HOMEWORK 8
GRA6039 ECONOMETRICS WITH PROGRAMMING
AUTUMN 2020

EMIL A. STOLTENBERG

Intro: Here is a short summary of some results mentioned in lecture today. I'll soon write this more fully out in the lecture notes. If $X_{1,1}, X_{1,2}, X_{2,1}, X_{2,2}$ are random variables, and

$$X = \begin{pmatrix} X_{1,1} & X_{1,2} \\ X_{2,1} & X_{2,2} \end{pmatrix} \quad \text{then} \quad \mathrm{E}\,[X] = \begin{pmatrix} \mathrm{E}\,[X_{1,1}] & \mathrm{E}\,[X_{1,2}] \\ \mathrm{E}\,[X_{2,1}] & \mathrm{E}\,[X_{2,2}] \end{pmatrix}.$$

The same applies to matrices of higher dimensions, and they need not be square. If $Y = (Y_1, \ldots, Y_n)$ is a vector of random variables, then its $n \times n$ covariance matrix is defined by

$$\mathrm{Var}(Y) = \mathrm{E}\,[(Y - \mathrm{E}\,[Y])(Y - \mathrm{E}\,[Y])^{\mathrm{t}}].$$

If A is a matrix containing constants only (not rv's), of dimensions such that AY makes sense, then

$$\mathrm{E}\,[AY] = A\,\mathrm{E}[Y], \quad \text{and} \quad \mathrm{Var}[AY] = A\mathrm{Var}(Y)A^{\mathrm{t}}.$$

Here is a useful lemma about linear combinations of normals. If $X \sim \mathrm{N}_n(\mu_1, \Gamma_1)$ and $Y \sim \mathrm{N}_n(\mu_2, \Gamma_2)$, all the elements of X are independent of all the elements of Y, and A and B are $p \times n$ matrices of constants, while c is a p vector of constants, then

$$AX + BY + c \sim \mathrm{N}_p(A\mu_1 + B\mu_2 + c, A\Gamma_1 A^{\mathrm{t}} + B\Gamma_2 B^{\mathrm{t}}). \tag{1}$$

Recall also that independence implies covariance equal to zero. When the random variables involved are normal, the reverse implication also holds: covariance equal to zero implies independence. So saying that $Z_1, \ldots, Z_n$ are i.i.d. standard normals, is the same as saying that

$$Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix} \sim \mathrm{N}_n(0, I_n),$$

where the 0 is supposed to be understood n dimensional column vector of zeros, and I_n is the $n \times n$ identity matrix.

Exercise 1. (Multiple regression in Matlab). In this exercise you are to build your own multiple regression procedure in Matlab. First, we repeat some of the matrix computations we did in class, then we implement it all in Matlab.

Consider the model

$$Y_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_{p-1} x_{i,p-1} + \varepsilon_i, \quad \text{for } i = 1, \ldots, n, \tag{2}$$

Date: October 15, 2020.

where $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. with expectation zero and variance σ^2, all the covariates are fixed numbers (not rv's), and $n > p$. Write

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}, \quad \text{and} \quad x_i = \begin{pmatrix} 1 \\ x_{i,1} \\ \vdots \\ x_{i,p-1} \end{pmatrix}, \quad \text{for } i = 1, \ldots, n,$$

for the column vector of regression coefficients and of covariates, respectively. The model in (2) can then be expressed as

$$Y_i = x_i^{\mathrm{t}}\beta + \varepsilon_i, \quad \text{for } i = 1, \ldots, n. \tag{3}$$

There is a third way to write our model. For that we need

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{n-1} \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{1,p-1} \\ 1 & x_{2,1} & \cdots & x_{2,p-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n-1,1} & \cdots & x_{n-1,p-1} \\ 1 & x_{n,1} & \cdots & x_{n,p-1} \end{pmatrix}, \quad \text{and} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{n-1} \\ \varepsilon_n \end{pmatrix},$$

where the $n \times p$ matrix X is often called the *design matrix*. With this notation, the model in (2) can be expressed as

$$Y = X\beta + \varepsilon, \tag{4}$$

where Y, X, β, and ε are as defined above, end the elements of ε are i.i.d. rv's with expectation zero and variance σ^2.

The least squares estimator $\widehat{\beta} = (\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_{p-1})^{\mathrm{t}}$ of $\beta = (\beta_0, \beta_1, \ldots, \beta_{p-1})^{\mathrm{t}}$ is the minimiser of the function

$$g(\beta) = g(\beta_0, \beta_1, \ldots, \beta_{p-1}) = \sum_{i=1}^{n} (Y_i - x_i^{\mathrm{t}}\beta)^2 = (Y - X\beta)^{\mathrm{t}}(Y - X\beta).$$

We will assume that the $p \times p$ matrix $X^{\mathrm{t}}X$ is invertible. (Assumption MLR.3 in Wooldridge (2019, p. 80) combined with the $n > p$ assumption imply invertibility of $X^{\mathrm{t}}X$).

(a) As in Lecture 8, start by finding the partial derivatives of $g(\beta_0, \beta_1, \ldots, \beta_{p-1})$, set these equal to zero, and convince yourself that the system of p equations you get can be expressed as

$$\sum_{i=1}^{n} x_i(Y_i - x_i^{\mathrm{t}}\beta) = X^{\mathrm{t}}(Y - X\beta) = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix},$$

dropping the -2 in front. *Hint*: $X^{\mathrm{t}}X = \sum_{i=1}^{n} x_i x_i^{\mathrm{t}}$.

(b) Solve the system of equations in (a) and obtain

$$\widehat{\beta}_n = (X^{\mathrm{t}}X)^{-1}X^{\mathrm{t}}Y.$$

Use this to show that

$$\mathrm{E}\,[\widehat{\beta}_n] = \beta, \quad \text{and} \quad \mathrm{Var}(\beta) = \sigma^2(X^{\mathrm{t}}X)^{-1}.$$

Note that this shows that $\widehat\beta_0, \widehat\beta_1, \ldots, \widehat\beta_{p-1}$ are all unbiased. Also, the variance $\mathrm{Var}(\widehat\beta_j)$ is the j'th diagonal element of $\sigma^2(X^{\mathrm t}X)^{-1}$.

(c) An unbiased estimator of σ^2 is

$$\widehat\sigma_n^2 = \frac{1}{n-p}\sum_{i=1}^n (Y_i - x_i^{\mathrm t}\widehat\beta_n)^2. \tag{5}$$

To show that this estimator is unbiased it helps to know that the trace of a matrix is the sum of its diagonal elements. For example,

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix}, \quad \text{then the trace of } A \text{ is} \quad \mathrm{tr}(A) = a_{1,1} + a_{2,2}.$$

Also, if A, B, and C are matrices

$$\mathrm{tr}(ABC) = \mathrm{tr}(BCA) = \mathrm{tr}(CAB).$$

Thus, for example

$$\sum_{i=1}^n x_i^{\mathrm t}(X^{\mathrm t}X)^{-1}x_i = \mathrm{tr}(X(X^{\mathrm t}X)^{-1}X^{\mathrm t}) = \mathrm{tr}(X^{\mathrm t}X(X^{\mathrm t}X)^{-1}) = \mathrm{tr}(I_p) = p.$$

Try to show that $\widehat\sigma_n^2$ is unbiased. One way of showing it is included in Appendix A.

(d) Download the dataset `hw8.txt`, read it into Matlab, and construct the design matrix. Here we work with the model in (2) with $p = 3$. Here is code

```
hw8 = readtable("hw8.txt");
y = hw8.y; x1 = hw8.x1;x2 = hw8.x2;
n = length(y);
X = [1 + zeros(n,1),x1,x2]; % The design matrix
p = length(X(1,:)); % Useful later
```

Implement the estimators for β and for σ^2 that you found above. To multiply two matrices in Matlab use $*$. For the transpose, use `transpose()`, and use `inv()` for the inverse.

(e) Before continuing, you should make scatter plots of the data, both `y` against `x1`, and `y` against `x2`.

(f) The standardised coefficients

$$\frac{\widehat\beta_j - \beta_j}{\mathrm{se}(\widehat\beta_j)} \sim \mathrm{N}(0,1), \quad \text{for } j = 0, 1, \ldots, p-1,$$

when $\varepsilon \sim \mathrm{N}_n(0, \sigma^2 I_n)$, or approximately so when the normality assumption is dropped, given that n is sufficiently big. Here, $\mathrm{se}(\widehat\beta_j)$ is the square root of $j+1$'th diagonal element of $\sigma^2(X^{\mathrm t}X)^{-1}$, where you need to estimate σ^2. In Matlab `diag()` gives you the diagonal elements of a square matrix.

The default p-values returned by Matlab are the p-values testing $H_0\colon \beta_j = 0$ against its two-sided alternative, for $j = 0, 1, \ldots, p-1$. That is, for the observed value of $\widehat\beta_j/\mathrm{se}(\widehat\beta_j)$, the p-value is

$$\Pr(|Z| \geq |\widehat\beta_j/\mathrm{se}(\widehat\beta_j)|) = \Pr(|Z| \geq |z_j|).$$

$\widehat{\beta}$	$\mathrm{se}(\widehat{\beta})$	z_j	$\Pr(Z \geq \lvert z_j \rvert)$
-0.405	0.103	-3.936	0.000
1.245	0.113	11.011	0.000
0.044	0.107	0.409	0.682

TABLE 1. Regression results from analysing the data in `hw8.txt`.

where Z is a standard normal random variable, and z_j is the observed value $\widehat{\beta}_j/\mathrm{se}(\widehat{\beta}_j)$. In Matlab Z is taken to have a t-distribution with $n-p$ degrees of freedom. You can use `normcdf(z)` for the cdf of the standard normal, and `tcdf(z,n-p)` for the cdf of the t-distribution with $n-p$ degrees of freedom. When $n-p$ is sufficiently big, `normcdf(z)` and `tcdf(z,n-p)` will be very close. They are both symmetric, i.e. `1-normcdf(z)` equals `normcdf(-z)`, and similarly for `tcdf(z,n-p)`.

(g) Do some Matlab coding (and googling) to make a nice table of your output that is sent to the terminal. Here is the table I get from running a regression on the `hw8.txt` dataset

```
betahat      se         z        p-value
-------     -----     ------     -------

-0.405      0.103     -3.936           0
 1.245      0.113     11.011           0
 0.044      0.107      0.409       0.682
```

This is nice to look at in the Matlab terminal, but on the group home exam you should format the table in you favourite text processing program (Word, Latex, etc.), and make it look somewhat like Table 1.

Exercise 2. (THE BIVARIATE NORMAL DISTRIBUTION). Let Z_1 and Z_2 be independent standard normal random variables. With $Z = (Z_1, Z_2)^{\mathrm{t}}$, this means that $Z \sim \mathrm{N}_2(0, I_2)$. Consider the matrix A and the vector μ, given by

$$A = \begin{pmatrix} \sigma_1 & 0 \\ \sigma_2\rho & \sigma_2(1-\rho^2)^{1/2} \end{pmatrix}, \quad \text{and} \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix},$$

where $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$, and $-1 < \rho < 1$. Define

$$\begin{pmatrix} X \\ Y \end{pmatrix} = AZ + \mu = \begin{pmatrix} \sigma_1 Z_1 \\ \sigma_2\{\rho Z_1 + (1-\rho^2)^{1/2} Z_2\} \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}.$$

(a) Use the fact in Eq. (1) to argue that

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathrm{N}_2\Big(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \Big),$$

that is, (X, Y) has a bivariate normal distribution. It suffices to find $\mathrm{E}[X]$, $\mathrm{E}[Y]$, $\mathrm{Var}(X)$, $\mathrm{Var}(Y)$, and $\mathrm{Cov}(X, Y)$.

(b) Show that

$$\mathrm{E}\left[Y \mid X\right] = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(X - \mu_1) = \mu_2 + \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}(X - \mu_1),$$

then you can argue that by symmetry,

$$\mathrm{E}\left[X \mid Y\right] = \mu_1 + \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)} = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(Y - \mu_2).$$

In this exercise and the ones that follow, the claims are trivial if $\rho = 0$, so we assume that $\rho \neq 0$.

(c) The conditional variance of Y given X is

$$\mathrm{Var}(Y \mid X) = \mathrm{E}\left[Y^2 \mid X\right] - (\mathrm{E}\left[Y \mid X\right])^2.$$

Show that with the bivariate normal (X, Y) from (a).

$$\mathrm{Var}(X \mid Y) = \sigma_1^2(1 - \rho^2), \quad \text{and} \quad \mathrm{Var}(Y \mid X) = \sigma_2^2(1 - \rho^2).$$

Hint: Use the rules of conditional expectation (see e.g. Wooldridge (2019, pp.700–704), or notes from Lecture 8) to show that

$$\mathrm{E}\left[Y^2 \mid X\right] = \mu_2^2 + 2\mu_2\rho\frac{\sigma_2}{\sigma_1}(X - \mu_1) + \rho^2\frac{\sigma_2^2}{\sigma_1^2}(X - \mu_1)^2 + \sigma_2^2(1 - \rho^2),$$

and use this to find the expression for $\mathrm{Var}(Y \mid X)$. Then argue that by symmetry we have the expression for $\mathrm{Var}(X \mid Y)$.

Exercise 3. (Omitted variable bias). Suppose that the true regression model is

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_1 X_{i,2} + \varepsilon_i, \quad \text{for } i = 1, \ldots, n, \tag{6}$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. standard normals independent of $X_{1,1}, \ldots, X_{n,1}$ and $X_{1,2}, \ldots, X_{n,2}$, while $(X_{1,1}, X_{1,2}), \ldots, (X_{n,1}, X_{n,2})$ is are i.i.d. bivariate normals with distribution

$$\begin{pmatrix} X_{i,1} \\ X_{i,2} \end{pmatrix} \sim \mathrm{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), \quad \text{for } i = 1, \ldots, n.$$

For some reason it is impossible to collect data on $X_{1,2}, \ldots, X_{n,2}$, so this variable is not in your dataset. You decide to estimate β_0 and β_1 by the estimators $\widetilde{\beta}_0$ and $\widetilde{\beta}_1$ defined as the minimisers of

$$g(\beta_0, \beta_1) = \sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_{i,1})^2. \tag{7}$$

Let $X_1^{(n)} = (X_{1,1}, \ldots, X_{n,1})$, and $X_2^{(n)} = (X_{1,2}, \ldots, X_{n,2})$, while $X^{(n)} = (X_1^{(n)}, X_2^{(n)})$. We'll use

$$\bar{X}_{n,1} = \frac{1}{n}\sum_{i=1}^{n} X_{i,1}, \quad \text{and} \quad \bar{X}_{n,2} = \frac{1}{n}\sum_{i=1}^{n} X_{i,2},$$

for the empirical means.

(a) Explain why

$$\mathrm{E}\left[Y_i \mid (X_{i,1}, X_{i,2})\right] = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2}.$$

(b) Use the Law of large numbers and Properties PLIM.1 and PLIM.2(i) to show that

$$\frac{1}{n}\sum_{i=1}^{n}(X_{i,1}-\bar{X}_{n,1})^2 \xrightarrow{p} 1.$$

(c) Use techniques similar to those in (b) to show that

$$\frac{1}{n}\sum_{i=1}^{n}(X_{i,1}-\bar{X}_{n,1})(X_{i,2}-\bar{X}_{n,2}) \xrightarrow{p} \rho.$$

(c) Derive the estimator for $\widetilde{\beta}_1$, and show that

$$\mathrm{E}\,[\widetilde{\beta}_1 \mid X^{(n)}] = \beta_1 + \beta_2\frac{\sum_{i=1}^{n}(X_{i,1}-\bar{X}_{n,1})(X_{i,2}-\bar{X}_{n,2})}{\sum_{i=1}^{n}(X_{i,1}-\bar{X}_{n,1})^2}.$$

This shows that $\widetilde{\beta}_1$ is biased, and this bias is what is referred to as *omitted variable bias*. The fact that $X_{i,2}$ was omitted in the estimation, create this bias.

(d) Explain why, when n is really big,

$$\mathrm{E}\,[\widetilde{\beta}_1 \mid X^{(n)}] \quad \text{will be close to} \quad \beta_1+\beta_2\rho,$$

with high probability. *Hint*: See (b) and (c).

(e) In view of (d), can you think of situations where omitting an independent variable from the estimation does not lead to omitted variable bias? In other words, situation where the model in (6) is the true model, but $\widetilde{\beta}_1$ is an unbiased, or nearly unbiased estimator.

1. Optional exercises

We will go through these exercises in a lecture or in a TA-session soon.

Exercise 4. (Instrumental variable estimation). Suppose that the model is the same as the one given in (6) in Ex. 3, and that the independent variables $X_{1,2},\ldots,X_{n,2}$ are still not available to us. There is, however, another variable we *can* collect data on, namely $Z_1,\ldots,Z_n$, and in fact, the $X_{i,1}$'s are functions of these Z_i's,

$$X_{i,1} = \gamma Z_i + \eta_i, \quad \text{for } i=1,\ldots,n,$$

where $\gamma \neq 0$; the $Z_1,\ldots,Z_n$ are i.i.d. standard normals and independent of the $\eta_1,\ldots,\eta_n$, independent of the $X_{1,2},\ldots,X_{n,2}$, and independent of the $\varepsilon_1,\ldots,\varepsilon_n$; while it is in fact the $(\eta_1,X_{1,2}),\ldots,(\eta_n,X_{n,2})$ that are i.i.d. bivariate normals with distribution

$$\begin{pmatrix}\eta_i\\ X_{i,2}\end{pmatrix} \sim \mathrm{N}_2\Big(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\Big), \quad \text{for } i=1,\ldots,n.$$

The $Z_1,\ldots,Z_n$ are what econometricians refer to as an *instrumental variable* (IV), and the idea in IV-estimation is to replace $X_{i,1}$ by their predicted values $\widehat{\gamma}_n Z_i$, where $\widehat{\gamma}_n$ is the minimiser of the function $g_1(\gamma)=\sum_{i=1}^{n}(X_{i,1}-\gamma Z_i)^2$. The instrumental variable-estimators for β_0 and β_1, say $\widehat{\beta}_{0,\mathrm{iv}}$ and $\widehat{\beta}_{1,\mathrm{iv}}$, are the minimisers of the function

$$g_2(\beta_0,\beta_1)=\sum_{i=1}^{n}(Y_i-\beta_0-\beta_1\widehat{\gamma}_n Z_i)^2.$$

Compare this function to the function given in (7), and notice how we have replaced each $X_{i,1}$ by $\widehat{\gamma}_n Z_i$.

(a) Read carefully through the assumptions we make about our model in this exercise.
(b) Derive an expression for $\widehat{\gamma}_n$, and show that $\widehat{\gamma}_n$ is consistent for γ. *Hint*: Use Properties PLIM.2(i) and PLIM.2(iii).
(c) Find an expression for the instrumental variable estimator of β_1.
(d) Make sure you understand why

$$\frac{1}{n}\sum_{i=1}^{n}(Z_i - \bar{Z}_n)\eta_i \xrightarrow{p} 0, \quad \frac{1}{n}\sum_{i=1}^{n}(Z_i - \bar{Z}_n)X_{i,2} \xrightarrow{p} 0, \quad \frac{1}{n}\sum_{i=1}^{n}(Z_i - \bar{Z}_n)\varepsilon_i \xrightarrow{p} 0,$$

and $(1/n)\sum_{i=1}^{n}(Z_i - \bar{Z}_n)^2 \to_p 1$.
(e) Use the expression you found in (c), and write

$$\widehat{\beta}_{1,\mathrm{iv}} = \beta_1\frac{\gamma}{\widehat{\gamma}_n} + \frac{\beta_1}{\widehat{\gamma}_n}\frac{\sum_{i=1}^{n}(Z_i - \bar{Z}_n)\{\beta_1\eta_i + \beta_2 X_{i,2} + \varepsilon_i\}}{\sum_{i=1}^{n}(Z_i - \bar{Z}_n)^2}.$$

(e) Use the expression in (e), combined with the results from (d) and the Properties PLIM.1 and PLIM.2 to show that

$$\widehat{\beta}_{1,\mathrm{iv}} \xrightarrow{p} \beta_1.$$

In other words, $\widehat{\beta}_{1,\mathrm{iv}}$ is consistent for β_1.

Exercise 5. (Testing the β). This exercise builds on Ex. 5 and Ex. 6 in Homework 7. Do those two exercises before you continue with this one. In this exercise we work with a stock S_{t_j} and an index C_{t_j} for $j = 0, 1, \ldots, n$. The model for these two is the same as in Ex. 6 in Homework 7, and all the notation is the same as well. In that exercise we showed that the estimator

$$\widehat{\beta}_n = \frac{\sum_{j=1}^{n}(X_{t_j} - X_{t_{j-1}})(Y_{t_j} - Y_{t_{j-1}})}{\sum_{j=1}^{n}(X_{t_j} - X_{t_{j-1}})^2},$$

is consistent for $\rho\sigma_S/\sigma_C$. We now want to test the null-hypothesis $\rho = 0$ against its two-sided alternative. You can think about what $\rho = 0$ means in the context of the Capital asset pricing model (CAPM). Since $\sigma_S > 0$ and $\sigma_C > 0$ we might as well just test,

$$H_0\colon \rho\sigma_S\sigma_C = 0, \quad \text{vs.} \quad H_A\colon \rho\sigma_S\sigma_C \neq 0.$$

To test this hypothesis we need to find the limit distribution of

$$\widehat{\mathrm{cov}}_n = \sum_{j=1}^{n}(X_{t_j} - X_{t_{j-1}})(Y_{t_j} - Y_{t_{j-1}}),$$

as $n \to \infty$. The estimator $\widehat{\mathrm{cov}}_n$ was defined in Ex. 6 of Homework 7, and in that exercise you also find an expression for it that you should use in this exercise. To find the limit distribution of $\widehat{\mathrm{cov}}_n$ we need to know about the Cramér–Slutsky rules. They are summarised in Appendix B.

(a) Explain why $\xi_{t_1}\eta_{t_1}, \ldots, \xi_{t_n}\eta_{t_n}$ are i.i.d. random variables, and use Ex. 2(b) to show that

$$\mathrm{E}\,[\xi_{t_j}\eta_{t_j}] = \rho.$$

Hint: $\mathrm{E}\,[\xi_{t_j}\eta_{t_j}] = \mathrm{E}\,\{\mathrm{E}\,[\xi_{t_j}\eta_{t_j} \mid \xi_{t_j}]\} = \mathrm{E}\,\{\xi_{t_j}\mathrm{E}\,[\eta_{t_j} \mid \xi_{t_j}]\}$.

(b) Show also that

$$\mathrm{Var}(\xi_{t_j}\eta_{t_j}) = 1 + \rho^2.$$

Hint: Write $\mathrm{Var}(\xi_{t_j}\eta_{t_j}) = \mathrm{E}\,[\xi_{t_j}^2\eta_{t_j}^2] - (\mathrm{E}\,[\xi_{t_j}\eta_{t_j}])^2$, and use that $\mathrm{E}\,[\xi_{t_j}^4] = 3$, and perhaps also that $\mathrm{E}\,[\xi_{t_j}^3] = 0$ (these are results about the moments of the normal distribution.

(c) Explain why (a) and (b) entail that

$$\frac{\sqrt{\Delta_n}(\sum_{j=1}^n \xi_{t_j}\eta_{t_j} - \rho)}{\sqrt{1+\rho^2}} \xrightarrow{d} \mathrm{N}(0,1).$$

Hint: See Theorem 5.5 (the CLT) in the Lecture notes.

(d) Use the Cramér–Slutsky rules to argue that

$$\frac{\Delta_n^{-1/2}(\widehat{\mathrm{cov}}_n - \rho\sigma_S\sigma_C)}{\sigma_S\sigma_C\sqrt{1+\rho^2}} \xrightarrow{d} \mathrm{N}(0,1),$$

as $n \to \infty$.

(e) Let $\widehat{\rho}_n$ be the estimator you found in Ex. 6(g). Importantly $\widehat{\rho}_n$ is consistent for ρ. Propose also consistent estimators $\widehat{\sigma}_{S,n}$ and $\widehat{\sigma}_{C,n}$ of σ_S and σ_C, respectively. Now, use the Cramér–Slutsky rules and Property PLIM.2 to argue that

$$\frac{\Delta_n^{-1/2}(\widehat{\mathrm{cov}}_n - \rho\sigma_S\sigma_C)}{\widehat{\sigma}_{S,n}\widehat{\sigma}_{C,n}\sqrt{1+\widehat{\rho}_n^2}} \xrightarrow{d} \mathrm{N}(0,1),$$

as $n \to \infty$.

(f) Consider the test that rejects H_0 if

$$\frac{\Delta_n^{-1/2}\widehat{\mathrm{cov}}_n}{\widehat{\sigma}_{S,n}\widehat{\sigma}_{C,n}\sqrt{1+\widehat{\rho}_n^2}} \leq -c_n \quad \text{or} \quad \frac{\Delta_n^{-1/2}\widehat{\mathrm{cov}}_n}{\widehat{\sigma}_{S,n}\widehat{\sigma}_{C,n}\sqrt{1+\widehat{\rho}_n^2}} \geq c_n,$$

for some critical value $c_n > 0$. Find c_n such that

$$\Pr(\text{Type I error}) \approx 0.01.$$

What kind of investor would choose this significance level?

Appendix A. $\widehat{\sigma}_n^2$ is unbiased

The estimator $\widehat{\sigma}_n^2$ is as defined in (5). Write

$$\begin{aligned}(n-p)\widehat{\sigma}_n^2 &= \sum_{i=1}^n (Y_i - x_i^{\mathrm{t}}\widehat{\beta}_n)^2 = \sum_{i=1}^n (Y_i - x_i^{\mathrm{t}}\widehat{\beta}_n)(Y_i - x_i^{\mathrm{t}}\widehat{\beta}_n) \\ &= \sum_{i=1}^n (Y_i - x_i^{\mathrm{t}}\widehat{\beta}_n)Y_i = \sum_{i=1}^n Y_i^2 - \sum_{i=1}^n x_i^{\mathrm{t}}Y_i\widehat{\beta}_n = \sum_{i=1}^n Y_i^2 - \widehat{\beta}_n^{\mathrm{t}}X^{\mathrm{t}}X\widehat{\beta}_n.\end{aligned}$$

Then take the expectation,

$$
\begin{aligned}
\mathrm{E}\,[(n-p)\widehat{\sigma}_n^2] &= \sum_{i=1}^n \mathrm{E}[\,Y_i^2] - \mathrm{E}\,[\widehat{\beta}_n^{\mathrm{t}} X^{\mathrm{t}} X \widehat{\beta}_n] \\
&= \sum_{i=1}^n (x_i^{\mathrm{t}}\beta)^2 + n\sigma^2 - \mathrm{E}\,\sum_{i=1}^n \widehat{\beta}_n^{\mathrm{t}} x_i x_i^{\mathrm{t}} \widehat{\beta}_n \\
&= \sum_{i=1}^n \beta^{\mathrm{t}} x_i x_i^{\mathrm{t}} \beta + n\sigma^2 - \mathrm{E}\,\sum_{i=1}^n \widehat{\beta}_n^{\mathrm{t}} x_i x_i^{\mathrm{t}} \widehat{\beta}_n \\
&= n\sigma^2 - \mathrm{E}\,\sum_{i=1}^n (\widehat{\beta}_n - \beta)^{\mathrm{t}} x_i x_i^{\mathrm{t}} (\widehat{\beta}_n - \beta) \\
&= n\sigma^2 - \sum_{i=1}^n \mathrm{E}\,(x_i^{\mathrm{t}}(\widehat{\beta}_n - \beta))^2 \\
&= n\sigma^2 - \sum_{i=1}^n \mathrm{Var}\,(x_i^{\mathrm{t}}(\widehat{\beta}_n - \beta)) = n\sigma^2 - \sum_{i=1}^n x_i^{\mathrm{t}} \mathrm{Var}\,(\widehat{\beta}_n - \beta) x_i \\
&= n\sigma^2 - \sigma^2 \sum_{i=1}^n x_i^{\mathrm{t}} (X^{\mathrm{t}} X)^{-1} x_i \\
&= n\sigma^2 - \sigma^2 \mathrm{tr}(X(X^{\mathrm{t}}X)^{-1} X^{\mathrm{t}}) = \sigma^2(n-p).
\end{aligned}
$$

This explains why one divides by $n-p$ when estimating the variance σ^2.

Appendix B. Cramér–Slutsky rules

If X_n converges in distribution to a random variable X, and Y_n converges in probability to a constant c, then

(i) $X_n + Y_n \to_d X + c$;
(ii) $X_n Y_n \to_d Xc$;
(iii) $X_n / Y_n \to_d X/c$ provided $c \neq 0$.

References

Wooldridge, J. M. (2019). *Introductory Econometrics: A Modern Approach. Seventh Edition.* Cengage Learning, Boston, MA.

Department of Economics, BI Norwegian Business School
Email address: emil.a.stoltenberg@bi.no