Omitted variables problem

Foundations of Statistical Learning - Homework 1

Edoardo Canti

Purpose

Set up a Monte Carlo experiment to understand the consequences of the omission of a relevant covariate

Data Generating Process

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon$$

$$\epsilon \sim N(0,1)$$

$$x_1, x_2 \sim N(\underline{0}, \Sigma)$$

scenario 1: $\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$

scenario 2: $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

Parameters Estimation

Model m0: $y_i = \beta_0 + \beta_1 x_{1i} + \epsilon$

```
m0 <- lm(y ~ X1)
```

Model m1: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon$

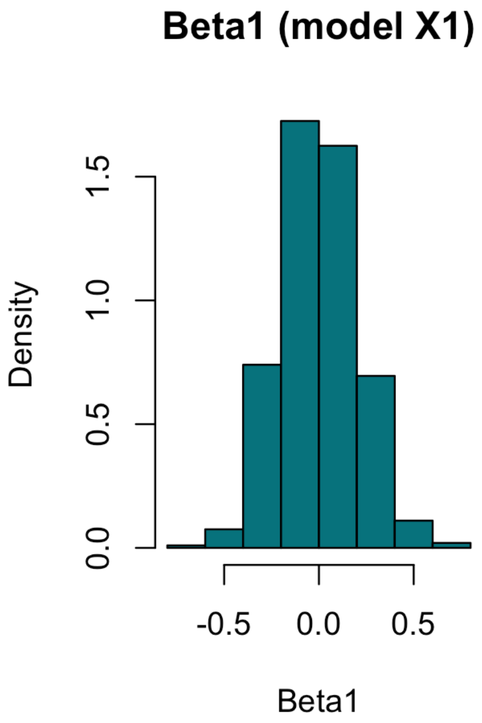
```
m1 <- lm(y ~ X1 + X2)
```

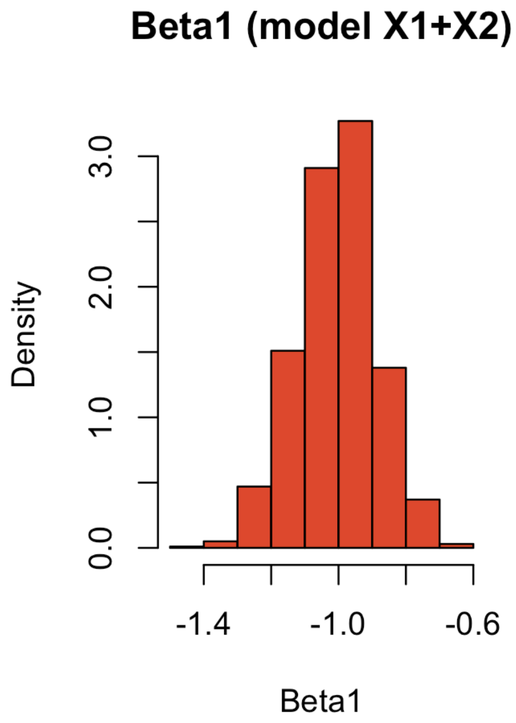
Experimental results

Number of samples (n_samples): 100

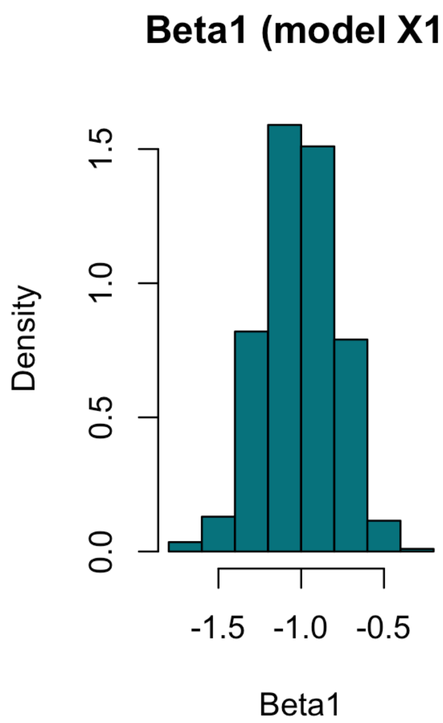
Number of MC simulations (sim_num): 1000

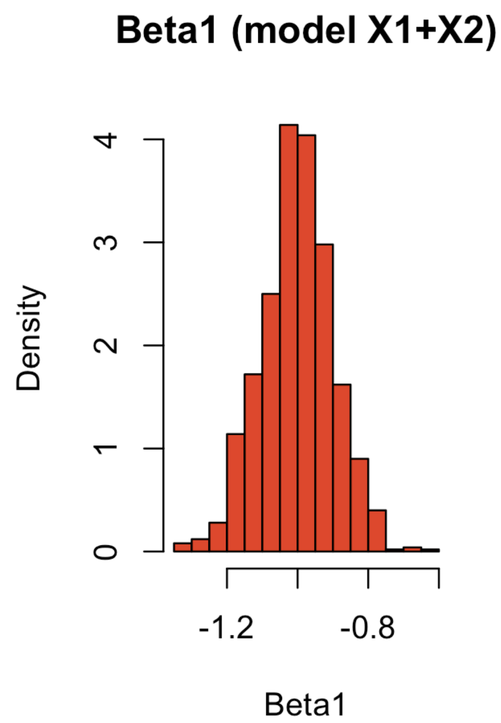
$\beta_0 = 1 \qquad \beta_1 = -1 \qquad \beta_2 = 2$





Scenario 1





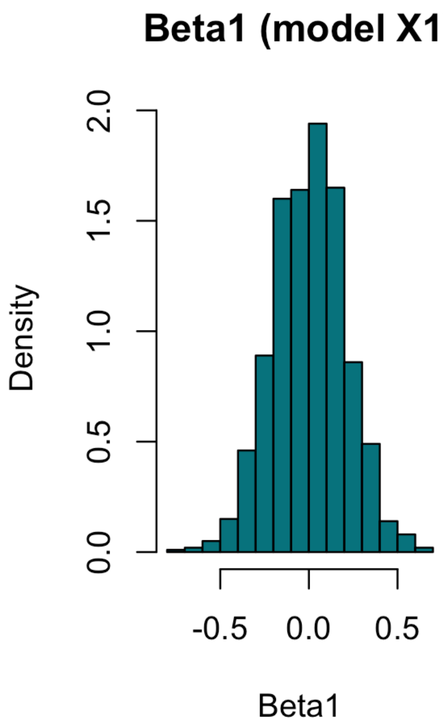
Scenario 2

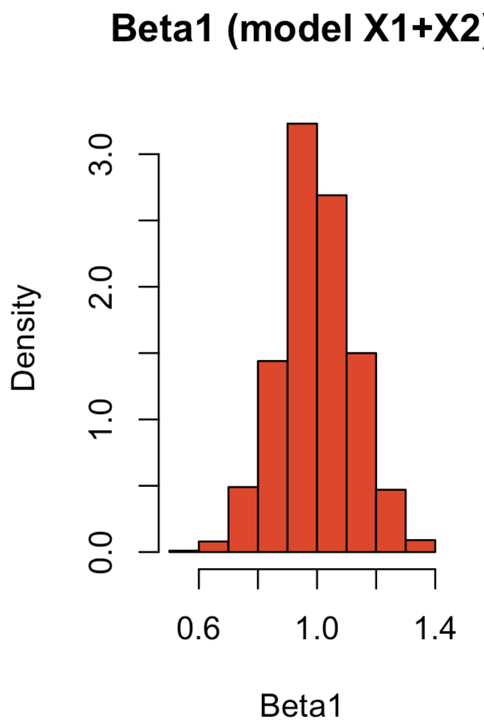
Experimental results on suggested dgp params

Number of samples (n_samples): 100

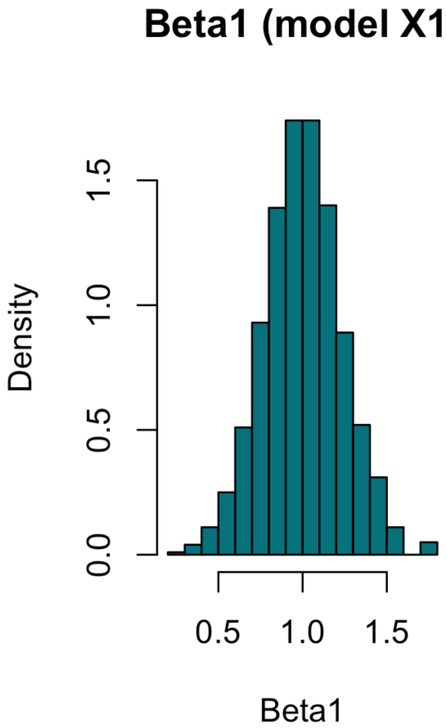
Number of MC simulations (sim_num): 1000

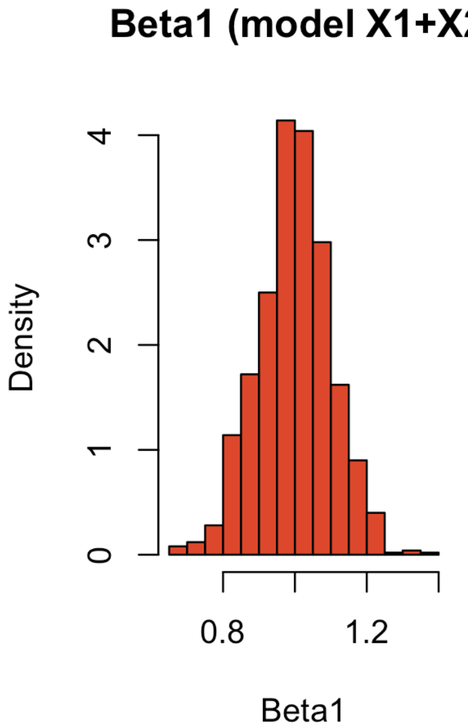
$\beta_0 = 1 \qquad \beta_1 = 1 \qquad \beta_2 = -2$





Scenario 1





Scenario 2