BI Norwegian Business School

Group exam: **GRA 6039 – Econometrics with programming**

With: **Emil Aas Stoltenberg**

Handed out: **November 13, 12:00.**

Due date: **November 20, 12:00.**

Permitted aids: **Lecture notes, books, Google, stackoverflow, etc.**

Impermissible aids: **People not in your group.**

Group size: **One, two, or three.**

Format for your answer: **A .pdf-file with text, mathematics, and Matlab code. You are required to write on a machine, Word, Latex, or the like. You do not need to hand in your .m-file, but the Matlab code you use to solve the exam should be included at the end of the .pdf-document you hand in.**

Instructions: **Brevity is beautiful. Be as concise as you can.**

This exam set contains three exercises and comprises five pages.

Exercise 1. The plot in Figure 1 shows data that stem from a model of the form

$$Y_i = g_K(x_i) + \varepsilon_i, \quad \text{for } i = 1, \dots, n,$$

where $\varepsilon_1, \dots, \varepsilon_n$ are independent random variables with mean zero and $\mathrm{Var}(\varepsilon_i) = \sigma_i^2$ for $i = 1, \dots, n$, and the $x_1, \dots, x_n$ are fixed numbers (not random variables), and

$$g_K(x) = \sum_{j=0}^{K} \beta_j x^j, \quad \text{for some } K \geq 0.$$

You can find these data in the file `ex1_data.txt`, and read them into Matlab by writing

```
data = readtable("ex1_data.txt");
```

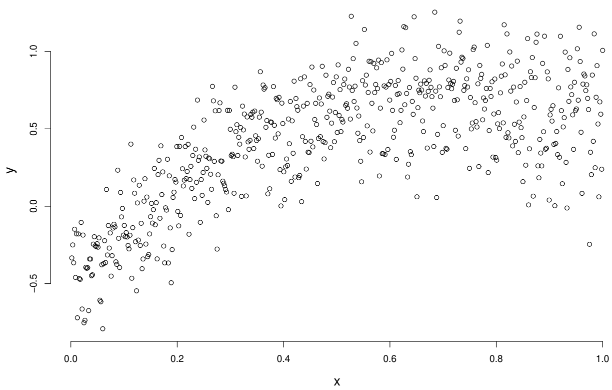

Figure 1. A plot of the data in the dataset `ex1_data.txt`.

(a) Pick a linear regression model that you think is suitable for these data. In other words, what K do you think is the correct one for $g_K(x)$? In two lines, explain why you chose the $g_K(x)$ function that you chose.

(b) To your chosen $g_K(x)$ function there is a corresponding design matrix X given by

$$X = \begin{pmatrix} 1 & x_1 & x_1^2 \cdots & x_1^K \\ 1 & x_2 & x_2^2 \cdots & x_2^K \\ & \vdots & & \\ 1 & x_n & x_n^2 \cdots & x_n^K \end{pmatrix}.$$

Please write down your X. Let $\widehat{\beta} = (\widehat{\beta}_0, \ldots, \widehat{\beta}_K)^{\mathrm{t}}$ and $Y = (Y_1, \ldots, Y_n)^{\mathrm{t}}$, with t standing for transpose. Provide an expression for the $(K+1) \times n$ matrix H that ensures that

$$\widehat{\beta} = HY,$$

minimises $h(\beta_0, \ldots, \beta_K) = \sum_{i=1}^n (Y_i - \sum_{j=0}^K \beta_j x_i^j)^2$, for the $g_K(x) = \sum_{j=0}^K \beta_j x^j$ you chose in (a).

(c) Show that $\widehat{\beta}$ is unbiased.

(d) Read the data in `ex1_data.txt` into Matlab and estimate β using the estimator $\widehat{\beta} = HY$. Make a little table where you present your estimates of $\widehat{\beta}_0, \ldots, \widehat{\beta}_K$, along with estimated standard errors computed under the assumption that $\mathrm{Var}(\varepsilon_1) = \cdots = \mathrm{Var}(\varepsilon_n)$.

(e) Let $\widehat{g}_K(x) = \sum_{j=0}^K \widehat{\beta}_j x^j$ be the estimate of your chosen $g_K(x)$ Reproduce the plot in Figure 1 and add the function $\widehat{g}_K(x)$ to this plot.

(f) Look closely at the spread of the points in the plot you made in (e). Might there be something problematic with the estimated standard errors in your table? If so, what?

Exercise 2. Let X be a random variable with probability density function (pdf)

$$f_\mu(x) = \frac{1}{2\mu}\left(\frac{x}{2}\right)^{1/\mu - 1}, \quad \text{for } 0 \leq x \leq 2,$$

and $f_\mu(x) = 0$ when x is not in $[0, 2]$, with $\mu > 0$.

(a) Find expressions for the expectation and the variance of X.

(b) Find an expression for the cumulative distribution function (cdf) $F_\mu(x) = \int_{-\infty}^x f_\mu(y)\,\mathrm{d}y$.

(c) Let $X_1, \ldots, X_n$ be independent and identically distributed (i.i.d.) random variables, each with the same distribution as X. Write down an expression for the log-likelihood function $\ell_n(\mu) = \sum_{i=1}^n \log f_\mu(X_i)$, and show that the maximum likelihood estimator for μ is

$$\widehat{\mu}_n = -\frac{1}{n}\sum_{i=1}^n \log(X_i/2),$$

by solving $\mathrm{d}\ell_n(\mu)/\mathrm{d}\mu = 0$. You do not need to check that $\widehat{\mu}_n$ is a maximiser.

(d) Define $Y_i = -\log(X_i/2)$ for $i = 1, \ldots, n$. Then $Y_1, \ldots, Y_n$ are i.i.d. random variables. Show that

$$\Pr(Y_1 \leq y) = 1 - \exp(-y/\mu), \quad \text{for } y > 0 \tag{1}$$

Find also the expectation and the variance of Y_1.

(e) Show that $\widehat{\mu}_n$ is an unbiased estimator for μ, and, using results given in the course, explain why

$$\Pr(\sqrt{n}(\widehat{\mu}_n - \mu)/\mu \leq x) \to \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)\,\mathrm{d}z,$$

as n tends to infinity, for any $x \in (-\infty, \infty)$.

(f) For some $\alpha \in (0,1)$, show that

$$\Big[\frac{\sqrt{n}\widehat{\mu}_n}{\sqrt{n}+\Phi^{-1}(1-\alpha/2)}, \frac{\sqrt{n}\widehat{\mu}_n}{\sqrt{n}+\Phi^{-1}(\alpha/2)}\Big],$$

is an approximate $(1-\alpha)\times 100$ percent confidence interval for μ, with $\Phi^{-1}(p)$ the inverse of the standard normal cdf $\Phi(x) = \int_{-\infty}^{x}(1/\sqrt{2\pi})\exp(-u^2/2)\,\mathrm{d}u$. In Matlab $\Phi^{-1}(p)$ can be found by `norminv(p)`, for $p \in [0,1]$.

(g) Set $\mu = 2$ and $\alpha = 0.05$. Simulate 1000 datasets with sample size $n = 53$. For each dataset, compute the confidence interval given in (f), and check if it contains μ. Count the number of confidence intervals that contains μ. Comment on what you find. In Matlab, to simulate independent observations from the distribution in (1), you can use the `exprnd()`-function.

Exercise 3. Consider the model

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \varepsilon_i, \quad \text{for } i = 1,\ldots,n, \tag{2}$$

where $\varepsilon_1,\ldots,\varepsilon_n$ are i.i.d. $\mathrm{N}(0,1)$, independent of the (X_i, Z_i) which are

$$\begin{pmatrix} X_i \\ Z_i \end{pmatrix} \sim \mathrm{N}_2\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma\rho \\ \sigma\rho & 1 \end{pmatrix}\Big), \quad \text{for } i = 1,\ldots,n,$$

and independent. Note that this is a model where the covariates are random variables. We will write $X = (X_1,\ldots,X_n)$, and $\mathrm{E}[W \mid X]$ for the conditional expectation of some random variable W given $X = (X_1,\ldots,X_n)$. Recall that $\mathrm{E}[h(X)W \mid X] = h(X)\mathrm{E}[W \mid X]$, for any real valued function h and random variable W. In particular,

$$\mathrm{E}[Y_i \mid X] = \beta_0 + \beta_1 X_i + \frac{\rho\beta_2}{\sigma} X_i, \tag{3}$$

for each i, using the fact that $\mathrm{E}[Z_i \mid X] = (\rho/\sigma)X_i$. You can use (3) in the following without proving it. In this exercise you may also need that if $\xi \sim \mathrm{N}(0,1)$, then

$$\mathrm{E}\,\xi^3 = 0, \quad \text{and} \quad \mathrm{E}\,\xi^4 = 3.$$

The goal of this exercise is to make inference on β_1, which is our parameter of interest. We will study a situation where, for some reason, we don't observe the $Z_1,\ldots,Z_n$, and therefore try to estimate β_1 by using the least squares estimator

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{\sum_{i=1}^n (X_i - \bar{X}_n)^2},$$

where $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$ and $\bar{Y}_n = (1/n)\sum_{i=1}^n Y_i$.

(a) We start by investigating the estimator $\widehat{\beta}_1$ by way of simulation. Set $n = 123$, $\sigma^2 = 1.208$, $\beta_0 = 0.432$, $\beta_1 = 1.234$, $\beta_2 = 2.467$, and $\rho = -0.567$. Simulate 1000 datasets $\{(X_1,Y_1),\ldots,(X_n,Y_n)\}$ from the model in (2) with these parameter values. For each data set, compute $\widehat{\beta}_1$. Make a histogram of the 1000 estimates, and add two vertical lines to you histogram. One vertical line indicating the mean of the simulated $\widehat{\beta}_1$, and one for the true value β_1. Comment on what you find. My histogram is in Figure 2. *Hint*: Look at Homework 8, Ex. 2 for how to simulate from a bivariate normal distribution.

(b) Much of what follows is easier if we express $\widehat{\beta}_1$ as

$$\widehat{\beta}_1 = \frac{1}{\sum_{j=1}^n (X_j - \bar{X}_n)^2} \sum_{i=1}^n (X_i - \bar{X}_n) Y_i.$$

Show that this equality is true.

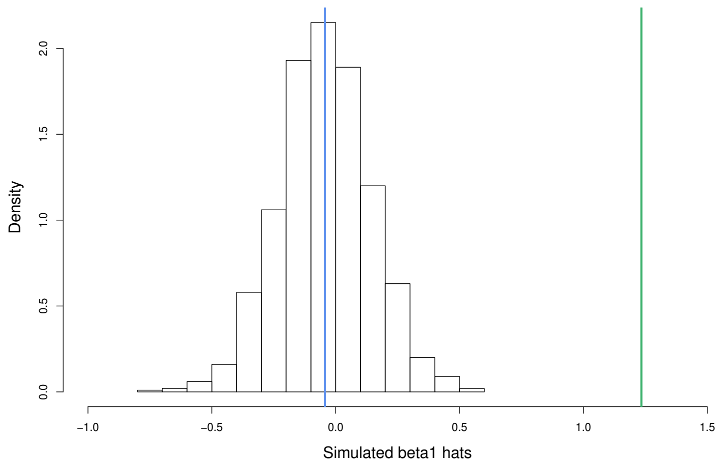

FIGURE 2. The histogram described in Ex. 3(a).

.

(c) Using, among other things, the linearity of conditional expectation, show that

$$\mathrm{E}\,[\widehat{\beta}_1 \mid X] = \beta_1 + \frac{\rho\beta_2}{\sigma}.$$

(d) Looking at the expectation in (c), under what conditions will $\widehat{\beta}_1$ be unbiased for β_1? And what is the implication of this insight when it comes to which independent variables we should, and which ones we need not, include in a regression model when we are interested in a particular regression coefficient?

(e) Suppose that in addition to (X_i, Y_i) for $i = 1, \ldots, n$, we obtain data on the variables $W_1, \ldots, W_n$, and that W_i is related to X_i by

$$X_i = bW_i + u_i, \quad \text{for } i = 1, \ldots, n,$$

where $u_1, \ldots, u_n$ are independent $\mathrm{N}(0, 1)$, with $\mathrm{Cov}(u_i, Z_i) = \rho$ for $i = 1, \ldots, n$, while $b \neq 0$. The $W_1, \ldots, W_n$ are independent $\mathrm{N}(0, 1)$, and are independent of the $u_1, \ldots, u_n$, $\varepsilon_1, \ldots, \varepsilon_n$, and the $Z_1, \ldots, Z_n$. Let

$$\widehat{b}_n = \frac{\sum_{i=1}^n W_i X_i}{\sum_{i=1}^n W_i^2},$$

be the estimator minimising the function $h(b) = \sum_{i=1}^n (X_i - bW_i)^2$. You do not need to show that $\widehat{b}_n$ is the minimiser of $h(b)$, but do show that $\widehat{b}_n$ is consistent for b.

(f) Let $\widetilde{\beta}_0, \widetilde{\beta}_1$ be the minimisers of $g_{\mathrm{fix}}(\beta_0, \beta_1) = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 \widehat{X}_i)^2$, where $\widehat{X}_i = \widehat{b}_n W_i$ for $i = 1, \ldots, n$, thus, from facts derived in the course, we get

$$\widetilde{\beta}_1 = \frac{\sum_{i=1}^n \{\widehat{X}_i - (1/n)\sum_{j=1}^n \widehat{X}_j\}(Y_i - \bar{Y}_n)}{\sum_{i=1}^n \{\widehat{X}_i - (1/n)\sum_{j=1}^n \widehat{X}_j\}^2},$$

a fact you need not show. Show that we can write

$$\widetilde{\beta}_1 = \frac{1}{\widehat{b}_n}(b\beta_1 + \beta_1 A_n + \beta_2 B_n + C_n),$$

where

$$A_n = \frac{\sum_{i=1}^n (W_i - \bar{W}_n) u_i}{\sum_{i=1}^n (W_i - \bar{W}_n)^2}, \quad B_n = \frac{\sum_{i=1}^n (W_i - \bar{W}_n) Z_i}{\sum_{i=1}^n (W_i - \bar{W}_n)^2}, \quad C_n = \frac{\sum_{i=1}^n (W_i - \bar{W}_n) \varepsilon_i}{\sum_{i=1}^n (W_i - \bar{W}_n)^2},$$

with $\bar{W}_n = (1/n)\sum_{i=1}^n W_i$. It can be shown that $(1/n)\sum_{i=1}^n (W_i - \bar{W}_n)^2 \to_p 1$ as $n \to \infty$, a fact you need not show. Show that

$$A_n \xrightarrow{p} 0.$$

as n tends to infinity.

(g) The argument you used to show that $A_n \to_p 0$ can be applied to B_n and C_n as well, so indeed $B_n \to_p 0$ and $C_n \to_p 0$, but this you need not prove. Show that

$$\widetilde{\beta}_1 \xrightarrow{p} \beta_1,$$

as n tends to infinity. This shows that $\widetilde{\beta}_1$ is a consistent estimator for β_1!

(h) The parameter ρ is, in real world applications, unknown. In this final exercise, let's compare the estimators $\widehat{\beta}_1$ and $\widetilde{\beta}_1$ in the case when ρ is small, that is, in situations where the bias you found in (c) might not be that pronounced. Except for ρ and b, which you should set to $\rho = -0.123$ and $b = 0.456$, keep the parameter values from the simulations in (a) unchanged. Simulate 1000 datasets $\{(X_1, W_1, Y_1), \ldots, (X_n, W_n, Y_n)\}$, and for each dataset compute $\widehat{\beta}_1$ and $\widetilde{\beta}_1$ and save the values you get. Based on the 1000 simulated values of $\widehat{\beta}_1$ and the 1000 simulated values of $\widetilde{\beta}_1$, estimate the variances $\mathrm{Var}(\widehat{\beta}_1)$, $\mathrm{Var}(\widetilde{\beta}_1)$, and the squared bias of both estimators,

$$\mathrm{bias}^2(\widehat{\beta}_1) = (\mathrm{E}\,[\widehat{\beta}_1] - \beta_1)^2, \quad \text{and} \quad \mathrm{bias}^2(\widetilde{\beta}_1) = (\mathrm{E}\,[\widetilde{\beta}_1] - \beta_1)^2,$$

as well as the mean squared errors

$$\mathrm{mse}(\widehat{\beta}_1) = \mathrm{E}\,[(\widehat{\beta}_1 - \beta_1)^2], \quad \text{and} \quad \mathrm{mse}(\widetilde{\beta}_1) = \mathrm{E}\,[(\widetilde{\beta}_1 - \beta_1)^2].$$

Summarise your findings in a table. Comment on what you find, and explain what you think causes this behaviour of the two estimators. The results I got from my simulations are summarised in Table 1. If you don't manage to do your own simulations, you can comment on those given in that table.

	variance	bias^2	mse
$\widehat{\beta}_1$	0.0482	0.0623	0.1105
$\widetilde{\beta}_1$	0.2999	0.0011	0.3010

TABLE 1. Results of simulations as described in Ex. 3(h). The estimates of the variance, bias^2, and the mean squared error are based on 1000 simulated datasets.