R300 – Advanced Econometric Methods

PROBLEM SET 5 - QUESTIONS

Due by Tue. November 16

1. Consider the classical linear regression model with two scalar regressors x_i and z_i. Set up the LM statistic for the null that the coefficient on z_i is zero. This is a test for an omitted variable.

2. Let x_i be binary with success probability $\theta \in (0, 1)$. For a sample of size n write down the LR, LM, and Wald test for the null $\theta = \theta_0$ and two-sided alternative $\theta \neq \theta_0$.

3. Consider the regression

$$y_i = x_i\beta + \varepsilon_i,$$

with ε_i mean-zero, homoskedastic, and independent of (scalar) x_i. Suppose, further, for simplicity that you know the variance of the errors. Consider a situation where each observation can be categorized into one of G mutually-exclusive groups; so $i \in g$ for one $g = 1, \ldots, G$ (An example is student i in classroom g or firm i in sector g.) Rather than (y_i, x_i), we observe averages at the group level, i.e, $\overline{y}_g = |g|^{-1} \sum_{i \in g} y_i$ and $\overline{x}_g = |g|^{-1} \sum_{i \in g} x_i$ for example, average test scores within a classroom.

(i) Show that the error in the regression

$$\overline{y}_g = \overline{x}_g\beta + \overline{\varepsilon}_g$$

is heteroskedastic, in general.

(ii) Under what condition(s) on the groups will $\overline{\varepsilon}_g$ be homoskedastic?

(iii) Does least-squares applied to the averaged regression lead to an unbiased estimator?

The following two questions are optional (as we have not quite covered the relevant material yet) but encouraged.

(iv) Is ordinary least-squares (semiparametrically) efficient here?

(v) If your answer to (iv) is negative can you give an alternative? (Think about constructing an estimator derived from a moment condition that restores the information equality. Check Gauss-Markov and weighted least squares.)

4. We have gathered survey data on health expenditure. For 1691 individuals we have access to the following variables

- eheal: expenditure on health-related products and services during the year (in British pound Sterling);
- income: total income during the year (in British pound Sterling);
- sex: gender dummy that equals one when individual is male and zero when female.

We presume that the conditional mean function takes the following form:

$$E(\text{eheal}|\text{income}, \text{sex}) = \beta_0 + \beta_1 \,\text{income} + \beta_2 \,\text{sex} + \beta_3 \,(\text{income} \times \text{sex}).$$

The coefficients in this model were estimated, and the results are given in Figure 1 below. The first column (Coef.) contains the point estimates, the second column (Std. Err.) provides the appropriate standard errors for statistical inference on these coefficients.

(i) Write down the equation for the conditional mean for males and for females separately.

(ii) Derive the average marginal effect of income on health expenditure for both males and females.

(iii) Formulate and test the hypothesis that the expected change in health expenditure resulting from a change in income is the same for males and females. That is, define H_0 and H_1, say which test procedure you use, and present your conclusion.

(iv) I tested the joint hypothesis H_0 : "$\beta_2 = 0$ and $\beta_3 = 0$" against "$H_1 : H_0$ is false". Interpret this null hypothesis in words.

v) The p-value of the test from the previous question was zero up to the fourth decimal digit. What does this mean?

Figure 1: Estimation results for Question 4

Source	SS	df	MS
Model	338619013	3	112873004
Residual	4.2621e+09	1687	2526417.94
Total	4.6007e+09	1690	2722299.45

Number of obs	=	1691
F(3, 1687)	=	44.68
Prob > F	=	0.0000
R-squared	=	0.0736
Adj R-squared	=	0.0720
Root MSE	=	1589.5

eheal	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
income	.0391367	.0068661	5.70	0.000	.0256697	.0526036
sex	-240.5175	145.4127	-1.65	0.098	-525.7259	44.69081
income_sex	-.038963	.0083186	-4.68	0.000	-.055279	-.0226471
_cons	869.0945	109.2712	7.95	0.000	654.773	1083.416