FINAL EXAM

Econometrics

Universidad Carlos III de Madrid

26/05/21

Write your name and group in each answer sheet. Answer all the questions in 2:30 hours.

1. (40%) Let $\{Y_i, X_{1i}, X_{2i}\}_{i=1}^n$ be observations independent and identically distributed as the random variables (Y, X_1, X_2) of some population, which maintain a causal relation according to the model

$$Y = X_1\beta_1 + X_2\beta_2 + u, \tag{1}$$

where u is an error with zero mean, variance σ^2, and independent of (X_1, X_2), and β_1 and β_2 are unknown parameters.

a. (1/5) Show that, if $\mathbb{E}(X_1X_2) = 0$ and $\mathbb{E}(X_1^2) > 0$,

$$\beta_1 = \frac{\mathbb{E}(X_1Y)}{\mathbb{E}(X_1^2)}.$$

SOLUTION:

$$\begin{aligned}
\frac{\mathbb{E}(X_1Y)}{\mathbb{E}(X_1^2)} &= \frac{\mathbb{E}(X_1(X_1\beta_1 + X_2\beta_2 + u))}{\mathbb{E}(X_1^2)} \\
&= \frac{\beta_1\mathbb{E}(X_1^2) + \beta_2\mathbb{E}(X_1X_2) + \mathbb{E}(X_1u)}{\mathbb{E}(X_1^2)} \\
&= \beta_1 \text{ because } \mathbb{E}(X_1X_2) = \mathbb{E}(X_1u) = 0.
\end{aligned}$$

b. (2/5) Derive an expression for the β_1 estimator in model (1) as the OLS coefficient in a simple regression where the explanatory variable is the residual in the OLS fit of X_1 on X_2 without constant.

SOLUTION: Exercise 3 of the Problem Set 5 part (d) solved in class. The OLS estimate of the coefficients

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \arg\min_{b_1,b_2} \sum_{i=1}^n (Y_i - b_1X_{1i} - b_2X_{2i}).$$

The FOC of OLS are

$$
\begin{aligned}
-2\sum_{i=1}^{n}\left(Y_i - X_{1i}\hat{\beta}_1 - X_{2i}\hat{\beta}_2\right)X_{1i} &= 0 \\
-2\sum_{i=1}^{n}\left(Y_i - X_{1i}\hat{\beta}_1 - X_{2i}\hat{\beta}_2\right)X_{2i} &= 0,
\end{aligned}
$$

and solving for $\hat{\beta}_1$, we obtain,

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\left(\sum_{i=1}^{n} X_{1i}Y_i\right)\left(\sum_{i=1}^{n} X_{2i}^2\right) - \left(\sum_{i=1}^{n} X_{2i}Y_i\right)\left(\sum_{i=1}^{n} X_{1i}X_{2i}\right)}{\left(\sum_{i=1}^{n} X_{1i}^2\right)\left(\sum_{i=1}^{n} X_{2i}^2\right) - \left(\sum_{i=1}^{n} X_{1i}X_{2i}\right)^2} \\
&= \frac{\left(\sum_{i=1}^{n} X_{2i}^2\right)\sum_{i=1}^{n} Y_i\left(X_{1i} - X_{2i}\frac{\sum_{i=1}^{n} X_{1i}X_{2i}}{\sum_{i=1}^{n} X_{2i}^2}\right)}{\left(\sum_{i=1}^{n} X_{2i}^2\right)\sum_{i=1}^{n} X_{1i}\left(X_{1i} - X_{2i}\frac{\sum_{i=1}^{n} X_{1i}X_{2i}}{\sum_{i=1}^{n} X_{2i}^2}\right)} \\
&= \frac{\sum_{i=1}^{n} Y_i\left(X_{1i} - X_{2i}\hat{\delta}\right)}{\sum_{i=1}^{n} X_{1i}\left(X_{1i} - X_{2i}\hat{\delta}\right)} \\
&= \frac{\sum_{i=1}^{n} Y_i\hat{e}_{1i}}{\sum_{i=1}^{n} X_{1i}\hat{e}_{1i}} = \frac{\sum_{i=1}^{n} Y_i\hat{e}_{1i}}{\sum_{i=1}^{n} \hat{e}_{1i}^2}.
\end{aligned}
$$

where

$$
\hat{e}_{1i} = X_{i1} - X_{2i}\hat{\delta} \text{ and } \hat{\delta} = \frac{\sum_{i=1}^{n} X_{1i}X_{2i}}{\sum_{i=1}^{n} X_{2i}^2},
$$

are the OLS residuals and the corresponding estimate of the slope in the regression of X_1 on X_2, respectively. Note that

$$
\sum_{i=1}^{n} X_{1i}\hat{e}_{1i} = \sum_{i=1}^{n}\left(\hat{\delta}X_{2i} + \hat{e}_{1i}\right)\hat{e}_{1i} = \sum_{i=1}^{n}\hat{e}_{1i}^2
$$

because $\sum_{i=1}^{n} X_{2i}\hat{e}_{1i} = 0$ from the FOC of this regression.

Therefore, $\hat{\beta}_1$ is the estimate of the slope in the model

$$
Y_i = \beta_1\hat{e}_{1i} + error.
$$

An alternative solution is to write both estimated regressions as

$$
Y_i = \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{e}_i, \quad \text{where } \sum_{i=1}^{n} X_{1i}\hat{e}_i = \sum_{i=1}^{n} X_{2i}\hat{e}_i = 0
$$

and

$$X_{1i} = \hat{\delta}_1 X_{2i} + \hat{e}_{1i}, \quad \text{where } \sum_{i=1}^{n} X_{2i}\hat{e}_{1i} = 0,$$

so that using the first expression

$$\begin{aligned}
\sum_{i=1}^{n} \hat{e}_{1i} Y_i &= \sum_{i=1}^{n} \hat{e}_{1i}\left(\hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{e}_i\right) \\
&= \hat{\beta}_1 \sum_{i=1}^{n} \hat{e}_{1i} X_{i1} + \hat{\beta}_2 \sum_{i=1}^{n} \hat{e}_{1i} X_{2i} + \sum_{i=1}^{n} \hat{e}_{1i}\hat{e}_i \\
&= \hat{\beta}_1 \sum_{i=1}^{n} \hat{e}_{1i}^2 + \hat{\beta}_2 \cdot 0 + \sum_{i=1}^{n} \left(X_{1i} - \hat{\delta}_1 X_{2i}\right)\hat{e}_i \\
&= \hat{\beta}_1 \sum_{i=1}^{n} \hat{e}_{1i}^2
\end{aligned}$$

because $\sum_{i=1}^{n} X_{1i}\hat{e}_i = \sum_{i=1}^{n} X_{2i}\hat{e}_i = 0$, and the expression for $\hat{\beta}_1$ follows at once.

c. (2/5) Suppose X_2 is not observable and is correlated with X_1. We must estimate β_1 in the model

$$Y_i = \beta_0 + X_{1i}\beta_1 + v_i, \ i = 1, ..., n \tag{2}$$

where $\beta_0 = \mu_{X_2}\beta_2$ and $v_i = \left(X_{2i} - \mu_{X_2}\right)\beta_2 + u_i$ ($\mu_{X_2} = \mathbb{E}(X_2)$). Suppose we have an instrument Z that satisfies the exogeneity and relevance conditions.

i. (1/3 of 1.c) Show that X_1 is an endogenous variable in model (2).

SOLUTION:

$$\begin{aligned}
Cov\left(X_{1i}, v_i\right) &= Cov\left(X_{1i}, \left(X_{2i} - \mu_{X_2}\right)\beta_2 + u_i\right) \\
&= \beta_2 Cov\left(X_{1i}, \left(X_{2i} - \mu_{X_2}\right)\right) + Cov\left(X_{1i}, u_i\right) \\
&= \beta_2 Cov\left(X_{1i}, X_{2i}\right) \\
&\neq 0
\end{aligned}$$

Note: $Cov\left(X_{1i}, \left(X_{2i} - \mu_{X_2}\right)\right) = \mathbb{E}\left(X_{1i}\left(X_{2i} - \mu_{X_2}\right)\right) - \mathbb{E}\left(X_{1i}\right)\mathbb{E}\left(X_{2i} - \mu_{X_2}\right)$ and $\mathbb{E}\left(X_{2i} - \mu_{X_2}\right) = 0$.

ii. (2/3 of 1.c) Express β_0, β_1 and v_i in terms of the coefficients and errors in the reduced forms of Y and X_1.

SOLUTION: Reduced forms:

$$\begin{aligned} X_{1i} &= \pi_0 + \pi_1 Z_i + e_i \\ Y_i &= \gamma_0 + \gamma_1 Z_i + w_i \end{aligned}$$

Now

$$Z_i = -\frac{\pi_0}{\pi_1} + \frac{1}{\pi_1} X_{1i} - \frac{1}{\pi_1} e_i,$$

substituting in the reduced form of Y,

$$Y_i = \left(\gamma_0 - \gamma_1 \frac{\pi_0}{\pi_1}\right) + \frac{\gamma_1}{\pi_1} X_{1i} + \left(w_i - \frac{\gamma_1}{\pi_1} e_i\right).$$

Therefore: $\beta_0 = \gamma_0 - \gamma_1 \pi_0 / \pi_1$, $\beta_1 = \gamma_1 / \pi_1$, and $v_i = w_i - e_i \gamma_1 / \pi_1$.

2. (30%) The "scrap rate" for a manufacturing firm is the number of defective items -products that must be discarded- out of every 100 produced. We are interested in using the scrap rate to measure the effect of worker training on productivity.

A sample of firms is used to obtain the following regression results,

$$\begin{aligned} \ln\widehat{(scrap_i)} &= \underset{(4.57)}{11.74} - \underset{(0.019)}{0.042} hrsemp_i - \underset{(0.370)}{0.951} \ln(sales_i) + \underset{(0.360)}{0.992} \ln(employ_i), \\ n &= 43, \; SCR = 65.91 \end{aligned} \tag{3}$$

where $hrsemp$ is the annual hours of training per employee, $sales$ is the annual firm sales (in dollars) and $employ$ is the number of firms employees. It is reported that $\widehat{Cov}\left(\hat{\beta}_{\ln(sales)}, \hat{\beta}_{\ln(employ)}\right) = -0.11$. Standard errors and covariance estimate are robust in the presence of heteroskedasticity.

a. (2/5) Somebody decided to slightly reformulate the above specification and obtained:

$$\begin{aligned} \ln\widehat{(scrap_i)} &= \underset{(4.57)}{11.74} - \underset{(0.019)}{0.042} hrsemp_i - \underset{(0.370)}{0.951} \ln\left(\frac{sales_i}{employ_i}\right) + \underset{(??)}{0.041} \ln(employ_i), \\ n &= 43, \; SCR = 65.91 \end{aligned} \tag{4}$$

Show that there is a one-to-one relationship between the parameters in (3) and (4) (1/2 of 2.a). Using this relationship and the information provided, calculate the omitted standard error of $\ln(employ)$ in (4) (1/2 of 2.a).

SOLUTION: We have

$$\ln\widehat{(scrap_i)} = \hat{\beta}_0 + \hat{\beta}_1 hrsemp_i + \hat{\beta}_2 \ln(sales_i) + \hat{\beta}_3 \ln(employ_i)$$

and defining $\hat{\theta} = \hat{\beta}_2 + \hat{\beta}_3$, we can write the model as

$$\ln\widehat{(scrap_i)} = \hat{\beta}_0 + \hat{\beta}_1 hrsemp_i + \hat{\beta}_2 \ln\left(\frac{sales_i}{employ_i}\right) + \hat{\theta} \ln(employ_i).$$

To calculate the standard error:

$$\begin{aligned} SE\left(\hat{\theta}\right) &= \sqrt{\widehat{Var}\left(\hat{\beta}_2\right) + \widehat{Var}\left(\hat{\beta}_3\right) + 2\widehat{Cov}\left(\hat{\beta}_2, \hat{\beta}_3\right)} \\ &= \sqrt{0.370^2 + 0.360^2 + 2 \cdot (-0.11)} \\ &= 0.21564 \end{aligned}$$

b. (1/5) How the (4)'s equation coefficients would change if *sales* is reported in thousands of dollars rather than in dollars? Provide the numerical values of the new estimated coefficients.

SOLUTION: The answer is the same in we analyze either model (3) or (4). When we change the units of measure, *sales* becomes in the model $sales_i^* = sales/1000$ and we have

$$\begin{aligned}
\ln\widehat{(scrap_i)} &= \hat{\beta}_0 + \hat{\beta}_1 hrsemp_i + \hat{\beta}_2 \ln\left(\frac{sales_i}{employ_i}\right) + \hat{\theta}\ln(employ_i) \\
&= \hat{\beta}_0 + \hat{\beta}_1 hrsemp_i + \hat{\beta}_2 \ln\left(\frac{1000 \cdot sales_i^*}{employ_i}\right) + \hat{\theta}\ln(employ_i) \\
&= \left(\hat{\beta}_0 + \ln(1000)\,\hat{\beta}_2\right) + \hat{\beta}_1 hrsemp_i + \hat{\beta}_2 \ln\left(\frac{sales_i^*}{employ_i}\right) \\
&\quad + \hat{\theta}\ln(employ_i)
\end{aligned}$$

The only coefficient that changes is the intercept, which now becomes

$$\left(\hat{\beta}_0 + \ln(1000)\cdot\hat{\beta}_2\right) = (11.74 + \ln(1000)\cdot(-0.951)) = 5.170\,7.$$

c. (2/5) Controlling for workers training ($hrsemp$) and for the sales to employees ratio ($sales/employ$), do bigger firms have larger statistically significant scrape rate? Establish the null and alternative hypotheses, the decision rule, and perform the test (3/4 of 2.c). How would you test that the $hrsemp$ and $\ln(sales)$ coefficients are identical but of different signs? (1/4 of 2.c). Critical values of the standard normal Z: $Z_{0.005} = 2.58$, $Z_{0.01} = 2.33$, $Z_{0.025} = 1.96$, $Z_{0.05} = 1.64$, $Z_{0.1} = 1.28$, where $\mathbb{P}(Z > Z_\alpha) = \alpha$.

SOLUTION: The hypothesis to be tested is

$$H_0 : \theta = 0 \text{ vs } H_1 : \theta > 0$$

where $\theta = \beta_2 + \beta_3$. We use the t statistic,

$$t = \frac{\hat{\theta}}{SE\left(\hat{\theta}\right)} = \frac{0.041}{0.21564} = 0.19013.$$

We do not reject H_0 at any reasonable significance level for a one-sided or two-sided alternative.

For the second test, the hypothesis to be tested is

$$H_0 : \beta_1 = -\beta_2 \text{ vs } H_1 : \beta_1 \neq -\beta_2$$

and we would compute the t statistic

$$t = \frac{\hat{\beta}_1 + \hat{\beta}_2}{SE\left(\hat{\beta}_1 + \hat{\beta}_2\right)}$$

with $\hat{\beta}_1 + \hat{\beta}_2 = 0.042 - 0.951 = -0.909$ and to calculate

$$SE\left(\hat{\beta}_1 + \hat{\beta}_2\right) = \sqrt{\widehat{Var}\left(\hat{\beta}_1\right) + \widehat{Var}\left(\hat{\beta}_2\right) + 2\widehat{Cov}\left(\hat{\beta}_1, \hat{\beta}_2\right)}$$

we would need to know $\widehat{Cov}\left(\hat{\beta}_1, \hat{\beta}_2\right)$ and proceed as usual: reject the null if $|t|$ is larger than the corresponding two-sided critical value, e.g. 1.96 for the 5% significance level.

3. (30%) Consider estimating the labour supply of married women. Labour demand provides the offered salary in terms of demanded hours. Once we impose the equilibrium condition, the two structural equations to be estimated are

$$\begin{aligned} hours &= \beta_{10} + \beta_{11}\ln(wage) + \beta_{12}educ + \beta_{13}age + \beta_{14}kidslt6 \\ &\quad + \beta_{15}nwifeinc + \beta_{16}(kidslt6 \times nwifeinc) + u_1 \end{aligned} \quad (5)$$

and

$$\ln(wage) = \beta_{20} + \beta_{21}hours + \beta_{22}exper + \beta_{23}exper^2 + u_2, \quad (6)$$

where age is the woman age, $educ$ years of education, $kidslt6$ the number of children less than 6, $nwifeinc$ is the income of the household in thousands of dollars, including husband salary, excluded the women wage, and $exper$ are the years of labour experience. We know that the control variables in the two equations ($educ$, age, $kidslt6$, $nwifeinc$ and $exper$) are independent of the errors (u_1 and u_2), which are independent with zero mean. Use the GRETL output at the end of the document to answer the questions.

a. (2/5) Which instrumental variables are available to estimate model (5) (labour supply) using $TSLS$? Explain with detail how to test that the available instruments are relevant: i) Provide the relevance condition and the null and alternative hypotheses; ii) explain how to compute the test statistic and iii) Provide the decision rule (1/2 of 3.a). Perform the test (1/2 of 3.a).

SOLUTION: The available instrumental variables are *exper* and $exper^2$, which are assumed to be exogenous. For their relevance, it must hold that: $\pi_6 \neq 0$ and/or $\pi_7 \neq 0$ in the reduced form equation

$$\begin{aligned} \ln(wage) &= \pi_0 + \pi_1 age + \pi_2 educ + \pi_3 kidslt6 + \pi_4 nwifeinc \\ &\quad + \pi_5 (kidslt6 \times nwifeinc) + \pi_6 exper + \pi_7 exper^2 + v \end{aligned}$$

We assume that explanatory variables in this equation do not exhibit perfect multicollinearity

i) We test that

$$H_0 : \pi_6 = \pi_7 = 0 \text{ vs } H_1 : \pi_6 \neq 0 \text{ and/or } \pi_7 \neq 0$$

through an F test.

ii) As we do not have available the variance-covariance matrix of coefficient estimates, but we know that the model is homoskedastic, we use the output from models 1 and 2, to compare the unrestricted and restricted models, respectively. In particular:

$$\begin{aligned} F &= \frac{1}{q}\frac{R^2_{unrestricted} - R^2_{restricted}}{(1 - R^2_{unrestricted})/n} \\ &= \frac{1}{2}\frac{0.163629 - 0.126543}{(1 - 0.163629)/428} \\ &= 9.4891. \end{aligned}$$

iii) To take a decision, we compare the value of F statistic with the critical value from a $\chi^2_2/2$, which is 3 at the 5% significance level. As $F = 9.4891 > 3$ we reject H_0 and conclude that the instruments are relevant. However, as $F < 10$, we could not conclude that the instruments are strong according to the rule of thumb which is applied in common practice.

b. (2/5) What is the mean difference in supplied hours for two women with identical characteristics except that one has two children less than 6 and her salary is the unique source of income, while the other woman does not have children less that six and she has a 20 thousand dollars household income additional to her salary? (1/2 of 3.b). Provide a confidence interval at 95% of confidence for such a difference and test whether the difference is significantly different from zero using the interval (1/2 of 3.b) Help: Critical values of the standard normal Z: $Z_{0.005} = 2.58$, $Z_{0.01} = 2.33$, $Z_{0.025} = 1.96$, $Z_{0.05} = 1.64$, $Z_{0.1} = 1.28$, where $\mathbb{P}(Z > Z_\alpha) = \alpha$.

SOLUTION: Using Model 3, the estimated difference is

$$\begin{aligned} \hat{\alpha} &= \left(2 \cdot \hat{\beta}_{14} + 0 \cdot \hat{\beta}_{15} + 2 \cdot 0 \cdot \hat{\beta}_{16}\right) - \left(0 \cdot \hat{\beta}_{14} + 20 \cdot \hat{\beta}_{15} + 0 \cdot 20 \cdot \hat{\beta}_{16}\right) \\ &= 2 \cdot \hat{\beta}_{14} - 20 \cdot \hat{\beta}_{15} \\ &= 2 \cdot (-240.214) - 20 \cdot (-10.6106) \\ &= -268.22 \end{aligned}$$

We conclude that the woman with no children and an additional household income of 20 thousand dollars to her salary, works 268.22 hours more on average than the woman with two children and whose salary is the only income in her household. The standard error of the estimate is

$$\begin{aligned} SE(\hat{\alpha}) &= \sqrt{\widehat{Var}\left(2 \cdot \hat{\beta}_{14} - 20 \cdot \hat{\beta}_{15}\right)} \\ &= \sqrt{2^2 \cdot \widehat{Var}\left(\hat{\beta}_{14}\right) + 20^2 \cdot \widehat{Var}\left(\hat{\beta}_{14}\right) - 2 \cdot 2 \cdot 20 \cdot \widehat{Cov}\left(\hat{\beta}_{14}, \hat{\beta}_{14}\right)} \\ &= \sqrt{2^2 \cdot 1.1107 \cdot 10^5 + 20^2 \cdot 52.837 - 2 \cdot 2 \cdot 20 \cdot 817.65} \\ &= 632.46. \end{aligned}$$

The 95% confidence interval is

$$\begin{aligned} \left[\hat{\alpha} - SE(\hat{\alpha}), \hat{\alpha} + SE(\hat{\alpha})\right] &= [-268.22 - 1.964 \cdot 632.46, \quad -268.22 + 1.964 \cdot 632.46] \\ &= [-1510.4, \quad 973.93]. \end{aligned}$$

Therefore, 0 is contained in the interval, and the null hypothesis that the difference in average worked hours between the two women is equal to zero cannot be rejected at the 5% significance level.

c. (1/5) Explain with detail how would you test that the instruments are exogenous: i) Establish the null and alternative hypotheses; ii) explain how to compute the test statistic and iii) provide the decision rule. (2/3 of 3.c). Perform the test at 5% of significance. (1/3 of 3.c). The critical values of the χ^2_q/q for $q = 1, \ldots, 5$ at 5% are $\chi^2_{1,0.05} = 3.84$, $\chi^2_{2,0.05}/2 = 3.00$,

$\chi^2_{3,0.05}/3 = 2.60$, $\chi^2_{4,0.05}/4 = 2.37$, $\chi^2_{5,0.05}/5 = 2.21$, respectively.

SOLUTION: i) The hypothesis to be tested is

$$H_0 : Cov\,(exper, u_1) = Cov\left(exper^2, u_1\right) = 0$$

vs

$$H_1 : Cov\,(exper, u_1) \neq 0 \; and/or \; Cov\left(exper^2, u_1\right) \neq 0$$

We would compute the residuals from model (5) estimated by 2SLS,

$$\begin{aligned} \hat{u}_i &= hours - \hat{\beta}_{10} - \hat{\beta}_{11} \ln{(wage)} - \hat{\beta}_{12} educ - \hat{\beta}_{13} age \\ &\quad - \hat{\beta}_{14} kidslt6 - \hat{\beta}_{15} nwifeinc - \hat{\beta}_{16}\,(kidslt6 \times nwifeinc) \end{aligned}$$

and compute the F statistic to test

$$H_0 : \gamma_1 = \gamma_2 = 0 \text{ vs } H_1 : \gamma_1 \neq 0 \text{ and/or } \gamma_2 \neq 0$$

in the model

$$\begin{aligned} \hat{u}_i &= \gamma_0 + \gamma_1 exper + \gamma_2 exper^2 + \gamma_3 educ + \gamma_4 age \\ &\quad + \gamma_5 kidslt6 + \gamma_6 nwifeinc + \gamma_7\,(kidslt6 \times nwifeinc) + e \end{aligned}$$

ii) The test statistic in this case is $J = 2F$ (one endogenous explanatory variable, $k = 1$, and 2 instruments, $m = 2$: $J = mF = 2F$) which is distributed approximately as a $\chi^2_{m-k} = \chi^2_1$ under H_0. iii) we will reject the null hypothesis when the value of the J statistic in the sample exceeds the critical value at the pre-specified significance level from a χ^2_1. We can check that $J = 2F = 2 \cdot 0.412728 = 0.825456$ in Model 4 and, therefore, the null hypothesis of exogeneity is not rejected at any reasonable significance level.

Model 1: OLS, using observations 1–428

Dependent variable: lwage

	Coefficient	Standard Error	t statistic	p value
const	−0.449268	0.285534	−1.5734	0.1164
age	−0.00269880	0.00520903	−0.5181	0.6047
educ	0.101004	0.0149790	6.7430	0.0000
kidslt6	0.00268457	0.163924	0.0164	0.9869
nwifeinc	0.00615072	0.00362019	1.6990	0.0901
nwifeincXkidslt6	−0.00322245	0.00795267	−0.4052	0.6855
exper	0.0414884	0.0132833	3.1233	0.0019
expersq	−0.000747477	0.000402880	−1.8553	0.0642

Average of dep. var	1.190173	Std. Dev. of dep. var.	0.723198
Sum Squared residuals	186.7847	Std. Error of regression	0.666877
R^2	0.163629	Adjusted R^2	0.149689
$F(7, 420)$	11.73846	p value (of F)	1.14e–13

Model 2: OLS, using observations 1–428

Dependent variable: lwage

	Coefficient	Standard Error	t statistic	p value
const	−0.434017	0.270691	−1.6034	0.1096
age	0.00503829	0.00455412	1.1063	0.2692
educ	0.107789	0.0151864	7.0977	0.0000
kidslt6	−0.0241129	0.166747	−0.1446	0.8851
nwifeinc	0.00315312	0.00356191	0.8852	0.3765
nwifeincXkidslt6	−0.00318766	0.00806527	−0.3952	0.6929

Average of dep. var	1.190173	Std. Dev. of dep. var.	0.723198
Sum Squared residuals	195.0670	Std. Error of regression	0.679885
R^2	0.126543	Adjusted R^2	0.116193
$F(5, 422)$	12.22748	p value (of F)	4.40e–11

Model 3: 2SLS, using observations 1–428

Dependent variable: hours

With Instruments: lwage

Instruments: const age educ kidslt6 nwifeinc nwifeincXkidslt6 exper expersq

	Coefficient	Standard Error	z	p value
const	2232.35	578.202	3.8608	0.0001
lwage	1643.67	471.901	3.4831	0.0005
age	−7.78173	9.40253	−0.8276	0.4079
educ	−184.111	59.2247	−3.1087	0.0019
kidslt6	−240.214	333.275	−0.7208	0.4711
nwifeinc	−10.6106	7.26892	−1.4597	0.1444
nwifeincXkidlt6	2.43423	16.1806	0.1504	0.8804

Average of dep. var	1302.930	Std. Dev. of dep. var.	776.2744
Sum Squared residuals	7.76e+08	Std. Error of regression	1358.086
R^2	0.000534	Adjusted R^2	-0.013710
$F(6, 421)$	2.866484	p value (of F)	0.009519

Covariance matrix of the coefficients in model 3

const	lwage	age	educ	kidslt6	nwifeinc	nwifXkids	
3.3432e+05	96652.	−4042.1	−21294.	−26188.	−225.98	833.62	const
	2.2269e+05	−1122.0	−24004.	5369.7	−702.17	709.86	lwage
		88.408	131.55	427.63	−3.2554	1.2589	age
			3507.6	−1403.5	16.223	−67.471	educ
				1.1107e+05	817.65	−4502.4	kidslt6
					52.837	−48.120	nwifeinc
						261.81	nwifXkids

Model 4: OLS, using observations 1–428

Dependent variable: uhat

IMPORTANT: uhat are the residuals corresponding to model 3.

	Coefficient	Standard Error	t statistic	p value
const	194.274	581.606	0.3340	0.7385
age	−3.68530	10.6103	−0.3473	0.7285
educ	0.190471	30.5108	0.0062	0.9950
kidslt6	16.7347	333.899	0.0501	0.9601
nwifeinc	1.19491	7.37399	0.1620	0.8713
nwifeincXkisdlt6	−1.50531	16.1989	−0.0929	0.9260
exper	−16.9634	27.0570	−0.6270	0.5310
expersq	0.673158	0.820630	0.8203	0.4125

Average of dep. var	6.44e–13	Std. Dev. of dep. var.	1348.511
Sum Squared residuals	7.75e+08	Std. Error of regression	1358.368
R^2	0.001962	Adjusted R^2	-0.014672
$F(7, 420)$	0.117922	p value (of F)	0.997131

Hypothesis test on model 4:

Null hypothesis: the regression parameters are zero for the variables exper and expersq

Test statistic: F(2,420)=0.412728, p value 0.66211