Econometrics
Universidad Carlos III de Madrid
Solutions Final Exam
May 27, 2013

1. [6 points/over 10] *We are interested in testing the impact of family size (number of children) on the amount of resources invested in children. Specifically, our depending variable, private, is a binary variable taking a value one in case parents had sent a child to a private school and zero otherwise. In order to answer this question we got access to US census data for the year 1980. In specific, we would like to estimate the following model:*

$$private = \beta_0 + \beta_1 kids + \beta_2 dropout + \beta_3 dropout * black + \beta_4 black + \beta_5 Age + u$$

where kids is the number of children in the household; droput is a binary variable taking a value one in case a mother has not finished high school and zero, otherwise; black is a binary variable taking a value one in case the mother is Afro-American and zero, otherwise, and Age represents the age of the mother in years. Our knowledge of the problem makes us believe that $Cov(kids, u) \neq 0$*. In order to address this problem we constructed two instruments (Angrist and Evans, AER 1998): mb2 is a binary variable taking a value one in case the mother has faced twins births and zero otherwise, and ssex that is a binary variable taking a value one in case the first two pregnancies of a woman end up being two boys or two girls.*

Table 1.

	(1)	(2)	(3)	(4)	(5)
	DEPENDENT VARIABLE				
	private	kids	private	private	vmc2e
VARIABLES	OLS	OLS	2SLS	OLS	OLS
ssex		0.077800			0.000580
		(0.001940)			(0.000744)
mb2		0.850000			-0.001480
		(0.009120)			(0.003670)
dropout	-0.047200	0.289000	-0.044200	-0.044200	-0.000005
	(0.001540)	(0.006160)	(0.001920)	(0.001920)	(0.001530)
dropout*black	0.008150	0.291000	0.011400	0.011400	0.000002
	(0.002420)	(0.017600)	(0.002720)	(0.002720)	(0.002420)
black	-0.049700	0.313000	-0.046400	-0.046400	0.000004
	(0.000966)	(0.003620)	(0.001570)	(0.001570)	(0.000941)
age	0.001110	0.019900	0.001200	0.001200	0.000001
	(0.000065)	(0.000151)	(0.000072)	(0.000072)	(0.000065)
kids	0.001130		-0.009360	-0.009360	
	(0.000408)		(0.004020)	(0.004020)	
v				0.010600	
				(0.004010)	
Constant	0.067200	1.813000	0.091600	0.091600	-0.000303
	(0.002480)	(0.005030)	(0.009610)	(0.009610)	(0.002310)
F-statistic for H0 that mb2 and ssex are jointly significant		5022.77			0.4
Observations	647565	647565	647565	647565	647565
R^2	0.00513	0.05283	0.00412	0.00412	0

On the top of each column is reported the dependent variable for each of the estimated models. OLS and 2SLS stands for Ordinary Least Squares and Two Stage Least Squares, respectively. OLS indicates that the model reported in that column has been estimated by OLS; 2SLS, that the model in that specific column has been estimated by 2SLS. v and vmc2e correspond to the residuals of the models estimated in column (2) and (3), respectively.

1. *What is the interpretation of β_1? [1 point/over 10]*

   [0/1: 0 points or 1 point, all right or all wrong, no fractions or decimal numbers] Since we have a linear probability model, β_1 indicates the change in the probability that a child would attend a private school when we increase the number of siblings in one (ceteris paribus).

2. *Are the instruments ssex and mb2 relevant instruments for kids? Explain. Do we have evidence of weak instruments? Explain. [1 point/over 10]*

   [0/1] In order to check the relevance of *ssex* and *mb*2, we should check the joint significance of these variables in the first stage (column 2) that is $H0 : \pi_{ssex} = \pi_{mb2} = 0$. The reported F statistic at the bottom of column 2 associated to this null is 5022.77 way over any critical value. That is, we not only have evidence that the instruments are relevant but also we can rule out the concern of weak instruments since we observe a F bigger than 10.

3. *Can we test the exogeneity of the instruments in this problem? Explain and provide an answer in case that exogeneity can be tested. Given the previous information, are ssex and mb2 valid instruments? [1 point/over 10]*

   [0/0.5 explanation; 0/0.5 test and checking on the weakness of the instruments] Given the fact that we have one endogenous variable ($k = 1$) and two excluded instruments ($m = 2$), the model of interest is over-identified. Therefore we can perform the over-identification test to check the exogeneity of the instruments. The over-identification test we learned over the course use the J statistic that has distribution χ^2_{m-k}. In specific, $J = mF$ where F corresponds to the F for the null hypothesis $H0 : \delta_{ssex} = \delta_{mb2} = 0$ in the model that has as dependent variable the residuals from MC2E (column 3) and regressors the excluded instruments (*ssex* and *mb*2) and the rest of the controls (*dropout*, $dropout * black$, *black*, *age*, and *Immigrant*) which is reported in column 5. In this way, the J-statist is 0.8 which is below the critical value of 3.8. That is we cannot reject the null, that is, we do not have evidence that the instruments are correlated with the error term.

   The validity of the instruments is defined by the relevance and exogeneity of these instruments. In the previous two questions we have shown that these two instruments are strongly correlated with the suspected endogenous variable (*kids*) and we have not been able to detect a correlation with the error term. Therefore we have supporting evidence about the validity of these two instruments.

4. *Do we have supporting evidence about the endogeneity of family size? [1 point/over 10]*

   [0/1] In order to check the endogeneity of *kids* we need to do Hausman which implies testing the null $H0 : \beta_v = 0$ in the auxiliar regression in column 4. The t-statistic for this null is approximately 2.6 which is larger than the critical value for a 5% significance level (1.96). That is, we reject the null so we have evidence supporting the endogeneity of *kids*.

5. *Using the appropriate model, do we have evidence that parents with fewer children invest more in the education of their children? [1 point/over 10]*

   [0/1] Given the evidence found in the previous question about the evidence of endogeneity of family size (plus the fact we have valid instruments) we should use for this and the following question that output reported in column 3. Given this output an additional child reduces the probability of attending a private school 0.93 percentage points. Moreover, we need to test the significance of this impact. The t-statistic for the null $H0 : \beta_1 = 0$ is approximately 2.3 which is larger than the critical value for a significance level of 5%. Therefore we have supporting evidence that families that have more children in average reduce their investment per-child measured by the probability of attending private school.

6. *Test that the impact that a mother has not finished high school is the same for Afro-American mothers and non Afro-American mothers. [1 point/over 10]*

   In order the impact that a mother has not finished high school would be the same between Afro-American and non Afro-American it should be the case that $\beta_3 = 0$. Using the appropriate model, column 3, the t-statistic for the null $H0 : \beta_3 = 0$ is approximately 4.2 which is larger than the critical value for a significance level of 5%. That is, we have evidence that the impact of education measured by the variable *dropout* is different between Afro-American and non Afro-American.

2. *[4 points/over 10] We are interested in explaining a worker's wage in terms of the number of years of education (educ) and years of experience (exper) using the following model:*

$$log\,(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + u,$$

where we assume that u satisfies the classical assumptions of OLS and is homoskedastic. The estimated parameters by OLS for a sample of $n = 935$ observations are displayed in the first column of Table 2.

Table 2. OLS estimates. Dependent variable: log($wage$)

Variables	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
educ	0.07778	0.05316	0.07815	0.07192	0.071984	0.05571
	(0.00669)	(0.02085)	(0.00653)	(0.00666)	(0.00677)	(0.00600)
exper	0.01977	0.00038	0.01829	0.01791	0.01678	
	(0.00330)	(0.01066)	(0.00330)	(0.00327)	(0.01389)	
$educ * married$		0.02813				
		(0.02194)				
$exper * married$		0.01952				
		(0.01120)				
married		-0.38069	0.20926	0.18886	0.18873	0.21311
		(0.36818)	(0.04272)	(0.04764)	(0.04763)	(0.04709)
black				-0.24167	-0.24128	-0.22500
				(0.08391)	(0.08417)	(0.08212)
$married * black$				0.03599	0.03543	0.01071
				(0.09387)	(0.09404)	(0.09224)
$exper^2$					4.86e-05	
					(0.00058)	
const	5.50271	5.85694	5.32796	5.46166	5.46653	5.86609
	(0.11427)	(0.34889)	(0.11574)	(0.12017)	(0.12914)	(0.09445)
Observations	935	935	935	935	935	935
R^2	0.13086	0.15705	0.15420	0.18131	0.18132	0.15417

Several extensions of this model were considered to address the effects of being married (with the binary variable married) and/or being black (with binary black) or possible nonlinearity on the effect of years of experience. Using the appropriate output from Table 2 answer the following questions:

1. *Test whether the wage regressions for married workers and unmarried workers are the same. Write down the estimated regression for each group and interpret the coefficient of educ. [1 point/over 10]*

   [0/0.5] For this test we have to consider the non restricted regression (Model 2)

   $$log\,(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 educ * married + \beta_4 exper * married + \beta_5 married + u$$

   and test

   $$H_0 \; : \; \beta_3 = \beta_4 = \beta_5 = 0$$
   $$H_1 \; : \; H_0 \;\; \text{false}$$

with the restricted regression (Model 1) with the F statistic under homoskedasticity,

$$\begin{aligned} F &= \frac{R^2_{nores} - R^2_{res}}{1 - R^2_{nores}} \frac{n-k-1}{q} \\ &= \frac{0.15705 - 0.13086}{1 - 0.15705} \frac{935-5-1}{3} = 9.6212 \end{aligned}$$

where the F statistic is distributed under the null as a $\chi^2_3/3$, with 5% critical value equal to $7.81/3 = 2.603$. That is, we can reject the null hypothesis that both regressions are the same.

[0/0.5] The two estimated regressions are then

$$\begin{aligned} \text{non married: } \log\widehat{(wage)} &= (5.85694+) + (0.05316)\, educ + (0.00038)\, exper \\ \text{married: } \log\widehat{(wage)} &= (5.85694 - 0.36818) + (0.05316 + 0.02813)\, educ + (0.00038 + 0.01952)\, expe \\ &= 5.4888 + 0.08129 educ + 0.0199 exper. \end{aligned}$$

2. *Based on a statistical test, do the effects of education and experience depend on the civil status? [1 point/over 10]*

   [0/1] For this test we have to consider again the non restricted regression (Model 2)

   $$log\,(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 educ * married + \beta_4 exper * married + \beta_5 married + u$$

   and test

   $$\begin{aligned} H_0 &: \beta_3 = \beta_4 = 0 \\ H_1 &: H_0 \text{ falsa} \end{aligned}$$

   with the restricted regression (Model 3) with the F statistic under homoskedasticity,

   $$\begin{aligned} F &= \frac{R^2_{nores} - R^2_{res}}{1 - R^2_{nores}} \frac{n-k-1}{q} \\ &= \frac{0.15705 - 0.15420}{1 - 0.15705} \frac{935-5-1}{2} = 1.5705, \end{aligned}$$

   where the F statistic is distributed under the null as a $\chi^2_2/2$, with 5% critical value equal to 3.00. That is, we can not reject the null hypothesis and we do not find evidence supporting that the effects of education and experience depend on the civil status.

3. *Given a number of years of education and experience, which is the group defined by the variables married and black with the lowest average* $\log(wage)$*? Does this group have a significantly lower wage than the group of single non-black workers? [Use Model 4 for this question] [1 point/over 10]*

   [0/0.5] For the first question we check the coefficients of *married*, *black* and *married* $*$ *black* in Model 4 ,

   $$E\left[\log(wage) \backslash educ, exper, married, black\right] = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_m married + \beta_b black + \beta_{mb} married * b$$

   and the different subgrups,

   $$\begin{aligned} E\left[\log(wage) \backslash educ, exper, married = 0, black = 0\right] &= \beta_0 + \beta_1 educ + \beta_2 exper \\ E\left[\log(wage) \backslash educ, exper, married = 1, black = 0\right] &= \beta_0 + \beta_1 educ + \beta_2 exper + \beta_m \\ E\left[\log(wage) \backslash educ, exper, married = 0, black = 1\right] &= \beta_0 + \beta_1 educ + \beta_2 exper + \beta_b \\ E\left[\log(wage) \backslash educ, exper, married = 1, black = 1\right] &= \beta_0 + \beta_1 educ + \beta_2 exper + \beta_m + \beta_b + \beta_{mb} \end{aligned}$$

for a fixed level of *educ* and *exper* and the combination which provides a lowest predicted value of $\log(wage)$ is $black = 1$ and $married = 0$.

[0/0.5] For the comparision we obtain

$$\begin{aligned} E\left[\log(wage) \mid educ, exper, married = 0, black = 1\right] &= \beta_0 + \beta_1 educ + \beta_2 exper + \beta_b \\ E\left[\log(wage) \mid educ, exper, married = 0, black = 0\right] &= \beta_0 + \beta_1 educ + \beta_2 exper \end{aligned}$$

so that the one-sided test required is

$$\begin{aligned} H_0 &: \beta_b = 0 \\ H_1 &: \beta_b < 0 \end{aligned}$$

which is performed with a t test,

$$t = \frac{\hat{\beta}_b}{e.e.\left(\hat{\beta}_b\right)} = \frac{-0.24167}{0.08391} = -2.8801$$

which is significant at the 1% and therefore the null hypothesis of wages equality is rejected in favour of black having a lower wage within the group of non-married.

4. *What we conclude about the possible nonlinearity of the relationship of* $\log(wage)$ *with respect to the years of experience? Can you conclude that years of experience has no significant effect on* $\log(wage)$ *in Model 5? Make two statistical tests to answer these questions. [1 p/ 10]*

   [0/0.5] For the first question we have to test in Model 5 if

   $$\begin{aligned} H_0 &: \beta_{exper^2} = 0 \\ H_1 &: \beta_{exper^2} \neq 0 \end{aligned}$$

   through a t test,

   $$t = \frac{\hat{\beta}_{exper^2}}{e.e.\left(\beta_{exper^2}\right)} = \frac{4.86e-05}{0.00058} = 0.084$$

   which is not significant at the 5% level, so that the quadratic (non linear) term is not significant once we have already included the linear term.

   [0/0.5] For the second question we have to test in Model 5 if

   $$\begin{aligned} H_0 &: \beta_{exper} = \beta_{exper^2} = 0 \\ H_1 &: H_0 \text{ false} \end{aligned}$$

   compared to the restricted regression (Model 6), with an F test under homoskedasticity,

   $$\begin{aligned} F &= \frac{R^2_{nores} - R^2_{res}}{1 - R^2_{nores}} \frac{n-k-1}{q} \\ &= \frac{0.18132 - 0.15417}{1 - 0.18132} \frac{935-6-1}{2} = 15.388, \end{aligned}$$

   where the F statistic is distributed under the null as a $\chi^2_2/2$, with 5% critical value equal to 3.00, so that we reject the null hypothesis that the years of experience have no effect on $\log(wage)$, despite the two t statistics are not individually significative.