Advanced Econometrics HWA1

Noel Waldenfjord

January 2025

1 1

1.1 1a

The given model is not in linear form. However, since we can change the expression to make all the beta coefficients appear in linear form, we can transform it into a linear regression model. This transformation is done by taking the logarithm on both sides as shown below:

$$y_i = e^{\beta_0} e^{\beta_1 x_i} x_i^{\beta_2} e^{\varepsilon_i} \tag{1}$$
$$\ln y_i = \ln(e^{\beta_0} e^{\beta_1 x_i} x_i^{\beta_2} e^{\varepsilon_i}) \tag{2}$$
$$\ln y_i = \beta_0 + \beta_1 x_i + \beta_2 \ln x_i + \varepsilon_i \tag{3}$$

As shown in equation (3) all parameters are in linear form hence the model is a linear regression model. This transformation is needed for us to compute the least square estimates using OLS. Other assumptions we do to use OLS is that we also assume that the error terms are homoskedastic, error are uncorrelated across observations, and that we have no multicollinearity..

To compute the least square estimates of the parameters $\boldsymbol{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}$ we use the following expression:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}' \ln \mathbf{y}) \tag{4}$$

after computing this in R by using the transformed y variable we get that $\hat{\beta}_0 = 5.362382$, $\hat{\beta}_1 = -0.4490397$ and $\hat{\beta}_2 = 1.114141$

1.2 1b

The fitted regression model for general cases is computed by

$$\hat{y}_i = \boldsymbol{X^T} \hat{\boldsymbol{\beta}} \tag{5}$$

Hence, in our case it becomes:

$$\hat{y}_i = e^{\hat{\beta}_0} e^{\hat{\beta}_1 x_i} x_i^{\hat{\beta}_2} \tag{6}$$

We plot the data in a scatterplot together with a fitted line for the fitted values of $\hat{y}$ Which is shown in figure 1

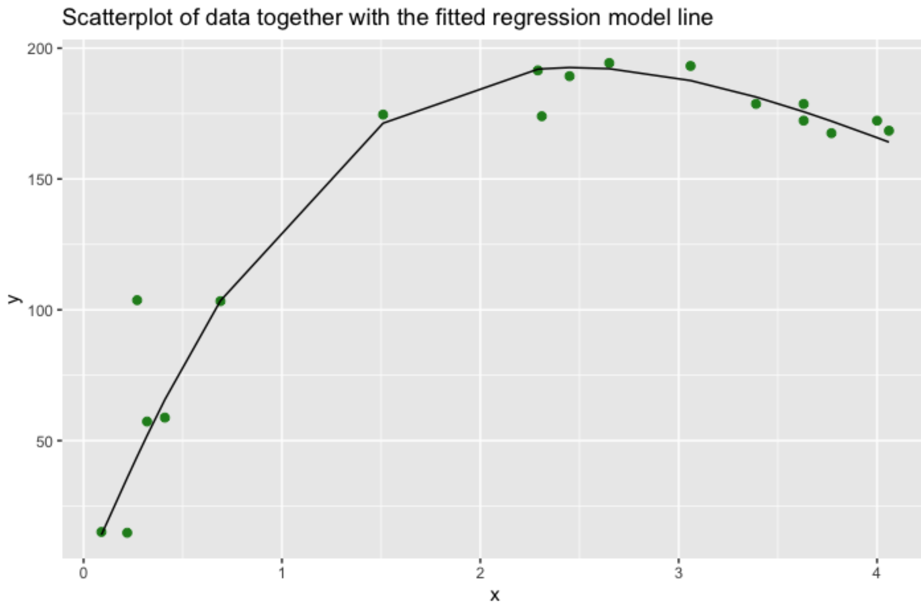

Figure 1: Scatterplot of data together with a fitted line for the fitted values of $\hat{y}$

1.3 1c

To find the value of x that maximizes
$\hat{y_i} = e^{\hat{\beta_0}} e^{\hat{\beta_1 x}} x^{\hat{\beta_2}}$
we can use maximum likelihood estimation since all the estimated coefficients are known

$$\hat{y_i} = e^{\hat{\beta}_0} e^{\hat{\beta}_1 x} x^{\hat{\beta}_2} \tag{7}$$

$$\log \hat{y_i} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 \tag{8}$$

$$\frac{d}{dx} \log \hat{y_i} = \hat{\beta}_1 + \frac{\hat{\beta}_2}{x} \tag{9}$$

$$\tag{10}$$

Setting the score function equal to zero gives:

$$x = -\frac{\hat{\beta}_2}{\hat{\beta}_1} \tag{11}$$

To check that it is a maximum we also compute second derivative:

$$\frac{d^2}{dx^2} \log \hat{y_i} = -\frac{\hat{\beta}_2}{x^2} \tag{12}$$

From the previous computed estimates of the coefficients we get that the x that maximizes $\hat{y_i}$ is 2.481163. And since $\hat{\beta}_2$ is positive the second derivative must be negative, hence it is indeed a maximum.

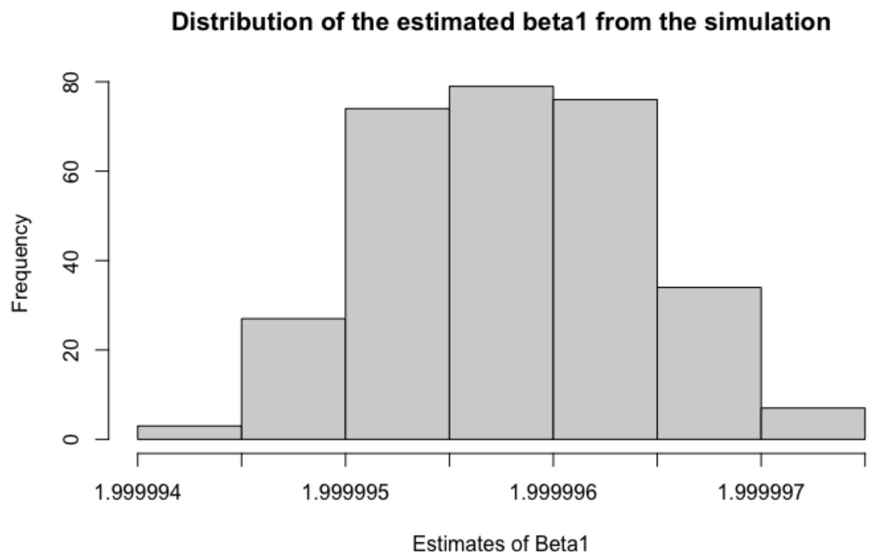

Figure 2: Histogram of estimated β_1^s

2 2

In the exercise it is given that the true data generating process is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \tag{13}$$

Such that $\boldsymbol{\varepsilon}$ is $N_n(\mathbf{0}, \sigma^2, \boldsymbol{I})$
The parameters of $\boldsymbol{\beta} = [\beta_0 \beta_1 \beta_2]$ and σ^2 is chosen as:

$$\beta_0 = 1 \ \beta_1 = 2, \ \beta_2 = 5 \ \sigma^2 = 1$$

The values of the beta coefficients are chosen arbitrarily. We will then consider a miss specified model for the fitting to the simulated data:

$$y = \beta_0 + \beta_1 x_1 + \varepsilon \tag{14}$$

Which is a model that lacks the variable x_2 and therefore should introduce omitted variable bias.

2.1 2a

After the simulation and estimating β_1 300 times we get the following histogram in Figure 2. And the mean is computed from all the estimated coefficients using formula (4) but without the logarithm of y.

$$\bar{\beta}_1^s = 1.99999579 \tag{15}$$

From the histogram we can see that $\bar{\beta}_1^s$ differs from the true β_1 since the true value is 2, however we cannot only by looking at the plot see if it is significant or not. To be sure that it is significant and that there is omtted variable bias we will therefore proceed by constructing a confidence interval around $\bar{\beta}_1^s$

2.2 2b

If we use the significance level of 0.05 for the confidence interval we can compute the upper and lower bound by the following expressions:

$$CI_U = \bar{\beta}_1^s + 1.96 * \mathrm{SE}(\bar{\beta}_1^s) \tag{16}$$
$$CI_L = \bar{\beta}_1^s - 1.96 * \mathrm{SE}(\bar{\beta}_1^s) \tag{17}$$

Where $\mathrm{SE}(\hat{\beta}_1^s)$ is computed by:

$$SE(\bar{\beta_1^s}) = \frac{\mathrm{SD}(\bar{\beta}_1^s)}{\sqrt{300}} \tag{18}$$

Hence, the final Confidence interval is

$$CI_L < \bar{\beta}_1^s < CI_U \tag{19}$$
$$= 1.99999572 < \bar{\beta}_1^s < 1.99999586 \tag{20}$$

As $\beta_1 = 2$, the confidence interval does not cover the true value and we can therefore conclude that $\bar{\beta}_1^s$ differs from true β_1 significantly and we have omitted variable bias.