Written Exam for the B.Sc. or M.Sc. in Economics Winter 2017

Økonometri 1/Econometrics 1

Take-home exam

January 7, 2017

SUGGESTED ANSWERS

The assessment of the take-home exam is based on the report. The STATA program is not assessed as such, but is used e.g. to clear up errors and ambiguities in the report and to check that there has not been collaboration across groups.

The assessment takes into account whether the report is overall consistent and whether it, within the specified framework, is able to answer the problems stated in the assignment. The report should not exceed the total page numbers stated in the assignment. Should the page number exceed this, it will influence the overall grading negatively.

For more information about the Econometrics 1 and the learning objectives of the course, please go to the course website: `http://kurser.ku.dk/course/akb08020u/2016-2017`.

The numerical results and conclusions in the suggested answers are based on GROUP-DATA5.dta. Full results and conclusions for other data sets can be obtained by running the STATA program X2017W_takehome.do with the relevant data file. Results and conclusions may differ for different data sets.

Problem 1:

1. The variables should as a minimum be described by a table indicating the average, standard deviation, min, max. Comments must address main features of the data and any divergent observations.

   Table 1: Key descriptive statistics, full sample

VARIABLES	(1) N	(2) mean	(3) p50	(4) sd	(5) min	(6) max
id	42,000	1,750	1,750	1,010	1	3,500
Dcouple	42,000	0.666	1	0.472	0	1
Income	42,000	22,616	13,618	29,045	585.0	410,900
Rebate	42,000	33.33	0	129.9	0	600
month	42,000	6.500	6.500	3.452	1	12
Constrained	42,000	0.114	0	0.318	0	1
C	42,000	1,937	1,187	2,485	49.96	63,186
dC	38,500	-1.854	0.00761	768.2	-61,297	61,423
DRebate	42,000	0.0667	0	0.249	0	1
RebateXConstrain	42,000	3.750	0	44.75	0	600

   There are 42,000 observations in the data set. 2/3 of the respondents are living as a couple. The average annual income level is 22,616USD, but there is a lot of heterogeneity as witnessed by the minimum and maximum observation. The income distribution is also skewed. The median income level is about 9,000 lower than the average. People received 33.33USD in rebate on average, but this is highly unevenly distributed. The average level of monthly spending is about 1,900 USD, and like income this distribution is also highly skewed with the median being about 750USD smaller than the average.

   It is also worth noting that there are 2,800 households receiving the rebate during May to September and that 2/3 of these received 600USD. Of the 2,800 who received the rebate 317 were recorded as constrained in the month where they received the rebate.

2. (a) β_1 is the fraction spent out of every USD paid out as a tax rebate.

   (b) The Ricardian Equivalence Theorem says that there should be no spending effect of a transitory increase in income. If households perceive the tax rebate as transitory income then the effect of the tax rebate on spending should be zero.

   (c) Results from estimating equation (1) by OLS are presented in Table 2.

Table 2: OLS estimation results of regression model (1)

	(1)	(2)	(3)
	OLS	2SLS	OLS
VARIABLES	eq. (1)	eq (1)	eq (2)
Rebate	0.1245***	0.1361***	0.0678**
	(0.0290)	(0.0301)	(0.0307)
Constrained			97.1665***
			(12.7148)
RebateXConstrain			0.5086***
			(0.0915)
Dcouple	-0.5841	-0.8386	-0.5196
	(8.3248)	(8.3268)	(8.3122)
Constant	-5.9915	-6.2457	-17.1603**
	(6.8050)	(6.8074)	(6.9482)
Observations	38,500	38,500	38,500
R-squared	0.0005	0.0005	0.0036

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The estimation results indicate that the rebate causes the spending to increase by 12.45 cents for every dollar of tax rebate that is paid out. The estimate is significantly different from zero at the 1 percent level. This is not consistent with the Ricardian Equivalence Theorem.

Problem 2 :

1. Using *DRebate* as an instrumental variable for *Rebate* singles out the part of the variation in *Rebate* that is related to the timing of the pay-out, i.e. disregarding the size of the tax rebate. The timing of the rebate payment is random and should hence be uncorrelated with the error term in equation (1). The estimate obtained from estimating equation (1) by IV/2SLS is presented in column (2). It is very close to the estimate obtained by OLS (column 1) suggesting that the OLS estimate actually represents the causal effect of the tax rebate.

2. The test for exogeneity of x is performed by running two regressions: (1) $Rebate_{it} = \alpha_0 + \alpha_1 DRebate_{it} + \alpha_2 Dcouple_{it} + \epsilon$, (2) $\Delta C_{it} = \beta_0 + \beta_1 Rebate_{it} + \beta_2 X_{it} + \gamma\hat{\epsilon} + u_{it}$. Both regressions are estimated by OLS. The null hypothesis of exogeneity is $H_0 : \gamma = 0$ and the alternative is $H_1 : \gamma \neq 0$. The test statistic is $t = \frac{\hat{\gamma}}{s.e.(\hat{\gamma})} = \frac{-0.1572}{0.1107} = -1.4201$ This is bigger than -1.96 and smaller than 1.96 and the null hypotheses cannot be rejected at the 5 percent level. There is thus no evidence that *Rebate* is an endogenous regressor.

Problem 3 :

1. α_0 is the intercept. It is an estimate of the spending growth for single respondents who are not constrained and who has not received a tax rebate. α_1 is the fraction spent out of every USD paid out as a tax rebate for respondents who are not constrained in the credit market. α_2 the spending growth for constrained respondents, who have not received a tax rebate payment, on top of the spending growth of unconstrained respondents. α_3 is the additional fraction on top of α_1 spent out of every USD paid out as a tax rebate for respondents who are constrained in the credit market. α_4 is the additional spending growth for couples.

2. Results from estimating equation (2) by OLS are presented in Table 2, column 3. The estimated parameters indicate that the spending effect among unconstrained respondents is about 7 cents per dollar of tax rebate, and this is smaller than when estimating the spending effect of the tax rebate using equation (1). However, the confidence interval for the estimate of α_1 includes the estimate obtained when estimating the effect using equation (1) suggesting that the estimated effect is in fact not significantly different from the average effect estimated based on equation (1). The estimate of α_2 shows that respondents who are constrained have a spending growth that is 97 USD bigger than unconstrained households even when they have not received a tax rebate. The estimate of α_3 shows that constrained individuals who receive a tax rebate spend 51 cents out of every dollar of tax rebate paid out on top of the 7 cents that unconstrained respondents spend.

3. To assess the importance of credit constraints for the magnitude of the response the following hypothesis is tested: $H_0 : \alpha_3 = 0$ against the alternative hypothesis $H_1 : \alpha_3 \neq 0$. Only one restriction is tested and the test can be performed as a t-test. $t = \frac{0.5086}{0.0915} = 5.5585$. That exceeds the critical value at the 5 percent significance level, which is 1.96, by far, and we thus reject the null hypotheses that credit constraints do not matter for the spending response to the tax rebate. This test can also be conducted as a one-sided test where the alternative hypothesis is $H_1 : \alpha_3 > 0$.

4. To test the hypothesis that there is no spending effect for people who are not credit constrained we test the hypothesis that there is no spending response to the tax rebate among people who received the payment but who are not constrained. The null hypothesis is $H_0 : \alpha_1 = 0$ against the alternative hypothesis $H_1 : \alpha_1 \neq 0$. One restriction is tested and the test can be performed as a t-test. $t = \frac{0.0678}{0.0307} = 2.2085$. That exceeds the the critical value at the 5 percent significance level, which is 1.96. However, it does not exceed the critical value at the 1 percent level, which is 2.576. There is thus some evidence against the Ricardian Equivalence Theorem.

Problem 4 :

1. Insert $y_i = y_i^* + \rho\epsilon_i$ and $x_i = x_i^* + \epsilon_i$ in equation (3) to get

$$y_i = \theta_0 + \theta_1 x_i + u_i + (\rho - \theta_1)\epsilon_i$$

write up the OLS estimator for θ_1

$$\begin{aligned}\widehat{\theta}_1 &= \frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})(\theta_0 + \theta_1 x_i + u_i + (\rho - \theta_1)\epsilon_i)}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \theta_1 + \frac{\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})(u_i + (\rho - \theta_1)\epsilon_i)}{\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2}\end{aligned}$$

Dividing the numerator and the denominator by n allows us to apply the law of large numbers. This says that the averages in the second part of the equation converge in probability to the population quantities. The probability limit of $\widehat{\theta}_1$ is

$$\begin{aligned}\text{plim}(\widehat{\theta}_1) &= \theta_1 + \frac{cov(x,u)}{var(x)} + (\rho - \theta_1)\frac{cov(x,\epsilon)}{var(x)} \\ &= \theta_1 + (\rho - \theta_1)\frac{cov(x,\epsilon)}{var(x)}\end{aligned}$$

The last step is because $cov(x,u) = 0$. Because $x_i = x_i^* + \epsilon_i$ and we have assumed that $cov(x^*,\epsilon) = 0$ we can write $cov(x,\epsilon) = E(x\epsilon) = E(x^*\epsilon) + E(\epsilon^2) = E(\epsilon^2) = \sigma_\epsilon^2$. Therefore,

$$\text{plim}(\widehat{\theta}_1) = \theta_1 + (\rho - \theta_1)\frac{\sigma_\epsilon^2}{(\sigma_{x^*}^2 + \sigma_\epsilon^2)}$$

2. If $\rho = 0$ then this reduces to the well known case with classical-errors-in-variables (CEV), where OLS is attenuated:

$$\text{plim}(\widehat{\theta}_1) = \theta_1 \frac{\sigma_{x^*}^2}{(\sigma_{x^*}^2 + \sigma_\epsilon^2)}$$

If $\rho < 0$ then the attenuation bias is even stronger.
If $0 < \rho < \theta_1$ attenuation bias is reduced relative to the CEV case, but is still

present.
If $\rho = \theta_1$ there is no bias altogether.
If $\rho > \theta_1$ the OLS estimator is positively biased.

Problem 5 (20%):

In this problem you are asked to perform a Monte Carlo (MC) study to verify the analytical results from problem 4. This means that the direction of the bias has to be verified for at least four values of ρ:

1. If $\rho < 0$ then the attenuation bias is even stronger.
2. If $0 < \rho < 0.5$ attenuation bias is reduced relative to the CEV case, but is still present.
3. If $\rho = 0.5$ there is no bias altogether.
4. If $\rho > 0.5$ the OLS estimator is positively biased.

We examine four values of ρ : $-0.5, 0.25, 0.5, 1$ and perform the MC study for each of these cases. The results are reported in Tables 3a-d.

Table 3a: Summary statistics from simulations rho=-0.5

VARIABLES	(1) N	(2) mean	(3) p50	(4) sd	(5) min	(6) max
θ_1	1,000	-0.00256	-0.00256	0.0585	-0.182	0.185

Table 3b: Summary statistics from simulations rho=0.25

VARIABLES	(1) N	(2) mean	(3) p50	(4) sd	(5) min	(6) max
θ_1	1,000	0.371	0.371	0.0492	0.225	0.511

Table 3c: Summary statistics from simulations rho=0.5

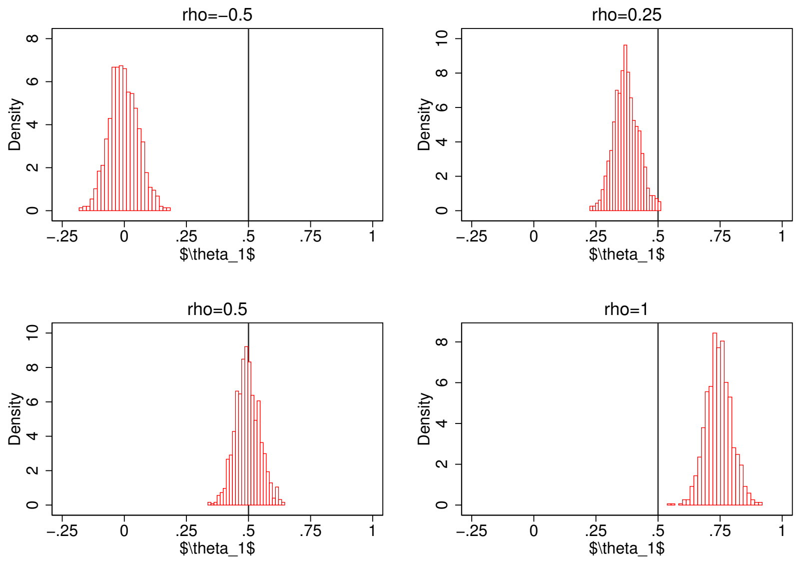
VARIABLES	(1) N	(2) mean	(3) p50	(4) sd	(5) min	(6) max
θ_1	1,000	0.495	0.495	0.0487	0.336	0.646

Table 3d: Summary statistics from simulations rho=1

VARIABLES	(1) N	(2) mean	(3) p50	(4) sd	(5) min	(6) max
θ_1	1,000	0.744	0.744	0.0519	0.537	0.919

The distributions of estimated θ_1 are displayed in figure 1.

Figure 1: Monte Carlo simulations for different values of rho



Monte Carlo simulations for different values of ρ confirm the analytical results from Problem 4.