UPPSALA UNIVERSITET

- Homework Assignment 1

Advanced Econometrics 2ST123

Homework Assignment 1

Author: Viktor Eriksson

Department of Statistics, Uppsala University

2026-01-26

Homework Assignment 1

Please note that this assignment is Mandatory. You are to work in groups, at most 4, with this assignment.

This assignment will require of you to use R for solving all programming related tasks.

What you should hand in is code as an R-file and separately a pdf file of your report. Your report contains your results, interpretations and explanations. Your R-file contains your code. I am supposed to be able to run your code without errors. You may assume that I already have the cps09mar.txt data set.

The assignment is handed in no later than 13/02/2026 at midnight. You have to hand in your assignment on studium!

Throughout all questions, you must supply full explanations and show how you came to your conclusions, formulas and mathematics included. Figures needs to be numbered and have title, captions and labeled axises. Your text-file should be maximum 6 pages, excluding cover page.

Homework Assignment 1

1) Consider the model $y_i = e^{\beta_0} e^{\beta_1 x_i} x_i^{\beta_2} e^{\varepsilon_i}$ where ε_i is an error term.

(a) Compute the least square estimate of the parameters $\boldsymbol{\beta} = [\beta_0 \ \ \beta_1 \ \ \beta_2]^T$. Do not use any in-built or package regression program. Program your own code based on formulas that uses matrix expressions of the data (get familiar with coding manually with matrices / data frames).

(b) Create a figure including a scatter plot of the data and a fitted regression model $e^{\widehat{\beta}_0} e^{\widehat{\beta_1 x}} x^{\widehat{\beta}_2}$ against x, all in the same figure. Use title, captions and labeled axises.

(c) Find analytically the value x that maximizes $\hat{y}_i = e^{\widehat{\beta}_0} e^{\widehat{\beta_1 x}} x^{\widehat{\beta}_2}$. Show solution.

UPPSALA UNIVERSITET

- Homework Assignment 1

Homework Assignment 1

Data set in assignment 1.

$$\begin{aligned}
\boldsymbol{x}^T &= [4.06,\ 3.63,\ 0.41,\ 2.45,\ 0.22,\ 3.39,\ 0.27,\ 0.69,\ 0.32,\\
&\quad 0.09,\ 4.00,\ 3.63, 2.65, 3.06, 3.77, 2.29, 2.31, 1.51]
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{y}^T &= [168.44,\ 178.66,\ 58.81,\ 189.29,\ 14.79,\ 178.70,\ 103.69,\ 103.30,\ 57.30,\\
&\quad 15.10,\ 172.28,\ 172.30,\ 194.32,\ 193.20,\ 167.50,\ 191.50,\ 173.98,\ 174.60]
\end{aligned}$$

Homework Assignment 1

(2) Investigate the consequences of model misspecification with a Monte Carlo simulation study. Do not use any in-built- or package regression program. Program your own code based on formulas that uses matrix expressions of the data. Follow the steps (i)-(v).

(i) *(Omitted Variable Bias)* Let the true data generating process be

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \quad \text{where} \quad \boldsymbol{\varepsilon} \sim \mathrm{N}_n(\mathbf{0}, \sigma^2 \boldsymbol{I}).$$

Let parameter values be $\boldsymbol{\beta} = [\beta_0 \;\; \beta_1 \;\; \beta_2 \;\; \beta_3] = [20, 0.03, 0.7, 1.4]$ and $\sigma^2 = 6$.

Homework Assignment 1

UPPSALA UNIVERSITET

- Homework Assignment 1

(ii) To generate simulated values for y a fixed data set for the explanatory variables $\boldsymbol{X}$ is used. Use cps09mar.txt data set. Choose earnings for x_1, hours for x_2 and education for x_3.

Generate simulated residuals. Enter everything into the true data generating process to generate simulated values for y.

(iii) Let the replication size be $s = 300$.

(iv) Consider the case of fitting, to simulated data, the misspecified model

$$y = \beta_0 + \beta_1 x_1 + \beta_3 x_3 + \varepsilon.$$

This model evidently lack the variable x_2.

UPPSALA UNIVERSITET

- Homework Assignment 1

Homework Assignment 1

(v) Estimate the parameters of misspecified model all $s = 300$ times with your simulated data sets, and plot the results in a histogram. What is the average $\bar{\beta}_1^s$? Does it differ significantly from true β_1? Compute the sample standard errors of the simulated histogram and compute confidence interval centered around $\bar{\beta}_1^s$. Does the confidence interval cover true β_1?

Notice that in this Monte Carlo simulation study then $\boldsymbol{X}$ is treated as a matrix of fixed constants. Therefore the only source of random variation in $\boldsymbol{y}$, according to the DGP, is from the error term $\boldsymbol{\varepsilon}$. Thus we view $\boldsymbol{y}$ as a transformation of $\boldsymbol{\varepsilon}$. This point of view was championed by the Norwegian researcher Trygve Haavelmo (1911-1999). The Dutch Jan Tinbergen, and the Norwegian Ragnar Frisch, are the founding fathers of econometrics. Frisch coined the word *econometrics*. Haavelmo was the doctoral student of Frisch. One of the main contributions of Haavelmo was incorporating statistical methodology into the field of econometrics. The philosophy of Haavelmo for regarding $\boldsymbol{y}$ as a transformation of ε, with $\boldsymbol{X}$ fixed, has not been challanged up to the present, as far as I myself am aware of.

Homework Assignment 1

UPPSALA UNIVERSITET

- Homework Assignment 1

(3) Consider the nonlinear regression model $y_i = m_i(\boldsymbol{\beta}) + e_i$ with i.i.d. innovation terms $e_i \sim \mathsf{N}(0, \sigma^2)$ for $n = 1, 2, \ldots, N$. You may assume that the variance σ^2 is known. The nonlinear function $m_i(\boldsymbol{\beta}) \equiv m(\boldsymbol{x}_i, \boldsymbol{\beta})$ is a model for the conditional expectation function $\mathbb{E}[y \,|\, \boldsymbol{x}]$ and as such it is a scalar-worth function dependent on the random p-vector $\boldsymbol{x}_i$ i.e. $m : \mathbb{R}^p \mapsto \mathbb{R}$. By stacking observations we get $\boldsymbol{y} = \boldsymbol{m}(\boldsymbol{\beta}) + \boldsymbol{e}$ where $\boldsymbol{y}$, $\boldsymbol{m}$ and $\boldsymbol{e}$ are column n-vectors. Define the $n \times p$ dimensional Jacobian matrix $\boldsymbol{J}(\boldsymbol{\beta}) = [J_{ij}(\boldsymbol{\beta})]$ where $J_{ij}(\boldsymbol{\beta})$ is the matrix element at the ith row and jth column such that:

$$J_{ij}(\boldsymbol{\beta}) = \frac{\partial m_i(\boldsymbol{\beta})}{\partial \beta_j}, \tag{1}$$

$$\boldsymbol{J}(\boldsymbol{\beta}) = \frac{\partial \boldsymbol{m}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T}. \tag{2}$$

Homework Assignment 1

(a) Show that $\boldsymbol{\beta}$ such that $\mathbb{E}[\boldsymbol{e}^T\boldsymbol{e}]$ is minimised, implies the orthogonality condition $\boldsymbol{J}(\boldsymbol{\beta})^T(\boldsymbol{y}-\boldsymbol{m}(\boldsymbol{\beta}))=\mathbf{0}$.

(b) Is the OLS estimator is a special case of the NLS estimator? Show it.

(c) Consider the nonlinear regression model $y_i = m_i(\boldsymbol{\beta}) + e_i$ with i.i.d. innovation terms $e_i \sim \mathrm{N}(0,\sigma^2)$ for $n = 1, 2, \ldots, N$ and where you may assume that the variance σ^2 is known. Suppose the model for the conditional expectation function is $m_i(\boldsymbol{\beta}) = \beta_0 \mathrm{exp}[\beta_1 x_{i1} + \beta_2 x_{i2}]$.
(i) Mathematically derive the Jacobian $\boldsymbol{J}(\boldsymbol{\beta})$.
(ii) Is the model identifiable? Assume that the matrix with rows $(1, x_{i1}, x_{i2})$ has full column rank. (Hint: Identifiability is necessary to guarantee the existence of unique parameter values. Identifiability means that $f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\beta}_1) = f(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\beta}_2) \Rightarrow \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$. But with normal distributed error terms with known homoskedastic innovation variance, the identifiability condition simplifies into $\mathbb{E}_{\boldsymbol{\beta}_1}(\boldsymbol{y}|\boldsymbol{x}) = \mathbb{E}_{\boldsymbol{\beta}_2}(\boldsymbol{y}|\boldsymbol{x}) \Rightarrow \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$.)

UPPSALA UNIVERSITET

- Homework Assignment 1