HG
Okt. 06

ECON 4130

Supplementary Exercises 5 - 6

Exercise 5

Let (X,Y) be two random variables with some joint distribution, $f(x,y)$, expected values, μ_X, μ_Y, and variances, σ_X^2, σ_Y^2 respectively. Suppose that the regression of *Y* with respect to *X* is linear and homoscedastic, i.e.,

(1) $\mathrm{E}(Y \mid x) = \mathrm{E}(Y \mid X = x) = \alpha + \beta x$ where α, β are constants
$\mathrm{var}(Y \mid x) = \mathrm{var}(Y \mid X = x) = \sigma^2$ (constant)

a. Show, using the law of double expectations (and the corresponding one for variances), that (1) implies

(2) $\mu_Y = \mathrm{E}(Y) = \alpha + \beta\mu_X$

(3) $\sigma_Y^2 = \mathrm{var}(Y) = \sigma^2 + \beta^2\sigma_X^2$

b. Show that

(4) $\mathrm{cov}(X,Y) = \beta\sigma_X^2$ (or $\beta = \dfrac{\mathrm{cov}(X,Y)}{\sigma_X^2}$)

[**Hint:** Note that $\mathrm{E}(XY) = \mathrm{E}\big[\mathrm{E}(XY \mid X)\big] = \mathrm{E}\big[X \cdot \mathrm{E}(Y \mid X)\big]$, which follows since the inner expectation, $\mathrm{E}(XY \mid X)$, is of the form, $h(X)$, where $h(x)$ is determined by the conditional distribution of XY when *X* is fixed to the value *x*. I.e., $h(x) = \mathrm{E}(XY \mid x) = \mathrm{E}(XY \mid X = x) = \mathrm{E}(xY \mid X = x) = x\mathrm{E}(Y \mid X = x) = x\mathrm{E}(Y \mid x)$, since *x* is a constant in the expectation. Hence, replacing *x* by the r.v. *X*, we get $h(X) = X \cdot \mathrm{E}(Y \mid X) = X(\alpha + \beta X)$ and so on ….]

c. Show that (3) and (4) imply that

(5) $\sigma^2 = \sigma_Y^2(1-\rho^2)$

where $\rho = \rho(X,Y) = \text{cov}(X,Y)/(\sigma_X \sigma_Y)$ is the correlation between *X* and *Y*.

[**Hint:** Solve (4) for β and substitute in (3).]

[**Note** that solving (5) with respect to ρ^2 gives an alternative interpretation of ρ: Interpreting σ^2 as measuring the part of Y which *is not* explained by the regression relation in (1), then $\rho^2 = (\sigma_Y^2 - \sigma^2)/\sigma_Y^2$ measures the part of the variation of *Y*, (i.e., σ_Y^2) which *is* explained.]

d. The model (1) may be reformulated as follows: Write

(6) $$Y = \alpha + \beta X + u$$

where the "error term" *u* is simply defined as $u = Y - \alpha - \beta X$. Show that (1) implies:

(7) $$\text{E}(u \mid x) = 0 \quad \text{and} \quad \text{var}(u \mid x) = \sigma^2$$

and therefore also

(8) $$\text{E}(u) = 0 \quad \text{and} \quad \text{var}(u) = \sigma^2$$

Show (as in **b.**) that (7) also implies that *u* and *X* are uncorrelated, i.e.,

(9) $$\text{cov}(u, X) = 0$$

e. Show the other way round, i.e. that (6) and (7) imply (1).

Exercise 6

Let $(X_i, Y_i, Z_i) \quad i = 1, 2, \ldots, n$ be *n iid* triples of rv's, having a common joint pdf, $f(x, y, z)$. (This implies that there is independence between variables from different triples, although there may be dependencies between X_i, Y_i, Z_i for the same *i*.)

To fix ideas imagine that *i* refers to household no. *i* in a random sample of households drawn from a certain large population. Assume further that for household *i*

Y_i is the observed expenditure (in a given period)
Z_i is "the true income" (not directly observable)
X_i is the observed income

We now assume a simple regression relationship between Y_i and Z_i

(1) $Y_i = \alpha + \beta Z_i + e_i \qquad i = 1, 2, \ldots, n$

where α, β are unknown constants and the error term, e_i, is assumed to satisfy

(2) $\mathrm{E}(e_i \mid z_i) = 0$ and $\mathrm{var}(e_i \mid z_i) = \sigma^2$ (implying $\mathrm{cov}(e_i, Z_i) = 0$ as in Ex.5d).

(1) and (2) constitutes our econometric model. The task is to estimate β from the information in the observed data ($(X_i, Y_i) \quad i = 1, 2, \ldots, n$). The problem here is that we don't know the values of Z_i (a non-observable variable is often called a *latent* variable in econometric literature). Instead we observe X_i which we assume is near Z_i but with some (random) error, expressed by the following assumption

(3) $X_i = Z_i + v_i$ where $\mathrm{E}(v_i \mid z_i) = 0$ and $\mathrm{var}(v_i \mid z_i) = \sigma_v^2$

Substituting (3) in (1), we get

$$Y_i = \alpha + \beta(X_i - v_i) + e_i = \alpha + \beta X_i + (e_i - \beta v_i)$$

Hence

(4) $Y_i = \alpha + \beta X_i + u_i$ where $u_i = e_i - \beta v_i$ is an error term.

a. Let μ_X, μ_Y, μ_Z denote expected values and $\sigma_X^2, \sigma_Y^2, \sigma_Z^2$ variances of *X,Y,Z* respectively. Exercise 5 shows that the error terms e_i, v_i, u_i all have expected value 0 (why?), which implies (why?) that $\mu_Y = \alpha + \beta \mu_X$.

b. Show that u_i and X_i are correlated, i.e. show that

(5) $\mathrm{cov}(u_i, X_i) = \mathrm{E}(u_i X_i) = -\beta \sigma_v^2$

c. We are interested to estimate β in particular. Using the ordinary least squares (OLS) method, we get the OLS estimator (derived in elementary econometrics):

$$\hat{\beta} = \frac{S_{XY}}{S_X^2} \quad \text{where} \quad S_{XY},\ S_X^2 \ \text{are the usual sample estimators for the}$$

covariance and variance respectively. As in example 3 and exercise 1, both in "Lecture notes to Rice chapter 5", we obtain (explain why):

$$\hat{\beta} \underset{n\to\infty}{\overset{P}{\rightarrow}} \frac{\text{cov}(X_i, Y_i)}{\sigma_X^2}$$

Now show that $\text{cov}(X, Y) = \beta(\sigma_X^2 - \sigma_v^2)$.

[**Hint:**
$\text{cov}(X_i, Y_i) = \text{E}(Y_i - \mu_Y)(X_i - \mu_X) = \text{E}(\alpha + \beta X_i + u_i - \alpha - \beta\mu_X)(X_i - \mu_X) =$
$= \cdots\cdots \text{fill in } \cdots\cdots = \beta(\sigma_X^2 - \sigma_v^2)$]

Hence show that

$$\hat{\beta} \underset{n\to\infty}{\overset{P}{\rightarrow}} \beta\left(1 - \frac{\sigma_v^2}{\sigma_X^2}\right)$$

Hence the OLS estimator $\hat{\beta}$ is an inconsistent estimator for β unless $\sigma_v^2 = \text{var}(v_i) = 0$ (in which case surely $v_i = 0$, i.e. $P(v_i = 0) = 1$; see **d.** below). If $\sigma_v^2 > 0$, the OLS estimator is biased in terms of probability limits. Since the bias, $1 - \frac{\sigma_v^2}{\sigma_X^2} < 1$, $\hat{\beta}$ tends to underestimate β. We have thus shown that the OLS estimator $\hat{\beta}$ in a simple regression model is consistent if and only if the explanatory variable can be observed without error.

d. We have used above the following property: Let *X* be a r.v. with $\mu = \text{E(X)}$. If $\text{var}(X) = 0$, then *X* must be constant and equal to μ (i.e. $P(X = \mu) = 1$).

This (intuitively obvious) result is slightly tricky to prove in the general case. Prove it, however, in the special case that *X* is a discrete rv that can take only finitely many possible values, $x_1, x_2, \ldots, x_k$, with pmf

$$p(x_i) = P(X = x_i) \quad \text{where all } p(x_i) > 0, \quad i = 1, 2, \ldots, k$$

[**Note:** The assumption (2) implies (as in Exercise 5)

(6) $\mathrm{E}(e_i) = 0, \quad \mathrm{var}(e_i) = \sigma^2, \quad \text{and} \quad \mathrm{cov}(e_i, Z_i) = 0$

Likewise, the assumption (3) implies

(7) $X_i = Z_i + v_i \quad \text{where} \quad \mathrm{E}(v_i) = 0, \quad \mathrm{var}(v_i) = \sigma_v^2, \quad \text{and} \quad \mathrm{cov}(v_i, Z_i) = 0.$

The assumptions (6) and (7) are slightly weaker than assumptions (2) and (3) respectively (i.e., we cannot prove (2) and (3) from (6) and (7) without extra assumptions. On the other hand, if we replace (2) and (3) by (6) and (7), we can still prove the limit results above by the same arguments as above. This is maybe the main reason why (6) and (7) are more common in econometric literature as assumptions in connection with the simple regression model than (2) and (3).]