ECON207 Session 3

SLR Standard Errors; Multiple Linear Regression

Anthony Tay

This Version: 31 Jul 2024

Session 3

- Recall key points from last session (simple linear regression) + filling a few gaps
- Standard error of OLS estimators
    - Derive s.e. formula under homoskedasticity
    - Demonstrate OLS efficiency under homoskedasticity
    - Derive s.e. formula under general heteroskedasticity
- Efficient estimation under heteroskedasticity
- Intro to **Multiple Linear Regression (MLR)** and OLS estimation of the MLR model

Session 3.1

Session 3.1 Recollection

- Recall key points from previous session (simple linear regression) + filling a few gaps

Recollection SLR

Assumptions $\{X_i, Y_i\}_{i=1}^n$ iid sample from population with $E(Y \mid X) = \beta_0 + \beta_1 X$

- $Y = \beta_0 + \beta_1 X + \epsilon\ ,\ \ E(\epsilon \mid X) = 0$
- $E(\epsilon) = 0,\ Cov(\epsilon, X) = E(X\epsilon) = 0,\ \beta_1 = \dfrac{Cov(X,Y)}{Var(X)}$
- Recall discussion about correlation vs causality

Simple Linear Regression Model: $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$

$$\text{OLS}: \quad \hat{\beta}_0^{ols},\ \hat{\beta}_1^{ols} = \underset{\hat{\beta}_0,\hat{\beta}_1}{\arg\min} \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$$

Recollection SLR

FOC:

$$\left.\begin{aligned} \left.\frac{\partial SSR}{\partial \hat{\beta}_0}\right|_{\hat{\beta}_0^{ols},\,\hat{\beta}_1^{ols}} &= -2\sum_{i=1}^{n}(Y_i - \hat{\beta}_0^{ols} - \hat{\beta}_1^{ols}X_i) = 0 \\ \left.\frac{\partial SSR}{\partial \hat{\beta}_1}\right|_{\hat{\beta}_0^{ols},\,\hat{\beta}_1^{ols}} &= -2\sum_{i=1}^{n}(Y_i - \hat{\beta}_0^{ols} - \hat{\beta}_1^{ols}X_i)X_i = 0 \end{aligned}\right\} \Rightarrow \begin{aligned} \hat{\beta}_0^{ols} &= \overline{Y} - \hat{\beta}_1^{ols}\overline{X} \\ \hat{\beta}_1^{ols} &= \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \end{aligned}$$

Different ways of writing $\hat{\beta}_1^{ols}$:

$$\hat{\beta}_1^{ols} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})Y_i}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \beta_1 + \frac{\sum_{i=1}^{n}(X_i - \overline{X})\epsilon_i}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

Recollection SLR

Unbiased: $E(\hat{\beta}_1^{ols}) = \beta_1$, Consistent: $\hat{\beta}_1^{ols} \xrightarrow{p} \beta_1$

- Nothing assumed re: $Var(\epsilon_i \mid X_1, \ldots, X_n)$. Irrelevant for unbiasedness/consistency
- Also: $\hat{\beta}_1^{ols}$ is a "Linear Estimator"

$$\hat{\beta}_1^{ols} = \frac{\sum_{i=1}^n (X_i - \overline{X})Y_i}{\sum_{i=1}^n (X_i - \overline{X})^2} = \sum_{i=1}^n \frac{(X_i - \overline{X})}{\sum_{i=1}^n (X_i - \overline{X})^2} Y_i = \sum_{i=1}^n w_i Y_i \text{ where } w_i = \frac{(X_i - \overline{X})}{\sum_{i=1}^n (X_i - \overline{X})^2}$$

- Can show: $\displaystyle\sum_{i=1}^n w_i = 0\,,\ \sum_{i=1}^n w_i X_i = 1\,,\ \sum_{i=1}^n w_i^2 = \frac{1}{\sum_{i=1}^n (X_i - \overline{X})^2}$

Recollection SLR

FOC can be written as

$$\sum_{i=1}^{n} \hat{\epsilon}_i^{ols} = 0 \quad \text{and} \quad \sum_{i=1}^{n} \hat{\epsilon}_i^{ols} X_i = 0 \quad \text{where}^* \quad \hat{\epsilon}_i^{ols} = Y_i - \hat{\beta}_0^{ols} - \hat{\beta}_1^{ols} X_i$$

- $\sum_{i=1}^{n} \hat{\epsilon}_i^{ols} = 0 \Rightarrow \overline{\hat{\epsilon}^{ols}} = 0$
- We describe $\sum_{i=1}^{n} \hat{\epsilon}_i^{ols} X_i = 0$ by saying $\hat{\epsilon}_i^{ols}$ and X_i are **orthogonal**
- Since $\overline{\hat{\epsilon}^{ols}} = 0$, the equation $\sum_{i=1}^{n} \hat{\epsilon}_i^{ols} X_i = 0$ implies sample covariance X_i and $\hat{\epsilon}_i^{ols}$ is zero

* I will use both $\hat{\epsilon}_i^{ols}$ and $\hat{\epsilon}_{i,ols}$ to denote OLS residuals

Recollection SLR

- The fitted values $\hat{Y}_i^{ols} = \hat{\beta}_0^{ols} + \hat{\beta}_1^{ols} X_i$ and the residuals $\hat{\epsilon}_i^{ols}$ are orthogonal

$$\sum_{i=1}^{n} \hat{Y}_i^{ols} \hat{\epsilon}_i^{ols} = \sum_{i=1}^{n} (\hat{\beta}_0^{ols} + \hat{\beta}_1^{ols} X_i) \hat{\epsilon}_i^{ols} = \hat{\beta}_0^{ols} \sum_{i=1}^{n} \hat{\epsilon}_i^{ols} + \hat{\beta}_1^{ols} \sum_{i=1}^{n} X_i \hat{\epsilon}_i^{ols} = 0$$

- Sample covariance of $\hat{Y}_i^{ols}$ and $\hat{\epsilon}_i^{ols}$ is zero
- Simple regression of Y_i on X_i, with intercept, breaks Y_i into two uncorrelated parts

$$Y_i = \hat{Y}_i^{ols} + \hat{\epsilon}_i^{ols}$$

$\hat{Y}_i^{ols}$ perfectly correlated with X_i, $\hat{\epsilon}_i^{ols}$ perfectly uncorrelated with X_i

Recollection SLR

- $\overline{Y} = \overline{\widehat{Y}^{ols}} + \overline{\widehat{\epsilon}^{ols}} = \overline{\widehat{Y}^{ols}}$
- smpl. var.(Y_i) = smpl. var.$(\hat{Y}_i^{ols})$ + smpl. var.$(\hat{\epsilon}^{ols})$

$$Y_i - \overline{Y} = \hat{Y}_i^{ols} - \overline{Y} + \hat{\epsilon}_i^{ols}$$

$$Y_i - \overline{Y} = \hat{Y}_i^{ols} - \overline{\hat{Y}^{ols}} + \hat{\epsilon}_i^{ols}$$

$$\left(Y_i - \overline{Y}\right)^2 = \left(\hat{Y}_i^{ols} - \overline{\hat{Y}^{ols}}\right)^2 + \hat{\epsilon}_{i,ols}^2 + 2\left(\hat{Y}_i^{ols} - \overline{\hat{Y}^{ols}}\right)\hat{\epsilon}_i^{ols}$$

$$\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 = \sum_{i=1}^{n}\left(\hat{Y}_i^{ols} - \overline{\hat{Y}^{ols}}\right)^2 + \sum_{i=1}^{n}\hat{\epsilon}_{i,ols}^2 + 2\sum_{i=1}^{n}\left(\hat{Y}_i^{ols} - \overline{\hat{Y}^{ols}}\right)\hat{\epsilon}_i^{ols}$$

$$\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 = \sum_{i=1}^{n}\left(\hat{Y}_i^{ols} - \overline{\hat{Y}^{ols}}\right)^2 + \sum_{i=1}^{n}\hat{\epsilon}_{i,ols}^2$$

Recollection SLR

$$\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)^2=\sum_{i=1}^{n}\left(\widehat{Y}_i^{ols}-\overline{\widehat{Y}^{ols}}\right)^2+\sum_{i=1}^{n}\widehat{\epsilon}_{i,ols}^2$$

Sum of Sq. Total (SST) = Sum of Sq. Explained (SSE) + Sum of Sq. Resiuals (SSR)

This suggests the following measure of goodness-of-fit

$$R^2=\frac{SSE}{SST}=1-\frac{SSR}{SST}=1-\frac{\sum_{i=1}^{n}\widehat{\epsilon}_{i,ols}^2}{\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)^2}$$

- $R^2=1 \Rightarrow \widehat{\epsilon}_i^{ols}=0$ for all i, i.e., perfect fit
- $R^2=0 \Rightarrow \widehat{Y}_i^{ols}=\overline{Y}$ for all i, i.e., $\widehat{\beta}_1^{ols}=0$

Recollection SLR

```r
library(tidyverse)
library(patchwork)
library(latex2exp)
dat <- read_csv("data\\earnings2019.csv", show_col_types=FALSE)
mdl1 <- lm(log(earn) ~ educ, data = dat)
dat$fitted <- fitted(mdl1); dat$residuals <- residuals(mdl1)
cat("\nDependent Variable: ln earn \n\n")
summary(mdl1)$coefficients %>% round(4)
cat("\nR-Squared:", summary(mdl1)$r.squared)
cat("\n\n(Illustration only)\nCov. fitted and residuals:",
    cov(dat$fitted, dat$residuals))
```

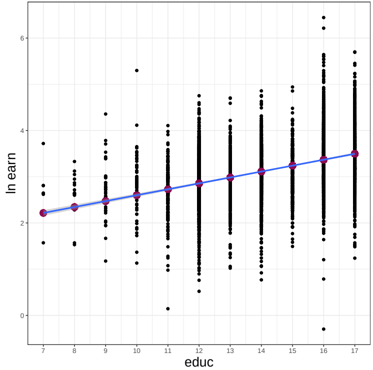
```
Dependent Variable: ln earn

            Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.320     0.0575 22.9399        0
educ           0.128     0.0040 32.1700        0

R-Squared: 0.1730933

(Illustration only)
Cov. fitted and residuals: 3.587746e-17
```

```r
mytheme <- theme_bw() +
  theme(aspect.ratio=1, axis.title=element_text(size=18))
ggplot(data=dat, aes(x=educ, y=log(earn))) + geom_point() +
  geom_point(aes(x=educ, y=fitted), size=4, color="deeppink4") +
  geom_smooth(method="lm", formula=y ~ x, se=TRUE) +
  ylab("ln earn") + scale_x_continuous(breaks=7:17) + mytheme
```



Session 3.2

Session 3.2 Estimator Standard Error

- Standard error of $\hat{\beta}_1^{ols}$
  - Derive s.e. formula under homoskedasticity
  - Demonstrate OLS efficiency under homoskedasticity
  - Derive s.e. formula under general heteroskedasticity

Estimator Standard Error

Now we turn to estimating the estimator standard errors

$$\begin{aligned}
\hat{\beta}_1^{ols} &= \beta_1 + \frac{\sum_{i=1}^n (X_i - \overline{X})\epsilon_i}{\sum_{i=1}^n (X_i - \overline{X})^2} \\
Var(\hat{\beta}_1^{ols} \mid X_1, \dots, X_n) &= Var\left(\frac{\sum_{i=1}^n (X_i - \overline{X})\epsilon_i}{\sum_{i=1}^n (X_i - \overline{X})^2} \;\middle|\; X_1, \dots, X_n \right) \\
&= \frac{\sum_{i=1}^n (X_i - \overline{X})^2 \, Var(\epsilon_i \mid X_1, \dots, X_n)}{\left(\sum_{i=1}^n (X_i - \overline{X})^2\right)^2}
\end{aligned}$$

Recall $Var(aX + bY) = a^2\, Var(X) + b^2\, Var(Y) + 2ab\, Cov(X, Y)$, last term zero if X, Y independent

Estimator Standard Error (Case 1)

Case 1: Population is $Y = \beta_0 + \beta_1 X + \epsilon$, $E(\epsilon \mid X) = 0$ and $Var(\epsilon \mid X) = \sigma^2$ and you have iid sample from the population. Then

$$Var(\epsilon_i \mid X_1, \dots, X_n) = \sigma^2 \text{ for all } i$$

$$\begin{aligned} Var(\hat{\beta}_1^{ols} \mid X_1, \dots, X_n) &= \frac{\sum_{i=1}^n (X_i - \overline{X})^2 \, Var(\epsilon_i \mid X_1, \dots, X_n)}{\left[\sum_{i=1}^n (X_i - \overline{X})^2\right]^2} \\ &= \frac{\sum_{i=1}^n (X_i - \overline{X})^2 \sigma^2}{\left[\sum_{i=1}^n (X_i - \overline{X})^2\right]^2} = \frac{\sigma^2 \sum_{i=1}^n (X_i - \overline{X})^2}{\left[\sum_{i=1}^n (X_i - \overline{X})^2\right]^2} \\ &= \frac{\sigma^2}{\sum_{i=1}^n (X_i - \overline{X})^2} \end{aligned}$$

Estimator Standard Error (Case 1)

- Recall $Var(Y) = E(Var(Y \mid X)) + Var(E(Y \mid X))$
- $Var(\hat{\beta}_1^{ols}) = E(Var(\hat{\beta}_1^{ols} \mid X_1, \ldots, X_n)) + Var(E(\hat{\beta}_1^{ols} \mid X_1, \ldots, X_n))$
- second term is zero, therefore

$$Var(\hat{\beta}_1^{ols}) = \sigma^2 E\left(\frac{1}{\sum_{i=1}^n (X_i - \overline{X})^2}\right)$$

Since $\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2 \xrightarrow{p} Var(X)$, we have $\sum_{i=1}^n (X_i - \overline{X})^2 \approx n\,Var(X)$ for n large

$$\text{i.e., } Var(\hat{\beta}_1^{ols}) \approx \frac{\sigma^2}{n\,Var(X)}$$

- increases with σ^2, decreases with $Var(X)$, goes to zero as $n \to \infty$

Estimator Standard Error (Case 1)

An unbiased estimator for σ^2 is

$$\widehat{\sigma^2} = \frac{1}{n-2}\sum_{i=1}^{n}\hat{\epsilon}_{i,ols}^2$$

We prove this later in a more general context

Have to divide by $n-2$ for unbiasedness because "two degrees of freedom" was lost in computing $\hat{\epsilon}_i^{ols}$ (have to estimate β_0 and β_1 first)

$$\widehat{Var}(\hat{\beta}_1^{ols} \mid X_1, \ldots, X_n) = \frac{\widehat{\sigma^2}}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \quad \text{and} \quad \text{s.e.}(\hat{\beta}_1 \mid X_1, \ldots, X_n) = \sqrt{\frac{\widehat{\sigma^2}}{\sum_{i=1}^{n}(X_i - \overline{X})^2}}$$

Estimator Standard Error (Case 1)

Recall $\hat{\beta}_1^{ols}$ is a linear estimator: $\sum_{i=1}^n w_i Y_i$ where $w_i = \dfrac{(X_i - \overline{X})}{\sum_{i=1}^n (X_i - \overline{X})^2}$

OLS is the linear unbiased estimator with the lowest variance

That is, if

$$\tilde{\beta}_1 = \sum_{i=1}^n (w_i + a_i) Y_i = \hat{\beta}_1^{ols} + \sum_{i=1}^n a_i Y_i$$

is some other linear estimator (a_i made up of constants and X_i, $i = 1, \dots, n$), then

$$Var(\tilde{\beta}_1 \mid X_1, \dots, X_n) \geq Var(\hat{\beta}_1^{ols} \mid X_1, \dots, X_n)$$

We prove this in a few steps

Estimator Standard Error (Case 1)

First note that

$$\tilde{\beta}_1 = \hat{\beta}_1^{ols} + \sum_{i=1}^{n} a_i Y_i = \hat{\beta}_1^{ols} + \sum_{i=1}^{n} a_i(\beta_0 + \beta_1 X_i + \epsilon_i)$$

$$= \hat{\beta}_1^{ols} + \beta_0 \sum_{i=1}^{n} a_i + \beta_1 \sum_{i=1}^{n} X_i a_i + \sum_{i=1}^{n} a_i \epsilon_i$$

$$E(\tilde{\beta}_1 \mid X_1, \dots, X_n) = \beta_1 + \beta_0 \sum_{i=1}^{n} a_i + \beta_1 \sum_{i=1}^{n} X_i a_i + \sum_{i=1}^{n} a_i E(\epsilon_i \mid X_1, \dots, X_n)$$

- last term is zero
- For $\tilde{\beta}_1$ to be unbiased, we require $\sum_{i=1}^{n} a_i = 0$ and $\sum_{i=1}^{n} X_i a_i = 0$

Estimator Standard Error (Case 1)

Then

$$Var(\tilde{\beta}_1 \mid X_1, \dots, X_n) = \sum_{i=1}^{n} (w_i + a_i)^2 \, Var(\epsilon_i \mid X_1, \dots, X_n) = \sum_{i=1}^{n} (w_i^2 + 2w_i a_i + a_i^2)\sigma^2$$

$$= \underbrace{\sigma^2 \sum_{i=1}^{n} w_i^2}_{Var(\hat{\beta}_1^{ols} \mid X_1, \dots, X_n)} + \underbrace{2\sigma^2 \sum_{i=1}^{n} w_i a_i}_{=0} + \underbrace{\sigma^2 \sum_{i=1}^{n} a_i^2}_{\geq 0}$$

- $\sum_{i=1}^{n} w_i a_i = \dfrac{\sum_{i=1}^{n} (X_i - \overline{X}) a_i}{\sum_{i=1}^{n} (X_i - \overline{X})^2} = \dfrac{\sum_{i=1}^{n} X_i a_i - \overline{X} \sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} (X_i - \overline{X})^2} = 0$
- $Var(\hat{\beta}_1^{ols} \mid X_1, \dots, X_n) = \sigma^2 / \sum_{i=1}^{n} (X_i - \overline{X})^2 = \sigma^2 \sum_{i=1}^{n} w_i^2$

Estimator Standard Error (Case 1)

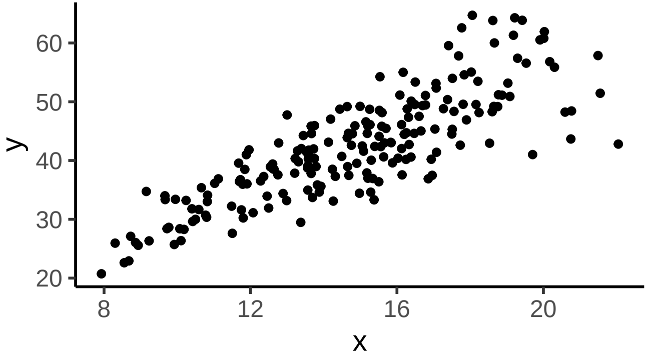
- $\hat{\beta}_1^{ols}$ has the smallest variance among all linear unbiased estimator
- same for $\hat{\beta}_0^{ols}$
- requires $Var(\epsilon_i \mid X_1, \dots, X_n)$ for all i ("Homoskedasticity")
- This is the famous "Gauss-Markov Theorem"
- Strong justification for using OLS to estimate linear regression models

However, "heteroskedasticity", i.e., $Var(\epsilon \mid X_1, \dots, x_n) = \sigma_i^2$ may be the more prevalent situation in economics

Estimator Standard Error (Case 2)

Example

```
df_het <- read_csv("data\\heterosk.csv", col_types=c("n","n","n"))
ggplot(data=df_het) + geom_point(aes(x=x,y=y), size=1) + theme_classic()
```



$$Y = \beta_0 + \beta_1 X + \epsilon$$

- $E(\epsilon \mid X) = 0$
- $Var(\epsilon \mid X) = \sigma^2 X^2$

may be more appropriate description of the population

Then we will have

$$Var(\epsilon_i \mid X_1, \ldots, X_n) \\ = \sigma^2 X_i^2$$

Estimator Standard Error (Case 2)

Case 2: Your sample follows $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ where $E(\epsilon_i \mid X_1, \dots, X_n) = 0$ and $Var(\epsilon_i \mid X_1, \dots, X_n) = \sigma_i^2$

$$\hat{\beta}_1^{ols} = \beta_1 + \frac{\sum_{i=1}^n (X_i - \overline{X})\epsilon_i}{\sum_{i=1}^n (X_i - \overline{X})^2}$$

$$Var(\hat{\beta}_1^{ols} \mid X_1, \dots, X_n) = Var\left(\frac{\sum_{i=1}^n (X_i - \overline{X})\epsilon_i}{\sum_{i=1}^n (X_i - \overline{X})^2} \,\middle|\, X_1, \dots, X_n \right)$$

$$= \frac{\sum_{i=1}^n (X_i - \overline{X})^2 \, Var(\epsilon_i \mid X_1, \dots, X_n)}{\left(\sum_{i=1}^n (X_i - \overline{X})^2\right)^2}$$

cannot be simplified without further assumptions regarding $Var(\epsilon_i \mid X_1, \dots, X_n)$

Estimator Standard Error (Case 2)

Heteroskedasticity-Robust / Het.-Consistent / White Standard Errors

It can be shown, under quite general conditions, that

$$\widehat{Var}_{HC}(\hat{\beta}_1^{ols}) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2\,\hat{\epsilon}_{i,ols}^2}{\left(\sum_{i=1}^{n}(X_i - \overline{X})^2\right)^2} \xrightarrow{p} Var(\hat{\beta}_1^{ols})$$

- This holds despite the fact that $E(\hat{\epsilon}_{i,ols}^2)$ is in general a biased and inconsistent estimator for $Var(\epsilon_i)$
- $\widehat{Var}_{HC}(\hat{\beta}_1^{ols})$ remains consistent even under homoskedasticity

Estimator Standard Error (Case 2)

- The default OLS s.e.

$$\widehat{Var}(\hat{\beta}_1^{ols} \mid X_1, \ldots, X_n) = \frac{\widehat{\sigma^2}}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

  was derived assuming homoskedasticity

- it is inconsistent for $Var(\hat{\beta}_1^{ols})$ under heteroskedasticity
- since heteroskedasticity-robust s.e. is consistent under both homoskedasticity and heteroskedasticity
  - maybe safer to always use the latter
  - default s.e. may be better estimate of s.e. under homoskedasticity

Estimator Standard Error (Case 2)

```
library(lmtest)      # lmtest::coeftest to calculate t-test
library(sandwich)    # using robust s.e. calculated with sandwich::vcovHC
cat("Assuming homoskedasticity (default)\n")
mdl2 <- lm(y ~ x, dat=df_het)  # Estimate OLS
summary(mdl2)$coefficients     # Print default coefficients and standard errors
cat("\nUsing heteroskedasticity-robust standard errors")
coeftest(mdl2, vcov=vcovHC, type="HC")  # Robust standard errors
```

```
Assuming homoskedasticity (default)
            Estimate Std. Error  t value     Pr(>|t|)
(Intercept) 6.285792  1.7817576  3.52786 5.207066e-04
x           2.430744  0.1177712 20.63955 3.034896e-51

Using heteroskedasticity-robust standard errors
t test of coefficients:

            Estimate Std. Error t value  Pr(>|t|)
(Intercept)  6.28579    1.86393  3.3723 0.0008963 ***
x            2.43074    0.13624 17.8414 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Estimator Standard Error (Case 2)

```
## Manual verification of the s.e. calculations
x <- df_het$x
y <- df_het$y
ehat <- mdl2$residuals   # residuals
N_2 <- mdl2$df.residual  # No of obs - no of parameters
s2hat <- sum(mdl2$residuals^2)/mdl2$df.residual # sigma^2 hat
x_xbar_2 <- (x-mean(x))^2  # squared dev. from mean
beta1hatse <- sqrt(s2hat/sum(x_xbar_2))  # default s.e.
beta1hatse_robust <- sqrt(sum(x_xbar_2*ehat^2)/(sum(x_xbar_2)^2)) #robust s.e.
cat("Verify default OLS s.e.:", beta1hatse)
cat("\nVerify OLS robust s.e.:", beta1hatse_robust)
```

```
Verify default OLS s.e.: 0.1177712
Verify OLS robust s.e.: 0.1362415
```

Session 3.3

Session 3.3 Weighted Least Squares

- Efficient estimation under heteroskedasticity
- Weighted Least Squares

Weighted Least Squares

- "Best Linear Unbiased" property of $\hat{\beta}_1^{ols}$ requires homoskedasticity
- If noise term is heteroskedastic
  - $\hat{\beta}_1^{ols}$ is still unbiased / consistent
  - remember to use heteroskedastic robust standard errors
  - OLS no longer minimum variance among all linear estimators

Is there a better estimator? (At least in the class of linear unbiased estimators?)

Weighted Least Squares

In some cases, there are more efficient options that are feasible

Suppose for all $i, j = 1, \dots, n$

- $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$
- $E(\epsilon_i \mid X_1, \dots, X_n) = 0$
- $E(\epsilon_i^2 \mid X_1, \dots, X_n) = \sigma_i^2 = \sigma^2 \eta(X_i)$ where $\eta(X_i)$ is known
- $E(\epsilon_i \epsilon_j \mid X_1, \dots, X_n) = 0$ for all $i \neq j$
- $\sum_{i=1}^{n} (X_i - \overline{X})^2 > 0$

e.g., $\eta(X_i) = X_i^2$, $\eta(X_i) = \exp(X_i)$, $\eta(X_i) = X_i$ (assuming $X_i > 0$)

Weighted Least Squares

WLS: weight observations so that weighted noise terms are no longer heteroskedastic

$$\frac{Y_i}{\sqrt{\eta(X_i)}} = \beta_0 \frac{1}{\sqrt{\eta(X_i)}} + \beta_1 \frac{X_i}{\sqrt{\eta(X_i)}} + \frac{\epsilon_i}{\sqrt{\eta(X_i)}}$$

$$Y_i^* = \beta_0 X_{i0}^* + \beta_1 X_{i1}^* + \epsilon_i^*$$

Note that ϵ_i^* is homoskedastic:

$$\begin{aligned} Var(\epsilon_i^* \mid X_1, \dots, X_n) &= Var\left(\epsilon_i / \sqrt{\eta(X_i)} \mid X_1, \dots, X_n\right) \\ &= \frac{1}{\eta(X_i)} Var(\epsilon_i \mid X_1, \dots, X_n) = \frac{1}{\eta(X_i)} \eta(X_i) \sigma^2 \\ &= \sigma^2 \end{aligned}$$

Weighted Least Squares

Then choose $\hat{\beta}_0$ and $\hat{\beta}_1$ to minimize

$$\begin{aligned}\sum_{i=1}^{n}(Y_i^* - \beta_0 X_{i0}^* + \beta_1 X_{i1}^*)^2 &= \sum_{i=1}^{n}\left(\frac{Y_i}{\sqrt{\eta(X_i)}} - \frac{1}{\sqrt{\eta(X_i)}}\hat{\beta}_0 - \hat{\beta}_1 \frac{X_i}{\sqrt{\eta(X_i)}}\right)^2 \\ &= \sum_{i=1}^{n}\frac{1}{\eta(X_i)}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)^2 \\ &= \sum_{i=1}^{n}\omega_i\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)^2\end{aligned}$$

where $\omega_i = 1/\eta(X_i)$, i.e., "Weighted Least Squares"

Weighted Least Squares

$$\text{WLS: } \hat{\beta}_0^{wls}, \hat{\beta}_1^{wls} = \underset{\hat{\beta}_0, \hat{\beta}_1}{\arg\min} \sum_{i=1}^{n} \omega_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$$

where the weights are $\omega_i = 1/\eta_i(x)$

Because

- WLS is OLS applied to a transformed regression model (transformed so that the errors are homoskedastic)
- None of the other assumptions affected

therefore OLS is “best linear unbiased”

Weighted Least Squares (Example)

Example: Suppose for all $i, j = 1, \dots, n$

- $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$
- $E(\epsilon_i \mid X_1, \dots, X_n) = 0$
- $E(\epsilon_i^2 \mid X_1, \dots, X_n) = \sigma_i^2 = \sigma^2 X_i^2$
- $E(\epsilon_i \epsilon_j \mid X_1, \dots, X_n) = 0$ for all $i \neq j$
- $\sum_{i=1}^n (X_i - \overline{X})^2 > 0$

Assume X_i is always positive

Weighted Least Squares (Example)

Weight each observation by $\dfrac{1}{\sqrt{\eta(X_i)}}$ (noisier observations get less weight)

Transformed regression is

$$\frac{Y_i}{X_i} = \beta_0 \frac{1}{X_i} + \beta_1 + \frac{\epsilon_i}{X_i}$$

Regress Y_i/X_i on a constant and $1/X_i$.

$$\hat{\beta}_1^{wls} = \overline{Y^*} - \hat{\beta}_0^{wls}\overline{X^*} \qquad \text{and} \qquad \hat{\beta}_0^{wls} = \frac{\sum_{i=1}^{n}(Y_i^* - \overline{Y^*})(X_i^* - \overline{X^*})}{\sum_{i=1}^{n}(X_i^* - \overline{X^*})^2}$$

Weighted Least Squares (Example)

For this example

- β_1 is intercept in transformed eq., but remains slope coefficient of original eq.
- β_0 is slope coef in transformed eq., but remains intercept term in the original eq.

You report

$$\hat{Y} = \hat{\beta}_0^{wls} + \hat{\beta}_1^{wls} X_i$$

(*you don't report transformed equation*)

Weighted Least Squares (Example)

WLS residuals are

$$\widehat{\epsilon}_i^{wls} = Y_i - \widehat{Y}_i^{wls} = Y_i - \widehat{\beta}_0^{wls} - \widehat{\beta}_1^{wls} X_i$$

For computing R^2 goodness-of-fit, you should use WLS residuals:

$$R^2_{wls} = 1 - \frac{\sum_{i=1}^{n} \widehat{\epsilon}^2_{i,wls}}{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}$$

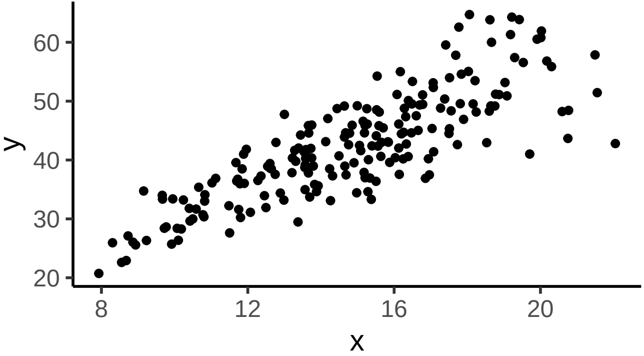
This R-squared will

- generally be less than the R-squared from OLS estimation (why?)
- and may even be negative (why?)

s.e. and test statistics *should* be based on the estimated transformed equation

Weighted Least Squares (Example)

Regression of y and x (with intercept) from data set **heterosk.csv**



Weighted Least Squares (Example)

Recall OLS estimates with default s.e.

```
             Estimate Std. Error  t value     Pr(>|t|)
(Intercept) 6.285792  1.7817576  3.52786 5.207066e-04
x           2.430744  0.1177712 20.63955 3.034896e-51
R-squared:  0.6826877
```

Obvious heteroskedasticity in this example (should not trust default s.e.)

OLS estimates with heteroskedasticity-robust s.e.

```
t test of coefficients:

             Estimate Std. Error t value  Pr(>|t|)
(Intercept)  6.28579    1.86393  3.3723 0.0008963 ***
x            2.43074    0.13624 17.8414 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Weighted Least Squares (Example)

Assume $Var(\epsilon_i \mid X_1, \ldots, X_n) = \sigma^2 X_i^2$ and run WLS

- fit appropriately transformed reg. eq. using OLS

```
df_het$ystar  <- df_het$y/df_het$x     # Transformed y
df_het$x0star <- 1/df_het$x            # Transformed intercept term
wls1 <- lm(ystar~x0star, data=df_het)
sum_wls1 <- summary(wls1)
coef(sum_wls1)                              # Print Coefficients
cat("R-squared: ",sum_wls1$r.squared,"\n")  # Print R-squared of Transformed Eq
```

```
            Estimate Std. Error   t value     Pr(>|t|)
(Intercept)  2.53166  0.1023702 24.730430 1.839271e-62
x0star       4.82571  1.4065451  3.430896 7.321779e-04
R-squared:  0.05611379
```

Weighted Least Squares (Example)

Estimated equation is

$$\begin{aligned} \hat{Y} = \underset{(1.4065)}{4.8257} + \underset{(0.1024)}{2.5317}\, X\,, \quad n = 200 \end{aligned}$$

- Make sure you put coefficient estimates in the right place
- Standard errors lower than in the OLS regression (not unexpected)
- R^2 in output above refers to the fit of the transformed equation (not very useful)

Weighted Least Squares (Example)

R^2 for the un-transformed regression is

```
ehat <- df_het$y - coef(wls1)[2] - coef(wls1)[1]*df_het$x
ssr <- sum(ehat^2)
sst <- sum((df_het$y - mean(df_het$y))^2)
R2 <- 1 - ssr/sst
cat("R-squared: ", R2,"\n")
```

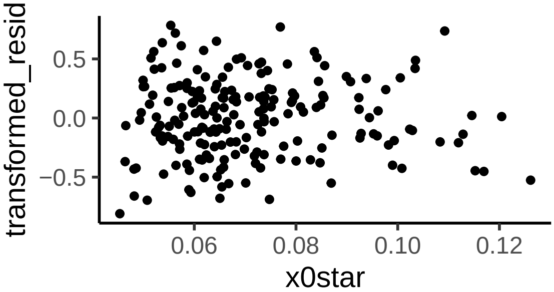
```
R-squared:  0.6814966
```

- R^2 from WLS regression is slightly lower than in the OLS regression
- Expected, since OLS minimizes SSR, equivalent to maximizing R^2

Weighted Least Squares (Example)

Finally, plot the residuals from the transformed regression against x_0^*:

- variance of the residuals of the weighted regression and x do not seem correlated
- assumption regarding the form of heteroskedasticity in ϵ_i appears reasonable



Weighted Least Squares (Example)

Can use `lm()` to carry out weighted least squares (option `weights` refer to weights on the squared residuals)

```
df_het$wt <- 1/df_het$x^2
wls2 <- lm(y~x,data=df_het, weights=wt)
sum_wls2 <- summary(wls2)
coef(sum_wls2)
cat("R-squared: ", sum_wls2$r.squared,"\n")
```

```
            Estimate Std. Error   t value     Pr(>|t|)
(Intercept)  4.82571  1.4065451  3.430896 7.321779e-04
x            2.53166  0.1023702 24.730430 1.839271e-62
R-squared:  0.755433
```

Weighted Least Squares (Example)

R^2 using `lm()` for WLS is different from what we previously obtained

"Weighted R-squared":

$$\text{weighted-}R^2 = \frac{\sum_{i=1}^{n} w_i(\hat{Y}_i^{wls} - \overline{Y}_{wls})^2}{\sum_{i=1}^{n} w_i(Y_i^{wls} - \overline{Y}_{wls})^2}\,.$$

where $\overline{Y}_{wls}$ is the weighted mean of $\{Y_i\}_{i=1}^{n}$, i.e., the WLS estimator of β_0 from the regression $Y_i = \beta_0 + \epsilon_i$ using same weights as in WLS estimator of the main equation

Weighted Least Squares

Weighted Least Squares for efficient estimation of heteroskedastic regression equations

- are usually more complicated than simple example given here
- require that the form of the heteroskedasticity be known or assumed
- the form of the heteroskedasticity may have parameters that need to be estimated

We leave WLS and the topic of heteroskedasticity for the moment, and return to discussion of *multiple* linear regression models

Session 3.4

Session 3.4 Multiple Linear Regression

- Intro to multiple linear regression and OLS estimation of the MLR model

Multiple Linear Regression

We start with the two-regressor multiple linear regression model

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \epsilon_i \,, \; i = 1, \dots, n$$

and consider estimation of the model by OLS

$$\hat{\beta}_0^{ols}, \hat{\beta}_1^{ols}, \hat{\beta}_2^{ols} = \underset{\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2}{\arg\min} \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i - \hat{\beta}_2 Z_i)^2$$

Extend to more regressors later

Multiple Linear Regression

Main result: if in population,

$$E(Y \mid X, Z) = \beta_0 + \beta_1 X + \beta_2 Z$$

or equivalently

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \epsilon\,, \;\; E(\epsilon \mid X, Z) = 0$$

and you have a representative iid sample $\{X_i, Y_i, Z_i\}_{i=1}^{n}$ from the population, then

$\widehat{\beta}_0^{ols}$, $\widehat{\beta}_1^{ols}$ and $\widehat{\beta}_2^{ols}$ are unbiased and consistent estimators for β_0, β_1 and β_2

We consider standard errors, hypotheses testing, etc. later

Multiple Linear Regression

- Predictive relationship involving more than one variable
- Allows for non-linearity in $E(Y \mid X)$, e.g.,
  - $E(Y \mid X) = \beta_0 + \beta_1 X \;\Rightarrow\; \frac{\partial}{\partial X} E(Y \mid X) = \beta_1$
  - $E(Y \mid X) = \beta_0 + \beta_1 X + \beta_2 X^2 \;\Rightarrow\; \frac{\partial}{\partial X} E(Y \mid X) = \beta_1 + 2\beta_2 X$
- Alleviates omitted variable problem for causal interpretation of coefficient
  - If $E(Y \mid X, Z) = \alpha_0 + \alpha_1 X + \alpha_2 Z$ where $\alpha_2 \neq 0$, and X and Z are correlated, then in $E(Y \mid X) = \gamma_0 + \gamma_1 X$, we have $\gamma_1 \neq \alpha_1$
    - $\hat{\beta}_1^{ols}$ in SLR $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ will be unbiased and consistent for γ_1
    - $\hat{\beta}_1^{ols}$ in MLR $Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \epsilon_i$ will be unbiased and consistent for α_1
- But, of course, must consider if there are yet other “confounding variables”

Multiple Linear Regression

- β_1 measures difference in $E(Y \mid X, Z)$ between two members of population with same $Z = z$ but different X

$$\text{Discrete: } \beta_1 = E(Y \mid X = x + 1, Z = z) - E(Y \mid X = x, Z = z)$$

$$\text{Continuous: } \beta_1 = \frac{\partial E(Y \mid X, Z)}{\partial X} \text{ which can be evaluated at any } X = x, Z = z$$

- β_1 and β_2 are called "partial regression coefficients"

OLS Estimation of the Multiple Linear Regression Model

Sample $\{Y_i, X_i, Z_i\}_{i=1}^n$

For any estimators $\widehat{\beta}_0$, $\widehat{\beta}_1$ and $\widehat{\beta}_2$ (whether or not obtained by OLS), define

- **Fitted values**: $\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_i + \widehat{\beta}_2 Z_i$
- **Residuals**: $\hat{\epsilon}_i = Y_i - \widehat{Y}_i = Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i - \widehat{\beta}_2 Z_i$

$$\begin{aligned}\widehat{\beta}_0^{ols}, \widehat{\beta}_1^{ols}, \widehat{\beta}_2^{ols} &= \text{argmin}_{\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2} SSR \\ &= \text{argmin}_{\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2} \sum_{i=1}^n (Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i - \widehat{\beta}_2 Z_i)^2\end{aligned}$$

OLS Estimation of the Multiple Linear Regression Model

OLS estimators can be found by solving FOC:

$$\left.\frac{\partial SSR}{\partial \hat{\beta}_0}\right|_{\hat{\beta}_0^{ols},\hat{\beta}_1^{ols},\hat{\beta}_2^{ols}} = -2\sum_{i=1}^{n}(Y_i - \hat{\beta}_0^{ols} - \hat{\beta}_1^{ols}X_i - \hat{\beta}_2^{ols}Z_i) = 0$$

$$\left.\frac{\partial SSR}{\partial \hat{\beta}_1}\right|_{\hat{\beta}_0^{ols},\hat{\beta}_1^{ols},\hat{\beta}_2^{ols}} = -2\sum_{i=1}^{n}(Y_i - \hat{\beta}_0^{ols} - \hat{\beta}_1^{ols}X_i - \hat{\beta}_2^{ols}Z_i)X_i = 0$$

$$\left.\frac{\partial SSR}{\partial \hat{\beta}_2}\right|_{\hat{\beta}_0^{ols},\hat{\beta}_1^{ols},\hat{\beta}_2^{ols}} = -2\sum_{i=1}^{n}(Y_i - \hat{\beta}_0^{ols} - \hat{\beta}_1^{ols}X_i - \hat{\beta}_2^{ols}Z_i)Z_i = 0$$

Can also write FOC as

$$\sum_{i=1}^{n}\hat{\epsilon}_i^{ols} = 0\,, \quad \sum_{i=1}^{n}\hat{\epsilon}_i^{ols}X_i = 0\,, \quad \text{and} \quad \sum_{i=1}^{n}\hat{\epsilon}_i^{ols}Z_i = 0\,.$$

OLS Estimation of the Multiple Linear Regression Model

Can solve FOC directly for $\hat{\beta}_0^{ols}, \hat{\beta}_1^{ols}, \hat{\beta}_2^{ols}$

We take an alternative indirect approach

- more illustrative approach
- highlight how including Z controls for its confounding effects

Focus on $\hat{\beta}_1^{ols}$

- can get the solution for $\hat{\beta}_2^{ols}$ by switching X_i with Z_i in the steps shown
- can obtain $\hat{\beta}_0^{ols}$ from $\hat{\beta}_0^{ols} = \overline{Y} - \hat{\beta}_1^{ols}\overline{X} - \hat{\beta}_1^{ols}\overline{Z}$

OLS Estimation of the Multiple Linear Regression Model

First recall (again!!)

- OLS estimation of SLR $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ breaks Y_i into two parts
$$Y_i = \widehat{Y}_i^{ols} + \widehat{\epsilon}_i^{ols}$$
where $\widehat{Y}_i^{ols} = \widehat{\beta}_0^{ols} + \widehat{\beta}_1^{ols} X_i$ is perfectly correlated with X_i and $\widehat{\epsilon}_i^{ols}$ is perfectly uncorrelated with X_i
- residuals $\widehat{\epsilon}_i^{ols}$ is that part of Y_i that is uncorrelated with X_i
- residuals $\widehat{\epsilon}_i^{ols}$ have sample mean equal to zero
- If $\overline{X} = 0$, then $\widehat{\beta}_1^{ols}$ takes the form $\widehat{\beta}_1^{ols} = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})Y_i}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \dfrac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}$

OLS Estimation of the Multiple Linear Regression Model

Solve for $\hat{\beta}_1^{ols}$ using the following "auxiliary" regressions:

- Regress X_i on Z_i, and collect residuals $r_{i,x|z}$

$$r_{i,x|z} = X_i - \hat{\delta}_0^{ols} - \hat{\delta}_1^{ols} Z_i \ , \ i = 1, 2, ..., n$$

  - $r_{i,x|z}$ contains movements in X_i that are perfectly uncorrelated with Z_i
  - $r_{i,x|z}$ has sample mean zero

- Regress Y_i on Z_i, and collect residuals $r_{i,y|z}$

$$r_{i,y|z} = Y_i - \hat{\alpha}_0^{ols} - \hat{\alpha}_1^{ols} Z_i \ , \ i = 1, 2, ..., n$$

  - $r_{i,y|z}$ contains movements in Y_i that are perfectly uncorrelated with Z_i
  - $r_{i,y|z}$ has sample mean zero

OLS Estimation of the Multiple Linear Regression Model

Consider regression of $r_{i,y|z}$ on $r_{i,x|z}$ using OLS

$$r_{i,y|z} = \gamma_0 + \gamma_1 r_{i,x|z} + u_i$$

we have

$$\hat{\gamma}_1^{ols} = \frac{r_{i,x|z} r_{i,y|z}}{r_{i,x|z}^2}$$

Consider numerator

$$\sum_{i=1}^{n} r_{i,x|z} r_{i,y|z} = \sum_{i=1}^{n} r_{i,x|z} (Y_i - \hat{\alpha}_0^{ols} - \hat{\alpha}_1^{ols} Z_i) = \sum_{i=1}^{n} r_{i,x|z} Y_i$$

OLS Estimation of the Multiple Linear Regression Model

If we had solved the MLR FOC for $\hat{\beta}_0^{ols}$, $\hat{\beta}_1^{ols}$ and $\hat{\beta}_2^{ols}$, we would have

$$Y_i = \hat{\beta}_0^{ols} + \hat{\beta}_1^{ols} X_i + \hat{\beta}_2^{ols} Z_i + \hat{\epsilon}_i^{ols}$$

Substitute this into $\sum_{i=1}^{n} r_{i,x|z} Y_i$ gives

$$\begin{aligned}\sum_{i=1}^{n} r_{i,x|z} r_{i,y|z} = \sum_{i=1}^{n} r_{i,x|z} Y_i &= \sum_{i=1}^{n} r_{i,x|z} (\hat{\beta}_0^{ols} + \hat{\beta}_1^{ols} X_i + \hat{\beta}_2^{ols} Z_i + \hat{\epsilon}_i) \\ &= \hat{\beta}_1^{ols} \sum_{i=1}^{n} r_{i,x|z}^2 + \sum_{i=1}^{n} r_{i,x|z} \hat{\epsilon}_i^{ols}\end{aligned}$$

OLS Estimation of the Multiple Linear Regression Model

Now we use MLR FOC. We have

$$\sum_{i=1}^{n} r_{i,x|z}\hat{\epsilon}_i^{ols} = \sum_{i=1}^{n} \hat{\epsilon}_i^{ols}(X_i - \hat{\delta}_0^{ols} - \hat{\delta}_1^{ols} Z_i) = \sum_{i=1}^{n} \hat{\epsilon}_i^{ols} X_i - \hat{\delta}_0 \sum_{i=1}^{n} \hat{\epsilon}_i^{ols} - \hat{\delta}_1 \sum_{i=1}^{n} \hat{\epsilon}_i^{ols} Z_i = 0$$

Therefore $\sum_{i=1}^{n} r_{i,x|z} r_{i,y|z} = \hat{\beta}_1^{ols} \sum_{i=1}^{n} r_{i,x|z}^2$ which gives

$$\hat{\gamma}_1^{ols} = \frac{\sum_{i=1}^{n} r_{i,x|z} r_{i,y|z}}{\sum_{i=1}^{n} r_{i,x|z}^2} = \hat{\beta}_1^{ols}$$

Likewise, we have $\hat{\beta}_2^{ols} = \dfrac{\sum_{i=1}^{n} r_{i,z|x} r_{i,y|z}}{\sum_{i=1}^{n} r_{i,z|x}^2}$

OLS Estimation of the Multiple Linear Regression Model

Derivation of $\hat{\beta}_1^{ols}$ shows how confounding factors are 'controlled' in a multiple regression analysis

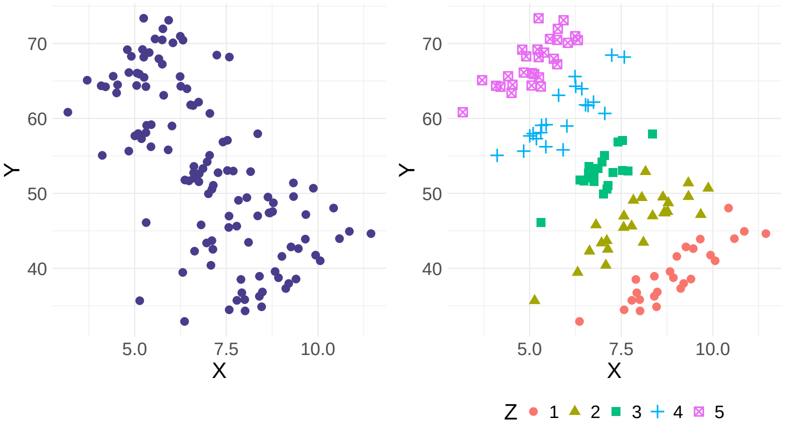
- Suppose we want to measure how Y_i is affected by X_i
- Suppose Z_i is an important determinant of Y_i that is correlated with X_i
- Omission of Z_i distorts measurement of the causal influence of X_i on Y_i
- Multiple regression
    - strips out all variation in Y_i and X_i that are correlated with Z_i
    - measures correlation in the remaining variation in Y_i and X_i

OLS Estimation of the Multiple Linear Regression Model

Example using data in **multireg_eg.csv**

```
df <- read_csv(
   ".\\data\\multireg_eg.csv",
   col_types = c("n","n","n"))
head(df, 8) # first 8 obs.
```

```
# A tibble: 8 x 3
      Z     X     Y
  <dbl> <dbl> <dbl>
1     1 10.1   41.0
2     2  7.11  43.7
3     1  9.20  38.0
4     5  6.04  70.1
5     2  9.67  47.2
6     4  7.58  68.2
7     2  8.11  43.5
8     3  7.11  50.6
```



OLS Estimation of the Multiple Linear Regression Model

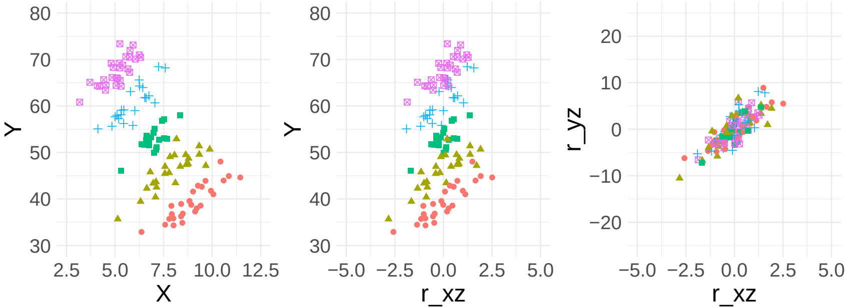
```
mdl1 <- lm(Y~X, data=df)
coef(summary(mdl1)); cat("R-squared:", summary(mdl1)$r.squared,"\n\n")
```

```
             Estimate Std. Error   t value     Pr(>|t|)
(Intercept) 82.892162  3.2234151 25.715634 3.394337e-50
X           -4.237024  0.4480503 -9.456581 3.882573e-16
R-squared: 0.431125
```

```
mdl2 <- lm(Y~X+Z, data=df)
coef(summary(mdl2)); cat("R-squared:", summary(mdl2)$r.squared,"\n\n")
```

```
             Estimate Std. Error    t value     Pr(>|t|)
(Intercept)  1.133354  2.0745669  0.5463086 5.858941e-01
X            3.122025  0.2048571 15.2400162 1.473978e-29
Z           10.110311  0.2369353 42.6711993 3.639969e-73
R-squared: 0.9656532
```

OLS Estimation of the Multiple Linear Regression Model



OLS Estimation of the Multiple Linear Regression Model

Another perspective: If

$$\hat{Y}_i^{ols} = \hat{\beta}_{01}^{ols} + \hat{\beta}_{yx}^{ols} X_i$$

$$\hat{Y}_i^{ols} = \hat{\beta}_{02}^{ols} + \hat{\beta}_{yz}^{ols} Z_i$$

$$\hat{X}_i^{ols} = \hat{\beta}_{03}^{ols} + \hat{\beta}_{xz}^{ols} Z_i$$

$$\hat{Z}_i^{ols} = \hat{\beta}_{04}^{ols} + \hat{\beta}_{zx}^{ols} X_i$$

then $\hat{\beta}_1^{ols} = \dfrac{\hat{\beta}_{yx}^{ols} - \hat{\beta}_{yz}^{ols}\hat{\beta}_{zx}^{ols}}{1 - \hat{\beta}_{xz}^{ols}\hat{\beta}_{zx}^{ols}}$

(proof omitted)

```
# Three different ways of
# calculating \hat\beta_1^{ols}

b_yx <- lm(Y~X, dat=df)$coef[2]
b_yz <- lm(Y~Z, dat=df)$coef[2]
b_xz <- lm(X~Z, dat=df)$coef[2]
b_zx <- lm(Z~X, dat=df)$coef[2]

b_1a <- lm(Y~X+Z, dat=df)$coef[2]
b_1b <- lm(r_yz~r_xz, dat=df)$coef[2]
b_1c <- (b_yx - b_yz*b_zx)/(1-b_xz*b_zx)

alt_b1s <- c(b_1a, b_1b, b_1c)
names(alt_b1s) <- c("Mtd 1", "Mtd 2", "Mtd 3")
alt_b1s
```

```
   Mtd 1    Mtd 2    Mtd 3
3.122025 3.122025 3.122025
```

OLS Estimation of the Multiple Linear Regression Model

Note: to derive $\hat{\beta}_1^{ols} = \dfrac{\sum_{i=1}^n r_{i,x|z} r_{i,y|z}}{\sum_{i=1}^n r_{i,x|z}^2}$, we must assume

- $\sum_{i=1}^n r_{i,x|z}^2 \neq 0$
- i.e., X_i and Z_i cannot be *perfectly* correlated (positively or negatively)

Also, to run auxiliary regression of X_i on Z_i in the first place

- There must be variation in Z_i
- Z_i cannot be equal to some constant value c for all i

Similarly, to derive $\hat{\beta}_2^{ols}$, we require variation in X_i

OLS Estimation of the Multiple Linear Regression Model

We summarize these requirements by saying:

$$c_1 + c_2 X_i + c_3 Z_i = 0 \text{ for all } i = 1, 2, \ldots, n \text{ iff } (c_1, c_2, c_3) = (0, 0, 0)$$

Examples of when this condition does not hold:

- If $X_i = c$ for all i, $(-c) + (1)X_i + (0)Z_i = -c + c + 0 = 0$
- If $Z_i = c$ for all i, $(-c) + (0)X_i + (1)Z_i = -c + 0 - c = 0$
- If X_i and Z_i are perfectly correlated, i.e., $X_i = \gamma_0 + \gamma_1 Z_i$ for all i, then

$$\gamma_0 + (-1)X_i + \gamma_1 Z_i = 0$$

In any of these cases, we cannot regress Y_i on X_i and Z_i

Properties of OLS Estimators (MLR)

Many of the algebraic properties carry over from the simple linear regression model

- FOC can be written as

$$\sum_{i=1}^{n} \hat{\epsilon}_i^{ols} = 0, \quad \sum_{i=1}^{n} X_i \hat{\epsilon}_i^{ols} = 0 \quad \text{and} \quad \sum_{i=1}^{n} Z_i \hat{\epsilon}_i^{ols} = 0$$

- Fitted values $\widehat{Y}_i^{ols}$ and residuals $\hat{\epsilon}_i^{ols}$ are also uncorrelated
- The fact that $\hat{\beta}_0^{ols} = \overline{Y} - \hat{\beta}_1^{ols}\overline{X} - \hat{\beta}_2^{ols}\overline{Z}$ means that the point $(\overline{X}, \overline{Y}, \overline{Z})$ lies on the sample regression function

Properties of OLS Estimators (MLR)

- $\overline{Y} = \overline{\hat{Y}^{ols}}$
- SST = SSE + SSR equality continues to hold in the multiple regression case

$$\begin{aligned}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 &= \sum_{i=1}^{n}\left(\hat{Y}_i - \overline{\hat{Y}^{ols}}\right)^2 + \sum_{i=1}^{n}\hat{\epsilon}_{i,ols}^2 \\ &= \sum_{i=1}^{n}\left(\hat{Y}_i - \overline{Y}\right)^2 + \sum_{i=1}^{n}\hat{\epsilon}_{i,ols}^2\end{aligned}$$

- We can use this to define the goodness-of-fit measure:

$$R^2 = 1 - \frac{SSR}{SST}$$

Properties of OLS Estimators (MLR)

- Let R_1^2 be R^2 from $Y_i = \widehat{\beta}_0^{ols} + \widehat{\beta}_1^{ols} X_i + \widehat{\epsilon}_i^{ols}$
- Let R_2^2 be R^2 from $Y_i = \widehat{\beta}_0^{ols} + \widehat{\beta}_1^{ols} X_i + \widehat{\beta}_2^{ols} Z_i + \widehat{\epsilon}_i^{ols}$
- We have $R_2^2 \geq R_1^2$
  - R^2 will never decrease as we add more variables to the regression
  - OLS minimizes SSR, and therefore maximizes R^2
- "Adjusted R^2" is sometimes used instead:

$$\text{Adj.-}R^2 = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)} = 1 - \frac{SSR}{SST}\frac{n-1}{n-k-1}$$

where k is no. of slope coefficients ($k = 2$ for the 2-regressor case)

Properties of OLS Estimators (MLR)

- Since

$$\hat{\beta}_1^{ols} = \frac{\sum_{i=1}^{n} r_{i,x|z} r_{i,y|z}}{\sum_{i=1}^{n} r_{i,x|z}^2} = \frac{\sum_{i=1}^{n} r_{i,x|z}(Y_i - \hat{\alpha}_0^{ols} - \hat{\alpha}_1^{ols} Z_i)}{\sum_{i=1}^{n} r_{i,x|z}^2} = \frac{\sum_{i=1}^{n} r_{i,x|z} Y_i}{\sum_{i=1}^{n} r_{i,x|z}^2}$$

- $\hat{\beta}_1^{ols}$ is a linear estimator, i.e.,

$$\hat{\beta}_1^{ols} = \sum_{i=1}^{n} w_i Y_i$$

where $w_i = \frac{r_{i,x|z}}{\sum_{i=1}^{n} r_{i,x|z}^2}$ involve only the regressors $\{X_i\}_{i=1}^{n}$ and $\{Z_i\}_{i=1}^{n}$

Properties of OLS Estimators (MLR)

- The weights have the following properties:

$$\sum_{i=1}^{n} w_i = 0$$

$$\sum_{i=1}^{n} w_i Z_i = \frac{\sum_{i=1}^{n} r_{i,x|z} Z_i}{\sum_{i=1}^{n} r_{i,x|z}^2} = 0$$

$$\sum_{i=1}^{n} w_i X_i = \frac{\sum_{i=1}^{n} r_{i,x|z} X_i}{\sum_{i=1}^{n} r_{i,x|z}^2} = 1$$

$$\sum_{i=1}^{n} w_i^2 = \frac{\sum_{i=1}^{n} r_{i,x|z}^2}{(\sum_{i=1}^{n} r_{i,x|z}^2)^2} = \frac{1}{\sum_{i=1}^{n} r_{i,x|z}^2}$$

Properties of OLS Estimators (MLR)

- If sample covariance of X_i and Z_i is zero, then
  - $\hat{\delta}_1^{ols}$ in auxiliary regression of X on Z is zero
  - $\hat{\delta}_0^{ols}$ is equal to $\overline{X}$
  - we have $r_{i,x|z} = X_i - \overline{X}$

Therefore, if sample covariance of X_i and Z_i is zero,

$$\hat{\beta}_1^{ols} = \frac{\sum_{i=1}^{n} r_{i,x|z} Y_i}{\sum_{i=1}^{n} r_{i,x|z}^2} = \frac{\sum_{i=1}^{n} (X_i - \overline{X}) Y_i}{\sum_{i=1}^{n} (X_i - \overline{X})^2}$$

$\hat{\beta}_1^{ols}$ in the multiple linear regression reduces to the OLS estimator for the X_i coefficient in the *simple* linear regression of Y on X

Properties of OLS Estimators (MLR)

Suppose that in population:

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \epsilon\,,\; E(\epsilon \mid X, Z) = 0\,,\; Var(\epsilon \mid X, Z) = \sigma^2$$

Suppose you have a representative iid sample $\{X_i, Y_i, Z_i\}_{i=1}^{n}$ from the population, and

$$c_1 + c_2 X_i + c_3 Z_i = 0 \text{ for all } i = 1, \dots, n \;\Leftrightarrow\; (c_1, c_2, c_3) = (0, 0, 0)$$

Properties of OLS Estimators (MLR)

These assumptions imply that

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \epsilon_i \,,\;\; i = 1, \ldots, n$$

such that

- $E(\epsilon_i \mid x, z) = 0$ for all $i = 1, ..., n,$
- $E(\epsilon_i^2 \mid x, z) = \sigma^2$ for all $i = 1, ..., n,$
- $E(\epsilon_i \epsilon_j \mid x, z) = 0$ for all $i \neq j,\; i, j = 1, ..., n$

where x denotes $\{X_1, X_2, ..., X_n\}$ and z denotes $\{Z_1, Z_2, ..., Z_n\}$

Properties of OLS Estimators (MLR)

OLS is unbiased, since

$$\hat{\beta}_1^{ols} = \sum_{i=1}^{n} w_i Y_i = \sum_{i=1}^{n} w_i(\beta_0 + \beta_1 X_i + \beta_2 Z_i + \epsilon_i) = \beta_1 + \sum_{i=1}^{n} w_i \epsilon_i$$

Taking conditional expectations gives

$$E(\hat{\beta}_1^{ols} \mid x, z) = \beta_1 + \sum_{i=1}^{n} w_i E(\epsilon_i \mid x, z) = \beta_1$$

It follows that the unconditional mean is $E(\hat{\beta}_1) = \beta_1$

Properties of OLS Estimators (MLR)

The conditional variance of $\hat{\beta}_1^{ols}$ under Assumption Set B is

$$Var(\hat{\beta}_1^{ols} \mid x, z) = Var\left(\beta_1 + \sum_{i=1}^{n} w_i \epsilon_i \;\middle|\; x, z\right) = \sum_{i=1}^{n} w_i^2 \, Var(\epsilon_i \mid x, z)$$
$$= \frac{\sigma^2}{\sum_{i=1}^{n} r_{i,x|z}^2}$$

Since the R^2 from the regression of X_i on Z_i is $R^2_{x|z} = 1 - \frac{\sum_{i=1}^{n} r_{i,x|z}^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$, we have

$$Var(\hat{\beta}_1^{ols} \mid x, z) = \frac{\sigma^2}{(1 - R^2_{x|z}) \sum_{i=1}^{n}(X_i - \overline{X})^2}$$

Properties of OLS Estimators (MLR)

These expressions show trade-offs involved in adding a second regressor:

Suppose the true data generating process is

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \epsilon\,,\;\; E(\epsilon \mid X, Z) = 0\,,\;\; Var(\epsilon \mid X, Z) = \sigma^2$$

but you write down the model

$$Y = \beta_0 + \beta_1 X + u$$

If X and Z are correlated, then

- X and u are correlated
- Regression of Y_i on X_i will give biased estimator for β_1

Properties of OLS Estimators (MLR)

By estimating the multiple linear regression

- You get an unbiased estimate of β_1 by controlling for Z **but**
- variance of the OLS estimator for β_1 changes from

$$Var(\widehat{\beta}_1^{ols} \mid x) = \frac{\sigma_u^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

to

$$Var(\widehat{\beta}_1^{ols} \mid x, z) = \frac{\sigma^2}{(1 - R_{x|z}^2)\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

Properties of OLS Estimators (MLR)

We have

- $\sigma^2 < \sigma_u^2$, reduces estimator variance (good!)
- However, denominator in the variance expression is smaller in the multiple linear regression case
- *reduced effective variation* in X
- which in turn increases the estimator variance (bad!)

Properties of OLS Estimators (MLR)

An unbiased estimator for σ^2 in the two-regressor case is

$$\widehat{\sigma^2} = \frac{1}{n-3}\sum_{i=1}^{n}\hat{\epsilon}^2_{i,ols}.$$

The SSR is divided by $n-3$ because

- three 'degrees-of-freedom' were used in computing $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$
- these are used in the computation of $\hat{\epsilon}_i$

Estimate $Var(\hat{\beta}_1^{ols} \mid x, z)$ using

$$\widehat{Var}(\hat{\beta}_1^{ols} \mid x, z) = \frac{\widehat{\sigma^2}}{(1-R^2_{x|z})\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

Roadmap

- (Previous) Session 1: Statistics Review
- (Previous) Session 2: Simple Linear Regression
- **This Session 3: Estimator Standard Errors; Multiple Linear Regression**
- *Next Session 4: Matrix Algebra*
- Session 5: OLS using Matrix Algebra
- Session 6: Hypothesis Testing
- Session 7: Prediction
- Session 8: Instrumental Variable Regression
- Session 9: Logistic and Other Regressions
- Session 10: Panel Data Regressions
- Session 11: Introduction to Time Series
- Session 12: Time Series Regressions