Advanced Econometrics

Jean-Paul Renne

2022-08-05

2

Contents

Chapter 1

Prerequisites

This is a *sample* book written in **Markdown**. You can use anything that Pandoc's Markdown supports, e.g., a math equation $a^2 + b^2 = c^2$.

The **bookdown** package can be installed from CRAN or Github:

```
install.packages("bookdown")
# or the development version
# devtools::install_github("rstudio/bookdown")
```

Remember each Rmd file contains one and only one chapter, and a chapter is defined by the first-level heading `#`.

To compile this example to PDF, you need XeLaTeX. You are recommended to install TinyTeX (which includes XeLaTeX): https://yihui.name/tinytex/.

Chapter 2

Introduction

You can label chapter and section titles using `{#label}` after them, e.g., we can reference Chapter 2. If you do not manually label them, there will be automatic labels anyway, e.g., Chapter 12.

Figures and tables with captions will be placed in `figure` and `table` environments, respectively.

```r
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19)
```

Reference a figure by its code chunk label with the `fig:` prefix, e.g., see Figure 2.1. Similarly, you can reference tables generated from `knitr::kable()`, e.g., see Table 2.1.

```r
knitr::kable(
  head(iris, 20), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

You can write citations, too. For example, we are using the **bookdown** package (Xie, 2022) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015).

Figure 2.1: Here is a nice figure!

Table 2.1: Here is a nice table!

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa

Chapter 3

Specification and standard assumptions

3.1 Specification

Definition 3.1. A linear regression is a model defined through:

$$y_i = \beta' \mathbf{x}_i + \varepsilon_i,$$

where $\mathbf{x}_i = [x_{i,1}, \dots, x_{i,K}]'$ is a vector of dimension $K \times 1$.

For entity i, the $x_{i,k}$'s, for $k \in \{1, \dots, K\}$, are explanatory variables. If one wants to have an intercept in the specification, then set $x_{i,1} = 1$ for all i, and β_1 then corresponds to the intercept.

Hypothesis 3.1 (Full rank)**.** There is no exact linear relationship among the independent variables (the $x_{i,k}$s, for a given $i \in \{1, \dots, n\}$).

Hypothesis 3.2 (Conditional mean-zero assumption)**.**

$$\mathbb{E}(\varepsilon|\mathbf{X}) = 0. \tag{3.1}$$

Note that, in Hypothesis 3.2, ε is a n-dimensional vector (where n is the sample size), and $\mathbf{X}$ is the matrix containing all explanatory variables, of dimension $n \times K$.

Proposition 3.1. *Under Hypothesis 3.2:*

    i. $\mathbb{E}(\varepsilon_i) = 0$*;*

    ii. the x_{ij}s and the ε_is are uncorrelated, i.e. $\forall i, j \quad \mathbb{C}orr(x_{ij}, \varepsilon_i) = 0$.

Proof. Let us prove (i) and (ii):

    i. By the law of iterated expectations:

$$\mathbb{E}(\varepsilon) = \mathbb{E}(\mathbb{E}(\varepsilon|\mathbf{X})) = \mathbb{E}(0) = 0.$$

    ii. $\mathbb{E}(x_{ij}\varepsilon_i) = \mathbb{E}(\mathbb{E}(x_{ij}\varepsilon_i|\mathbf{X})) = \mathbb{E}(x_{ij}\underbrace{\mathbb{E}(\varepsilon_i|\mathbf{X})}_{=0})) = 0.\square$

$\square$

Hypothesis 3.3 (Homoskedasticity)**.**

$$\forall i, \quad \mathbb{V}ar(\varepsilon_i|\mathbf{X}) = \sigma^2.$$

Panel (b) of Figure 3.1 corresponds to a situation of heteroskedasticity. Let us be more specific. In the two plots, we have $X_i \sim \mathcal{N}(0,1)$ and $\varepsilon_i^* \sim \mathcal{N}(0,1)$. In Panel (a) (homoskedasticity): $Y_i = 2 + 2X_i + \varepsilon_i^*$. In Panel (b) (heteroskedasticity): $Y_i = 2 + 2X_i + \left(2\mathbb{1}_{\{X_i<0\}} + 0.2\mathbb{1}_{\{X_i\geq0\}}\right)\varepsilon_i^*$.

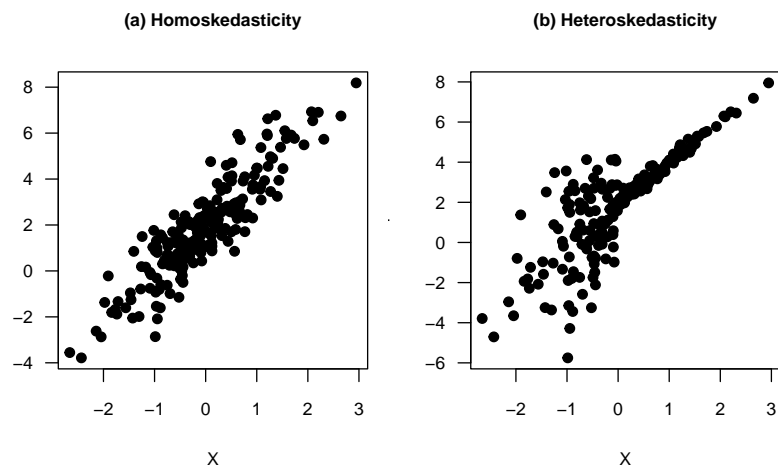**(a) Homoskedasticity**   **(b) Heteroskedasticity**

Figure 3.1: This is the caption

```r
# load data into R
data(Salaries, package = "carData")
# six first rows of the data
head(Salaries)
```

```
##           rank discipline yrs.since.phd yrs.service  sex salary
## 1        Prof          B            19          18 Male 139750
## 2        Prof          B            20          16 Male 173200
## 3    AsstProf          B             4           3 Male  79750
## 4        Prof          B            45          39 Male 115000
## 5        Prof          B            40          41 Male 141500
## 6   AssocProf          B             6           6 Male  97000
```

```r
# Regression:
eq <- lm(salary~.,data=Salaries)
summary(eq)
```

```
##
## Call:
## lm(formula = salary ~ ., data = Salaries)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -65248 -13211  -1775  10384  99592
##
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   65955.2     4588.6  14.374  < 2e-16 ***
## rankAssocProf 12907.6     4145.3   3.114  0.00198 **
## rankProf      45066.0     4237.5  10.635  < 2e-16 ***
## disciplineB   14417.6     2342.9   6.154 1.88e-09 ***
## yrs.since.phd   535.1      241.0   2.220  0.02698 *
## yrs.service    -489.5      211.9  -2.310  0.02143 *
## sexMale        4783.5     3858.7   1.240  0.21584
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22540 on 390 degrees of freedom
## Multiple R-squared:  0.4547, Adjusted R-squared:  0.4463
## F-statistic:  54.2 on 6 and 390 DF,  p-value: < 2.2e-16
```

```
par(mfrow=c(1,1))
par(plt=c(.2,.95,.2,.95))
plot(salary/1000~yrs.since.phd,pch=19,xlab="years since PhD",ylab="Salary",data=Salari
abline(lm(salary/1000~yrs.since.phd,data=Salaries),col="red",lwd=2)
```
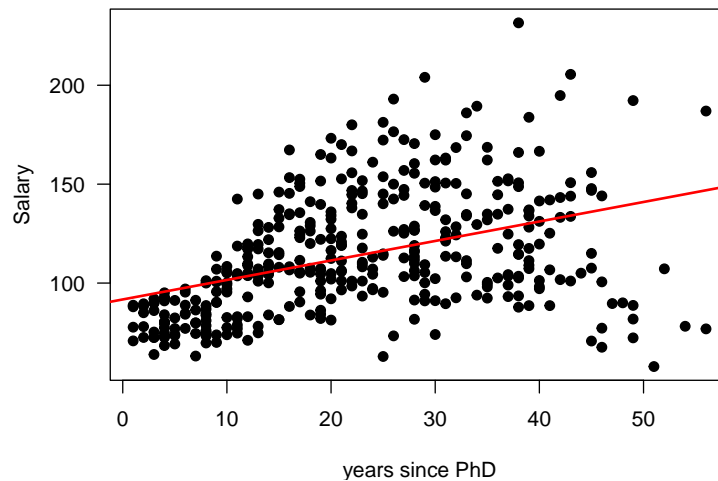

Figure 3.2: Salary versus years after PhD

Hypothesis 3.4 (Uncorrelated residuals)**.**

$$\forall i \neq j, \quad \mathbb{C}ov(\varepsilon_i, \varepsilon_j | \mathbf{X}) = 0.$$

Proposition 3.2. *If 3.3 and 3.4 hold, then:*

$$\mathbb{V}ar(\varepsilon|\mathbf{X}) = \sigma^2 Id,$$

where Id is the $n \times n$ identity matrix.

Hypothesis 3.5 (Normal distribution)**.**

$$\forall i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2).$$

3.2 Least square estimation

For a given vector of coefficients $\mathbf{b} = [b_1, \dots, b_K]'$, the sum of square residuals is:

$$f(\mathbf{b}) = \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{K} x_{i,j} b_j \right)^2 = \sum_{i=1}^{n} (y_i - \mathbf{x}_i' \mathbf{b})^2.$$

Minimizing the sum of squared residuals amounts to minimizing:

$$f(\mathbf{b}) = (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b}).$$

We have:

$$\frac{\partial f}{\partial \mathbf{b}}(\mathbf{b}) = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\mathbf{b}.$$

Necessary first-order condition (FOC):

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}. \qquad (3.2)$$

Under Assumption 3.1, $\mathbf{X}'\mathbf{X}$ is invertible. Hence:

$$\boxed{\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.}$$

Vector $\mathbf{b}$ minimises the sum of squared residuals. (f is a non-negative quadratic function, it admits a minimum.)

The estimated residuals are:

$$\mathbf{e} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{M}\mathbf{y} \tag{3.3}$$

where $\mathbf{M} := \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is called the **residual maker** matrix. Let us further define a **projection matrix** by $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. These matrices $\mathbf{M}$ and $\mathbf{P}$ are such that:

- $\mathbf{M}\mathbf{X} = \mathbf{0}$: if one regresses one of the explanatory variables on $\mathbf{X}$, the residuals are null.
- $\mathbf{M}\mathbf{y} = \mathbf{M}\varepsilon$ (because $\mathbf{y} = \mathbf{X}\beta + \varepsilon$ and $\mathbf{M}\mathbf{X} = \mathbf{0}$).
- The fitted values are:

$$\hat{\mathbf{y}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{P}\mathbf{y}, \tag{3.4}$$

  i.e., $\hat{\mathbf{y}}$ is the projection of the vector $\mathbf{y}$ onto the vectorial space spanned by the columns of $\mathbf{X}$.
- It can be shown that each column $\tilde{\mathbf{x}}_k$ of $\mathbf{X}$ is orthogonal to $\mathbf{e}$. $\Rightarrow$ If intercepts are included in the regression ($x_{i,1} \equiv 1$), the average of the residuals is null.

Here are some properties of $\mathbf{M}$ and $\mathbf{P}$:

- $\mathbf{M}$ is symmetric ($\mathbf{M} = \mathbf{M}'$) and **idempotent** ($\mathbf{M} = \mathbf{M}^2 = \mathbf{M}^k$ for $k > 0$).
- $\mathbf{P}$ is symmetric and idempotent.
- $\mathbf{P}\mathbf{X} = \mathbf{X}$.
- $\mathbf{P}\mathbf{M} = \mathbf{M}\mathbf{P} = 0$.
- $\mathbf{y} = \mathbf{P}\mathbf{y} + \mathbf{M}\mathbf{y}$ (decomposition of $\mathbf{y}$ into two orthogonal parts).

Proposition 3.3 (Properties of the OLS estimator)**.** *We have:*

i. *Under Assumptions 3.1 and 3.2, the OLS estimator is linear and unbiased.*

ii. *Under Hypotheses 3.1 to 3.4, the conditional covariance matrix of* $\mathbf{b}$ *is:*
$\mathbb{V}ar(\mathbf{b}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.

Proof. Under Hypothesis 3.1, $\mathbf{X}'\mathbf{X}$ can be inverted. We have:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \;.$$

   i. Let us consider the expectation of the last term, i.e. $\mathbb{E}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$. Using the law of iterated expectations, we obtain:

$$\mathbb{E}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = \mathbb{E}(\mathbb{E}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'|\mathbf{X}]) = \mathbb{E}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{E}[|\mathbf{X}]).$$

   By Hypothesis 3.2, we have $\mathbb{E}[|\mathbf{X}] = 0$. Hence $\mathbb{E}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = 0$ and result (i) follows.

  ii. $\mathbb{V}ar(\mathbf{b}|\mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{E}(\varepsilon\varepsilon'|\mathbf{X})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$. By Prop. 3.2, if 3.3 and 3.4 hold, then we have $\mathbb{E}(\varepsilon\varepsilon'|\mathbf{X}) = \mathbb{V}ar(\varepsilon|\mathbf{X}) = \sigma^2 Id$. The result follows. $\square$

$\square$

3.3 Bivariate case

Consider a bivariate situation, where we regress y_i on a constant and an explanatory variable w_i. We have $K = 2$, and $\mathbf{X}$ is a $n \times 2$ matrix whose i^{th} row is $[x_{i,1}, x_{i,2}]$, with $x_{i,1} = 1$ (to account for the intercept) and with $w_i = x_{i,2}$ (say).

We have:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} n & \sum_i w_i \\ \sum_i w_i & \sum_i w_i^2 \end{bmatrix},$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{n\sum_i w_i^2 - (\sum_i w_i)^2} \begin{bmatrix} \sum_i w_i^2 & -\sum_i w_i \\ -\sum_i w_i & n \end{bmatrix},$$

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \frac{1}{n\sum_i w_i^2 - (\sum_i w_i)^2} \begin{bmatrix} \sum_i w_i^2 \sum_i y_i - \sum_i w_i \sum_i w_i y_i \\ -\sum_i w_i \sum_i y_i + n \sum_i w_i y_i \end{bmatrix}$$

$$= \frac{1}{\frac{1}{n}\sum_i (w_i - \bar{w})^2} \begin{bmatrix} \frac{\bar{y}}{n}\sum_i w_i^2 - \frac{\bar{w}}{n}\sum_i w_i y_i \\ \frac{1}{n}\sum_i (w_i - \bar{w})(y_i - \bar{y}) \end{bmatrix}.$$

It can be seen that the second element of $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is:

$$b_2 = \frac{\overline{\mathbb{C}ov(W,Y)}}{\overline{\mathbb{V}ar(W)}},$$

where $\overline{\mathbb{C}ov(W,Y)}$ and $\overline{\mathbb{V}ar(W)}$ are sample estimates.

Since there is a constant in the regression, we have $b_1 = \bar{y} - b_2\bar{w}$.

3.4 Gauss Markow Theorem

Theorem 3.1 (Gauss-Markov Theorem)**.** *Under Assumptions 3.1 to 3.4, for any vector w, the minimum-variance linear unbiased estimator of $w'\beta$ is $w'\mathbf{b}$, where $\mathbf{b}$ is the least squares estimator. (BLUE: Best Linear Unbiased Estimator.)*

Proof. Consider $\mathbf{b}^* = C\mathbf{y}$, another linear unbiased estimator of β. Since it is unbiased, we must have $\mathbb{E}(C\mathbf{y}|\mathbf{X}) = \mathbb{E}(C\mathbf{X}\beta + C\varepsilon|\mathbf{X}) = \beta$. We have $\mathbb{E}(C\varepsilon|\mathbf{X}) = C\mathbb{E}(\varepsilon|\mathbf{X}) = 0$ (by 3.2).

Therefore $\mathbf{b}^*$ is unbiased if $\mathbb{E}(C\mathbf{X})\beta = \beta$. This has to be the case for any β, which implies that we must have $C\mathbf{X} = \mathbf{I}.\backslash$

Let us compute $\mathbb{V}ar(\mathbf{b}^*|\mathbf{X})$. For this, we introduce $D = C - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, which is such that $D\mathbf{y} = \mathbf{b}^* - \mathbf{b}$. The fact that $C\mathbf{X} = \mathbf{I}$ implies that $D\mathbf{X} = \mathbf{0}$.

We have $\mathbb{V}ar(\mathbf{b}^*|\mathbf{X}) = \mathbb{V}ar(C\mathbf{y}|\mathbf{X}) = \mathbb{V}ar(C\varepsilon|\mathbf{X}) = \sigma^2 CC'$ (by Assumptions 3.3 and 3.4, see Prop. 3.2). Using $C = D + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and exploiting the fact that $D\mathbf{X} = \mathbf{0}$ leads to:

$$\mathbb{V}ar(\mathbf{b}^*|\mathbf{X}) = \sigma^2 \left[(D + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')(D + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')'\right] = \mathbb{V}ar(\mathbf{b}|\mathbf{X}) + \sigma^2 DD'.$$

Therefore, we have $\mathbb{V}ar(w'\mathbf{b}^*|\mathbf{X}) = w'\mathbb{V}ar(\mathbf{b}|\mathbf{X})w + \sigma^2 w'DD'w \geq w'\mathbb{V}ar(\mathbf{b}|\mathbf{X})w = \mathbb{V}ar(w'\mathbf{b}|\mathbf{X})$. $\square$ $\square$

3.5 Frish-Waugh

Consider the linear least square regression of $\mathbf{y}$ on $\mathbf{X}$. We introduce the notations:

- $\mathbf{b}^{\mathbf{y}/\mathbf{X}}$: OLS estimates of β,
- $\mathbf{M}^{\mathbf{X}}$: residual-maker matrix of any regression on $\mathbf{X}$,
- $\mathbf{P}^{\mathbf{X}}$: projection matrix of any regression on $\mathbf{X}$.

Consider the case where we have two sets of explanatory variables: $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$. With obvious notations: $\mathbf{b}^{\mathbf{y}/\mathbf{X}} = [\mathbf{b}_1', \mathbf{b}_2']'$.

Theorem 3.2 (Frisch-Waugh Theorem)**.** *We have:*

$$\mathbf{b}_2 = \mathbf{b}^{\mathbf{M}^{\mathbf{X}_1}\mathbf{y}/\mathbf{M}^{\mathbf{X}_1}\mathbf{X}_2}.$$

Proof. The minimization of the least squares leads to (these are first-order conditions, see Eq. (3.2)):

$$
\begin{bmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1'\mathbf{y} \\ \mathbf{X}_2'\mathbf{y} \end{bmatrix}.
$$

Use the first-row block of equations to solve for $\mathbf{b}_1$ first; it comes as a function of $\mathbf{b}_2$. Then use the second set of equations to solve for $\mathbf{b}_2$, which leads to:

$$
\mathbf{b}_2 = [\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)\mathbf{X}_1'\mathbf{X}_2]^{-1}\mathbf{X}_2'(Id - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)\mathbf{X}_1')\mathbf{y} = [\mathbf{X}_2'\mathbf{M}^{\mathbf{X}_1}\mathbf{X}_2]^{-1}\mathbf{X}_2'\mathbf{M}^{\mathbf{X}_1}\mathbf{y}.
$$

Using the fact that $\mathbf{M}^{\mathbf{X}_1}$ is idempotent and symmetric leads to the result. $\square$
$\square$

This suggests a second way of estimating $\mathbf{b}_2$:

1. Regress Y on X_1, regress X_2 on X_1.
2. Regress the former residuals on the latter.

```
Data <- read.csv("https://raw.githubusercontent.com/jrenne/Data4courses/master/parapluie/data4par
dummies <- as.matrix(Data[,4:14])
eq_all <- lm(parapluie~precip+dummies,data=Data)
summary(eq_all)$coefficients
```

```
##                Estimate Std. Error    t value      Pr(>|t|)
## (Intercept)  -2.9534478 3.40601605 -0.8671268 0.3893855798
## precip        0.1300055 0.03594492  3.6167985 0.0006192876
## dummiesX1    -8.5990682 3.30046623 -2.6054101 0.0115966151
## dummiesX2   -13.3290386 3.30657128 -4.0310755 0.0001613897
## dummiesX3    -7.9829582 3.25949898 -2.4491366 0.0173091333
## dummiesX4    -3.3923533 3.27605582 -1.0354992 0.3046614769
## dummiesX5    -3.7038158 3.25710094 -1.1371511 0.2600724511
## dummiesX6    -3.3606412 3.27334327 -1.0266694 0.3087670632
## dummiesX7    -7.3158812 3.28491529 -2.2271141 0.0297682836
## dummiesX8    -7.7172773 3.26128164 -2.3663327 0.0212677987
## dummiesX9    -4.6491997 3.26005024 -1.4261129 0.1591057652
## dummiesX10   -5.1091987 3.25143961 -1.5713651 0.1214457733
## dummiesX11    1.9807700 3.25942144  0.6077060 0.5457145714
```

```
deseas_parapluie <- lm(parapluie~dummies,data=Data)$residuals
deseas_precip     <- lm(precip~dummies,data=Data)$residuals
eq_frac <- lm(deseas_parapluie~deseas_precip)
summary(eq_frac)$coefficients
```

```
##                     Estimate Std. Error       t value     Pr(>|t|)
## (Intercept)   -3.265931e-16 0.60898084 -5.362946e-16 1.0000000000
## deseas_precip  1.300055e-01 0.03300004  3.939557e+00 0.0001907741
```

When b_2 is scalar (and then $\mathbf{X}_2$ is of dimension $n \times 1$), Theorem 3.2 leads to:

$$b_2 = \frac{\mathbf{X}_2' M^{\mathbf{X}_1} \mathbf{y}}{\mathbf{X}_2' M^{\mathbf{X}_1} \mathbf{X}_2} \quad \text{(partial regression coefficient).}$$

3.6 Goodness of fit

Define the total variation in y as the sum of squared deviations:

$$TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2.$$

We have:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e} = \hat{\mathbf{y}} + \mathbf{e}$$

In the following, we assume that the regression includes a constant (i.e. for all i, $x_{i,1} = 1$). Denote by $\mathbf{M}^0$ the matrix that transforms observations into deviations from sample means. Using that $\mathbf{M}^0 \mathbf{e} = \mathbf{e}$ and that $\mathbf{X}' \mathbf{e} = 0$, we have:

$$\underbrace{\mathbf{y}' \mathbf{M}^0 \mathbf{y}}_{\text{Total sum of sq.}} = (\mathbf{Xb} + \mathbf{e})' \mathbf{M}^0 (\mathbf{Xb} + \mathbf{e})$$

$$= \underbrace{\mathbf{b}' \mathbf{X}' \mathbf{M}^0 \mathbf{Xb}}_{\text{"Explained" sum of sq.}} + \underbrace{\mathbf{e}' \mathbf{e}}_{\text{Sum of sq. residuals}}$$

$$TSS = Expl.SS + SSR.$$

$$\boxed{\text{Coefficient of determination} = \frac{Expl.SS}{TSS} = 1 - \frac{SSR}{TSS} = 1 - \frac{\mathbf{e}' \mathbf{e}}{\mathbf{y}' \mathbf{M}^0 \mathbf{y}}.} \quad (3.5)$$
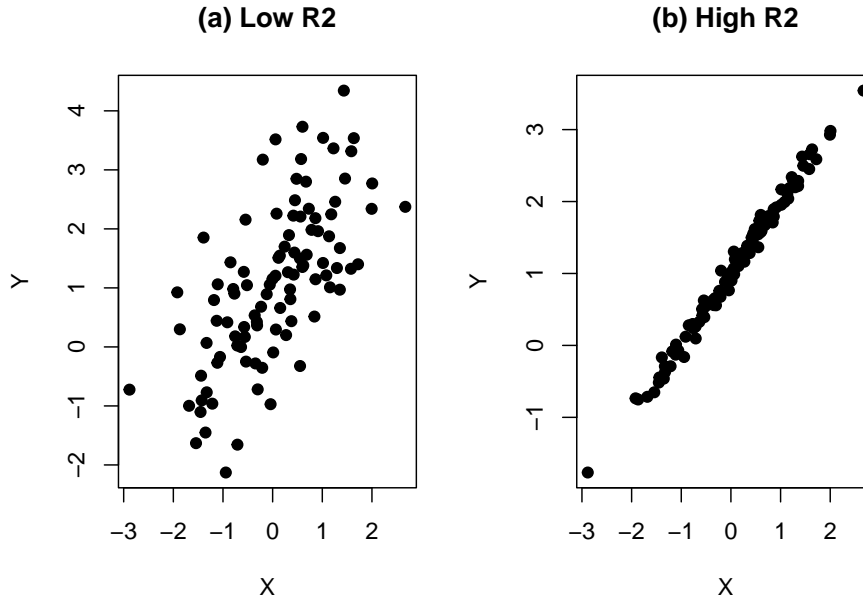
It can be shown [Greene, 2012, Section 3.5] that:

$$\text{Coefficient of determination} = \frac{[\sum_{i=1}^{n} (y_i - \bar{y})(\hat{y}_i - \bar{y})]^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2 \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}.$$

$\Rightarrow R^2$ is the sample squared correlation between y and the (regression-implied) y's predictions.

```
par(mfrow=c(1,2))
par(plt=c(.3,.95,.2,.85))
N <- 100
eps <- rnorm(N)
X <- rnorm(N)
Y <- 1 + X + eps
plot(X,Y,pch=19,main="(a) Low R2")
Y <- 1 + X + .1*eps
plot(X,Y,pch=19,main="(b) High R2")
```


The **partial correlation** between y and z, controlling for some variables $\mathbf{X}$ is the sample correlation between y^* and z^*, where the latter two variables are the residuals in regressions of y on $\mathbf{X}$ and of z on $\mathbf{X}$, respectively.

This correlation is denoted by $r_{yz}^{\mathbf{X}}$. By definition, we have:

$$r_{yz}^{\mathbf{X}} = \frac{\mathbf{z}^{*\prime}\mathbf{y}^*}{\sqrt{(\mathbf{z}^{*\prime}\mathbf{z}^*)(\mathbf{y}^{*\prime}\mathbf{y}^*)}}. \tag{3.6}$$

Proposition 3.4 (Change in SSR when a variable is added). *We have:*

$$\mathbf{u}'\mathbf{u} = \mathbf{e}'\mathbf{e} - c^2(\mathbf{z}^{*\prime}\mathbf{z}^*) \qquad (\leq \mathbf{e}'\mathbf{e}) \tag{3.7}$$

where (i) $\mathbf{u}$ and $\mathbf{e}$ are the residuals in the regressions of $\mathbf{y}$ on $[\mathbf{X}, \mathbf{z}]$ and of $\mathbf{y}$ on $\mathbf{X}$, respectively, (ii) c is the regression coefficient on $\mathbf{z}$ in the former regression and where $\mathbf{z}^*$ are the residuals in the regression of $\mathbf{z}$ on $\mathbf{X}$.

Proof. The OLS estimates $[\mathbf{d}', \mathbf{c}]'$ in the regression of $\mathbf{y}$ on $[\mathbf{X}, \mathbf{z}]$ satisfies (first-order cond., Eq. (3.2))

$$\left[\begin{array}{cc} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{z} \\ \mathbf{z}'\mathbf{X} & \mathbf{z}'\mathbf{z} \end{array} \right] \left[\begin{array}{c} \mathbf{d} \\ \mathbf{c} \end{array} \right] = \left[\begin{array}{c} \mathbf{X}'\mathbf{y} \\ \mathbf{z}'\mathbf{y} \end{array} \right].$$

Hence, in particular $\mathbf{d} = \mathbf{b} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{z}c$, where $\mathbf{b}$ is the OLS of $\mathbf{y}$ on $\mathbf{X}$. Substituting in $\mathbf{u} = \mathbf{y} - \mathbf{X}\mathbf{d} - \mathbf{z}c$, we get $\mathbf{u} = \mathbf{e} - \mathbf{z}^*c$. We therefore have:

$$\mathbf{u}'\mathbf{u} = (\mathbf{e} - \mathbf{z}^*c)(\mathbf{e} - \mathbf{z}^*c) = \mathbf{e}'\mathbf{e} + c^2(\mathbf{z}^{*'}\mathbf{z}^*) - 2c\mathbf{z}^{*'}\mathbf{e}. \tag{3.8}$$

Now $\mathbf{z}^{*'}\mathbf{e} = \mathbf{z}^{*'}(\mathbf{y} - \mathbf{X}\mathbf{b}) = \mathbf{z}^{*'}\mathbf{y}$ because $\mathbf{z}^*$ are the residuals in an OLS regression on $\mathbf{X}$. Since $c = (\mathbf{z}^{*'}\mathbf{z}^*)^{-1}\mathbf{z}^{*'}\mathbf{y}^*$ (by an application of Theorem 3.2), we have $(\mathbf{z}^{*'}\mathbf{z}^*)c = \mathbf{z}^{*'}\mathbf{y}^*$ and, therefore, $\mathbf{z}^{*'}\mathbf{e} = (\mathbf{z}^{*'}\mathbf{z}^*)c$. Inserting this in Eq. (3.8) leads to the results. $\square$ $\hfill\square$

Proposition 3.5 (Change in the coefficient of determination when a variable is added)**.** *Denoting by R_W^2 the coefficient of determination in the regression of $\mathbf{y}$ on some variable $\mathbf{W}$, we have:*

$$R_{\mathbf{X},\mathbf{z}}^2 = R_{\mathbf{X}}^2 + (1 - R_{\mathbf{X}}^2)(r_{yz}^{\mathbf{X}})^2,$$

where $r_{yz}^{\mathbf{X}}$ is the coefficient of partial correlation.

Proof. Let's use the same notations as in Prop. @ref{prp:chgeR2}. Theorem 3.2 implies that $c = (\mathbf{z}^{*'}\mathbf{z}^*)^{-1}\mathbf{z}^{*'}\mathbf{y}^*$. Using this in Eq. (3.7) gives $\mathbf{u}'\mathbf{u} = \mathbf{e}'\mathbf{e} - (\mathbf{z}^{*'}\mathbf{y}^*)^2/(\mathbf{z}^{*'}\mathbf{z}^*)$. Using the definition of the partial correlation (Eq. (3.6)), we get $\mathbf{u}'\mathbf{u} = \mathbf{e}'\mathbf{e}\left(1 - (r_{yz}^{\mathbf{X}})^2\right)$. The results is obtained by dividing both sides of the previous equation by $\mathbf{y}'\mathbf{M}_0\mathbf{y}$. $\square$ $\hfill\square$

The previous theorem shows that we necessarily increase the R^2 if we add variables, **even if they are irrelevant**.

The **adjusted** R^2, denoted by $\bar{R}^2$, is a fit measure that penalizes large numbers of regressors:

$$\boxed{\bar{R}^2 = 1 - \frac{\mathbf{e}'\mathbf{e}/(n-K)}{\mathbf{y}'\mathbf{M}^0\mathbf{y}/(n-1)} = 1 - \frac{n-1}{n-K}(1 - R^2).}$$

3.7 Inference and Prediction

Under the normality assumption (Assumption 3.5), we know the distribution of **b** (conditional on **X**). Indeed, $(\mathbf{b}|\mathbf{X}) \equiv (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is multivariate Gaussian:

$$\mathbf{b}|\mathbf{X} \sim \mathcal{N}(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}). \tag{3.9}$$

Problem: In practice, we do not know σ^2 (**population parameter**).

Proposition 3.6. *Under 3.1 to 3.4, an unbiased estimate of σ^2 is given by:*

$$s^2 = \frac{\mathbf{e}'\mathbf{e}}{n-K}. \tag{3.10}$$

(It is sometimes denoted by σ^2_{OLS}.)

Proof. $\mathbb{E}(\mathbf{e}'\mathbf{e}|\mathbf{X}) = \mathbb{E}(\varepsilon'\mathbf{M}\varepsilon|\mathbf{X}) = \mathbb{E}(\text{Tr}(\varepsilon'\mathbf{M}\varepsilon)|\mathbf{X})) = \text{Tr}(\mathbf{M}\mathbb{E}(\varepsilon\varepsilon'|\mathbf{X})) = \sigma^2\text{Tr}(\mathbf{M})$. (Note that we have $\mathbb{E}(\varepsilon\varepsilon'|\mathbf{X}) = \sigma^2 Id$ by Assumptions 3.3 and 3.4, see Prop. 3.2.) Finally: $\text{Tr}(\mathbf{M}) = n - \text{Tr}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = n - \text{Tr}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}) = n - \text{Tr}(Id_{K\times K})$. $\square$ $\qquad\square$

Two results will prove important to perform hypothesis testing:

   i. We know the distribution of s^2 (Prop. 3.7).
   ii. s^2 and **b** are independent random variables (Prop. 3.8).

Proposition 3.7. *Under 3.1 to 3.5, we have:* $\dfrac{s^2}{\sigma^2}|\mathbf{X} \sim \chi^2(n-K)/(n-K)$.

Proof. We have $\mathbf{e}'\mathbf{e} = \varepsilon'\mathbf{M}\varepsilon$. $\mathbf{M}$ is an idempotent symmetric matrix. Therefore it can be decomposed as PDP' where D is a diagonal matrix and P is an orthogonal matrix. As a result $\mathbf{e}'\mathbf{e} = (P'\varepsilon)'D(P'\varepsilon)$, i.e. $\mathbf{e}'\mathbf{e}$ is a weighted sum of independent squared Gaussian variables (the entries of $P'\varepsilon$ are independent because they are Gaussian – under 3.5 – and uncorrelated). The variance of each of these i.i.d. Gaussian variable is σ^2. Because $\mathbf{M}$ is an idempotent symmetric matrix, its eigenvalues are either 0 or 1, and its rank equals its trace. Further, its trace is equal to $n-K$ (see proof of Eq. (3.10)). Therefore D has $n-K$ entries equal to 1 and K equal to 0. Hence, $\mathbf{e}'\mathbf{e} = (P'\varepsilon)'D(P'\varepsilon)$ is a sum of $n-K$ squared independent Gaussian variables of variance σ^2. Therefore $\frac{\mathbf{e}'\mathbf{e}}{\sigma^2} = (n-K)\frac{s^2}{\sigma^2}$ is a sum of $n-k$ squared i.i.d. standard normal variables. $\square$ $\qquad\square$

Proposition 3.8. *Under Hypotheses 3.1 to 3.5,* **b** *and* s^2 *are independent.*

Proof. We have $\mathbf{b} = \beta + [\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}\varepsilon$ and $s^2 = \varepsilon'\mathbf{M}\varepsilon/(n-K)$. Hence **b** is an affine combination of ε and s^2 is a quadratic combination of the same Gaussian shocks. One can write s^2 as $s^2 = (\mathbf{M}\varepsilon)'\mathbf{M}\varepsilon/(n-K)$ and **b** as $\beta + \mathbf{T}\varepsilon$. Since $\mathbf{TM} = 0$, $\mathbf{T}\varepsilon$ and $\mathbf{M}\varepsilon$ are independent (because two uncorrelated Gaussian variables are independent), therefore **b** and s^2, which are functions of respective independent variables, are independent. $\square$ $\square$

Under Hypotheses 3.1 to 3.5, let us consider b_k, the kth entry of **b**:

$$b_k | \mathbf{X} \sim \mathcal{N}(\beta_k, \sigma^2 v_k),$$

where v_k is the kth component of the diagonal of $(\mathbf{X}'\mathbf{X})^{-1}$.

Besides, we have (Prop. 3.7):

$$\frac{(n-K)s^2}{\sigma^2} | \mathbf{X} \sim \chi^2(n-K).$$

As a result (using Props. 3.7 and 3.8), we have:

$$t_k = \frac{\frac{b_k - \beta_k}{\sqrt{\sigma^2 v_k}}}{\sqrt{\frac{(n-K)s^2}{\sigma^2(n-K)}}} = \frac{b_k - \beta_k}{\sqrt{s^2 v_k}} \sim t(n-K), \qquad (3.11)$$

where $t(n-K)$ denotes a t distribution with $n-K$ degrees of freedom.

Remark: $\frac{b_k - \beta_k}{\sqrt{\sigma^2 v_k}} | \mathbf{X} \sim \mathcal{N}(0,1)$ and $\frac{(n-K)s^2}{\sigma^2} | \mathbf{X} \sim \chi^2(n-K)$. These two distributions do not depend on $\mathbf{X} \Rightarrow$ the *marginal distribution* of t_k is also t.

Note that $s^2 v_k$ is not exactly the conditional variance of b_k: The variance of b_k conditional on $\mathbf{X}$ is $\sigma^2 v_k$. However $s^2 v_k$ is an unbiased estimate of $\sigma^2 v_k$ (by Prop. 3.6).

The previous result (Eq. (3.11)) can be extended to any linear combinations of elements of **b** (Eq. (3.11) is for its kth component only).

Let us consider $\alpha'\mathbf{b}$, the OLS estimate of $\alpha'\beta$. From Eq. (3.9), we have:

$$\alpha'\mathbf{b} | \mathbf{X} \sim \mathcal{N}(\alpha'\beta, \sigma^2 \alpha'(\mathbf{X}'\mathbf{X})^{-1}\alpha).$$

Therefore:

$$\frac{\alpha'\mathbf{b} - \alpha'\beta}{\sqrt{\sigma^2 \alpha'(\mathbf{X}'\mathbf{X})^{-1}\alpha}} | \mathbf{X} \sim \mathcal{N}(0,1).$$

Using the same approach as the one used to derive Eq. (3.11), one can show that Props. 3.7 and 3.8 imply that:

$$\frac{\alpha'\mathbf{b} - \alpha'\beta}{\sqrt{s^2\alpha'(\mathbf{X'X})^{-1}\alpha}} \sim t(n - K). \tag{3.12}$$

```r
par(plt=c(.2,.95,.2,.95))
xx <- seq(-3.5,3.5,by=.01)
plot(xx,dnorm(xx),xlab="X",ylab="",type="l",lwd=2)
lines(xx,dt(xx,df=3),col="red",lwd=2)
lines(xx,dt(xx,df=7),col="red",lwd=2,lty=2)
lines(xx,dt(xx,df=20),col="blue",lwd=3,lty=3)
legend("topright",
       c("N(0,1)","t(3)","t(7)","t(20)"),
       lty=c(1,1,2,3), # gives the legend appropriate symbols (lines)
       lwd=c(2,2,2,3), # line width
       col=c("black","red","red","blue"))
```
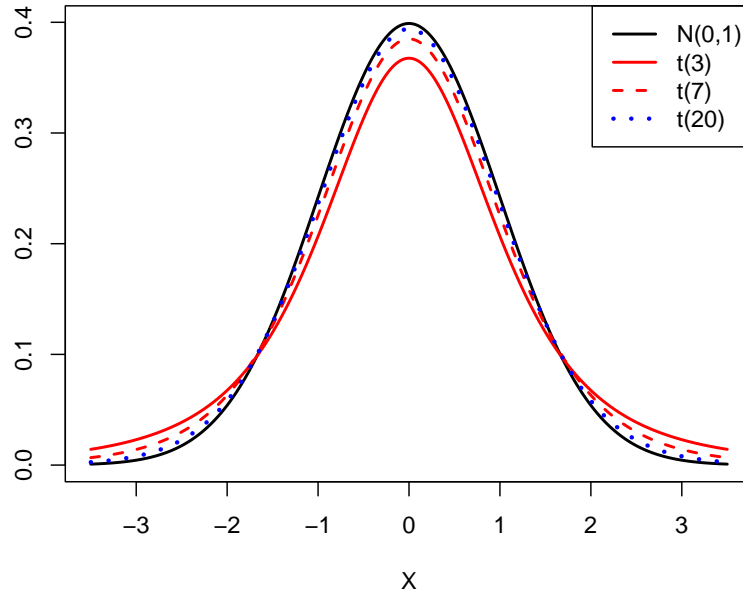

Figure 3.3: The chart shows that the higher the degree of freedom, the closer the distribution of $t(u)$ gets to the normal distribution.

3.8   Confidence interval of β_k

Assume we want to compute a (symmetrical) confidence interval $[I_{d,1-\alpha}, I_{u,1-\alpha}]$ that is such that $\mathbb{P}(\beta_k \in [I_{d,1-\alpha}, I_{u,1-\alpha}]) = 1 - \alpha$. In particular, we want to have: $\mathbb{P}(\beta_k < I_{d,1-\alpha}) = \frac{\alpha}{2}$.

For this purpose, we make use of $t_k = \frac{b_k - \beta_k}{\sqrt{s^2 v_k}} \sim t(n - K)$ (Eq. (3.11)).

We have:

$$\mathbb{P}(\beta_k < I_{d,1-\alpha}) = \frac{\alpha}{2} \Leftrightarrow$$

$$\mathbb{P}\left(\frac{b_k - \beta_k}{\sqrt{s^2 v_k}} > \frac{b_k - I_{d,1-\alpha}}{\sqrt{s^2 v_k}}\right) = \frac{\alpha}{2} \Leftrightarrow \mathbb{P}\left(t_k > \frac{b_k - I_{d,1-\alpha}}{\sqrt{s^2 v_k}}\right) = \frac{\alpha}{2} \Leftrightarrow$$

$$1 - \mathbb{P}\left(t_k \leq \frac{b_k - I_{d,1-\alpha}}{\sqrt{s^2 v_k}}\right) = \frac{\alpha}{2} \Leftrightarrow \frac{b_k - I_{d,1-\alpha}}{\sqrt{s^2 v_k}} = \Phi^{-1}_{t(n-K)}\left(1 - \frac{\alpha}{2}\right),$$

where $\Phi_{t(n-K)}(\alpha)$ is the c.d.f. of the $t(n - K)$ distribution (Table @ref{tab:Studenttable}).

Doing the same for $I_{u,1-\alpha}$, we obtain:

$$[I_{d,1-\alpha}, I_{u,1-\alpha}] =$$
$$\left[b_k - \Phi^{-1}_{t(n-K)}\left(1 - \frac{\alpha}{2}\right)\sqrt{s^2 v_k}, b_k + \Phi^{-1}_{t(n-K)}\left(1 - \frac{\alpha}{2}\right)\sqrt{s^2 v_k}\right].$$

3.8.1   Example

The following example is based on the HRS dataset (Health and Retirement Study). We only consider only a subset of this large dataset, focusing on a few variables, and for year 2018 (wave 14). This R script builds the reduced dataset.

```
reducedHRS <- read.csv("https://raw.githubusercontent.com/jrenne/Data4courses/master/HI
eq <- lm(riearn~raedyrs+ragey_b+I(ragey_b^2)+rfemale,data=reducedHRS)
print(summary(eq))
```

```
##
## Call:
## lm(formula = riearn ~ raedyrs + ragey_b + I(ragey_b^2) + rfemale,
##     data = reducedHRS)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -82512 -29447  -8144  18083 394724
##
```

```
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.336e+05  3.245e+04  -4.116 3.92e-05 ***
## raedyrs       5.216e+03  2.384e+02  21.876  < 2e-16 ***
## ragey_b       4.758e+03  1.086e+03   4.382 1.20e-05 ***
## I(ragey_b^2) -4.441e+01  9.097e+00  -4.882 1.09e-06 ***
## rfemale      -1.499e+04  1.498e+03 -10.007  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 50130 on 4540 degrees of freedom
## Multiple R-squared:  0.1173, Adjusted R-squared:  0.1165
## F-statistic: 150.8 on 4 and 4540 DF,  p-value: < 2.2e-16
```

The last two columns give the test statistic and p-values associated to the test whose null hypothesis is:

$$H_0 : \beta_k = 0.$$

The **t-statistics**, that is $b_k / \sqrt{s^2 v_k}$, is the test statistic of the test. Under H_0, the t-statistic is $t(n - K)$ (see Eq. (3.11)). Hence, the **critical region** for the test of size α is:

$$\left] -\infty, -\Phi^{-1}_{t(n-K)} \left(1 - \frac{\alpha}{2} \right) \right] \cup \left[\Phi^{-1}_{t(n-K)} \left(1 - \frac{\alpha}{2} \right), +\infty \right[.$$

The **p-value** is defined as the probability that $|Z| > |t|$, where t is the (computed) t statistics and where $Z \sim t(n - K)$. That is, the p-value is given by $2(1 - \Phi_{t(n-K)}(|t_k|))$.

See this webpage for details regarding the link between critical regions, p-value, and test outcomes.

3.9   Set of linear restrictions

We consider the following model:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon, \quad \varepsilon \sim i.i.d.\mathcal{N}(0, \sigma^2).$$

and we want to test for the *joint* validity of a set of restrictions involving the components of β in a linear way.

Set of linear restrictions:

$$
\begin{aligned}
r_{1,1}\beta_1 + \cdots + r_{1,K}\beta_K &= q_1 \\
\vdots \quad\quad\quad &\quad \vdots \\
r_{J,1}\beta_1 + \cdots + r_{J,K}\beta_K &= q_J,
\end{aligned}
\tag{3.13}
$$

that can be written in matrix form:

$$\mathbf{R}\beta = \mathbf{q}. \tag{3.14}$$

Defin the **Discrepancy vector $\mathbf{m} = \mathbf{Rb} - \mathbf{q}$**. Under the null hypothesis:

$$
\begin{aligned}
\mathbb{E}(\mathbf{m}|\mathbf{X}) &= \mathbf{R}\beta - \mathbf{q} = 0 \quad \text{and} \\
\mathbb{V}ar(\mathbf{m}|\mathbf{X}) &= \mathbf{R}\mathbb{V}ar(\mathbf{b}|\mathbf{X})\mathbf{R}'.
\end{aligned}
$$

Under Hypotheses 3.1 to 3.4, $\mathbb{V}ar(\mathbf{m}|\mathbf{X}) = \sigma^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'$ (see Prop. 3.3).

Consider the test:

$$\boxed{H_0 : \mathbf{R}\beta - \mathbf{q} = 0 \text{ against } H_1 : \mathbf{R}\beta - \mathbf{q} \neq 0.} \tag{3.15}$$

We could perform a **Wald test**. Under 3.1 to 3.5 –we need the normality assumption– and under H_0, it can be shown that we have:

$$W = \mathbf{m}'\mathbb{V}ar(\mathbf{m}|\mathbf{X})^{-1}\mathbf{m} \sim \chi^2(J). \tag{3.16}$$

However, σ^2 is unknown. Hence we cannot compute W.

We can however approximate it be replacing σ^2 by s^2. The distribution of this new statistic is not $\chi^2(J)$ any more; it is an $\mathcal{F}$ **distribution**, and the test is called F **test**.

Proposition 3.9. *Under Hypotheses 3.1 to 3.5 and if Eq. (3.15) holds, we have:*

$$F = \frac{W}{J}\frac{\sigma^2}{s^2} = \frac{\mathbf{m}'(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}\mathbf{m}}{s^2 J} \sim \mathcal{F}(J, n-K), \tag{3.17}$$

where $\mathcal{F}$ is the distribution of the F-statistic.

Proof. According to Eq. (3.16), $W/J \sim \chi^2(J)/J$. Moreover, the denominator (s^2/σ^2) is $\sim \chi^2(n-K)$. Therefore, F is the ratio of a r.v. distributed as $\chi^2(J)/J$ and another distributed as $\chi^2(n-K)/(n-K)$. It remains to verify that these r.v. are independent.

Under H_0, we have $\mathbf{m} = \mathbf{R}(\mathbf{b} - \beta) = \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon$. Therefore $\mathbf{m}'(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}\mathbf{m}$ is of the form $\varepsilon'\mathbf{T}\varepsilon$ with $\mathbf{T} = \mathbf{D}'\mathbf{C}\mathbf{D}$ where $\mathbf{D} = \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{C} = (\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}$. Under Hypotheses 3.1 to 3.4, the covariance between $\mathbf{T}\varepsilon$ and $\mathbf{M}\varepsilon$ is $\sigma^2\mathbf{T}\mathbf{M} = \mathbf{0}$. Therefore, under 3.5, these variables are Gaussian variables with 0 covariance. Hence they are independent. $\square$ $\hfill\square$

Remark: For large $n - K$, the $\mathscr{F}_{J,n-K}$ distribution converges to $\mathscr{F}_{J,\infty} = \chi^2(J)/J$.

Proposition 3.10. *The F-statistic defined by Eq. (3.17) is also equal to:*

$$F = \frac{(R^2 - R_*^2)/J}{(1 - R^2)/(n - K)} = \frac{(SSR_{restr} - SSR_{unrestr})/J}{SSR_{unrestr}/(n - K)}, \qquad (3.18)$$

*where R_*^2 is the coef. of determination (Eq. **??**eq:RR2)) of the "restricted regression''* (SSR: sum of squared residuals.)

Proof. Let's denote by $\mathbf{e}_* = \mathbf{y} - \mathbf{X}\mathbf{b}_*$ the vector of residuals associated to the *restricted regression* (i.e. $\mathbf{R}\mathbf{b}_* = \mathbf{q}$). We have $\mathbf{e}_* = \mathbf{e} - \mathbf{X}(\mathbf{b}_* - \mathbf{b})$. Using $\mathbf{e}'\mathbf{X} = 0$, we get $\mathbf{e}'_*\mathbf{e}_* = \mathbf{e}'\mathbf{e} + (\mathbf{b}_* - \mathbf{b})'\mathbf{X}'\mathbf{X}(\mathbf{b}_* - \mathbf{b}) \geq \mathbf{e}'\mathbf{e}$.

By Prop. @ref(prp:constrained_LS), we know that $\mathbf{b}_* - \mathbf{b} = -(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\{\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\}^{-1}(\mathbf{R}\mathbf{b} - \mathbf{q})$. Therefore:

$$\mathbf{e}'_*\mathbf{e}_* - \mathbf{e}'\mathbf{e} = (\mathbf{R}\mathbf{b} - \mathbf{q})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{q}).$$

This implies that the F statistic defined in Prop. 3.9 is also equal to:

$$\frac{(\mathbf{e}'_*\mathbf{e}_* - \mathbf{e}'\mathbf{e})/J}{\mathbf{e}'\mathbf{e}/(n - K)}.\square$$

$\square$

The null hypothesis H_0 (Eq. (3.15)) of the F-test is rejected if F –defined by Eq. (3.17) or (3.18)– is higher than $\mathscr{F}_{1-\alpha}(J, n - K)$. (Hence, this test is a one-sided test.)

3.10   Common pitfalls

3.10.1   Multicollinearity

Consider the model: $y_i = \beta_1 x_{i,1} + \beta_2 x_{i,2} + \varepsilon_i$, where all variables are zero-mean and $\mathbb{V}ar(\varepsilon_i) = \sigma^2$. We have

$$\mathbf{X}'\mathbf{X} = \left[\begin{array}{cc} \sum_i x_{i,1}^2 & \sum_i x_{i,1}x_{i,2} \\ \sum_i x_{i,1}x_{i,2} & \sum_i x_{i,2}^2 \end{array} \right],$$

therefore:

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{\sum_i x_{i,1}^2 \sum_i x_{i,2}^2 - (\sum_i x_{i,1} x_{i,2})^2} \begin{bmatrix} \sum_i x_{i,2}^2 & -\sum_i x_{i,1} x_{i,2} \\ -\sum_i x_{i,1} x_{i,2} & \sum_i x_{i,1}^2 \end{bmatrix}.$$

The inverse of the upper-left parameter of $(\mathbf{X}'\mathbf{X})^{-1}$ is:

$$\sum_i x_{i,1}^2 - \frac{(\sum_i x_{i,1} x_{i,2})^2}{\sum_i x_{i,2}^2} = \sum_i x_{i,1}^2 (1 - correl_{1,2}^2), \qquad (3.19)$$

where $correl_{1,2}$ is the sample correlation between $\mathbf{x}_1$ and $\mathbf{x}_2$.

Hence, the closer to one $correl_{1,2}$, the higher the variance of b_1 (recall that the variance of b_1 is the upper-left component of $\sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$).

3.10.2 Omitted variables

Consider the following model, called "True model'':

$$\mathbf{y} = \underbrace{\mathbf{X}_1}_{n \times K_1} \underbrace{\beta_1}_{K_1 \times 1} + \underbrace{\mathbf{X}_2}_{n \times K_2} \underbrace{\beta_2}_{K_2 \times 1} + \varepsilon$$

Then, if one computes $\mathbf{b}_1$ by regressing $\mathbf{y}$ on $\mathbf{X}_1$ only, we get:

$$\mathbf{b}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y} = \beta_1 + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\beta_2 + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\varepsilon.$$

Hence, we obtain the omitted-variable formula:

$$\boxed{\mathbb{E}(\mathbf{b}_1|\mathbf{X}) = \beta_1 + \underbrace{(\mathbf{X}_1'\mathbf{X}_1)^{-1}(\mathbf{X}_1'\mathbf{X}_2)}_{K_1 \times K_2} \beta_2}$$

(each column of $(\mathbf{X}_1'\mathbf{X}_1)^{-1}(\mathbf{X}_1'\mathbf{X}_2)$ are the OLS regressors obtained when regressing the columns of $\mathbf{X}_2$ on $\mathbf{X}_1$).

Example 3.1. Consider the "true model'':

$$wage_i = \beta_0 + \beta_1 edu_i + \beta_2 ability_i + \varepsilon_i, \quad \varepsilon_i \sim i.i.d. \, \mathcal{N}(0, \sigma_\varepsilon^2) \qquad (3.20)$$

Further, we assume that the *edu* variable is correlated to the *ability*. Specifically:

$$edu_i = \alpha_0 + \alpha_1 ability_i + \eta_i, \quad \eta_i \sim i.i.d. \, \mathcal{N}(0, \sigma_\eta^2).$$

Assume we mistakingly run the regression omitting the *ability* variable:

$$wage_i = \gamma_0 + \gamma_1 edu_i + \xi_i. \qquad (3.21)$$

It can be seen that $\xi_i = \varepsilon_i - (\beta_2/\alpha_1)\eta_i \sim i.i.d. \, \mathcal{N}(0, \sigma_\varepsilon^2 + (\beta_2/\alpha_1)^2 \sigma_\eta^2)$ and that the population regression coefficient is $\gamma_1 = \beta_1 + \beta_2/\alpha_1 \neq \beta_1$.

Example 3.2. Let us use the California Test Score dataset (in the package `AER`). Assume we want to measure the effect of the students-to-teacher ratio ('`str`') on student test scores (`testscr`). The following regressions show that the effect is lower when controls are added.

```
library(AER); data("CASchools")
CASchools$str <- CASchools$students/CASchools$teachers
CASchools$testscr <- .5 * (CASchools$math + CASchools$read)
eq <- lm(testscr~str,data=CASchools)
summary(eq)$coefficients
```

```
##                Estimate Std. Error    t value       Pr(>|t|)
## (Intercept) 698.932949  9.4674911 73.824516 6.569846e-242
## str          -2.279808  0.4798255 -4.751327  2.783308e-06
```

```
eq <- lm(testscr~str+lunch,data=CASchools)
summary(eq)$coefficients
```

```
##                Estimate Std. Error    t value       Pr(>|t|)
## (Intercept) 702.9113020 4.70024626 149.547760  0.000000e+00
## str          -1.1172255 0.24035528  -4.648225  4.498554e-06
## lunch        -0.5997501 0.01676439 -35.775242 3.709097e-129
```

```
eq <- lm(testscr~str+lunch+english,data=CASchools)
summary(eq)$coefficients
```

```
##                Estimate Std. Error    t value      Pr(>|t|)
## (Intercept) 700.1499572 4.68568672 149.423126 0.000000e+00
## str          -0.9983090 0.23875428  -4.181324 3.535873e-05
## lunch        -0.5473454 0.02159885 -25.341418 2.303048e-86
## english      -0.1215735 0.03231728  -3.761872 1.928369e-04
```

3.10.3 Irrelevant variable

Consider the *True model*:

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \varepsilon,$$

while the *Estimated model* is:

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \varepsilon$$

The estimates are unbiased. However, adding irrelevant explanatory variables increases the variance of the estimate of β_1 (compared to the case where one

uses the correct explanatory variables). This is the case unless the correlation between $\mathbf{X}_1$ and $\mathbf{X}_2$ is null, see Eq. (3.19).

In other words, the estimator is *inefficient*, i.e., there exists an alternative consistent estimator whose variance is lower. The inefficiency problem can have serious consequences when testing hypotheses of type $H_0 : \beta_1 = 0$ due to the loss of power, so we might infer that they are no relevant variables when they truly are (Type-II error; False Negative).

3.11   Large Sample Properties

Even if we relax the normality assumption (Hypothesis 3.5), we can approximate the finite-sample behavior of the estimators by using *large-sample* or *asymptotic properties*.

To begin with, we proceed under Hypothesis 3.1 to 3.4. (We will see later how to deal with –partial– relaxations of Hypothesis 3.3 and 3.4.)

Under regularity assumptions, under 3.1 to 3.4, even if the residuals are not normally-distributed, the least square estimators can be *asymptotically normal* and inference can be performed as in small samples when 3.1 to 3.5 hold. This derives from Prop. 3.11 (below). The F-test (Prop. **??**prp:Ftest)) and the t-test can then be performed.

───────────────

Proposition 3.11. *Under Assumptions 3.1 to 3.4, and assuming further that:*

$$Q = plim_{n \to \infty} \frac{\mathbf{X}'\mathbf{X}}{n}, \qquad (3.22)$$

and that the $(\mathbf{x}_i, \varepsilon_i)$s are independent (across entities i), we have:

$$\sqrt{n}(\mathbf{b} - \beta) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 Q^{-1}\right). \qquad (3.23)$$

───────────────

Proof. Since $\mathbf{b} = \beta + \left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}\left(\frac{\mathbf{X}'\varepsilon}{n}\right)$, we have: $\sqrt{n}(\mathbf{b}-\beta) = \left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}\left(\frac{1}{\sqrt{n}}\right)\mathbf{X}'\varepsilon$. Since $f : A \to A^{-1}$ is a continuous function (for $A \neq \mathbf{0}$), $plim_{n\to\infty}\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1} = \mathbf{Q}^{-1}$. Let us denote by V_i the vector $\mathbf{x}_i\varepsilon_i$. Because the $(\mathbf{x}_i, \varepsilon_i)$s are independent, the V_is are independent as well. Their covariance matrix is $\sigma^2\mathbb{E}(\mathbf{x}_i\mathbf{x}_i') = \sigma^2 Q$. Applying the multivariate central limit theorem on the V_is gives $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\varepsilon_i\right) = \left(\frac{1}{\sqrt{n}}\right)\mathbf{X}'\varepsilon \xrightarrow{d} \mathcal{N}(0, \sigma^2 Q)$. An application of Slutsky's theorem then leads to the results. □   □

In practice, σ^2 is estimated with $\frac{\mathbf{e}'\mathbf{e}}{n-K}$ (Eq. (3.10)) and $\mathbf{Q}^{-1}$ with $\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}$.

Eqs. (3.22) and (3.23) respectively correspond to convergences in probability and in distribution.

3.12 Instrumental Variables

Here, we want to relax Hypothesis 3.2 –conditional mean zero assumption, implying in particular that $\mathbf{x}_i$ and ε_i are uncorrelated.

We consider the following model:

$$y_i = \mathbf{x_i}'\beta + \varepsilon_i, \quad \text{where } \mathbb{E}(\varepsilon_i) = 0 \text{ and } \mathbf{x_i} \not\perp \varepsilon_i. \tag{3.24}$$

Definition 3.2. The L-dimensional random variable $\mathbf{z}_i$ is a **valid set of instruments** if:

    a. $\mathbf{z}_i$ is correlated to $\mathbf{x}_i$;
    b. we have $\mathbb{E}(\varepsilon|\mathbf{Z}) = 0$ and
    c. the orthogonal projections of the $\mathbf{x}_i$s on the $\mathbf{z}_i$s are not multicollinear.

Example. Let us make the assumption $\mathbf{x}_i \not\perp \varepsilon_i$ (in (3.24)) more precise:

$$\mathbb{E}(\varepsilon_i) = 0 \quad \text{and} \quad \mathbb{E}(\varepsilon_i \mathbf{x_i}) = \gamma. \tag{3.25}$$

By the law of large numbers, $\text{plim}_{n\to\infty}\mathbf{X}'\varepsilon/n = \gamma$. If $\mathbf{Q}_{xx} := \text{plim } \mathbf{X}'\mathbf{X}/n$, the OLS estimator is not consistent because

$$\mathbf{b} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon \xrightarrow{p} \beta + \mathbf{Q}_{xx}^{-1}\gamma \neq \beta.$$

If $\mathbf{z}_i$ is a valid set of instruments, we have:

$$\text{plim}\left(\frac{\mathbf{Z}'\mathbf{y}}{n}\right) = \text{plim}\left(\frac{\mathbf{Z}'(\mathbf{X}\beta + \varepsilon)}{n}\right) = \text{plim}\left(\frac{\mathbf{Z}'\mathbf{X}}{n}\right)\beta$$

Indeed, by the law of large numbers, $\frac{\mathbf{Z}'\varepsilon}{n} \xrightarrow{p} \mathbb{E}(\mathbf{z}_i\varepsilon_i) = 0$.

If $L = K$, the matrix $\frac{\mathbf{Z}'\mathbf{X}}{n}$ is of dimension $K \times K$ and we have:

$$\left[\text{plim}\left(\frac{\mathbf{Z}'\mathbf{X}}{n}\right)\right]^{-1}\text{plim}\left(\frac{\mathbf{Z}'\mathbf{y}}{n}\right) = \beta.$$

By continuity of the inverse funct.: $\left[\text{plim}\left(\frac{\mathbf{Z}'\mathbf{X}}{n}\right)\right]^{-1} = \text{plim}\left(\frac{\mathbf{Z}'\mathbf{X}}{n}\right)^{-1}$. The Slutsky Theorem further implies that

$$\text{plim}\left(\frac{\mathbf{Z}'\mathbf{X}}{n}\right)^{-1}\text{plim}\left(\frac{\mathbf{Z}'\mathbf{y}}{n}\right) = \text{plim}\left(\left(\frac{\mathbf{Z}'\mathbf{X}}{n}\right)^{-1}\frac{\mathbf{Z}'\mathbf{y}}{n}\right).$$

Hence $\mathbf{b}_{iv}$ is consistent if it is defined by:

$$\boxed{\mathbf{b}_{iv} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}.}$$

Proposition 3.12. *If $\mathbf{z}_i$ is a L-dimensional random variable that constitutes a valid set of instruments (see Def. 3.2) and if $L = K$, then the asymptotic distribution of $\mathbf{b}_{iv}$ is:*

$$\mathbf{b}_{iv} \xrightarrow{d} \mathcal{N}\left(\beta, \frac{\sigma^2}{n}\left[Q_{xz}Q_{zz}^{-1}Q_{zx}\right]^{-1}\right)$$

where plim $\mathbf{Z}'\mathbf{Z}/n =: \mathbf{Q}_{zz}$, plim $\mathbf{Z}'\mathbf{X}/n =: \mathbf{Q}_{zx}$, plim $\mathbf{X}'\mathbf{Z}/n =: \mathbf{Q}_{xz}$.

Proof. The proof is very similar to that of Prop. 3.11, the starting point being that $\mathbf{b}_{iv} = \beta + (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\varepsilon$.   □   □

When $L = K$, we have:

$$\left[Q_{xz}Q_{zz}^{-1}Q_{zx}\right]^{-1} = Q_{zx}^{-1}Q_{zz}Q_{xz}^{-1}$$

In practice, to estimate $\mathbb{V}ar(\mathbf{b}_{iv}) = \frac{\sigma^2}{n}Q_{zx}^{-1}Q_{zz}Q_{xz}^{-1}$, we replace σ^2 by:

$$s_{iv}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mathbf{x}_i'\mathbf{b}_{iv})^2$$

And when $L > K$? Idea: First regress $\mathbf{X}$ on the space spanned by $\mathbf{Z}$ and then regress $\mathbf{y}$ on the fitted values $\hat{\mathbf{X}} := \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}$. That is $\mathbf{b}_{iv} = (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\mathbf{y}$:

$$\boxed{\mathbf{b}_{iv} = [\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}.} \tag{3.26}$$

In this case, Prop. 3.12 still holds, with $\mathbf{b}_{iv}$ given by Eq. (3.26).

$\mathbf{b}_{iv}$ is also the result of the regression of $\mathbf{y}$ on $\mathbf{X}^*$, where the columns of $\mathbf{X}^*$ are the (othogonal) projections of those of $\mathbf{X}$ on $\mathbf{Z}$, i.e. $\mathbf{X}^* = \mathbf{P}^{\mathbf{Z}}\mathbf{X}$ (using the

notations introduced in Eq. (3.4)). Hence the other names of this estimator: **Two-Stage Least Squares (TSLS)**.

If the instruments do not properly satisfy Condition (a) in Def. 3.2 (i.e. if $\mathbf{x}_i$ and $\mathbf{z}_i$ are only loosely related), the instruments are said to be **weak**.

This problem is for instance discussed in Stock and Yogo (2003). See also Stock and Watson pp.,489-490.

The Hausman test can be used to test if IV necessary. IV techniques are required if $\operatorname{plim}_{n\to\infty}\mathbf{X}'\varepsilon/n \neq 0$. Hausman (1978) proposes a test of the efficiency of estimators. Under the null hypothesis two estimators, $\mathbf{b}_0$ and $\mathbf{b}_1$, are consistent but $\mathbf{b}_0$ is (asymptotically) efficient relative to $\mathbf{b}_1$. Under the alternative hypothesis, $\mathbf{b}_1$ (IV in the present case) remains consistent but not $\mathbf{b}_0$ (OLS in the present case).

The test statistic is:

$$H = (\mathbf{b}_1 - \mathbf{b}_0)' MPI (\mathbb{V}ar(\mathbf{b}_1) - \mathbb{V}ar(\mathbf{b}_0))(\mathbf{b}_1 - \mathbf{b}_0),$$

where MPI is the Moore-Penrose pseudo-inverse. Under the null hypothesis, $H \sim \chi^2(q)$, where q is the rank of $\mathbb{V}ar(\mathbf{b}_1) - \mathbb{V}ar(\mathbf{b}_0)$.

Example 3.3. Estimation of price elasticity

See e.g. WHO and estimation of tobacco price elasticity of demand.

We want to estimate what is the effect on demand of an *exogenous increase* in prices of cigarettes (say).

The model is:

$$
\begin{aligned}
\underset{\log(\text{demand})}{q_t^d} &= \alpha_0 + \alpha_1 \underset{\log(\text{price})}{\underbrace{\times p_t}} + \alpha_2 \underset{\text{income}}{\underbrace{\times w_t}} + \varepsilon_t^d \\
\underset{\log(\text{supply})}{q_t^s} &= \gamma_0 + \gamma_1 \times p_t + \gamma_2 \underset{\text{cost factors}}{\underbrace{\times \mathbf{y}_t}} + \varepsilon_t^s,
\end{aligned}
$$

where $\mathbf{y}_t$, w_t, $\varepsilon_t^s \sim \mathcal{N}(0,\sigma_s^2)$ and $\varepsilon_t^d \sim \mathcal{N}(0,\sigma_d^2)$ are independent.

Equilibrium: $q_t^d = q_t^s$. This implies that prices are **endogenous**:

$$p_t = \frac{\alpha_0 + \alpha_2 w_t + \varepsilon_t^d - \gamma_0 - \gamma_2 \mathbf{y}_t - \varepsilon_t^s}{\gamma_1 - \alpha_1}.$$

In particular we have $\mathbb{E}(p_t \varepsilon_t^d) = \frac{\sigma_d^2}{\gamma_1 - \alpha_1} \neq 0 \Rightarrow$ Regressing by OLS q_t^d on p_t gives biased estimates (see Eq. (3.25)).

Estimation of the price elasticity of cigarette demand. Instrument: real tax on cigarettes arising from the state's general sales tax. Presumption: in states with a larger general sales tax, cigarette prices are higher, but the general tax is not determined by other forces affecting ε_t^d.
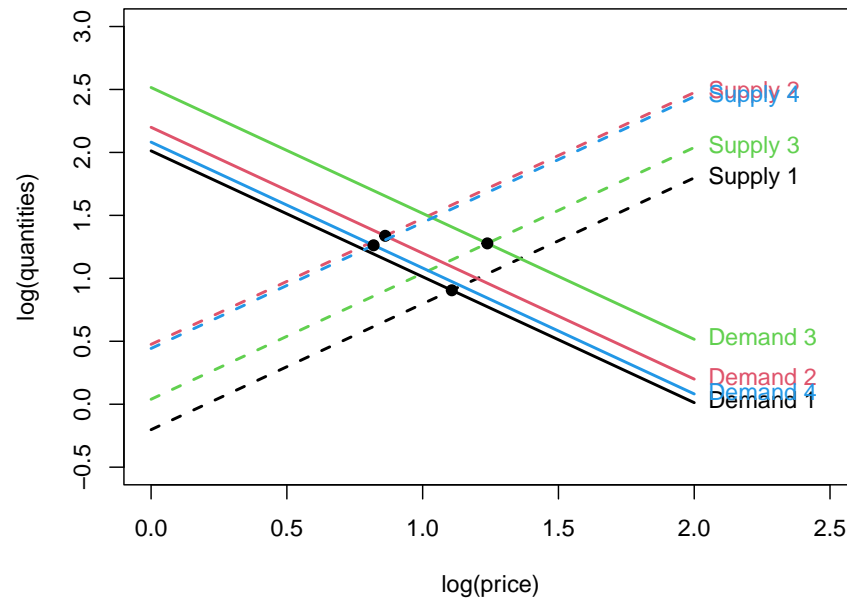
Figure 3.4: This figure illustrates the situation prevailing when estimating a price-elasticity (and the price is endogenous).

```
data("CigarettesSW", package = "AER")
CigarettesSW$rprice  <- with(CigarettesSW, price/cpi)
CigarettesSW$rincome <- with(CigarettesSW, income/population/cpi)
CigarettesSW$tdiff   <- with(CigarettesSW, (taxs - tax)/cpi)

## model
fm <- ivreg(log(packs) ~ log(rprice) + log(rincome) | log(rincome) + tdiff + I(tax/cpi),
            data = CigarettesSW, subset = year == "1995")
eq.no.IV <- lm(log(packs) ~ log(rprice) + log(rincome),
               data = CigarettesSW, subset = year == "1995")
summary(fm, vcov = sandwich, diagnostics = TRUE)
```

```
##
## Call:
## ivreg(formula = log(packs) ~ log(rprice) + log(rincome) | log(rincome) +
##     tdiff + I(tax/cpi), data = CigarettesSW, subset = year ==
##     "1995")
##
## Residuals:
##       Min        1Q     Median        3Q       Max
## -0.6006931 -0.0862222 -0.0009999  0.1164699  0.3734227
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.8950     0.9288  10.654 6.89e-14 ***
## log(rprice)  -1.2774     0.2417  -5.286 3.54e-06 ***
## log(rincome)  0.2804     0.2458   1.141     0.26
##
## Diagnostic tests:
##                  df1 df2 statistic p-value
## Weak instruments   2  44   228.738  <2e-16 ***
## Wu-Hausman         1  44     3.823  0.0569 .
## Sargan             1  NA     0.333  0.5641
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1879 on 45 degrees of freedom
## Multiple R-Squared: 0.4294,  Adjusted R-squared: 0.4041
## Wald test: 17.25 on 2 and 45 DF,  p-value: 2.743e-06
```

```
summary(eq.no.IV)
```

```
##
## Call:
```

```
## lm(formula = log(packs) ~ log(rprice) + log(rincome), data = CigarettesSW,
##     subset = year == "1995")
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.59077 -0.07856 -0.00149  0.11860  0.35442
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   10.3420     1.0227  10.113 3.66e-13 ***
## log(rprice)   -1.4065     0.2514  -5.595 1.24e-06 ***
## log(rincome)   0.3439     0.2350   1.463     0.15
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1873 on 45 degrees of freedom
## Multiple R-squared:  0.4327, Adjusted R-squared:  0.4075
## F-statistic: 17.16 on 2 and 45 DF,  p-value: 2.884e-06
```

```
fm2 <- ivreg(log(packs) ~ log(rprice) | tdiff, data = CigarettesSW, subset = year == "
anova(fm, fm2)
```

```
## Analysis of Variance Table
##
## Model 1: log(packs) ~ log(rprice) + log(rincome) | log(rincome) + tdiff +
##     I(tax/cpi)
## Model 2: log(packs) ~ log(rprice) | tdiff
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     45 1.5880
## 2     46 1.6668 -1 -0.078748 1.3815  0.246
```

```
library(sem)
data("CollegeDistance", package = "AER")
simple.ed.1s<- lm(education ~ urban + gender + ethnicity + unemp + distance,
                data = CollegeDistance)
CollegeDistance$ed.pred<- predict(simple.ed.1s)
simple.ed.2s<- lm(wage ~ urban + gender + ethnicity + unemp + ed.pred ,
                data = CollegeDistance)

simple.comp<- encomptest(wage ~ urban + gender + ethnicity + unemp + ed.pred ,
                    wage ~ urban + gender + ethnicity + unemp + education ,
                    data = CollegeDistance)
fsttest<- encomptest(education ~ tuition + gender + ethnicity + urban ,
                   education ~ distance ,
                   data = CollegeDistance)
```

```
eqOLS <- lm(wage ~ urban + gender + ethnicity + unemp + education,
            data=CollegeDistance)

summary(eqOLS)
```

```
##
## Call:
## lm(formula = wage ~ urban + gender + ethnicity + unemp + education,
##     data = CollegeDistance)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.3484 -0.8408  0.1808  0.8119  3.9875
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       8.641490   0.157008  55.039   <2e-16 ***
## urbanyes          0.070117   0.044727   1.568   0.1170
## genderfemale     -0.085242   0.037069  -2.300   0.0215 *
## ethnicityafam    -0.556056   0.052167 -10.659   <2e-16 ***
## ethnicityhispanic -0.544007  0.048670 -11.177   <2e-16 ***
## unemp             0.133101   0.006711  19.834   <2e-16 ***
## education         0.005369   0.010362   0.518   0.6044
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.268 on 4732 degrees of freedom
## Multiple R-squared:  0.1098, Adjusted R-squared:  0.1087
## F-statistic: 97.27 on 6 and 4732 DF,  p-value: < 2.2e-16
```

```
summary(simple.ed.2s)
```

```
##
## Call:
## lm(formula = wage ~ urban + gender + ethnicity + unemp + ed.pred,
##     data = CollegeDistance)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1692 -0.8294  0.1502  0.8482  3.9537
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -0.359032   1.412087  -0.254  0.79931
```

```
## urbanyes            0.046144   0.044691   1.033  0.30188
## genderfemale       -0.070753   0.036978  -1.913  0.05576 .
## ethnicityafam      -0.227240   0.072984  -3.114  0.00186 **
## ethnicityhispanic  -0.351291   0.057021  -6.161 7.84e-10 ***
## unemp               0.139163   0.006748  20.622  < 2e-16 ***
## ed.pred             0.647099   0.100592   6.433 1.38e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.263 on 4732 degrees of freedom
## Multiple R-squared:  0.1175, Adjusted R-squared:  0.1163
## F-statistic:    105 on 6 and 4732 DF,  p-value: < 2.2e-16
```

```
eqTSLS <- tsls(wage ~ urban + gender + ethnicity + unemp + education,
               ~ urban + gender + ethnicity + unemp + distance,
             data=CollegeDistance)

eqTSLS <- ivreg(wage ~ urban + gender + ethnicity + unemp + education|
                urban + gender + ethnicity + unemp + distance,
              data=CollegeDistance)

summary(eqTSLS, vcov = sandwich, diagnostics = TRUE)
```

```
##
## Call:
## ivreg(formula = wage ~ urban + gender + ethnicity + unemp + education |
##     urban + gender + ethnicity + unemp + distance, data = CollegeDistance)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.20896 -1.14578 -0.02361  1.33303  4.77571
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -0.35903    1.91755  -0.187   0.8515
## urbanyes           0.04614    0.05926   0.779   0.4362
## genderfemale      -0.07075    0.04974  -1.422   0.1550
## ethnicityafam     -0.22724    0.09539  -2.382   0.0172 *
## ethnicityhispanic -0.35129    0.07577  -4.636 3.64e-06 ***
## unemp              0.13916    0.00934  14.899  < 2e-16 ***
## education          0.64710    0.13691   4.727 2.35e-06 ***
##
## Diagnostic tests:
##                  df1  df2 statistic  p-value
## Weak instruments   1 4732     50.19 1.60e-12 ***
```

```
## Wu-Hausman           1 4731      40.30 2.38e-10 ***
## Sargan               0   NA         NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.706 on 4732 degrees of freedom
## Multiple R-Squared: -0.6118, Adjusted R-squared: -0.6138
## Wald test: 57.08 on 6 and 4732 DF,  p-value: < 2.2e-16
```

3.13 General Regression Model

We want to relax the assumption according to which the disturbances are uncorrelated with each other (Hypothesis @ref(hyp:noncorrel_resid)) or the homoskedasticity Hypothesis 3.3.

We replace the latter two assumptions by the general formulation:

$$\mathbb{E}(\varepsilon\varepsilon'|\mathbf{X}) \;=\; \Sigma. \tag{3.27}$$

Note that Eq. ((3.27)) is more general than Hypothesis 3.3 and @ref(hyp:noncorrel_resid) because the diagonal entries of Σ may be different (not the case under Hypothesis 3.3), and the non-diagonal entries of Σ can be $\neq 0$ (contrary to Hypothesis 3.4).

Definition 3.3. Hypothesis 3.1 and 3.2, together with Eq. (3.27), form the **General Regression Model** (GRM) framework.

Note that a regression model where Hypotheses 3.1 to 3.4 hold is a specific case of the GRM framework.

The GRM context notably allows to model **heteroskedasticity** and **autocorrelation**.

- Heteroskedasticity:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & \sigma_n^2 \end{bmatrix}. \tag{3.28}$$

- Autocorrelation:

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho_{2,1} & \dots & \rho_{n,1} \\ \rho_{2,1} & 1 & & \vdots \\ \vdots & & \ddots & \rho_{n,n-1} \\ \rho_{n,1} & \rho_{n,2} & \dots & 1 \end{bmatrix}. \tag{3.29}$$

Example 3.4. Autocorrelation is, in particular, a recurrent problem when time-series data are used (see Section **??**).

In a time-series context, subscript i refers to a date. Assume for instance that:

$$y_i = \mathbf{x}_i'\beta + \varepsilon_i \tag{3.30}$$

with

$$\varepsilon_i = \rho\varepsilon_{i-1} + v_i, \quad v_i \sim \mathcal{N}(0, \sigma_v^2). \tag{3.31}$$

In this case, we are in the GRM context, with:

$$\Sigma = \frac{\sigma_v^2}{1 - \rho^2} \begin{bmatrix} 1 & \rho & \cdots & \rho^{n-1} \\ \rho & 1 & & \vdots \\ \vdots & & \ddots & \rho \\ \rho^{n-1} & \rho^{n-2} & \cdots & 1 \end{bmatrix}. \tag{3.32}$$

3.13.1   Generalized Least Squares

Assume Σ is known ("feasible GLS''). Because Σ is symmetric positive, it admits a spectral decomposition of the form $\Sigma = \mathbf{C}\Lambda\mathbf{C}'$, where $\mathbf{C}$ is an orthogonal matrix (i.e. $\mathbf{CC}' = Id$) and Λ is a diagonal matrix (the diagonal entries are the eigenvalues of Σ).

We have $\Sigma = (\mathbf{PP}')^{-1}$ with $\mathbf{P} = \mathbf{C}\Lambda^{-1/2}$.

Consider the transformed model:

$$\mathbf{P}'\mathbf{y} = \mathbf{P}'\mathbf{X}\beta + \mathbf{P}'\varepsilon \quad \text{or} \quad \mathbf{y}^* = \mathbf{X}^*\beta + \varepsilon^*.$$

The variance of ε^* is $\mathbf{I}$. In the transformed model, OLS is BLUE (Gauss-Markow Theorem 3.1).

The **Generalized least squares** estimator of β is:

$$\boxed{\mathbf{b}_{GLS} = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{y}}. \tag{3.33}$$

We have:

$$\mathbb{V}ar(\mathbf{b}_{GLS}|\mathbf{X}) = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}.$$

When Σ is unknown, the GLS estimator is said to be *infeasible.* Some structure is required. Assume Σ admits a parametric form $\Sigma(\theta)$. The estimation becomes *feasible* (FGLS) if one replaces $\Sigma(\theta)$ by $\Sigma(\hat{\theta})$.

If $\hat{\theta}$ is a consistent estimator of θ, then the FGLS is asymptotically efficient (see Example **??**).

By contrast, when Σ has no obvious structure: the OLS (or IV) is the only estimator available. It remains unbiased, consistent, and asymptotically normally

distributed, but not efficient. Standard inference procedures are not appropriate any longer.

Autocorrelation in the time-series context. Consider the case presented in Example 3.4. Because the OLS estimate $\mathbf{b}$ of β is consistent, the estimates e_is of the ε_is also are. Consistent estimators of ρ and σ_v are then obtained by regressing the e_is on the e_{i-1}s. Using these estimates in Eq. (3.32) provides a consistent estimate of Σ.

See Cochrane and Orcutt (2012).

Proposition 3.13. *Conditionally on* $\mathbf{X}$*, we have:*

$$\mathbb{V}ar(\mathbf{b}|\mathbf{X}) = \frac{1}{n}\left(\frac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1}\left(\frac{1}{n}\mathbf{X}'\Sigma\mathbf{X}\right)\left(\frac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1}. \qquad (3.34)$$

Under Hypothesis 3.5, since $\mathbf{b}$ *is linear in* ε*, we have:*

$$\mathbf{b}|\mathbf{X} \sim \mathcal{N}\left(\beta, (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\Sigma\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}\right). \qquad (3.35)$$

Note that the variance of the estimator is not $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ any more, so using $s^2(\mathbf{X}'\mathbf{X})^{-1}$ for inference may be misleading.

Proposition 3.14. *If* $plim\,(\mathbf{X}'\mathbf{X}/n)$ *and* $plim\,(\mathbf{X}'\Sigma\mathbf{X}/n)$ *are finite positive definite matrices, then* $plim\,(\mathbf{b}) = \beta$*.*

Proof. We have $\mathbb{V}ar(\mathbf{b}) = \mathbb{E}[\mathbb{V}ar(\mathbf{b}|\mathbf{X})] + \mathbb{V}ar[\mathbb{E}(\mathbf{b}|\mathbf{X})]$. Since $\mathbb{E}(\mathbf{b}|\mathbf{X}) = \beta$, $\mathbb{V}ar[\mathbb{E}(\mathbf{b}|\mathbf{X})] = 0$. Eq. (3.34) implies that $\mathbb{V}ar(\mathbf{b}|\mathbf{X}) \to 0$. Hence $\mathbf{b}$ converges in mean square and therefore in probability. $\square$ $\qquad\qquad$ $\square$

Proposition 3.15. *If* $Q_{xx} = plim\,(\mathbf{X}'\mathbf{X}/n)$ *and* $Q_{x\Sigma x} = plim\,(\mathbf{X}'\Sigma\mathbf{X}/n)$ *are finite positive definite matrices, then:*

$$\sqrt{n}(\mathbf{b} - \beta) \xrightarrow{d} \mathcal{N}(0, Q_{xx}^{-1}Q_{x\Sigma x}Q_{xx}^{-1}).$$

Proposition 3.16. *If regressors and IV variables are "well-behaved'', then:*

$$\mathbf{b}_{iv} \overset{a}{\sim} \mathcal{N}(\beta, \mathbf{V}_{iv}),$$

where

$$\mathbf{V}_{iv} = \frac{1}{n}(\mathbf{Q}^*) \; plim \; \left(\frac{1}{n}\mathbf{Z}'\Sigma\mathbf{Z}\right)(\mathbf{Q}^*)',$$

with

$$\mathbf{Q}^* = [\mathbf{Q}_{xz}\mathbf{Q}_{zz}^{-1}\mathbf{Q}_{zx}]^{-1}\mathbf{Q}_{xz}\mathbf{Q}_{zz}^{-1}.$$

For practical purposes, one needs to have estimates of Σ in Props. 3.13, 3.15 or 3.16.

Idea: instead of estimating Σ (dimension $n \times n$) directly, one can estimate $\frac{1}{n}\mathbf{X}'\Sigma\mathbf{X}$, of dimension $K \times K$ (or $\frac{1}{n}\mathbf{Z}'\Sigma\mathbf{Z}$ in the IV case). Indeed, this is this expression ($\mathbf{X}'\Sigma\mathbf{X}$) that eventually appears in the formulas – for instance in Eq. (3.34).

We have:

$$\frac{1}{n}\mathbf{X}'\Sigma\mathbf{X} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{i,j}\mathbf{x}_i\mathbf{x}_j'. \tag{3.36}$$

Robust estimation of asymptotic covariance matrices look for estimates of the previous matrix. Their computation is based on the fact that if $\mathbf{b}$ is consistent, then the e_is are consistent (pointwise) estimators of the ε_is.

Example 3.5. Heteroskedasticity.

This is the case of Eq. (3.28).

We then need to estimate $\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2\mathbf{x}_i\mathbf{x}_i'$. White (1980): Under general conditions:

$$\text{plim}\left(\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2\mathbf{x}_i\mathbf{x}_i'\right) = \text{plim}\left(\frac{1}{n}\sum_{i=1}^{n}e_i^2\mathbf{x}_i\mathbf{x}_i'\right). \tag{3.37}$$

The estimator of $\frac{1}{n}\mathbf{X}'\Sigma\mathbf{X}$ therefore is:

$$\frac{1}{n}\mathbf{X}'\mathbf{E}^2\mathbf{X},$$

where $\mathbf{E}$ is an $n \times n$ diagonal matrix whose diagonal elements are the estimated residuals e_i.

Illustration: Figure 3.2.

Let us illustrate the influence of heteroskedasticity using simulations.

We consider the following model:

$$y_i = x_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, x_i^2).$$

where the x_is are i.i.d. $t(4)$.

Here is a simulated sample $(n = 200)$ of this model:

```
n <- 200
x <- rt(n,df=5)
y <- x + x*rnorm(n)
plot(x,y,pch=19)
```
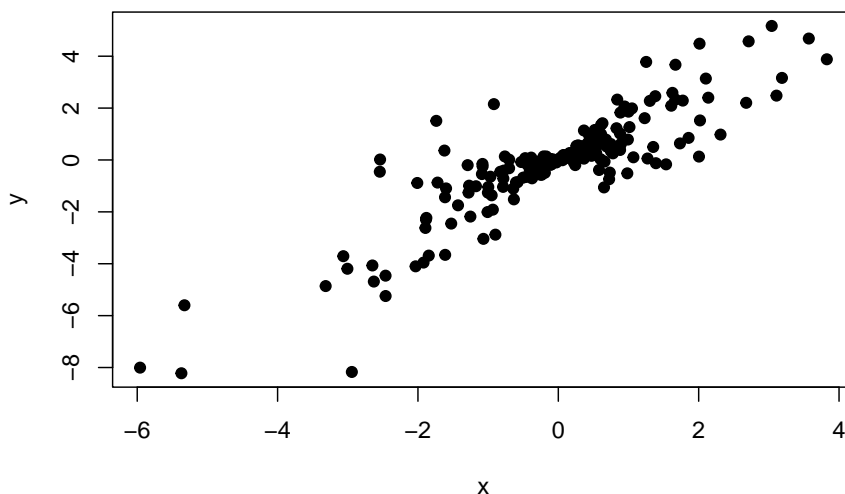

We simulate 1000 samples of the same model with $n = 200$. For each sample, we compute the OLS estimate of β (=1). Using these 1000 estimates of b, we construct an approximated *(kernel-based) distribution of this OLS estimator* (in red on the figure).

For each of the 1000 OLS estimations, we employ *the standard OLS variance formula $(s^2(\mathbf{X}'\mathbf{X})^{-1})$* to estimate the variance of b. The blue curve is a normal distribution centred on 1 and whose variance is the average of the 1000 previous variance estimates.

The variance of the simulated b is of 0.040 (that is the *true* one); the average of the estimated variances based on the standard OLS formula is of 0.005 (*bad*

estimate); the average of the estimated variances based on the White robust covariance matrix is of 0.030 (better estimate).

The standard OLS formula for the variance of b overestimates the precision of this estimator.

For almost 50% of the simulations, 1 is not included in the 95% confidence interval of β when the computation of the interval is based on the standard OLS formula for the variance of b.

When the White robust covariance matrix is used, 1 is not in the 95% confidence interval of β for less than 10% of the simulations.

```r
n <- 200
N <- 1000
XX <- matrix(rt(n*N,df=5),n,N)
YY <- matrix(XX + XX*rnorm(n),n,N)
all_b        <- NULL
all_V_OLS    <- NULL
all_V_White <- NULL
for(j in 1:N){
  Y <- matrix(YY[,j],ncol=1)
  X <- matrix(XX[,j],ncol=1)
  b <- solve(t(X)%*%X) %*% t(X)%*%Y
  e <- Y - X %*% b
  S <- 1/n * t(X) %*% diag(c(e^2)) %*% X
  V_OLS   <- solve(t(X)%*%X) * var(e)
  V_White <- 1/n * (solve(1/n*t(X)%*%X)) %*% S %*% (solve(1/n*t(X)%*%X))

  all_b        <- c(all_b,b)
  all_V_OLS    <- c(all_V_OLS,V_OLS)
  all_V_White <- c(all_V_White,V_White)
}
plot(density(all_b))
abline(v=mean(all_b),lty=2)
abline(v=1)
x <- seq(0,2,by=.01)
lines(x,dnorm(x,mean = 1,sd = mean(sqrt(all_V_OLS))),col="blue")
lines(x,dnorm(x,mean = 1,sd = mean(sqrt(all_V_White))),col="red")
```

density.default(x = all_b)

N = 1000   Bandwidth = 0.03209
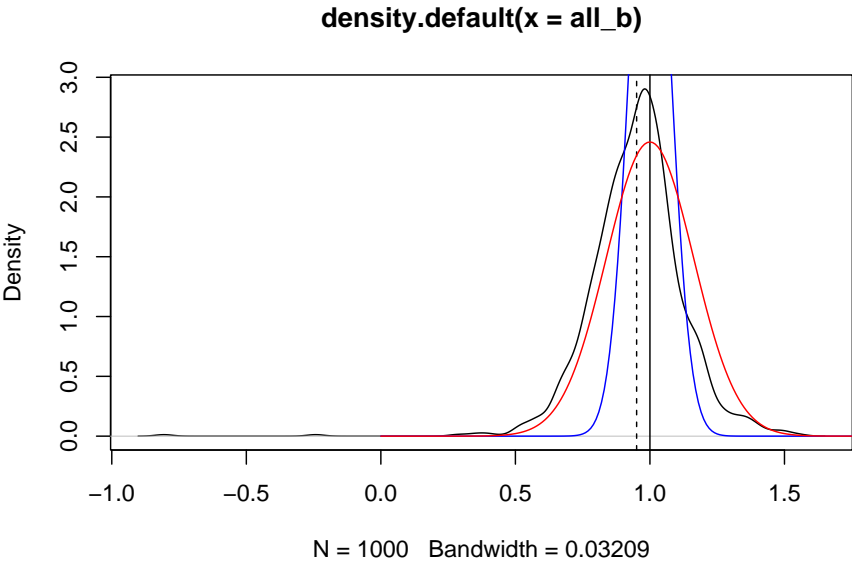
Chapter 4

Appendix

4.1 Statitical Tables

Table 4.1: Quantiles of the $\mathcal{N}(0,1)$ distribution. If a and b are respectively the row and column number; then the corresponding cell gives $\mathbb{P}(0 < X \leq a + b)$, where $X \sim \mathcal{N}(0,1)$.

	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5000	0.6179	0.7257	0.8159	0.8849	0.9332	0.9641	0.9821	0.9918	0.9965
0.1	0.5040	0.6217	0.7291	0.8186	0.8869	0.9345	0.9649	0.9826	0.9920	0.9966
0.2	0.5080	0.6255	0.7324	0.8212	0.8888	0.9357	0.9656	0.9830	0.9922	0.9967
0.3	0.5120	0.6293	0.7357	0.8238	0.8907	0.9370	0.9664	0.9834	0.9925	0.9968
0.4	0.5160	0.6331	0.7389	0.8264	0.8925	0.9382	0.9671	0.9838	0.9927	0.9969
0.5	0.5199	0.6368	0.7422	0.8289	0.8944	0.9394	0.9678	0.9842	0.9929	0.9970
0.6	0.5239	0.6406	0.7454	0.8315	0.8962	0.9406	0.9686	0.9846	0.9931	0.9971
0.7	0.5279	0.6443	0.7486	0.8340	0.8980	0.9418	0.9693	0.9850	0.9932	0.9972
0.8	0.5319	0.6480	0.7517	0.8365	0.8997	0.9429	0.9699	0.9854	0.9934	0.9973
0.9	0.5359	0.6517	0.7549	0.8389	0.9015	0.9441	0.9706	0.9857	0.9936	0.9974
1	0.5398	0.6554	0.7580	0.8413	0.9032	0.9452	0.9713	0.9861	0.9938	0.9974
1.1	0.5438	0.6591	0.7611	0.8438	0.9049	0.9463	0.9719	0.9864	0.9940	0.9975
1.2	0.5478	0.6628	0.7642	0.8461	0.9066	0.9474	0.9726	0.9868	0.9941	0.9976
1.3	0.5517	0.6664	0.7673	0.8485	0.9082	0.9484	0.9732	0.9871	0.9943	0.9977
1.4	0.5557	0.6700	0.7704	0.8508	0.9099	0.9495	0.9738	0.9875	0.9945	0.9977
1.5	0.5596	0.6736	0.7734	0.8531	0.9115	0.9505	0.9744	0.9878	0.9946	0.9978
1.6	0.5636	0.6772	0.7764	0.8554	0.9131	0.9515	0.9750	0.9881	0.9948	0.9979
1.7	0.5675	0.6808	0.7794	0.8577	0.9147	0.9525	0.9756	0.9884	0.9949	0.9979
1.8	0.5714	0.6844	0.7823	0.8599	0.9162	0.9535	0.9761	0.9887	0.9951	0.9980
1.9	0.5753	0.6879	0.7852	0.8621	0.9177	0.9545	0.9767	0.9890	0.9952	0.9981
2	0.5793	0.6915	0.7881	0.8643	0.9192	0.9554	0.9772	0.9893	0.9953	0.9981
2.1	0.5832	0.6950	0.7910	0.8665	0.9207	0.9564	0.9778	0.9896	0.9955	0.9982
2.2	0.5871	0.6985	0.7939	0.8686	0.9222	0.9573	0.9783	0.9898	0.9956	0.9982
2.3	0.5910	0.7019	0.7967	0.8708	0.9236	0.9582	0.9788	0.9901	0.9957	0.9983
2.4	0.5948	0.7054	0.7995	0.8729	0.9251	0.9591	0.9793	0.9904	0.9959	0.9984
2.5	0.5987	0.7088	0.8023	0.8749	0.9265	0.9599	0.9798	0.9906	0.9960	0.9984
2.6	0.6026	0.7123	0.8051	0.8770	0.9279	0.9608	0.9803	0.9909	0.9961	0.9985
2.7	0.6064	0.7157	0.8078	0.8790	0.9292	0.9616	0.9808	0.9911	0.9962	0.9985
2.8	0.6103	0.7190	0.8106	0.8810	0.9306	0.9625	0.9812	0.9913	0.9963	0.9986
2.9	0.6141	0.7224	0.8133	0.8830	0.9319	0.9633	0.9817	0.9916	0.9964	0.9986

Table 4.2: Quantiles of the Student-t distribution. The rows correspond to different degrees of freedom (ν, say); the columns correspond to different probabilities (z, say). The cell gives q that is s.t. $\mathbb{P}(-q < X < q) = z$, with $X \sim t(\nu)$.

	0.05	0.1	0.75	0.9	0.95	0.975	0.99	0.999
1	0.079	0.158	2.414	6.314	12.706	25.452	63.657	636.619
2	0.071	0.142	1.604	2.920	4.303	6.205	9.925	31.599
3	0.068	0.137	1.423	2.353	3.182	4.177	5.841	12.924
4	0.067	0.134	1.344	2.132	2.776	3.495	4.604	8.610
5	0.066	0.132	1.301	2.015	2.571	3.163	4.032	6.869
6	0.065	0.131	1.273	1.943	2.447	2.969	3.707	5.959
7	0.065	0.130	1.254	1.895	2.365	2.841	3.499	5.408
8	0.065	0.130	1.240	1.860	2.306	2.752	3.355	5.041
9	0.064	0.129	1.230	1.833	2.262	2.685	3.250	4.781
10	0.064	0.129	1.221	1.812	2.228	2.634	3.169	4.587
20	0.063	0.127	1.185	1.725	2.086	2.423	2.845	3.850
30	0.063	0.127	1.173	1.697	2.042	2.360	2.750	3.646
40	0.063	0.126	1.167	1.684	2.021	2.329	2.704	3.551
50	0.063	0.126	1.164	1.676	2.009	2.311	2.678	3.496
60	0.063	0.126	1.162	1.671	2.000	2.299	2.660	3.460
70	0.063	0.126	1.160	1.667	1.994	2.291	2.648	3.435
80	0.063	0.126	1.159	1.664	1.990	2.284	2.639	3.416
90	0.063	0.126	1.158	1.662	1.987	2.280	2.632	3.402
100	0.063	0.126	1.157	1.660	1.984	2.276	2.626	3.390
200	0.063	0.126	1.154	1.653	1.972	2.258	2.601	3.340
500	0.063	0.126	1.152	1.648	1.965	2.248	2.586	3.310

Table 4.3: Quantiles of the χ^2 distribution. The rows correspond to different degrees of freedom; the columns correspond to different probabilities.

	0.05	0.1	0.75	0.9	0.95	0.975	0.99	0.999
1	0.004	0.016	1.323	2.706	3.841	5.024	6.635	10.828
2	0.103	0.211	2.773	4.605	5.991	7.378	9.210	13.816
3	0.352	0.584	4.108	6.251	7.815	9.348	11.345	16.266
4	0.711	1.064	5.385	7.779	9.488	11.143	13.277	18.467
5	1.145	1.610	6.626	9.236	11.070	12.833	15.086	20.515
6	1.635	2.204	7.841	10.645	12.592	14.449	16.812	22.458
7	2.167	2.833	9.037	12.017	14.067	16.013	18.475	24.322
8	2.733	3.490	10.219	13.362	15.507	17.535	20.090	26.124
9	3.325	4.168	11.389	14.684	16.919	19.023	21.666	27.877
10	3.940	4.865	12.549	15.987	18.307	20.483	23.209	29.588
20	10.851	12.443	23.828	28.412	31.410	34.170	37.566	45.315
30	18.493	20.599	34.800	40.256	43.773	46.979	50.892	59.703
40	26.509	29.051	45.616	51.805	55.758	59.342	63.691	73.402
50	34.764	37.689	56.334	63.167	67.505	71.420	76.154	86.661
60	43.188	46.459	66.981	74.397	79.082	83.298	88.379	99.607
70	51.739	55.329	77.577	85.527	90.531	95.023	100.425	112.317
80	60.391	64.278	88.130	96.578	101.879	106.629	112.329	124.839
90	69.126	73.291	98.650	107.565	113.145	118.136	124.116	137.208
100	77.929	82.358	109.141	118.498	124.342	129.561	135.807	149.449
200	168.279	174.835	213.102	226.021	233.994	241.058	249.445	267.541
500	449.147	459.926	520.950	540.930	553.127	563.852	576.493	603.446

Table 4.4: Quantiles of the $\mathcal{F}$ distribution. The columns and rows correspond to different degrees of freedom (resp. n_1 and n_2). The different panels correspond to different probabilities (α) The corresponding cell gives z that is s.t. $\mathbb{P}(X \leq z) = \alpha$, with $X \sim \mathcal{F}(n_1, n_2)$.

	1	2	3	4	5	6	7	8	9	10
alpha = 0.9										
5	4.060	3.780	3.619	3.520	3.453	3.405	3.368	3.339	3.316	3.297
10	3.285	2.924	2.728	2.605	2.522	2.461	2.414	2.377	2.347	2.323
15	3.073	2.695	2.490	2.361	2.273	2.208	2.158	2.119	2.086	2.059
20	2.975	2.589	2.380	2.249	2.158	2.091	2.040	1.999	1.965	1.937
50	2.809	2.412	2.197	2.061	1.966	1.895	1.840	1.796	1.760	1.729
100	2.756	2.356	2.139	2.002	1.906	1.834	1.778	1.732	1.695	1.663
500	2.716	2.313	2.095	1.956	1.859	1.786	1.729	1.683	1.644	1.612
alpha = 0.95										
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348
50	4.034	3.183	2.790	2.557	2.400	2.286	2.199	2.130	2.073	2.026
100	3.936	3.087	2.696	2.463	2.305	2.191	2.103	2.032	1.975	1.927
500	3.860	3.014	2.623	2.390	2.232	2.117	2.028	1.957	1.899	1.850
alpha = 0.99										
5	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158	10.051
10	10.044	7.559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849
15	8.683	6.359	5.417	4.893	4.556	4.318	4.142	4.004	3.895	3.805
20	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.457	3.368
50	7.171	5.057	4.199	3.720	3.408	3.186	3.020	2.890	2.785	2.698
100	6.895	4.824	3.984	3.513	3.206	2.988	2.823	2.694	2.590	2.503
500	6.686	4.648	3.821	3.357	3.054	2.838	2.675	2.547	2.443	2.356

Chapter 5

Macro data

Use `haven` package to read dta data.

5.1 Blabla

Below is an example borrowed from Petersen.

```
## Petersen's data
data("PetersenCL", package = "sandwich")
m <- lm(y ~ x, data = PetersenCL)

## clustered covariances
## one-way
vcovCL(m, cluster = ~ firm)
```

```
##               (Intercept)            x
## (Intercept)  4.490702e-03 -6.473517e-05
## x           -6.473517e-05  2.559927e-03
```

```
vcovCL(m, cluster = PetersenCL$firm) ## same
```

```
##               (Intercept)            x
## (Intercept)  4.490702e-03 -6.473517e-05
## x           -6.473517e-05  2.559927e-03
```

```
## one-way with HC2
vcovCL(m, cluster = ~ firm, type = "HC2")
```

```
##               (Intercept)             x
## (Intercept)  4.494487e-03 -6.592912e-05
## x            -6.592912e-05  2.568236e-03
```

```
## two-way
vcovCL(m, cluster = ~ firm + year)
```

```
##               (Intercept)             x
## (Intercept)  4.233313e-03 -2.845344e-05
## x            -2.845344e-05  2.868462e-03
```

```
vcovCL(m, cluster = PetersenCL[, c("firm", "year")]) ## same
```

```
##               (Intercept)             x
## (Intercept)  4.233313e-03 -2.845344e-05
## x            -2.845344e-05  2.868462e-03
```

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:
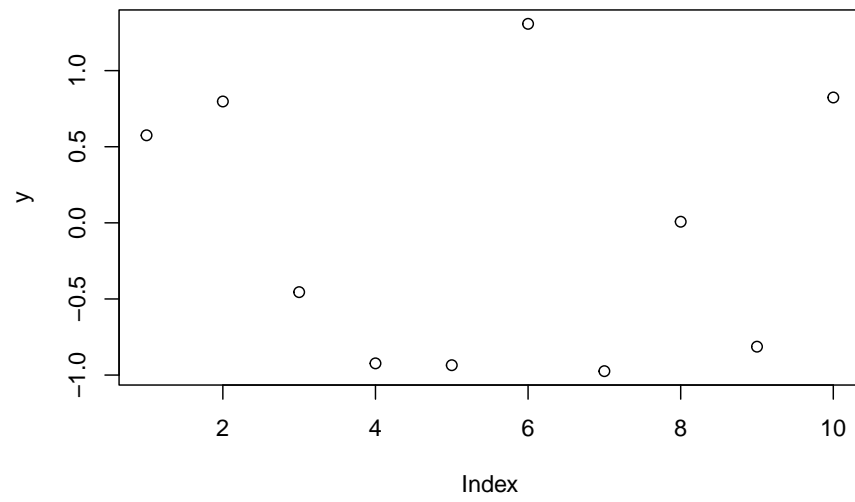
Chapter 6

Loading the dataset

```r
1+1
```

```
## [1] 2
```

```r
f <- function(x){
  2*x + 2
}
y <- rnorm(10)
plot(y)
```

Chapter 7

Question 7

```
summary(cars)
```

```
##      speed          dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Theorem 7.1. *This is a* `theorem` *environment that can contain **any** Markdown syntax.*

Proof. Nice proof. □

How is Theorem 7.1?

Chapter 8

Including Plots

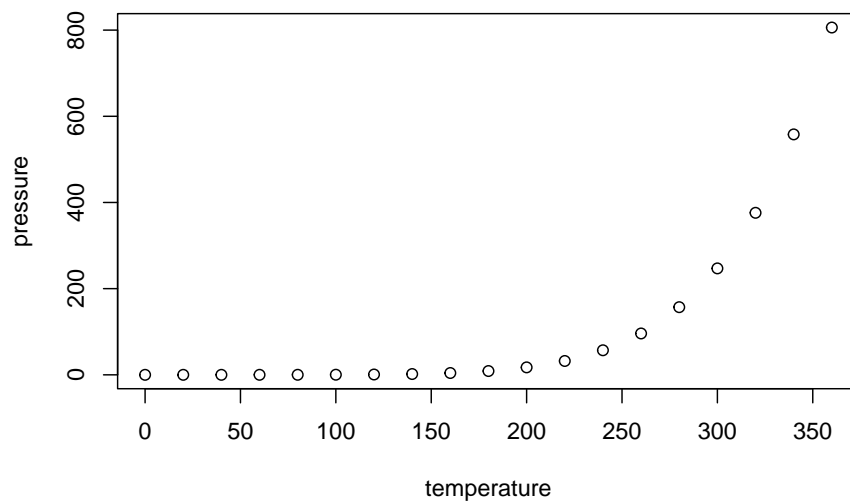You can also embed plots, for example:

Figure 8.1: This is the caption

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot. Figure 8.1 is nice!

$$F(x) \quad = \quad 1 \tag{8.1}$$
$$f(y) \quad = \quad 2 \tag{8.2}$$
$$\int_{x=0}^{\infty} f(x)dx \quad = \quad 0 \tag{8.3}$$

This was equation (8.4). See also Section 8.

$$F(x) = 1 \tag{8.4}$$

```
f <- function(x){3*x}
f(3)
```

```
## [1] 9
```

- This is a list.
- This website is useful.
- The analysis builds on **?**

```
a <- 3
print(a)
```

```
## [1] 3
```

```
2*a
```

```
## [1] 6
```

Chapter 9

Question 7

Create some important variables for the reference person such as dummy variables for female ($female$, 1 if female, 0 otherwise); race ($black$, 1 if black); retired ($retired$, 1 if labour force participation $= 5$); disabled (1 if labour force participation $= 6$); not married ($notmar$, 1 if marital status is different from being married); full-time work ($workft$, 1 if works full time); pat-time work ($workpt$, 1 if works part time).

```
1+1
```

```
## [1] 2
```

Chapter 10

Appendix

Proposition 10.1. *We have:*

$$f_{X|Y}(x, y) = \frac{f_{XY}(x, y)}{f_Y(y)}.$$

Proof. We have:

$$
\begin{aligned}
f_{X|Y}(x, y) &= \lim_{\varepsilon \to 0} \frac{\mathbb{P}(x < X \le x + \varepsilon | Y = y)}{\varepsilon} \\
&= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \mathbb{P}(x < X \le x + \varepsilon | y < Y \le y + \varepsilon) \\
&= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \frac{\mathbb{P}(x < X \le x + \varepsilon, y < Y \le y + \varepsilon)}{\mathbb{P}(y < Y \le y + \varepsilon)} \\
&= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \frac{\varepsilon^2 f_{XY}(x, y)}{\varepsilon f_Y(y)}. \square
\end{aligned}
$$

$\square$

Chapter 11

References

Chapter 12

Methods

We describe our methods in this chapter.

Math can be added in body using usual syntax like this

12.1 math example

p is unknown but expected to be around 1/3. Standard error will be approximated

$$SE = \sqrt{(\frac{p(1-p)}{n})} \approx \sqrt{\frac{1/3(1-1/3)}{300}} = 0.027$$

You can also use math in footnotes like this[1].

We will approximate standard error to 0.027[2]

---

[1] where we mention $p = \frac{a}{b}$

[2] p is unknown but expected to be around 1/3. Standard error will be approximated

$$SE = \sqrt{(\frac{p(1-p)}{n})} \approx \sqrt{\frac{1/3(1-1/3)}{300}} = 0.027$$

Chapter 13

Applications

Some *significant* applications are demonstrated in this chapter.

13.1   Example one

13.2   Example two

Chapter 14

Final Words

We have finished a nice book.

Bibliography

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.

Xie, Y. (2022). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.27.