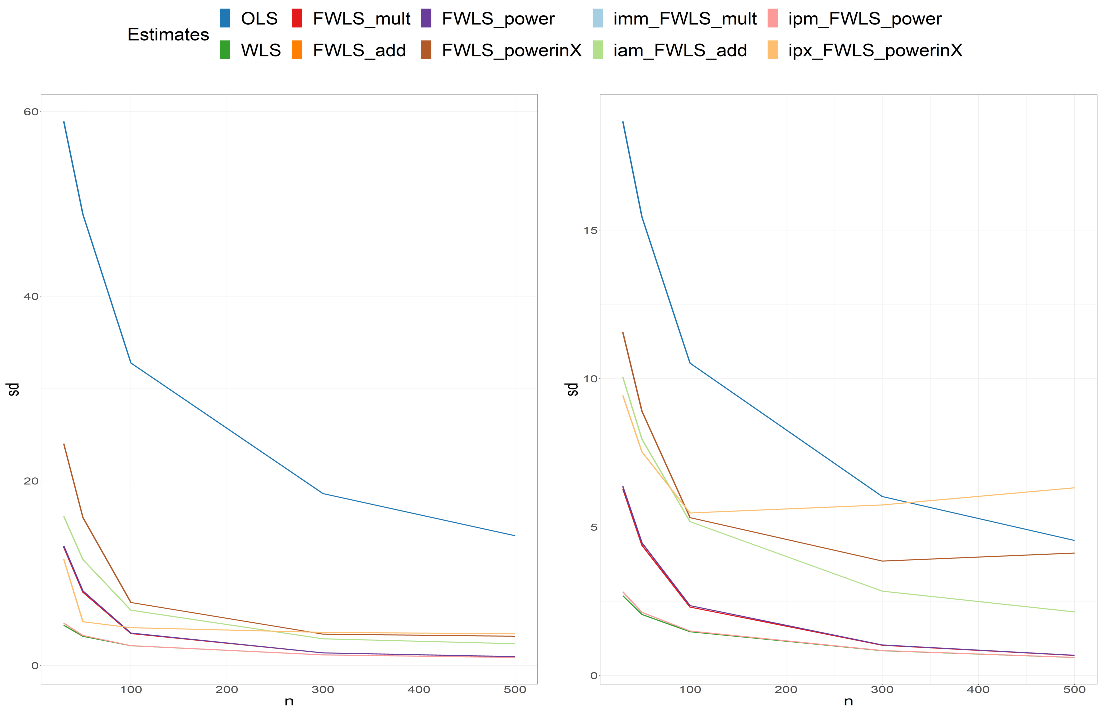


Figure 4.4: Study 2: Generated data around $2000 + 20X_i$ line (blue solid line).

On the graphs present in Figure 4.5 we showcase the dependence of the sd on the sample size n for all 10 candidate estimators, both for the intercept β_0 in Figure (a) and the slope parameter β_1 in Figure (b).



(a) Intercept β_1.

(b) Slope parameter β_1.

Figure 4.5: Study 2: Dependence of the sd on the sample size n for the candidate estimators.

In Study 2, we notice that the iterative and one-step (non-iterative) multiplicative models significantly outperformed the other FWLS estimates. That is the expected outcome, as multiplicative models assume $\mathsf{Var}[\epsilon_i|X_i] = \exp[\alpha X_i]$. The simulation results also show that the usage of the OLS estimator performs badly even for large data samples. Finally, we observe that the iterative power in X_i model performed worse than the OLS estimator for $n > 300$ in terms of β_1 estimation, whereas the one-step power in X_i model is comparable to the OLS estimator at around $n = 500$. This shows us that misspecified FWLS estimators may prove to be a worse option than the OLS estimator.

4.3 Study 3

In the third case, the data are generated from

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i = 100 + 20X_i + \epsilon_i,$$

where

$$\epsilon_i|X_i \sim \mathcal{N}(0, 100 + 50 \cdot X_i).$$

The following graph 4.6 shows us an example of a data sample of 500 observations generated by the Study 2 configuration.

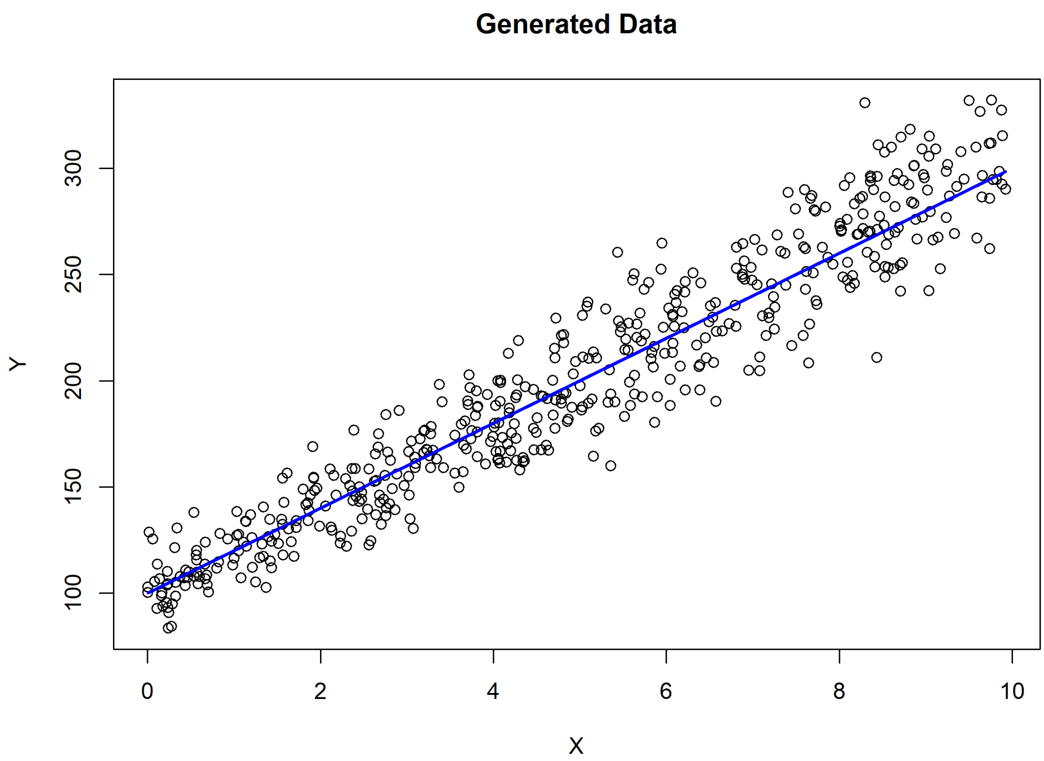
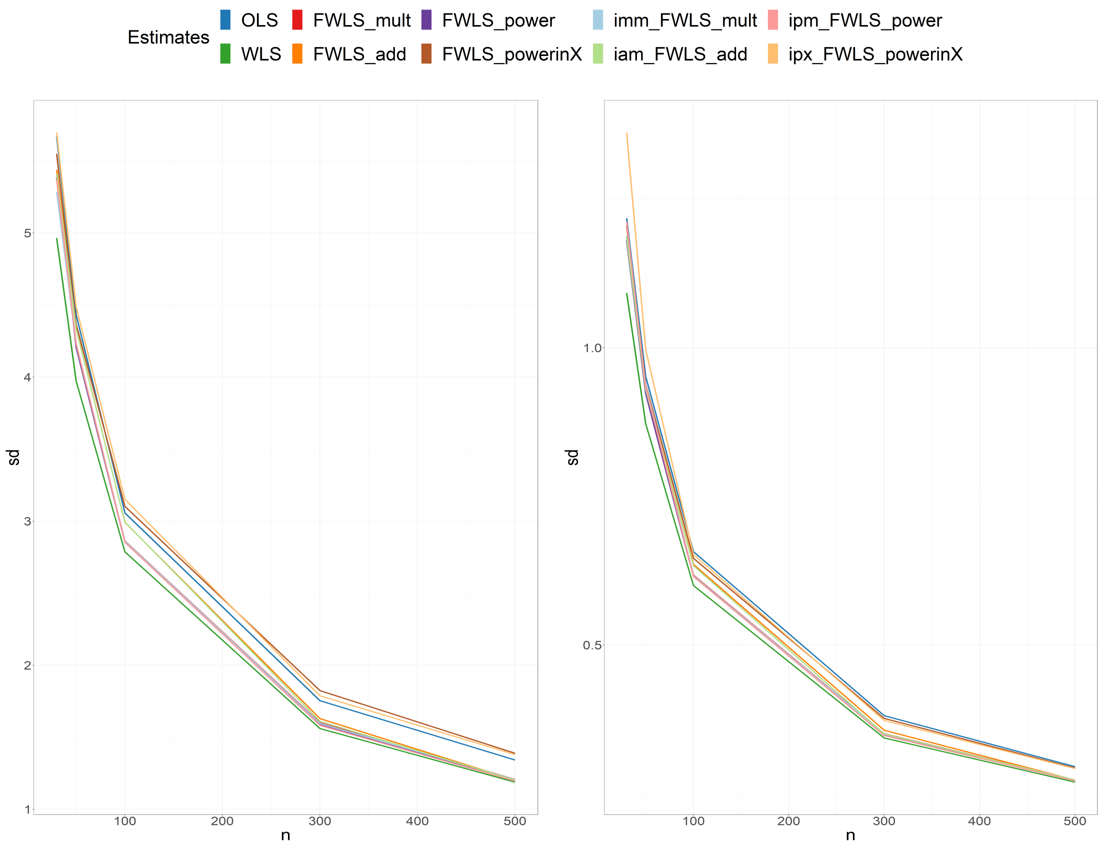


Figure 4.6: Study 3: Generated data around $100 + 20X_i$ line (blue solid line).

On the graphs present in Figure 4.7 we showcase the dependence of the *sd* on the sample size n for all 10 candidate estimators, both for the intercept β_0 in Figure (a) and the slope parameter β_1 in Figure (b).



(a) Intercept β_1.

(b) Slope parameter β_1.

Figure 4.7: Study 3: Dependence of the *sd* on the sample size n for the candidate estimators.

Overall the FWLS estimators are very comparable, except for the Power in X_i models which are comparable to the OLS estimator. This tells us that in regards to Study 3 configuration, it doesn't make much difference whether we misspecify the FWLS estimator or not. Such a result is expected, as the presence of heteroscedasticity in this scenario is notably minimal, in contrast to previous cases. This observation is supported by the graphical representation of simulated data in Figure 4.6. Consequently, the impact of heteroscedasticity on the outcome is significantly attenuated.

4.4 Study 4

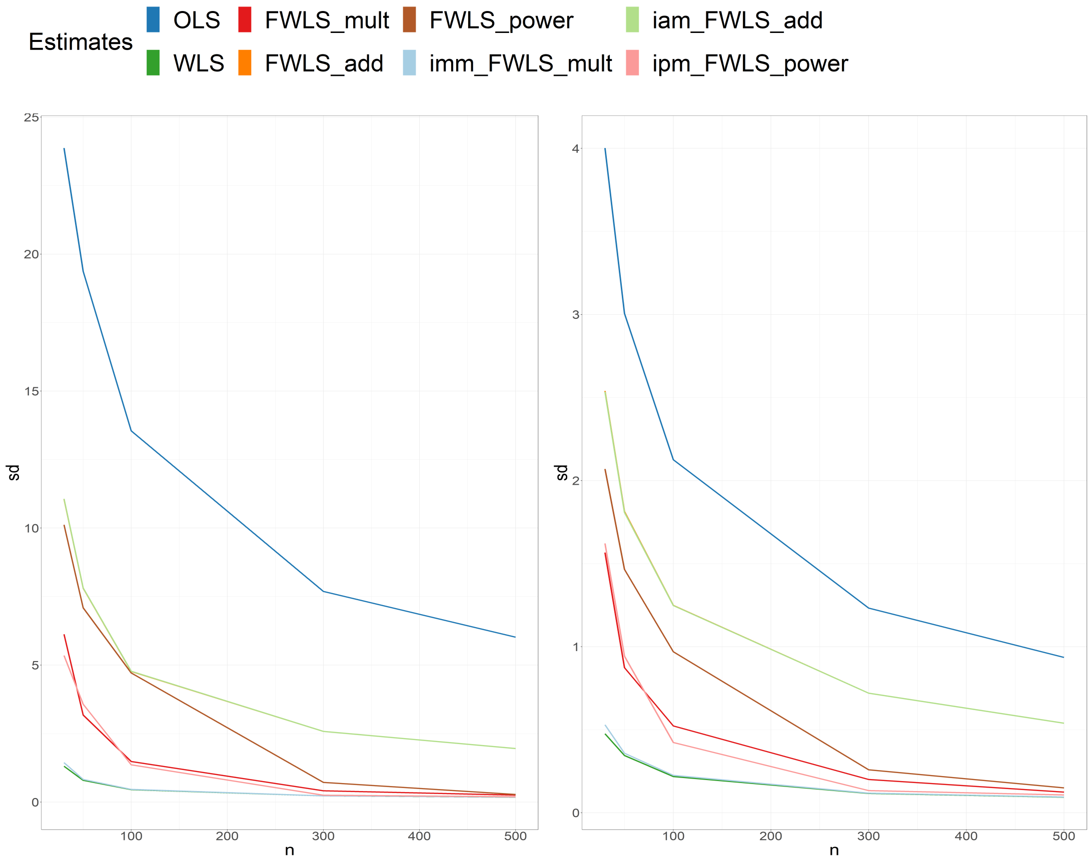
In the fourth case, let us use the model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i = 1000 + 5X_{i1} + 3X_{i2} + \epsilon_i,$$

where

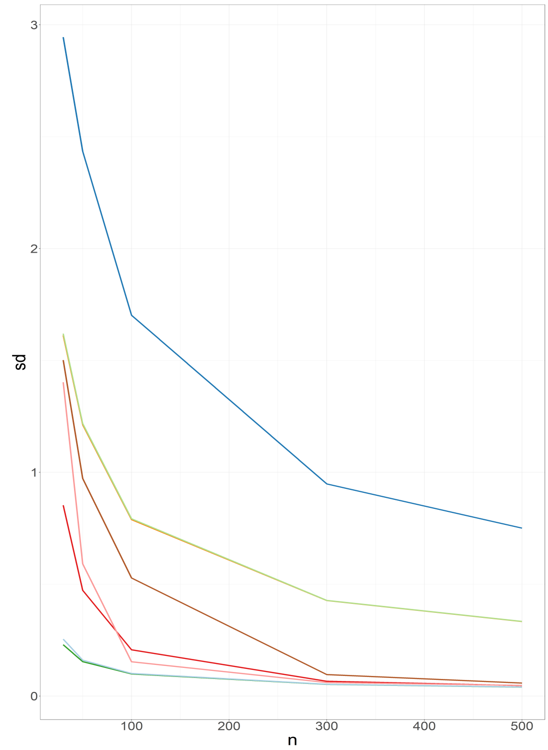
$$\epsilon_i | \mathbf{X}_i \sim \mathcal{N}(0, \exp[-2 + 1 \cdot X_{i1} + 0.3 \cdot X_{i2}]).$$

On the graphs in Figure 4.8 we showcase the dependence of the *sd* on the sample size n for 8 candidate estimators, for the intercept β_0 in Figure (a), the slope parameter β_1 in Figure (b), and the slope parameter β_2 in Figure (c).



(a) Intercept β_0.

(b) Slope parameter β_1.



(c) Slope parameter β_2.

Figure 4.8: Study 4: Dependence of the *sd* on the sample size n for the candidate estimators.

Note that in this study we do not employ Power in X_i models as they cannot be applied to multiple regressors at once. The same holds for all the remaining simulation sections.

As in Study 2, we can notice that the iterative multiplicative model has the best efficiency, which is again expected as the multiplicative model assumes $\exp[\boldsymbol{\alpha}^\top \mathbf{X}_i]$. From $n > 300$ it is being matched by the one-step multiplicative model and the iterative power model. We observe that the OLS estimator performed the worst, with almost the double *sd* of the least efficient FWLS estimator, which is an iterative additive model. These results are expected, as misspecifying a multiplicative form of conditioned variance by constant or linear form should deviate significantly.

Study 4 showcases that an iterative misspecified model may produce a better or comparable result to the of one-step model assuming the correct form.

4.5 Study 5

In the fifth case, let us have the same MLR model as in the fourth study

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i = 1000 + 5X_{i1} + 3X_{i2} + \epsilon_i,$$

where

$$\epsilon_i | \mathbf{X}_i \sim \mathcal{N}(0, \exp[-2 + 1 \cdot X_{i1} + 0.3 \cdot X_{i2}]).$$

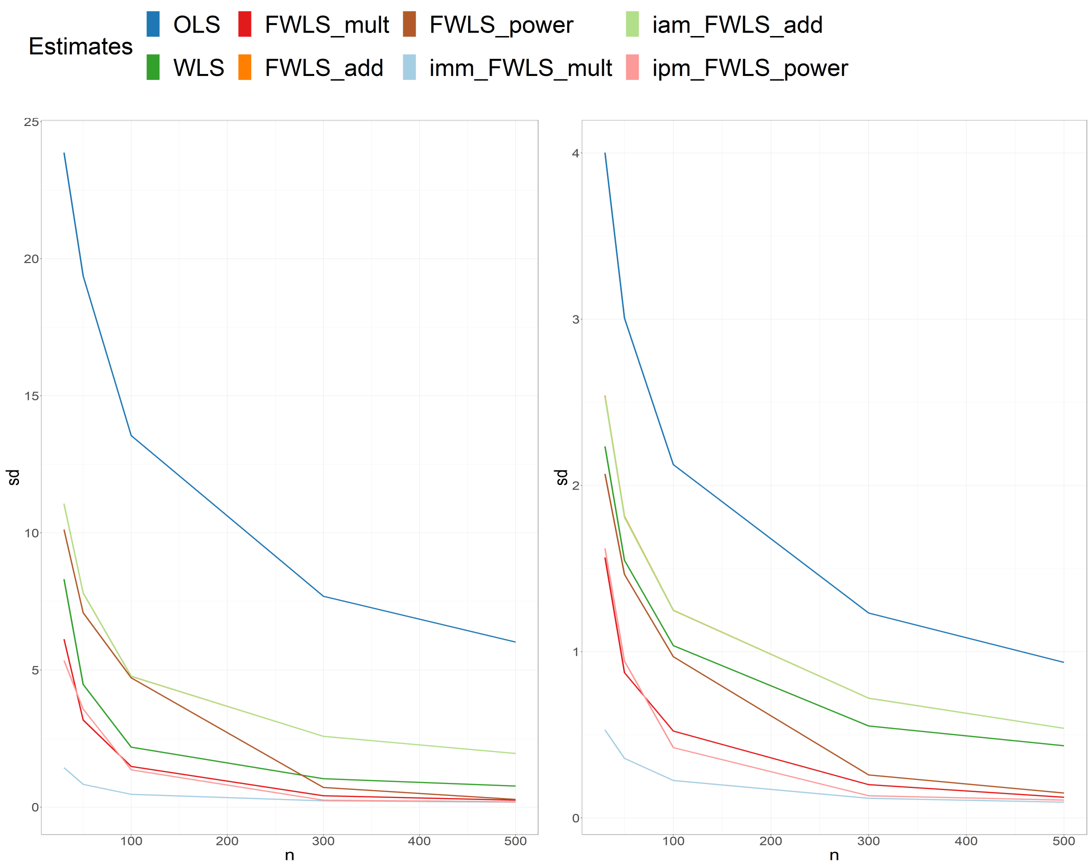
However, in this case, we will misspecify the weights of the WLS method to be

$$\epsilon_i | X_i \sim \mathcal{N}(0, \exp[-2 + 0.3 \cdot X_{i1} + 1 \cdot X_{i2}]).$$

On the graphs present in Figure 4.9 we showcase the dependence of the *sd* on the sample size n for 8 candidate estimators, for the intercept β_0 in Figure (a), the slope parameter β_1 in Figure (b), and the slope parameter β_2 in Figure (c).

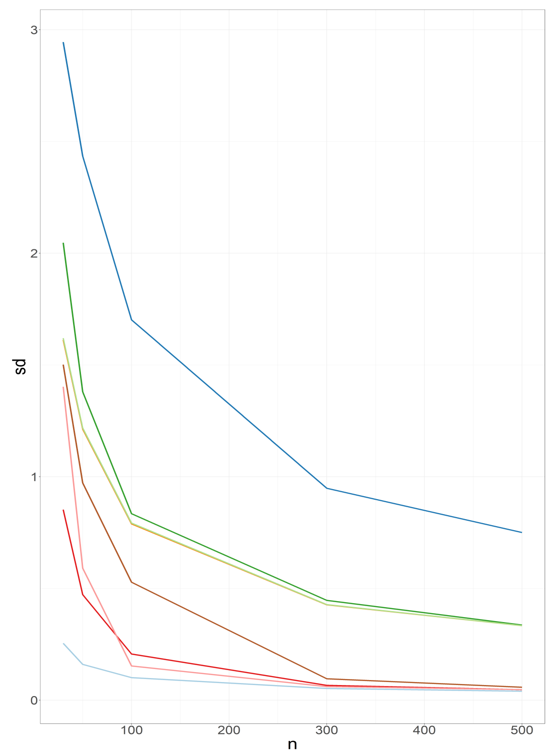
This study demonstrates that choosing an incorrect form of heteroscedasticity leads to a WLS estimate that is less efficient than the FWLS estimates.

The practical takeaway from Study 5 is that when there's no guarantee of accurately determining the correct form of heteroscedasticity, it may be preferable to utilize FWLS estimators, especially in cases where it is expected that the form has an exponential or higher order of magnitude.



(a) Intercept β_0.

(b) Slope parameter β_1.



(c) Slope parameter β_2.

Figure 4.9: Study 5: Dependence of the sd on the sample size n for the candidate estimators.

4.6 Study 6

In the last case, the data are generated from

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i = 200 + 6X_{i1} + 4X_{i2} + 3X_{i3} + \epsilon_i,$$

where we assume

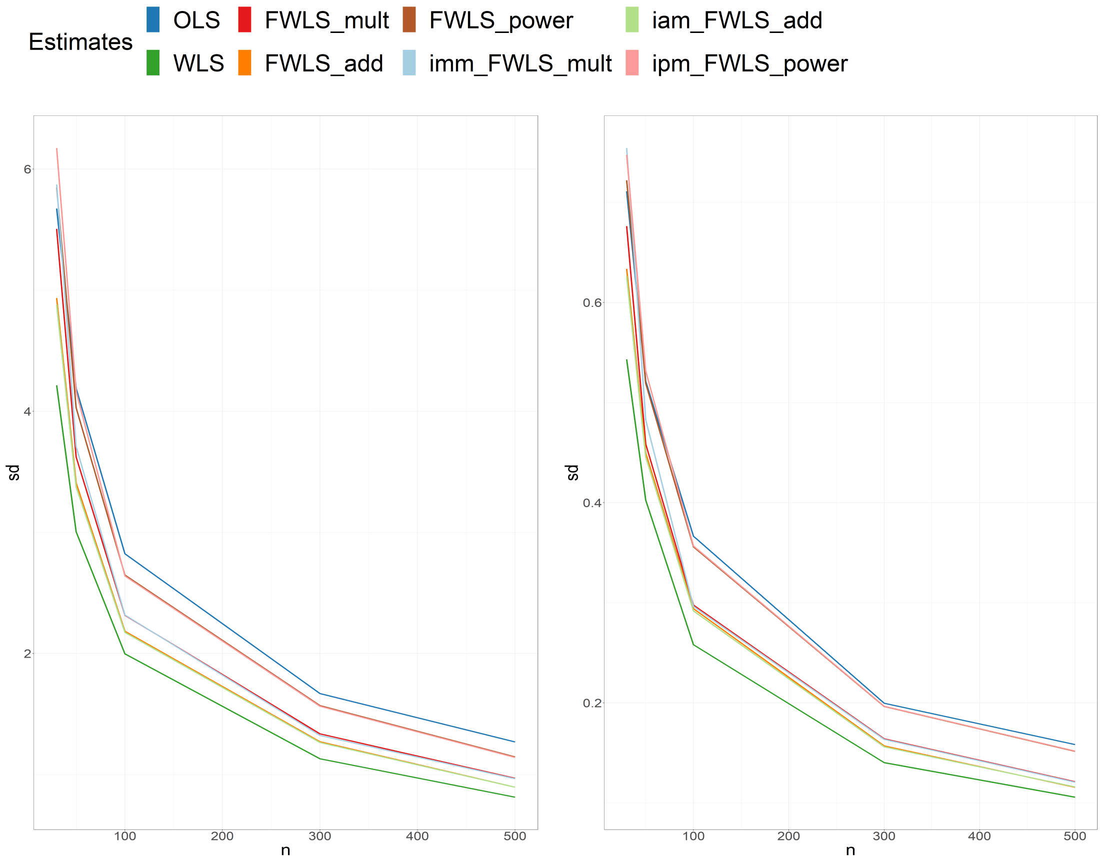
$$\epsilon_i | X_i \sim \mathcal{N}(0, \exp[\alpha_1 \cdot X_{i1}] + \alpha_2 \cdot X_{i3}).$$

We examine the change in performance between multiplicative and additive models by making changes in $\boldsymbol{\alpha}$.

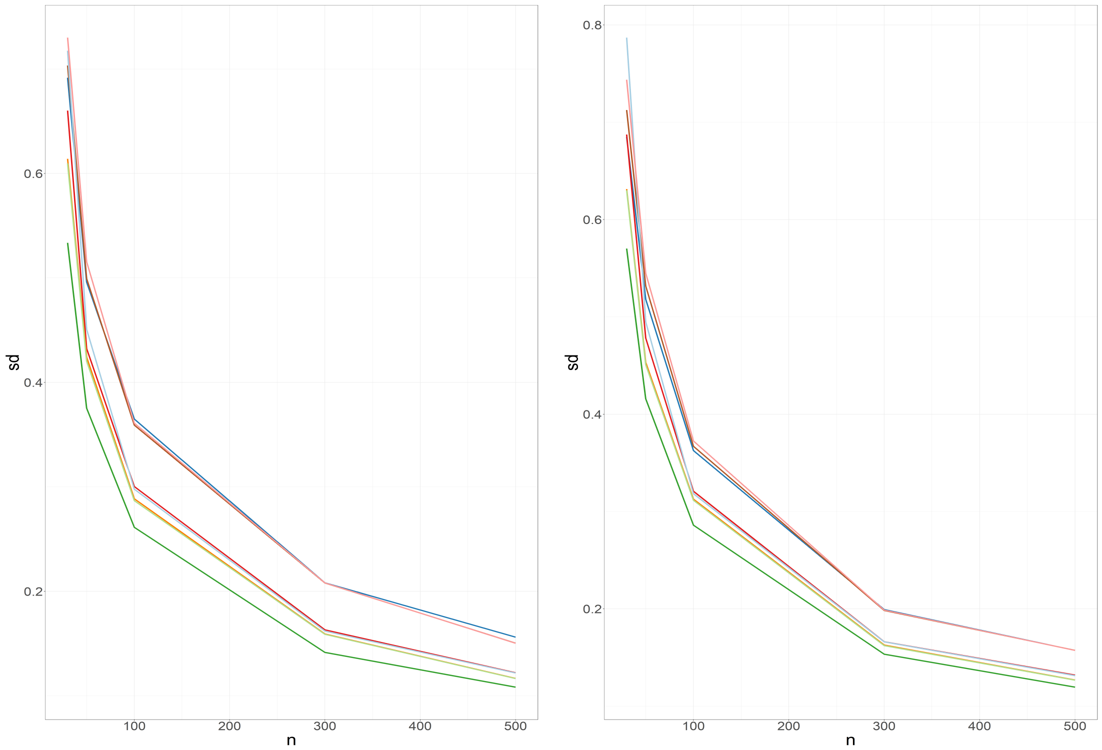
On the graphs present in Figure 4.10 and Figure 4.11 we showcase the dependence of the *sd* on the sample size n for 8 candidate estimators, for the intercept β_0 in Figures (a), the slope parameter β_1 in Figures (b), the slope parameter β_2 in Figures (c), and the slope parameter β_3 in Figures (d).

In this study we can observe that adjusting $\boldsymbol{\alpha} = (0.2, 20)^\top$ to $\boldsymbol{\alpha} = (1, 4)^\top$ led to multiplicative models surpassing additive models in efficiency. Furthermore, we notice the effect of amplifying the multiplicative term on the conditional variance significantly deteriorates the performance of the OLS estimator, as well as of the power models. That is due to the orders of magnitude faster growth of the multiplicative term.

Attachments A.1 includes results for studies, where $\boldsymbol{\alpha} = (0.4, 16)^\top$ (6.2), $\boldsymbol{\alpha} = (0.6, 12)^\top$ (6.3), and $\boldsymbol{\alpha} = (0.8, 8)^\top$ (6.4).

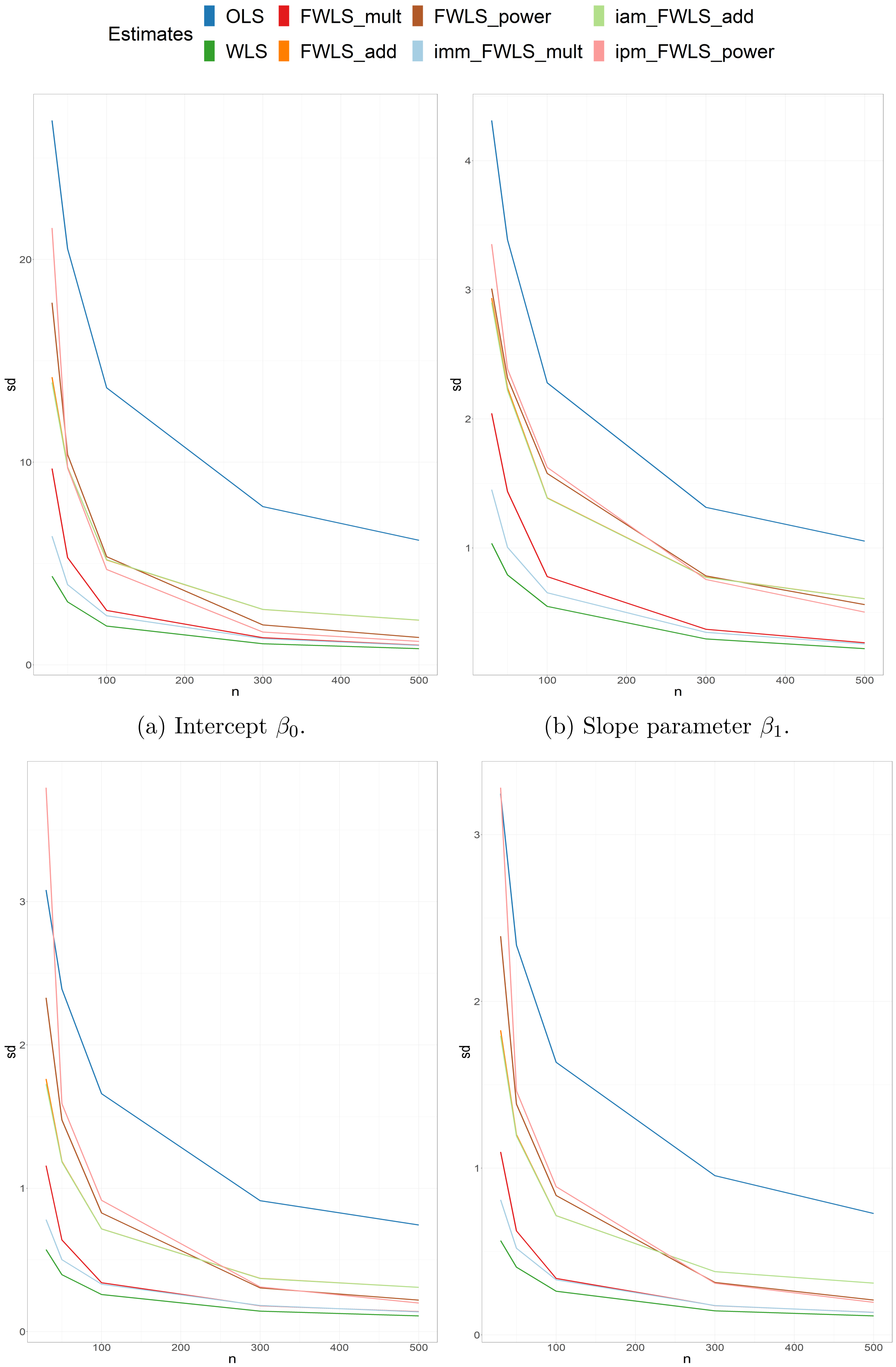


(a) Intercept β_0. (b) Slope parameter β_1.



(c) Slope parameter β_2. (d) Slope parameter β_3.

Figure 4.10: Study 6: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.2, 20)^\top$.



(a) Intercept β_0.

(b) Slope parameter β_1.

(c) Slope parameter β_2.

(d) Slope parameter β_3.

Figure 4.11: Study 6: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (1, 4)^\top$.

4.7 Conclusion to simulation studies

In conclusion, we observed the varying behavior of the OLS and WLS estimators, and FWLS estimators obtained via iterative and non-iterative two-step estimation, depending on the study specifications.

As a result, we provide recommendations regarding the selection between the OLS and FWLS methods based on specific model characteristics, while highlighting cases where the OLS method could produce highly misleading results. Furthermore, we emphasize the impact of misspecified heteroscedasticity, highlighting the risks associated with relying on the WLS or a FWLS estimator assuming conditioned variance of a different order of magnitude.

Conclusion

In the theoretical part, we introduced the weighted least squares (WLS) estimator, and proved its superior efficiency over the ordinary least squares (OLS) under the assumption of heteroscedasticity with a known form. Moreover, we introduced the feasible weighted least squares (FWLS) estimator as an alternative to the WLS, requiring only partial knowledge of heteroscedasticity's structure.

Through the simulations, we observed the varying behavior of FWLS estimators obtained via two-step estimation, contingent upon the model specifications. Consequently, we provide recommendations concerning the choice between OLS and FWLS methods based on the specific model characteristics. Additionally, we demonstrate that an incorrectly specified WLS estimate may exhibit inferior performance compared to the OLS model, thereby highlighting the preferable utilization of FWLS models. This underscores the importance of considering heteroscedasticity's nuances and the potential advantages of alternative estimation techniques in empirical research.

Bibliography

Greene, William H. (2003). *Econometric Analysis*. 5th. New Jersey: Prentice Hall. ISBN: 0-13-066189-9.

Harvey, A. C. (1976). "Estimating Regression Models with Multiplicative Heteroscedasticity". In: *The Econometric Society* 44, pp. 461–465.

Heij, Christiaan et al. (2004). *Econometric Methods with Applications in Business and Economics*. New York: Oxford University Press. ISBN: 0–19–926801–0.

Romano, Joseph and Michael Wolf (2016). "Resurrecting weighted least squares". In: *Journal of Econometrics*, pp. 1–19.

Shukla, Sumit Kumar (2019). "Kaggle Video Games Sales Data". In: *Data.World*. URL: `https://data.world/sumitrock/video-games-sales`.

Wooldridge, Jeffrey M. (2013). *Introductory Econometrics: A Modern Approach*. 5th. Mason: Cengage Learning. ISBN: 978-1-111-53104-1.

List of Figures

1.1 Illustration of the idea of the ordinary least squares estimation method. 6

2.1 Estimating user ratings Y_i from video game sales X_i with the OLS estimate $\widehat{\boldsymbol{\beta}}$. 11

3.1 Data showcasing heteroscedasticity 16

4.1 Study 1: Generated data around $100 + 20X_i$ line (blue solid line). 28
4.2 Study 1: Dependence of the *sd* on the sample size n for the candidate estimators. 29
4.3 Study 1: Dependence of the bias on the sample size n for the candidate estimators. 30
4.4 Study 2: Generated data around $2000 + 20X_i$ line (blue solid line). 32
4.5 Study 2: Dependence of the *sd* on the sample size n for the candidate estimators. 32
4.6 Study 3: Generated data around $100 + 20X_i$ line (blue solid line). 33
4.7 Study 3: Dependence of the *sd* on the sample size n for the candidate estimators. 34
4.8 Study 4: Dependence of the *sd* on the sample size n for the candidate estimators. 35
4.9 Study 5: Dependence of the *sd* on the sample size n for the candidate estimators. 37
4.10 Study 6: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.2, 20)^\top$. 39
4.11 Study 6: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (1, 4)^\top$. 40

List of Tables

4.1 Study 1: Dependence of the *sd* on the sample size n for the candidate estimators, for parameter β_0 29
4.2 Study 1: Dependence of the *sd* on the sample size n for the candidate estimators, for parameter β_1 29
4.3 Study 1: Dependence of the bias on the sample size n for the candidate estimators, for parameter β_0 31
4.4 Study 1: Dependence of the bias on the sample size n for the candidate estimators, for parameter β_1 31

A.1 Study 2: Dependence of the *sd* on the sample size n for the candidate estimators, for parameter β_0 47
A.2 Study 2: Dependence of the *sd* on the sample size n for the candidate estimators, for parameter β_1 48
A.3 Study 3: Dependence of the *sd* on the sample size n for the candidate estimators, for parameter β_0 48
A.4 Study 3: Dependence of the *sd* on the sample size n for the candidate estimators, for parameter β_1 48
A.5 Study 4: Dependence of the *sd* on the sample size n for the candidate estimators, for parameter β_0 48
A.6 Study 4: Dependence of the *sd* on the sample size n for the candidate estimators, for parameter β_1 49
A.7 Study 4: Dependence of the *sd* on the sample size n for the candidate estimators, for parameter β_2 49
A.8 Study 5: Dependence of the *sd* on the sample size n for the candidate estimators, for parameter β_0 49
A.9 Study 5: Dependence of the *sd* on the sample size n for the candidate estimators, for parameter β_1 49
A.10 Study 5: Dependence of the *sd* on the sample size n for the candidate estimators, for parameter β_2 50
A.11 Study 6.1: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.2, 20)^\top$, for parameter β_0 . . . 50
A.12 Study 6.1: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.2, 20)^\top$, for parameter β_1 . . . 50
A.13 Study 6.1: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.2, 20)^\top$, for parameter β_2 . . . 50
A.14 Study 6.1: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.2, 20)^\top$, for parameter β_3 . . . 51
A.15 Study 6.2: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.4, 16)^\top$, for parameter β_0 . . . 51
A.16 Study 6.2: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.4, 16)^\top$, for parameter β_1 . . . 51
A.17 Study 6.2: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.4, 16)^\top$, for parameter β_2 . . . 51
A.18 Study 6.2: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.4, 16)^\top$, for parameter β_3 . . . 52

A.19 Study 6.3: Dependence of the sd on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.6, 12)^\top$, for parameter β_0 . . . 52
A.20 Study 6.3: Dependence of the sd on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.6, 12)^\top$, for parameter β_1 . . . 52
A.21 Study 6.3: Dependence of the sd on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.6, 12)^\top$, for parameter β_2 . . . 52
A.22 Study 6.3: Dependence of the sd on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.6, 12)^\top$, for parameter β_3 . . . 53
A.23 Study 6.4: Dependence of the sd on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.8, 8)^\top$, for parameter β_0 . . . 53
A.24 Study 6.4: Dependence of the sd on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.8, 8)^\top$, for parameter β_1 . . . 53
A.25 Study 6.4: Dependence of the sd on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.8, 8)^\top$, for parameter β_2 . . . 53
A.26 Study 6.4: Dependence of the sd on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.8, 8)^\top$, for parameter β_3 . . . 54
A.27 Study 6.5: Dependence of the sd on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (1, 4)^\top$, for parameter β_0 54
A.28 Study 6.5: Dependence of the sd on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (1, 4)^\top$, for parameter β_1 54
A.29 Study 6.5: Dependence of the sd on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (1, 4)^\top$, for parameter β_2 54
A.30 Study 6.5: Dependence of the sd on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (1, 4)^\top$, for parameter β_3 55

A. Attachments

A.1 First Attachment

Simulations result in tables for each study:

We estimate $\boldsymbol{\beta}$ by 10 candidate estimators:

- OLS = Ordinary least squares (2.5),
- WLS = Weighted least squares (3.13),
- mult = FWLS, using two-step estimation with a multiplicative model (3.21),
- add = FWLS, using two-step estimation with an additive model (3.24),
- pwr = FWLS, using two-step estimation with power model (3.26),
- PinX = FWLS, using two-step estimation with Power in X_i model (3.28),
- imm = FWLS, using iterative (3.4.2) two-step estimation with a multiplicative model (3.21),
- iam = FWLS, using iterative (3.4.2) two-step estimation with an additive model (3.24),
- ipm = FWLS, using iterative (3.4.2) two-step estimation with power model (3.26),
- ipx = FWLS, using iterative (3.4.2) two-step estimation with power in X_i model (3.28).

Power in X_i models are only applicable to the SLR model. Consequently, we don't measure them in studies under the MLR model.

n	OLS	WLS	mult	add	pwr	PinX	imm	iam	ipm	ipx
30	58.95	4.358	12.83	16.167	12.96	24.04	4.588	16.17	4.582	11.57
50	48.881	3.164	7.948	11.48	8.076	16.04	3.254	11.483	3.251	4.75
100	32.771	2.140	3.464	5.988	3.519	6.828	2.155	5.988	2.154	4.099
300	18.625	1.149	1.365	2.893	1.369	3.395	1.155	2.893	1.154	3.590
500	14.056	0.873	0.959	2.360	0.961	3.170	0.876	2.360	0.876	3.429

Table A.1: Study 2: Dependence of the sd on the sample size n for the candidate estimators, for parameter β_0

n	OLS	WLS	mult	add	pwr	PinX	imm	iam	ipm	ipx
30	18.669	2.690	6.293	10.054	6.374	11.57	2.819	10.054	2.82	9.44
50	15.443	2.056	4.386	7.960	4.470	8.900	2.124	7.960	2.129	7.531
100	10.524	1.475	2.305	5.183	2.355	5.319	1.498	5.183	1.499	5.477
300	6.029	0.837	1.020	2.841	1.031	3.855	0.842	2.841	0.842	5.747
500	4.548	0.608	0.677	2.142	0.682	4.125	0.611	2.142	0.611	6.326

Table A.2: Study 2: Dependence of the *sd* on the sample size n for the candidate estimators, for parameter β_1

n	OLS	WLS	mult	add	pwr	PinX	imm	iam	ipm	ipx
30	5.670	4.965	5.389	5.439	5.385	5.549	5.285	5.409	5.370	5.696
50	4.433	3.973	4.222	4.339	4.206	4.367	4.230	4.332	4.234	4.488
100	3.057	2.787	2.863	2.992	2.853	3.102	2.861	2.994	2.853	3.151
300	1.754	1.561	1.600	1.630	1.587	1.824	1.597	1.611	1.583	1.788
500	1.342	1.189	1.208	1.203	1.198	1.389	1.206	1.203	1.197	1.380

Table A.3: Study 3: Dependence of the *sd* on the sample size n for the candidate estimators, for parameter β_0

n	OLS	WLS	mult	add	pwr	PinX	imm	iam	ipm	ipx
30	1.218	1.092	1.181	1.187	1.181	1.206	1.188	1.187	1.212	1.362
50	0.952	0.872	0.930	0.929	0.923	0.939	0.937	0.929	0.934	0.997
100	0.657	0.600	0.617	0.636	0.615	0.646	0.617	0.633	0.615	0.652
300	0.381	0.344	0.351	0.357	0.348	0.377	0.350	0.351	0.348	0.373
500	0.295	0.269	0.273	0.273	0.272	0.293	0.273	0.273	0.272	0.292

Table A.4: Study 3: Dependence of the *sd* on the sample size n for the candidate estimators, for parameter β_1

n	OLS	WLS	mult	add	pwr	imm	iam	ipm
30	23.870	1.310	6.127	11.055	10.120	1.440	11.069	5.348
50	19.370	0.796	3.177	7.809	7.090	0.830	7.815	3.575
100	13.549	0.455	1.482	4.769	4.714	0.466	4.776	1.361
300	7.689	0.227	0.415	2.581	0.719	0.228	2.582	0.252
500	6.021	0.181	0.255	1.956	0.280	0.182	1.957	0.197

Table A.5: Study 4: Dependence of the *sd* on the sample size n for the candidate estimators, for parameter β_0

n	OLS	WLS	mult	add	pwr	imm	iam	ipm
30	4.002	0.476	1.567	2.540	2.070	0.530	2.535	1.622
50	3.006	0.345	0.874	1.814	1.465	0.359	1.809	0.943
100	2.126	0.218	0.523	1.249	0.970	0.225	1.248	0.423
300	1.233	0.117	0.201	0.720	0.259	0.118	0.720	0.134
500	0.936	0.094	0.125	0.539	0.150	0.095	0.539	0.108

Table A.6: Study 4: Dependence of the *sd* on the sample size n for the candidate estimators, for parameter β_1

n	OLS	WLS	mult	add	pwr	imm	iam	ipm
30	2.945	0.229	0.853	1.612	1.502	0.254	1.620	1.402
50	2.434	0.154	0.472	1.213	0.972	0.160	1.218	0.591
100	1.702	0.099	0.207	0.789	0.528	0.101	0.792	0.153
300	0.948	0.052	0.066	0.427	0.096	0.053	0.428	0.061
500	0.750	0.040	0.047	0.333	0.058	0.040	0.333	0.047

Table A.7: Study 4: Dependence of the *sd* on the sample size n for the candidate estimators, for parameter β_2

n	OLS	WLS	mult	add	pwr	imm	iam	ipm
30	23.870	8.316	6.127	11.055	10.120	1.440	11.069	5.348
50	19.370	4.476	3.177	7.809	7.090	0.830	7.815	3.575
100	13.549	2.188	1.482	4.769	4.714	0.466	4.776	1.361
300	7.689	1.036	0.415	2.581	0.719	0.228	2.582	0.252
500	6.021	0.768	0.255	1.956	0.280	0.182	1.957	0.197

Table A.8: Study 5: Dependence of the *sd* on the sample size n for the candidate estimators, for parameter β_0

n	OLS	WLS	mult	add	pwr	imm	iam	ipm
30	4.002	2.235	1.567	2.540	2.070	0.530	2.535	1.622
50	3.006	1.550	0.874	1.814	1.465	0.359	1.809	0.943
100	2.126	1.037	0.523	1.249	0.970	0.225	1.248	0.423
300	1.233	0.553	0.201	0.720	0.259	0.118	0.720	0.134
500	0.936	0.434	0.125	0.539	0.150	0.095	0.539	0.108

Table A.9: Study 5: Dependence of the *sd* on the sample size n for the candidate estimators, for parameter β_1

n	OLS	WLS	mult	add	pwr	imm	iam	ipm
30	2.945	2.047	0.853	1.612	1.502	0.254	1.620	1.402
50	2.434	1.380	0.472	1.213	0.972	0.160	1.218	0.591
100	1.702	0.835	0.207	0.789	0.528	0.101	0.792	0.153
300	0.948	0.447	0.066	0.427	0.096	0.053	0.428	0.061
500	0.750	0.336	0.047	0.333	0.058	0.040	0.333	0.047

Table A.10: Study 5: Dependence of the *sd* on the sample size n for the candidate estimators, for parameter β_2

n	OLS	WLS	mult	add	pwr	imm	iam	ipm
30	5.674	4.215	5.506	4.935	5.848	5.872	4.881	6.173
50	4.190	3.006	3.623	3.408	4.029	3.709	3.376	4.162
100	2.823	1.997	2.316	2.184	2.648	2.318	2.175	2.640
300	1.670	1.131	1.337	1.271	1.570	1.324	1.266	1.565
500	1.270	0.814	0.970	0.897	1.147	0.966	0.896	1.142

Table A.11: Study 6.1: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.2, 20)^\top$, for parameter β_0

n	OLS	WLS	mult	add	pwr	imm	iam	ipm
30	0.711	0.543	0.676	0.634	0.722	0.754	0.626	0.748
50	0.522	0.403	0.459	0.450	0.519	0.483	0.446	0.532
100	0.367	0.258	0.298	0.294	0.356	0.296	0.292	0.357
300	0.200	0.140	0.164	0.157	0.196	0.164	0.156	0.197
500	0.158	0.106	0.121	0.116	0.152	0.121	0.116	0.152

Table A.12: Study 6.1: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.2, 20)^\top$, for parameter β_1

n	OLS	WLS	mult	add	pwr	imm	iam	ipm
30	0.692	0.534	0.660	0.614	0.703	0.718	0.610	0.730
50	0.496	0.375	0.432	0.424	0.499	0.451	0.420	0.516
100	0.365	0.261	0.300	0.289	0.359	0.298	0.287	0.361
300	0.208	0.142	0.163	0.159	0.208	0.162	0.159	0.208
500	0.156	0.108	0.122	0.117	0.150	0.122	0.117	0.150

Table A.13: Study 6.1: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.2, 20)^\top$, for parameter β_2

n	OLS	WLS	mult	add	pwr	imm	iam	ipm
30	0.685	0.570	0.687	0.631	0.713	0.787	0.630	0.744
50	0.518	0.416	0.478	0.453	0.532	0.495	0.451	0.545
100	0.363	0.286	0.321	0.313	0.367	0.318	0.312	0.373
300	0.199	0.153	0.166	0.163	0.198	0.166	0.162	0.199
500	0.157	0.120	0.132	0.127	0.157	0.131	0.127	0.157

Table A.14: Study 6.1: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.2, 20)^\top$, for parameter β_3

n	OLS	WLS	mult	add	pwr	imm	iam	ipm
30	5.431	4.124	5.257	4.775	5.519	5.815	4.716	5.788
50	4.032	3.087	3.617	3.441	3.840	3.720	3.411	3.870
100	2.843	1.986	2.434	2.284	2.596	2.421	2.263	2.586
300	1.636	1.096	1.285	1.213	1.464	1.276	1.207	1.455
500	1.256	0.849	0.974	0.904	1.113	0.970	0.900	1.108

Table A.15: Study 6.2: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.4, 16)^\top$, for parameter β_0

n	OLS	WLS	mult	add	pwr	imm	iam	ipm
30	0.668	0.545	0.675	0.612	0.677	0.758	0.608	0.709
50	0.510	0.426	0.478	0.456	0.500	0.499	0.455	0.503
100	0.356	0.289	0.315	0.308	0.346	0.313	0.307	0.346
300	0.203	0.159	0.174	0.168	0.195	0.173	0.168	0.194
500	0.154	0.118	0.130	0.125	0.146	0.129	0.125	0.146

Table A.16: Study 6.2: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.4, 16)^\top$, for parameter β_1

n	OLS	WLS	mult	add	pwr	imm	iam	ipm
30	0.666	0.541	0.658	0.609	0.671	0.731	0.606	0.699
50	0.486	0.389	0.457	0.435	0.485	0.466	0.433	0.486
100	0.348	0.252	0.306	0.292	0.334	0.307	0.290	0.336
300	0.196	0.146	0.162	0.158	0.188	0.162	0.158	0.188
500	0.158	0.113	0.127	0.124	0.151	0.127	0.123	0.151

Table A.17: Study 6.2: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.4, 16)^\top$, for parameter β_2

n	OLS	WLS	mult	add	pwr	imm	iam	ipm
30	0.658	0.558	0.674	0.608	0.665	0.740	0.605	0.709
50	0.492	0.407	0.462	0.440	0.499	0.485	0.438	0.508
100	0.337	0.279	0.314	0.302	0.334	0.318	0.301	0.336
300	0.193	0.156	0.168	0.165	0.187	0.168	0.165	0.187
500	0.151	0.124	0.131	0.128	0.147	0.131	0.128	0.147

Table A.18: Study 6.2: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.4, 16)^\top$, for parameter β_3

n	OLS	WLS	mult	add	pwr	imm	iam	ipm
30	6.384	4.439	5.845	5.303	5.815	6.406	5.256	5.980
50	4.635	3.070	3.871	3.620	3.778	3.885	3.614	3.824
100	3.288	2.155	2.630	2.633	2.626	2.607	2.608	2.594
300	1.870	1.109	1.294	1.352	1.378	1.283	1.351	1.373
500	1.483	0.880	1.017	1.050	1.074	1.014	1.051	1.072

Table A.19: Study 6.3: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.6, 12)^\top$, for parameter β_0

n	OLS	WLS	mult	add	pwr	imm	iam	ipm
30	0.854	0.653	0.826	0.757	0.813	0.881	0.752	0.839
50	0.648	0.489	0.593	0.569	0.588	0.615	0.568	0.601
100	0.465	0.355	0.405	0.402	0.423	0.407	0.401	0.422
300	0.263	0.200	0.218	0.221	0.233	0.218	0.221	0.234
500	0.198	0.142	0.157	0.159	0.167	0.156	0.159	0.167

Table A.20: Study 6.3: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.6, 12)^\top$, for parameter β_1

n	OLS	WLS	mult	add	pwr	imm	iam	ipm
30	0.742	0.557	0.701	0.660	0.689	0.779	0.658	0.715
50	0.564	0.397	0.480	0.471	0.482	0.489	0.471	0.483
100	0.413	0.282	0.338	0.345	0.342	0.336	0.343	0.341
300	0.233	0.153	0.175	0.182	0.190	0.175	0.182	0.191
500	0.181	0.121	0.137	0.143	0.150	0.137	0.143	0.150

Table A.21: Study 6.3: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.6, 12)^\top$, for parameter β_2

n	OLS	WLS	mult	add	pwr	imm	iam	ipm
30	0.763	0.592	0.731	0.684	0.722	0.829	0.682	0.742
50	0.583	0.437	0.526	0.515	0.530	0.542	0.512	0.545
100	0.405	0.299	0.347	0.349	0.362	0.347	0.351	0.361
300	0.233	0.164	0.183	0.193	0.195	0.183	0.193	0.195
500	0.171	0.123	0.135	0.139	0.146	0.135	0.139	0.146

Table A.22: Study 6.3: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.6, 12)^\top$, for parameter β_3

n	OLS	WLS	mult	add	pwr	imm	iam	ipm
30	11.85	4.926	7.058	7.556	7.952	7.105	7.446	7.519
50	8.362	3.282	4.299	4.760	4.734	4.103	4.694	4.377
100	5.834	2.074	2.669	3.368	2.97	2.565	3.359	2.818
300	3.294	1.116	1.350	1.966	1.423	1.338	1.97	1.409
500	2.562	0.891	1.064	1.726	1.107	1.057	1.683	1.10

Table A.23: Study 6.4: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.8, 8)^\top$, for parameter β_0

n	OLS	WLS	mult	add	pwr	imm	iam	ipm
30	1.835	0.909	1.266	1.379	1.456	1.207	1.360	1.445
50	1.376	0.672	0.857	0.990	1.056	0.808	0.980	1.015
100	0.926	0.444	0.553	0.664	0.698	0.512	0.658	0.683
300	0.561	0.255	0.293	0.397	0.382	0.290	0.397	0.380
500	0.421	0.192	0.216	0.303	0.289	0.215	0.295	0.287

Table A.24: Study 6.4: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.8, 8)^\top$, for parameter β_1

n	OLS	WLS	mult	add	pwr	imm	iam	ipm
30	1.450	0.603	0.842	0.952	1.056	0.848	0.941	1.068
50	1.009	0.446	0.560	0.644	0.651	0.571	0.641	0.623
100	0.699	0.276	0.342	0.453	0.406	0.343	0.450	0.407
300	0.400	0.150	0.185	0.288	0.210	0.184	0.289	0.208
500	0.312	0.122	0.149	0.252	0.168	0.149	0.243	0.167

Table A.25: Study 6.4: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.8, 8)^\top$, for parameter β_2

n	OLS	WLS	mult	add	pwr	imm	iam	ipm
30	1.359	0.612	0.828	0.907	1.027	0.833	0.897	1.032
50	1.048	0.439	0.552	0.667	0.676	0.548	0.661	0.627
100	0.731	0.303	0.354	0.454	0.436	0.355	0.452	0.429
300	0.394	0.162	0.188	0.291	0.216	0.191	0.291	0.216
500	0.310	0.125	0.145	0.237	0.174	0.146	0.228	0.173

Table A.26: Study 6.4: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (0.8, 8)^\top$, for parameter β_3

n	OLS	WLS	mult	add	pwr	imm	iam	ipm
30	25.24	4.569	8.934	13.701	16.78	6.089	13.60	20.62
50	19.77	3.162	5.078	9.052	10.533	3.920	9.007	9.907
100	13.85	2.052	2.801	5.651	5.832	2.510	5.664	5.026
300	7.795	1.026	1.292	2.764	1.779	1.254	2.768	1.555
500	6.284	0.790	0.976	2.242	1.250	0.955	2.242	1.128

Table A.27: Study 6.5: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (1, 4)^\top$, for parameter β_0

n	OLS	WLS	mult	add	pwr	imm	iam	ipm
30	4.288	1.108	2.182	2.979	3.020	1.485	2.946	3.355
50	3.229	0.791	1.307	2.081	2.366	0.964	2.063	2.587
100	2.289	0.524	0.773	1.428	1.604	0.609	1.421	1.592
300	1.326	0.301	0.376	0.784	0.798	0.352	0.781	0.774
500	1.028	0.227	0.272	0.591	0.551	0.262	0.590	0.527

Table A.28: Study 6.5: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (1, 4)^\top$, for parameter β_1

n	OLS	WLS	mult	add	pwr	imm	iam	ipm
30	3.139	0.588	1.135	1.782	2.223	0.781	1.755	3.157
50	2.402	0.399	0.630	1.179	1.460	0.514	1.171	1.882
100	1.720	0.263	0.355	0.764	0.804	0.332	0.763	0.862
300	0.966	0.141	0.176	0.395	0.304	0.175	0.395	0.313
500	0.733	0.107	0.131	0.302	0.210	0.130	0.302	0.203

Table A.29: Study 6.5: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (1, 4)^\top$, for parameter β_2

n	OLS	WLS	mult	add	pwr	imm	iam	ipm
30	2.965	0.570	1.067	1.669	2.143	0.767	1.650	2.933
50	2.315	0.406	0.594	1.159	1.347	0.499	1.148	1.650
100	1.585	0.265	0.338	0.697	0.804	0.329	0.696	0.890
300	0.940	0.151	0.181	0.393	0.309	0.184	0.392	0.317
500	0.723	0.108	0.132	0.301	0.213	0.133	0.300	0.207

Table A.30: Study 6.5: Dependence of the *sd* on the sample size n for the candidate estimators, where $\boldsymbol{\alpha} = (1, 4)^\top$, for parameter β_3