Regression and Linear Models (37252)

Lecture 8 - Generalised Least Squares

Lecturer: Joanna Wang
Notes adopted from Scott Alexander

School of Mathematical and Physical Sciences, UTS

Autumn 2024

Acknowledgement

These notes are based, in part, on earlier versions prepared by Dr Ed Lidums and Prof. James Brown.

8. Lecture outline

Topics:

- review of ordinary least squares (OLS)
- generalised least squares (GLS)
- weighted least squares (WLS)
    - example 1
    - example 2
        - OLS model
        - WLS model
    - example 3
        - OLS model
        - WLS model
    - example 4
        - OLS model
        - WLS model

See Chapter 9 of Draper and Smith (1998).

8. Review of Ordinary Least Squares

Multiple regression uses sample data $(\boldsymbol{X}, \boldsymbol{Y})$ that can be described as

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{\epsilon} \tag{1}$$

with

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \ \boldsymbol{X} = \begin{pmatrix} 1 & X_{1,1} & \cdots & X_{1,m} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n,1} & \cdots & X_{n,m} \end{pmatrix}, \ \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_m \end{pmatrix} \text{ and } \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

The method we have been using has required imposing the critical condition

$$\boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}) \tag{2}$$

implying the zero-mean, normally-distributed errors

1. have constant variance σ^2
2. are uncorrelated with each other.

8. Review of Ordinary Least Squares

The models we have been fitting have been based on

$$\mathbb{E}[\boldsymbol{Y}|\boldsymbol{X}] = \boldsymbol{X}\beta$$

which follows from the assumption $\mathbb{E}[\epsilon] = \mathbf{0}$.

The fitted data approximation is

$$\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\beta} \tag{3}$$

with the least squares parameter approximation of $\boldsymbol{\beta}$ given by

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}. \tag{4}$$

This estimate exists as long as the inverse $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ exists, which can be guaranteed if the columns of $\boldsymbol{X}$ are linearly independent.

The result is the model

$$\hat{Y}|x_1, \ldots, x_m = \hat{\beta}_0 + \sum_{j=1}^{m} \hat{\beta}_j x_j \tag{5}$$

where $\hat{\beta}_j$ is the $(j+1)$-th component of $\hat{\boldsymbol{\beta}}$.

8. Review of Ordinary Least Squares

Necessary for the statistical analysis we have applied has been the decomposition of the sum of squares

$$SST = SSR + SSE \tag{6}$$

where, letting $\mathbf{1}$ be square matrix of ones, the total sum of squares

$$SST = \sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \boldsymbol{Y}^T\boldsymbol{Y} - \frac{1}{n}\boldsymbol{Y}^T\mathbf{1}\boldsymbol{Y}, \tag{7}$$

the total sum of squares due to the regression

$$SSR = \sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2 = \hat{\beta}^T\boldsymbol{X}^T\boldsymbol{Y} - \frac{1}{n}\boldsymbol{Y}^T\mathbf{1}\boldsymbol{Y} \tag{8}$$

and the sum of squared errors

$$SSE = \hat{\boldsymbol{\epsilon}}^T\hat{\boldsymbol{\epsilon}}. \tag{9}$$

The residual

$$\hat{\boldsymbol{\epsilon}} = \boldsymbol{Y} - \hat{\boldsymbol{Y}} \tag{10}$$

is the estimate of $\boldsymbol{\epsilon}$.

8. Review of Ordinary Least Squares

Much of our work has been analysis of the residuals $\hat{\epsilon}$ to determine if their behaviour is consistent with the assumption $\epsilon \sim \mathrm{N}(\mathbf{0}, \sigma^2 \boldsymbol{I})$.

This has involved not only testing for normality, but also serial correlation and non-constant variance.

Where non-compliance has been identified, we have attempted to rectify the situation through transformation of the sample data $(\boldsymbol{X}, \boldsymbol{Y})$ and/or the addition of additional predictors.

Sometimes, despite best efforts, it is not possible to salvage an **Ordinary Least Squares (OLS)** model.

Where the problems are due to correlation and **heteroscedasticity** (non-constant variance) in the residuals an alternative method is available, the method of **Generalised Least Squares (GLS)**.

8. Generalised Least Squares

Again consider the model

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{\epsilon} \tag{11}$$

but relax the assumption in (2) to

$$\boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2\boldsymbol{\Sigma}) \tag{12}$$

where Σ is a symmetric, positive-definite matrix; i.e. for any $\boldsymbol{z} \in \mathbb{R}^n$

$$\boldsymbol{z}^T\boldsymbol{\Sigma}\boldsymbol{z} > 0. \tag{13}$$

The form of $\boldsymbol{\Sigma}$ is quite flexible and permits

1. unequal diagonal entries describing heteroscedasticity
2. non-zero off-diagonal entries describing non-zero correlation.

The condition (13) allows for the matrix factorisation

$$\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^T \tag{14}$$

where $\boldsymbol{\Gamma}$ is invertible (e.g. via Cholesky decomposition).

8. Generalised Least Squares

Multiplying (11) by the inverse of $\boldsymbol{\Gamma}$ gives

$$\boldsymbol{\Gamma}^{-1}\boldsymbol{Y} = \boldsymbol{\Gamma}^{-1}\boldsymbol{X}\beta + \boldsymbol{\Gamma}^{-1}\boldsymbol{\epsilon} \tag{15}$$

which after the re-labeling

$$\mathfrak{Y} = \boldsymbol{\Gamma}^{-1}\boldsymbol{Y}, \quad \mathfrak{X} = \boldsymbol{\Gamma}^{-1}\boldsymbol{X} \quad \text{and} \quad \mathfrak{e} = \boldsymbol{\Gamma}^{-1}\boldsymbol{\epsilon} \tag{16}$$

becomes

$$\mathfrak{Y} = \mathfrak{X}\beta + \mathfrak{e}. \tag{17}$$

By (12), the error term appearing above has the properties

$$\mathbb{E}[\mathfrak{e}] = \mathbb{E}[\boldsymbol{\Gamma}^{-1}\boldsymbol{e}] = \boldsymbol{\Gamma}^{-1}\,\mathbb{E}[\boldsymbol{e}] = \mathbf{0}$$

and

$$\begin{aligned}
\text{covar}(\mathfrak{e}) &= \text{covar}(\boldsymbol{\Gamma}^{-1}\boldsymbol{\epsilon}) = \boldsymbol{\Gamma}^{-1}\,\text{covar}(\boldsymbol{\epsilon})(\boldsymbol{\Gamma}^{-1})^{T} \\
&= \sigma^2\boldsymbol{\Gamma}^{-1}\boldsymbol{\Sigma}(\boldsymbol{\Gamma}^{-1})^{T} = \sigma^2\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{T}(\boldsymbol{\Gamma}^{-1})^{T} = \sigma^2\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}(\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma})^{T} \\
&= \sigma^2\boldsymbol{I}.
\end{aligned}$$

8. Generalised Least Squares

Such transformations of normally distributed RVs preserves their normality so the error term in (17)

$$\mathfrak{e} \sim \mathrm{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}). \tag{18}$$

We see that the factorisation of $\boldsymbol{\Sigma}$ in (14) and use in the data transformations (16) has turned the GLS problem (11) into the OLS problem (17).

The solution to the transformed problem is the approximation

$$\hat{\mathfrak{Y}} = \mathfrak{X}\hat{\beta} \tag{19}$$

with least squares estimate of $\boldsymbol{\beta}$,

$$\hat{\boldsymbol{\beta}} = (\mathfrak{X}^T \mathfrak{X})^{-1} \mathfrak{X}^T \mathfrak{Y}, \tag{20}$$

obtained in the usual fashion.

Of course, the estimate exists only if the inverse $(\mathfrak{X}^T \mathfrak{X})^{-1}$ exists, which can be guaranteed if the columns of $\mathfrak{X}$ are linearly independent.

8. Generalised Least Squares

To transform the solution back to the original coordinate system left-multiply (19) by $\mathbf{\Gamma}$ so that

$$\mathbf{\Gamma}\hat{\mathfrak{Y}} = \mathbf{\Gamma}\mathfrak{X}\hat{\beta}$$

which by (16) becomes

$$\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\beta} \tag{21}$$

with

$$\hat{\beta} = (\mathfrak{X}^T\mathfrak{X})^{-1}\mathfrak{X}^T\mathfrak{Y} = (\boldsymbol{X}^T\mathbf{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\mathbf{\Sigma}^{-1}\boldsymbol{Y} \tag{22}$$

following after various matrix manipulations.

The result is the fitted model

$$\hat{Y}|x_1,\ldots,x_m = \hat{\beta}_0 + \sum_{j=1}^{m}\hat{\beta}_j x_j \tag{23}$$

where $\hat{\beta}_j$ is the $(j+1)$-th component of $\hat{\beta}$ described in (22).

8. Generalised Least Squares

As with every other regression model we have built, the fitted GLS model requires further analysis.

The model must be tested for statistical significance using F-tests and/or T-tests on the individual parameters.

Potential points of influence need to be identified using Cook's D or DFITs.

Comparison in terms of R^2_{adj} needs to be conducted where multiple models are being considered.

Most importantly, residual analysis need to be undertaken. This is done in terms of the errors $\mathfrak{e}$ in the transformed model (17) according to the assumption (18).

The assumption of the residuals in the transformed GLS model is the same for those in the OLS models we have been studying - we look for departures from this assumption in exactly the same way.

8. Weighed Least Squares

Weighted least squares (WLS) is a particular case of GLS where the assumption on the errors

$$\boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2 \boldsymbol{\Sigma}) \tag{24}$$

involves a covariance matrix of the form

$$\boldsymbol{\Sigma} = \begin{pmatrix} \frac{1}{w_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{w_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{w_n} \end{pmatrix}$$

so that the variance of the i-th error $\epsilon_i = \frac{\sigma^2}{w_i}$ for all $i \in \{1, \ldots, n\}$.

The correlation between the residuals (ϵ_i, ϵ_k), $i \neq k$, is zero, as indicated by the zero off-diagonal components in $\boldsymbol{\Sigma}$.

So the residuals may have non-constant variance, but they must still be independent.

8. Weighed Least Squares

The decomposition of $\mathbf{\Sigma}$ in (14)

$$\mathbf{\Sigma} = \mathbf{\Gamma}\mathbf{\Gamma}^T$$

is trivial with

$$\mathbf{\Gamma} = \begin{pmatrix} \frac{1}{\sqrt{w_1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{w_2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{w_n}} \end{pmatrix}$$

as is the inverse

$$\mathbf{\Sigma}^{-1} = \begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{pmatrix}$$

used to find $\hat{\beta}$ in (22).

8. Weighed Least Squares

For WLS models, the components of (16) have the simpler forms

$$\mathfrak{X} = \begin{pmatrix} \sqrt{w_1} & \sqrt{w_1}X_{1,1} & \cdots & \sqrt{w_1}X_{1,m} \\ \sqrt{w_2} & \sqrt{w_2}X_{2,1} & \cdots & \sqrt{w_2}X_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{w_n} & \sqrt{w_n}X_{n,1} & \cdots & \sqrt{w_n}X_{n,m} \end{pmatrix},$$

$$\mathfrak{Y} = \begin{pmatrix} \sqrt{w_1}Y_1 \\ \sqrt{w_2}Y_2 \\ \vdots \\ \sqrt{w_n}Y_n \end{pmatrix} \quad \text{and} \quad \mathfrak{e} = \begin{pmatrix} \sqrt{w_1}\epsilon_1 \\ \sqrt{w_2}\epsilon_2 \\ \vdots \\ \sqrt{w_n}\epsilon_n \end{pmatrix},$$

the last of which can be used to modify the residuals produced by R before conducting the usual analysis (R only produces raw residuals when the WLS option is used).

8. Weighed Least Squares

The major difficulty in WLS, indeed in GLS, is identification of the matrix $\boldsymbol{\Sigma}$.

This issue did not arise in our earlier study of OLS as by definition $\boldsymbol{\Sigma} = \boldsymbol{I}$.

There are a variety of estimation techniques available, which we do not go into here.

An alternative is to start with an OLS model and then inspect the residuals for patterns suggestive of particular forms of $\boldsymbol{\Sigma}$.

In this way, WLS regression is seen as a version of the data transformation techniques under OLS described in the previous lecture.

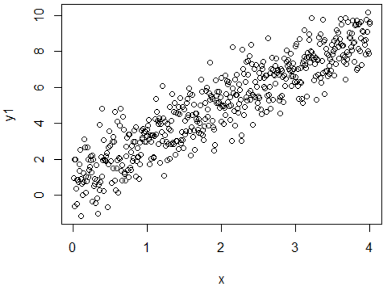
The difference is that the form of transformation is suggested by inspection of the residuals, not the sample data.

8. Weighted least squares - example 1

Recall the toy example from Lecture 7, where we generated sample data (X_i, Y_i), $i \in \{1, \ldots, 500\}$, according to the rule

$$Y_i = 1 + 2X_i + \epsilon_i, \quad \boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{I}).$$

The sample data is plotted below.



(Data for all examples is available in "WLS.sav" on Canvas).

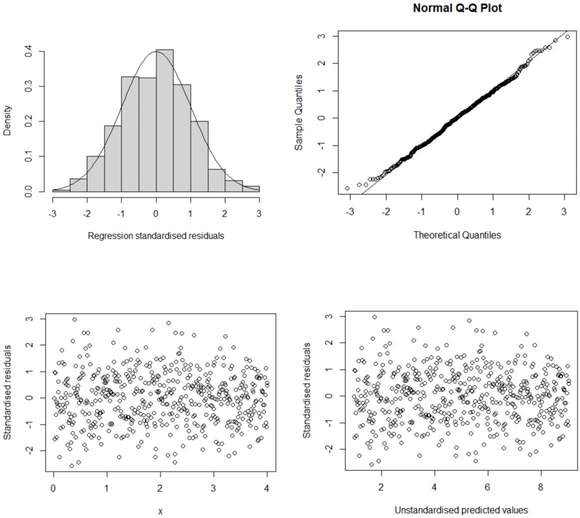
8. Weighted least squares - example 1

In this case there is no need for WLS and we can build an OLS model.

```
> mod1 <- lm(y1 ~ x, data = dat)
> summary(mod1)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.94916    0.09303   10.20   <2e-16 ***
x            2.02275    0.04022   50.29   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’

Residual standard error: 1.038 on 498 degrees of freedom
Multiple R-squared:  0.8355,Adjusted R-squared:  0.8352
F-statistic:  2529 on 1 and 498 DF,  p-value: < 2.2e-16

> durbinWatsonTest(mod1)
 lag Autocorrelation D-W Statistic p-value
   1     -0.03014957      2.059866   0.548
 Alternative hypothesis: rho != 0
```

8. Weighted least squares - example 1

Given the nature of the toy data, we expect no problems with the residuals.

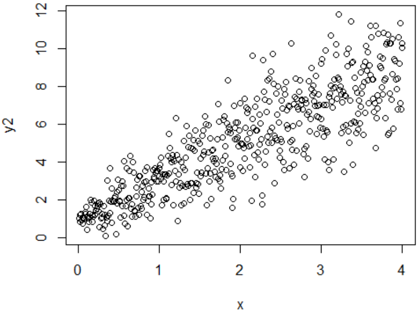


8. Weighted least squares - example 2

Now let the errors $\boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$ where

$$\boldsymbol{\Sigma} = \begin{pmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_n \end{pmatrix}.$$

The sample data is plotted below.



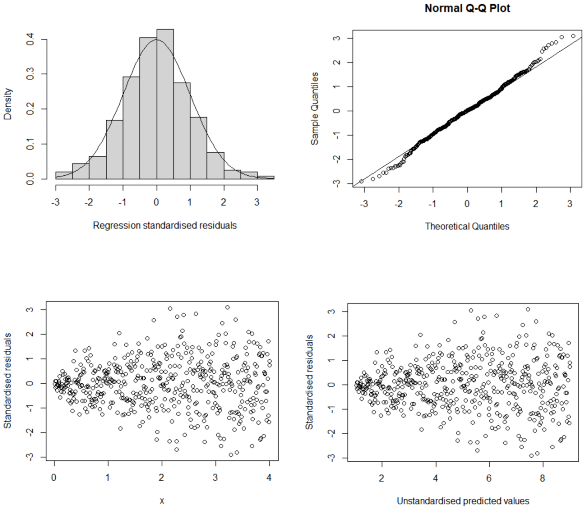
8. Weighted least squares - example 2 (OLS model)

For an OLS model, R produces the following output.

```
> mod2 <- lm(y2 ~ x, data = dat)
> summary(mod2)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.02124    0.12646   8.076 5.09e-15 ***
x            1.99546    0.05468  36.497  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’

Residual standard error: 1.412 on 498 degrees of freedom
Multiple R-squared:  0.7279,Adjusted R-squared:  0.7273
F-statistic:  1332 on 1 and 498 DF,  p-value: < 2.2e-16

> durbinWatsonTest(mod2)
 lag Autocorrelation D-W Statistic p-value
   1     -0.01506471      2.029039   0.728
 Alternative hypothesis: rho != 0
```

8. Weighted least squares - example 2 (OLS model)

Increasing variance of the residuals is revealed in the scatter plots.



8. Weighted least squares - example 2 (WLS model)

The covariance matrix can be composed as

$$\begin{aligned}
\mathbf{\Sigma} &= \begin{pmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_n \end{pmatrix} \\
&= \begin{pmatrix} \sqrt{X_1} & 0 & \cdots & 0 \\ 0 & \sqrt{X_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{X_n} \end{pmatrix} \begin{pmatrix} \sqrt{X_1} & 0 & \cdots & 0 \\ 0 & \sqrt{X_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{X_n} \end{pmatrix} \\
&= \mathbf{\Gamma}\mathbf{\Gamma}^T
\end{aligned}$$

and the WLS model weights

$$w_i = \frac{1}{X_i}.$$

8. Weighted least squares - example 2 (WLS model)

Using these weights, R produces the following output.

```
> wt2 <- 1/dat$x
> mod2_wls <- lm(y2 ~ x, data = dat, weights = wt2)
> summary(mod2_wls)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.95155    0.04237   22.46   <2e-16 ***
x            2.03024    0.03900   52.05   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.037 on 498 degrees of freedom
Multiple R-squared:  0.8448,Adjusted R-squared:  0.8444
F-statistic:  2710 on 1 and 498 DF,  p-value: < 2.2e-16

> durbinWatsonTest(mod2_wls)
 lag Autocorrelation D-W Statistic p-value
   1     -0.01417072      2.027391   0.774
 Alternative hypothesis: rho != 0
```

8. Weighted least squares - example 2 (WLS model)

Unfortunately, R does not produce the weighted residuals, so these must be calculated.

The weighted raw residuals are given by

$$\hat{\boldsymbol{\mathfrak{e}}} = \boldsymbol{\Gamma}^{-1}\hat{\boldsymbol{\epsilon}}$$

or component wise

$$\hat{\mathfrak{e}}_i = \frac{1}{\sqrt{X_i}}\hat{\epsilon}_i.$$

From this the Student T version of the standardised weighted residuals $\hat{\mathfrak{t}}_i$ can be calculated as

$$\hat{\mathfrak{t}}_i = \frac{\hat{\mathfrak{e}}_i - \bar{\hat{\mathfrak{e}}}}{S_{\hat{\mathfrak{e}}}},$$

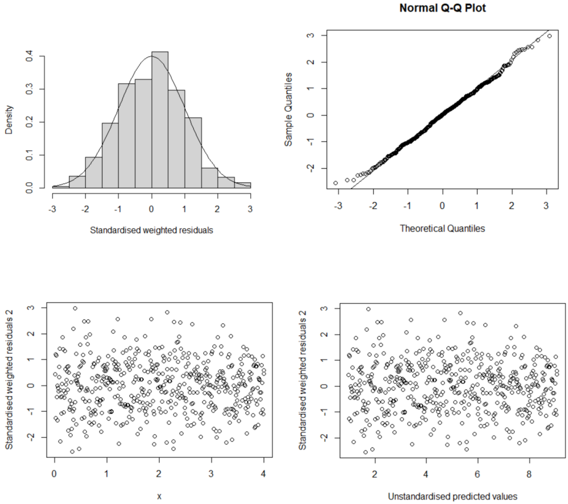
with the sample mean $\bar{\hat{\mathfrak{e}}}$ and standard deviation $S_{\hat{\mathfrak{e}}}$ available from R.

8. Weighted least squares - example 2 (WLS model)

```
> summary(mod2_wls$residuals)
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-4.12904 -0.91325  0.03236  0.00000  0.83919  4.32077
> sd(mod2_wls$residuals)
[1] 1.410866
> weighted_mod2_resid<-mod2_wls$residuals*sqrt(wt2)
> summary(weighted_mod2_resid)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-2.628982 -0.736788  0.029813  0.001275  0.728310  3.088846
> sd(weighted_mod2_resid)
[1] 1.036405
> std_weighted_mod2_resid<-
(weighted_mod2_resid-mean(weighted_mod2_resid))/sd(weighted_mod2_resid)
> summary(std_weighted_mod2_resid)
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-2.53786 -0.71214  0.02754  0.00000  0.70150  2.97912
> sd(std_weighted_mod2_resid)
[1] 1
```

8. Weighted least squares - example 2 (WLS model)

The Student T version of the standardised residuals, $\hat{t}_i$, behave as assumed, as can be seen from the plots below.

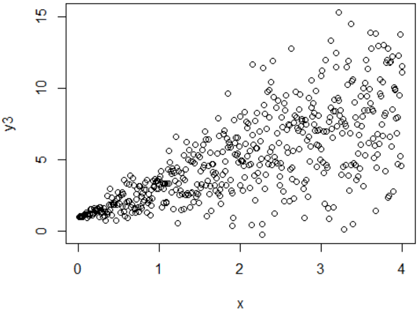


8. Weighted least squares - example 3

Now let the errors $\boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$ where

$$\boldsymbol{\Sigma} = \begin{pmatrix} X_1^2 & 0 & \cdots & 0 \\ 0 & X_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_n^2 \end{pmatrix}.$$

The sample data is plotted below.



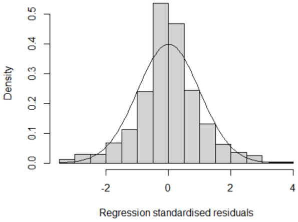
8. Weighted least squares - example 3 (OLS model)

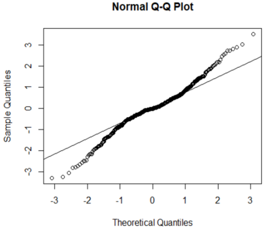
For an OLS model, R produces the following output.

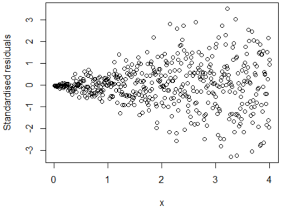
```
> summary(mod3)
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.08011    0.20117   5.369 1.22e-07 ***
x            1.96992    0.08698  22.648  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’

Residual standard error: 2.246 on 498 degrees of freedom
Multiple R-squared:  0.5074,Adjusted R-squared:  0.5064
F-statistic: 512.9 on 1 and 498 DF,  p-value: < 2.2e-16

> durbinWatsonTest(mod3)
 lag Autocorrelation D-W Statistic p-value
   1     -0.02138337      2.040964   0.736
 Alternative hypothesis: rho != 0
```

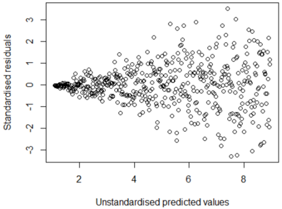
8. Weighted least squares - example 3 (OLS model)

All of the residual plots reveal violations of assumptions.









8. Weighted least squares - example 3 (WLS model)

The covariance matrix can be composed as

$$
\begin{aligned}
\boldsymbol{\Sigma} &= \begin{pmatrix} X_1^2 & 0 & \cdots & 0 \\ 0 & X_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_n^2 \end{pmatrix} \\
&= \begin{pmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_n \end{pmatrix} \begin{pmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_n \end{pmatrix} \\
&= \boldsymbol{\Gamma}\boldsymbol{\Gamma}^T
\end{aligned}
$$

and the WLS model weights

$$w_i = \frac{1}{X_i^2}.$$

8. Weighted least squares - example 3 (WLS model)

Using these weights, R produces the following output.

```
> wt3 <- 1/(dat$x)^2
> mod3_wls <- lm(y3 ~ x, data = dat, weights = wt3)
> summary(mod3_wls)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.99477    0.00667  149.15   <2e-16 ***
x            2.00364    0.04779   41.92   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '

Residual standard error: 1.038 on 498 degrees of freedom
Multiple R-squared:  0.7792,Adjusted R-squared:  0.7788
F-statistic:  1758 on 1 and 498 DF,  p-value: < 2.2e-16

> durbinWatsonTest(mod3_wls)
 lag Autocorrelation D-W Statistic p-value
   1     -0.02096808      2.040219   0.668
 Alternative hypothesis: rho != 0
```

8. Weighted least squares - example 3 (WLS model)

Again, R does not produce the weighted residuals, so these must be calculated.

The weighted raw residuals are given by

$$\hat{\mathfrak{e}} = \boldsymbol{\Gamma}^{-1}\hat{\boldsymbol{\epsilon}}$$

or component wise

$$\hat{\mathfrak{e}}_i = \frac{1}{X_i}\hat{\epsilon}_i.$$

The Student T version of the standardised weighted residuals, $\hat{\mathfrak{t}}_i$, can be calculated as

$$\hat{\mathfrak{t}}_i = \frac{\hat{\mathfrak{e}}_i - \bar{\hat{\mathfrak{e}}}}{S_{\hat{\mathfrak{e}}}},$$

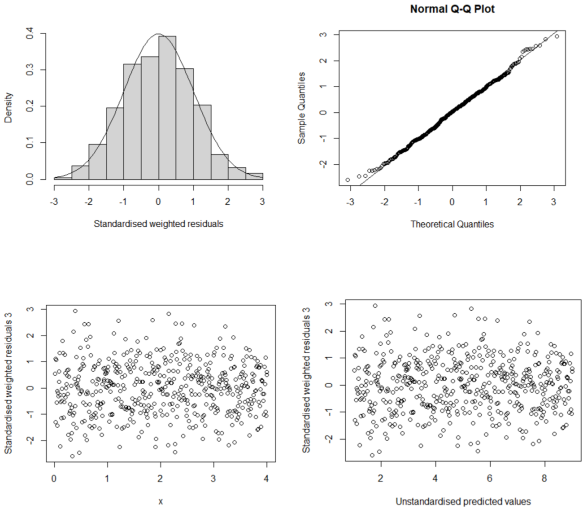
with the sample mean $\bar{\hat{\mathfrak{e}}}$ and standard deviation $S_{\hat{\mathfrak{e}}}$ available from R.

8. Weighted least squares - example 3 (WLS model)

```
> summary(mod3_wls$residuals)          # raw residuals
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-7.39285 -1.04621  0.02688  0.01777  1.15229  7.82954
> sd(mod3_wls$residuals)
[1] 2.243883
> weighted_mod3_resid<-mod3_wls$residuals*sqrt(wt3)
> summary(weighted_mod3_resid)
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-2.68311 -0.73086  0.03609  0.00000  0.70158  3.04010
> sd(weighted_mod3_resid)
[1] 1.037148
> std_weighted_mod3_resid<-
(weighted_mod3_resid-mean(weighted_mod3_resid))/sd(weighted_mod3_resid)
> summary(std_weighted_mod3_resid)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-2.5870 -0.7047  0.0348  0.0000  0.6764  2.9312
> sd(std_weighted_mod3_resid)
[1] 1
```

8. Weighted least squares - example 3 (WLS model)

The Student T version of the standardised residuals, $\hat{t}_i$, behave as assumed, as can be seen from the plots below.

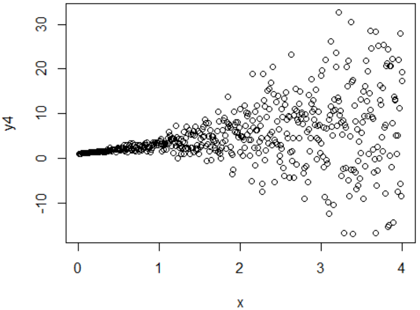


8. Weighted least squares - example 4

Finally let the errors $\boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \mathbf{\Sigma})$ where

$$\mathbf{\Sigma} = \begin{pmatrix} X_1^4 & 0 & \cdots & 0 \\ 0 & X_2^4 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_n^4 \end{pmatrix}.$$

The sample data is plotted below.



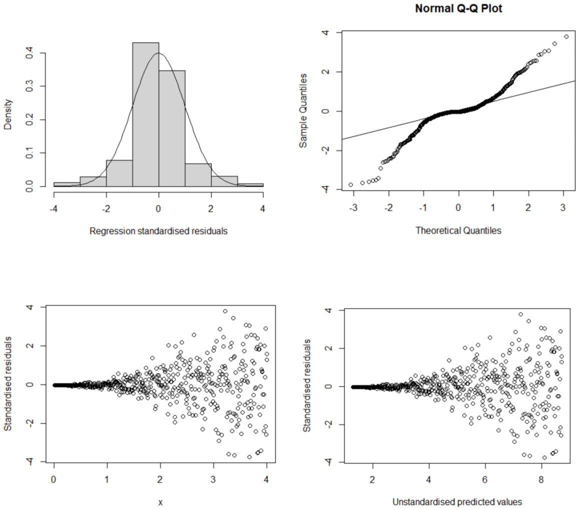
8. Weighted least squares - example 4 (OLS model)

For an OLS model, R produces the following output.

```
> mod4 <- lm(y4 ~ x, data = dat)
> summary(mod4)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.2761     0.6004   2.125    0.034 *
x             1.8620     0.2596   7.173 2.68e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’
1

Residual standard error: 6.703 on 498 degrees of freedom
Multiple R-squared:  0.09364,Adjusted R-squared:  0.09182
F-statistic: 51.45 on 1 and 498 DF,  p-value: 2.677e-12

> durbinWatsonTest(mod4)
 lag Autocorrelation D-W Statistic p-value
   1     -0.05586149      2.108398   0.264
 Alternative hypothesis: rho != 0
```

8. Weighted least squares - example 4 (OLS model)

All of the residual plots reveal violations of assumptions.



8. Weighted least squares - example 4 (WLS model)

The covariance matrix can be composed as

$$
\begin{aligned}
\boldsymbol{\Sigma} &= \begin{pmatrix} X_1^4 & 0 & \cdots & 0 \\ 0 & X_2^4 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_n^4 \end{pmatrix} \\
&= \begin{pmatrix} X_1^2 & 0 & \cdots & 0 \\ 0 & X_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_n^2 \end{pmatrix} \begin{pmatrix} X_1^2 & 0 & \cdots & 0 \\ 0 & X_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_n^2 \end{pmatrix} \\
&= \boldsymbol{\Gamma}\boldsymbol{\Gamma}^T
\end{aligned}
$$

and the WLS model weights

$$
w_i = \frac{1}{X_i^4}.
$$

8. Weighted least squares - example 4 (WLS model)

Using these weights, R produces the following output.

```
> wt4 <- 1/(dat$x)^4
> mod4_wls <- lm(y4 ~ x, data = dat, weights = wt4)
> summary(mod4_wls)
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.0001116  0.0001474  6782.9   <2e-16 ***
x           1.9846842  0.0149593   132.7   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.038 on 498 degrees of freedom
Multiple R-squared:  0.9725,Adjusted R-squared:  0.9724
F-statistic: 1.76e+04 on 1 and 498 DF,  p-value: < 2.2e-16

> durbinWatsonTest(mod4_wls)
 lag Autocorrelation D-W Statistic p-value
   1     -0.05528315       2.10741   0.252
 Alternative hypothesis: rho != 0
```

8. Weighted least squares - example 4 (WLS model)

R does not produce the weighted residuals, so these must be calculated.

The weighted raw residuals are given by

$$\hat{\mathfrak{e}} = \mathbf{\Gamma}^{-1}\hat{\boldsymbol{\epsilon}}$$

or component wise

$$\hat{\mathfrak{e}}_i = \frac{1}{X_i^2}\hat{\epsilon}_i.$$

The Student T version of the standardised weighted residuals, $\hat{\mathfrak{t}}_i$, can be calculated as

$$\hat{\mathfrak{t}}_i = \frac{\hat{\mathfrak{e}}_i - \bar{\hat{\mathfrak{e}}}}{S_{\hat{\mathfrak{e}}}},$$

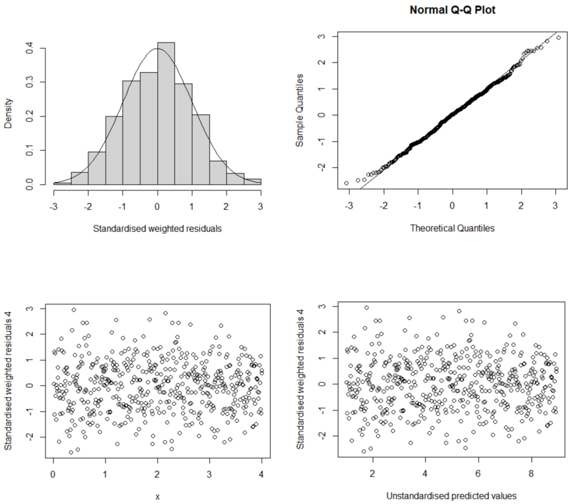
with the sample mean $\bar{\hat{\mathfrak{e}}}$ and standard deviation $S_{\hat{\mathfrak{e}}}$ available from R.

8. Weighted least squares - example 4 (WLS model)

```
> summary(mod4_wls$residuals)          # raw residuals
      Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
-25.118555  -1.651478   0.007525   0.030229   2.334118  25.249810
> sd(mod4_wls$residuals)
[1] 6.697527
> weighted_mod4_resid<-mod4_wls$residuals*sqrt(wt4)
> summary(weighted_mod4_resid)
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-2.65044 -0.71983  0.04747  0.01504  0.73823  3.06874
> sd(weighted_mod4_resid)
[1] 1.036449
> std_weighted_mod4_resid<-
(weighted_mod4_resid-mean(weighted_mod4_resid))/sd(weighted_mod4_resid)
> summary(std_weighted_mod4_resid)
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-2.57174 -0.70902  0.03129  0.00000  0.69776  2.94631
> sd(std_weighted_mod4_resid)
[1] 1
```

8. Weighted least squares - example 4 (WLS model)

The Student T version of the standardised residuals, $\hat{t}_i$, behave as assumed, as can be seen from the plots below.



References I

Draper, N. R. and Smith, H. (1998). *Applied regression analysis*. Wiley-Interscience, Somerset, US.