Universidad Carlos III de Madrid
Econometrics I
Final Exam, Convocatoria Ordinaria, Course 2009-2010.

January 18, 2010

Total time allowed for the exam: 2 hours and 30 minutes.

Important remark: Some information contained in the outputs is redundant. The exam contains 7 pages.

Question 1 Using the data in WAGE2.RAW, the following model has been estimated

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \beta_4 married + \beta_5 black + \beta_6 south + \beta_7 urban + U.$$

In GRETL outputs provided below, we estimate the above regression model and three other models using the Ordinary Least Squares (OLS) method. Use these outputs to answer the following questions.

a. (0,5 points) Holding other factors fixed, in the context of the underlying model (model 1 in the outputs provided below), what is the approximate difference between the monthly wage of black individuals and of those who are not? Is this difference statistically significant?

b. (0,5 points) Add the variables $exper^2$ and $tenure^2$ to the previous equation and show that they are jointly statistically insignificant even at 10% significance level.

c. (0,5 points) Extend the initial model such that the returns to education depend on the race of the individual. Write down the corresponding model and test the significance of this dependence.

d. (0,5 points) Using the initial model, allow wage to differ among four groups of individuals: married blacks, single blacks, married nonblacks , and single nonblacks. Express the model as concisely as possible and provide the estimated value of the wage differential between married black and married nonblack with the same values of the remaining explanatory variables.

Output 1 (Model 1): OLS, using the observations 1-935
Dependent Variable: $log(wage)$

	Coefficient	Std. Error	t-Statistic	Prob.
const	5,39550	0,113225	47,6529	0,0000
educ	0,0654307	0,00625040	10,4683	0,0000
exper	0,0140430	0,00318519	4,4089	0,0000
tenure	0,0117473	0,00245297	4,7890	0,0000
married	0,199417	0,0390502	5,1067	0,0000
black	−0,188350	0,0376666	−5,0004	0,0000
south	−0,0909037	0,0262485	−3,4632	0,0006
urban	0,183912	0,0269583	6,8221	0,0000

Sum squared resid	123,8185	S.E. of regression	0,365471
R-squared	0,252558	Adj. R-squared	0,246914

Output 2 (Model 2): OLS, using the observations 1-935

Dependent Variable: $log(wage)$

	Coefficient	Std. Error	t-Statistic	Prob.
const	5,35868	0,125914	42,5581	0,0000
educ	0,0642761	0,00631148	10,1840	0,0000
exper	0,0172146	0,0126138	1,3647	0,1727
tenure	0,0249291	0,00812966	3,0664	0,0022
married	0,198547	0,0391103	5,0766	0,0000
black	−0,190664	0,0377011	−5,0572	0,0000
south	−0,0912153	0,0262356	−3,4768	0,0005
urban	0,185424	0,0269585	6,8781	0,0000
*exper*ˆ2	−0,000113801	0,000531871	−0,2140	0,8306
*tenure*ˆ2	−0,000796448	0,000471013	−1,6909	0,0912

Sum squared resid	123,4210	S.E. of regression	0,365278
R-squared	0,254958	Adj. R-squared	0,247709

Output 3 (Model 3): OLS, using the observations 1-935

Dependent Variable: $log(wage)$

	Coefficient	Std. Error	t-Statistic	Prob.
const	5,37482	0,114703	46,8587	0,0000
educ	0,0671153	0,00642769	10,4416	0,0000
exper	0,0138259	0,00319063	4,3333	0,0000
tenure	0,0117870	0,00245289	4,8054	0,0000
married	0,198908	0,0390474	5,0940	0,0000
black	0,0948093	0,255399	0,3712	0,7106
south	−0,0894495	0,0262769	−3,4041	0,0007
urban	0,183852	0,0269547	6,8208	0,0000
$educ * black$	−0,0226237	0,0201827	−1,1209	0,2626

Mean dependent var	6,779004	S.D. dependent var	0,421144
Sum squared resid	123,6507	S.E. of regression	0,365420
R-squared	0,253571	Adj. R-squared	0,247122
$F(8, 926)$	39,32158	Prob (F-statistic)	$4{,}35\ 10^{-54}$

Output 4 (Model 4): OLS, using the observations 1-935

Dependent Variable: $log(wage)$

	Coefficient	Std. Error	t-Statistic	Prob.
const	5,40379	0,114122	47,3509	0,0000
educ	0,0654751	0,00625302	10,4710	0,0000
exper	0,0141462	0,00319103	4,4331	0,0000
tenure	0,0116628	0,00245795	4,7449	0,0000
married	0,188915	0,0428777	4,4059	0,0000
black	−0,240820	0,0960229	−2,5079	0,0123
south	−0,0919894	0,0263212	−3,4949	0,0005
urban	0,184350	0,0269778	6,8334	0,0000
$married * black$	0,0613538	0,103275	0,5941	0,5526

Mean dependent var	6,779004	S.D. dependent var	0,421144
Sum squared resid	123,7714	S.E. of regression	0,365599
R-squared	0,252842	Adj. R-squared	0,246388
$F(8, 926)$	39,17047	Prob (F-statistic)	$6{,}78\ \ 0^{-54}$

Question 2 Consider a Simultaneous Equations Model of the form of "Supply and Demand", where the same dependent variable Y_1 (typically, the "quantity") appears on the left-hand side of each equation:

$$\begin{aligned} Y_1 &= \alpha_1 Y_2 + \beta_1 Z_1 + U_1 \\ Y_1 &= \alpha_2 Y_2 + \beta_2 Z_2 + U_2. \end{aligned}$$

As usual the Z's are exogenous.

a. (0,5 points) If $\alpha_1 \neq 0$, $\alpha_2 \neq 0$ and $\alpha_1 \neq \alpha_2$, find the reduced form of Y_1. In this case, does Y_2 have a reduced form?

b. (0,5 points) Is it reasonable to assume that $\alpha_1 \neq \alpha_2$ in a Supply and Demand system?

c. (0,5 points) Which equation can never be identified when $\beta_2 = 0$?

Question 3 Consider a Linear Probability Model (LPM) to explain the probability of granting a mortgage:

$$E(Y|X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

where:
Y : takes the value 1 if the person is NOT granted the mortgage;
$X_1 = DEUDA =$ debt ratio per unit (excluding the mortgage loan considered) of the applicant;
$X_2 = NEG =$ takes the value 1 if the applicant is black and zero if it is white.

A sample of 2380 mortgage applications in Boston (USA) has been used to estimate the above model using OLS. The heteroskedasticity-robust standard errors are given in the brackets [.].

$$\hat{Y} = \underset{\substack{(0{,}029) \\ [0{,}040]}}{-0,091} + \underset{\substack{(0{,}089) \\ [0{,}129]}}{0,559} X_1 + \underset{\substack{(0{,}025) \\ [0{,}125]}}{0,177} X_2.$$

a. (0,5 points) Interpret the coefficient β_1. What is the probability that the mortgage was granted to a white applicant with a debt ratio of 30%?

b. (0,5 points) Is there heteroscedasticity in this model? Provide the value of the statistic to test whether being white is significant in the granting of a mortgage and perform the test at the 1% significance level.

Question 4 The following equation explains the weekly hours of television watched by a child depending on his (her) age, mother's education, father's education and number of siblings:

$$tvhours^* = \beta_0 + \beta_1 age + \beta_2 age^2 + \beta_3 mothereduc + \beta_4 fathereduc + \beta_5 sibs + U.$$

We are concerned that $tvhours^*$ is measured with an error in our survey. Suppose that $tvhours$ are the weekly hours of television stated in the survey.

a. (0,5 points) What must be satisfied in this application so that the Ordinary Least Squares estimator is unbiased?

b. (0,5 points) What if the *mothereduc* variable was also measured with an error under the assumptions of the previous part?

Question 5 There is evidence that women's work behavior is particularly determined by their fertility decisions. To assess the impact of fertility (measured by the number of children) on the hours worked, we consider the following specification

$$\begin{aligned} \texttt{HRS} = \; & \beta_0 + \beta_1 \texttt{WHITE} + \beta_2 \texttt{BLACK} + \beta_3 \texttt{HISPAN} \\ & + \beta_4 \texttt{HIGHSC} + \beta_5 \texttt{UNIV} + \beta_6 \texttt{AGE} + \beta_7 \texttt{AGE2} \\ & + \beta_8 \texttt{SPOUSE} + \beta_9 (\texttt{SPOUSE} \times \texttt{AGE}) + \beta_{10} (\texttt{SPOUSE} \times \texttt{AGE2}) \\ & + \beta_{11} \texttt{NCHILD} + \texttt{U} \end{aligned} \tag{E.1}$$

where, for every woman:

$\texttt{HRS}$ = number of hours worked per week;
$\texttt{WHITE}$ = binary variable which takes the value 1 if the woman is white and zero otherwise;
$\texttt{BLACK}$ = binary variable which takes the value 1 if the woman is black and zero otherwise;
$\texttt{HISPAN}$ = binary variable which takes the value 1 if the woman is Hispanic and zero otherwise;
$\texttt{HIGHSC}$ = binary variable which takes the value 1 if the woman has only completed secondary education and zero otherwise;
$\texttt{UNIV}$ = binary variable which takes the value 1 if the woman has a university degree and zero otherwise;
$\texttt{AGE}$ = age in years;
$\texttt{AGE2}$ = age squared ;
$\texttt{SPOUSE}$ = binary variable which takes the value 1 if the woman has a spouse who lives at home and zero otherwise;
$\texttt{NCHILD}$ = number of children under 18 years old living at home.

Note: There are four mutually exclusive ethnic groups: white, black, Hispanic and Asian. Only three levels of education (mutually exclusive): Primary education or less, secondary education, university education.

To estimate this equation, we use data from 69852 women with children from a U.S. survey conducted in 1980.

We also know that fertility decisions are correlated with unobserved characteristics which, in turn, affect employment decisions. For example, those women that are better prepared for the labor market may not only have higher wages on average (and therefore an opportunity cost of assuming the highest household chores), but also higher personal costs associated with childcare. Therefore, we would expect

$$Cov(\texttt{U}, \texttt{NCHILD}) \neq 0.$$

while the remaining variables on the right-hand side of (E.1) **are uncorrelated** with unobservable variables (U).

In addition to the above variables, we have the information on whether women had experienced multiple births (MB), i.e., if she delivered twins, triplets, quadruplets or quintuplets at a given pregnancy. We can therefore define the variable for multiple births (***Multiple Births***) MB:

$\texttt{MB}$ = binary variable which takes the value 1 if the woman has had multiple births and zero otherwise.

We also know that $Cov(\texttt{MB}, \texttt{U}) = 0$.

The estimation results corresponding to different models are provided below:

a. (0,5 points) Assuming that $Cov(\texttt{MB}, \texttt{U}) = 0$, test if MB is a valid instrument. Test at the 1% significance level if NCHILD is exogenous.

b. (0,5 points) Assuming that $Cov(\texttt{NCHILD}, \texttt{U}) \neq 0$ and that MB is a valid instrument, obtain the estimated value of the average decrease in the hours of work caused by having an extra child holding race, level of education, age and marital status fixed.

Output 1				
Dependent Variable: HRS				
Method: Ordinary Least Squares				
Sample: 69852				
Included observations: 69852				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
$WHITE$	1.3224	0.7901	1.67	0.094
$BLACK$	3.3607	0.8005	4.20	0.000
$HISPAN$	−0.5399	1.0007	−0.54	0.590
$HIGHSC$	4.4627	0.1770	25.22	0.000
$UNIV$	4.3868	0.2005	21.88	0.000
AGE	1.4505	0.0288	50.37	0.000
$AGE2$	−0.0207	0.0006	−36.04	0.000
$SPOUSE$	0.4258	0.4343	0.98	0.327
$SPOUSE * AGE$	−0.5668	0.0335	−16.90	0.000
$SPOUSE * AGE2$	0.0101	0.0006	15.42	0.000
$NCHILD$	−1.8506	0.0610	−30.32	0.000
C	0.2952	0.8568	0.34	0.730
R-squared		0.1344		
Adjusted R-squared		0.1343		
S.E. of regression		335.4738		

Output 2				
Dependent Variable: $NCHILD$				
Method: Ordinary Least Squares				
Sample: 69852				
Included observations: 69852				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
$WHITE$	−0.3527	0.0482	−7.32	0.000
$BLACK$	0.2822	0.0488	5.78	0.000
$HISPAN$	0.0638	0.0610	1.05	0.296
$HIGHSC$	−0.4758	0.0106	−44.68	0.000
$UNIV$	−0.5453	0.0121	−45.22	0.000
AGE	0.0801	0.0017	46.26	0.000
$AGE2$	−0.0012	0.00003	−35.08	0.000
$SPOUSE$	0.2622	0.0265	9.90	0.000
$SPOUSE * AGE$	−0.0095	0.0020	−4.65	0.000
$SPOUSE * AGE2$	0.0001	0.00004	3.42	0.001
MB	1.2192	0.0255	47.79	0.000
C	1.5957	0.0519	30.73	0.000
R-squared		0.1568		
Adjusted R-squared		0.1566		

Output 3				
Dependent Variable: HRS				
Method: Two-Stage Least Squares				
Sample: 69852				
Included observations: 69852				
Instrument list: MB				
Variable Variable	Coefficient	Std. Error	t-Statistic	Prob.
$WHITE$	1.7045	0.8019	2.13	0.034
$BLACK$	3.0750	0.8077	3.81	0.000
$HISPAN$	−0.5925	1.0029	−0.59	0.555
$HIGHSC$	4.9467	0.2398	20.63	0.000
$UNIV$	4.9483	0.2746	18.02	0.000
AGE	1.3697	0.0395	34.69	0.000
$AGE2$	−0.0195	0.0007	−27.66	0.000
$SPOUSE$	0.1529	0.4446	0.34	0.731
$SPOUSE*AGE$	−0.5575	0.0337	−16.52	0.000
$SPOUSE*AGE2$	0.0100	0.0007	15.15	0.000
$NCHILD$	−0.8362	0.3437	−2.43	0.015
C	−1.3616	1.0209	−1.33	0.182
R-squared		0.1310		
Adjusted R-squared		0.1308		

Output 4				
Dependent Variable: HRS				
Method: Ordinary Least Squares				
Sample: 69852				
Included observations: 69852				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
$WHITE$	1.7045	0.8002	2.13	0.033
$BLACK$	3.07502	0.8061	3.81	0.000
$HISPAN$	−0.5925	1.0008	−0.59	0.554
$HIGHSC$	4.9467	0.2393	20.67	0.000
$UNIV$	4.9483	0.2741	18.05	0.000
AGE	1.3697	0.0394	34.76	0.000
$AGE2$	−0.0195	0.0007	−27.71	0.000
$SPOUSE$	0.1529	0.4437	0.34	0.730
$SPOUSE*AGE$	−0.5575	0.0337	−16.55	0.000
$SPOUSE*AGE2$	0.0100	0.0007	15.19	0.000
$NCHILD$	−0.8362	0.3430	−2.44	0.015
RES	−1.0475	0.3486	−3.01	0.003
C	−1.3616	1.0188	−1.34	0.181
R-squared		0.1345		
Adj. R-squared		0.1344		

(**Remark:** RES are the residuals from Output 2)

CRITICAL VALUES

$Z_{0,025} = 1,96$	$Z_{0,05} = 1,645$	$Z_{0,01} = 2,326$	$Z_{0,005} = 2,576$
$Z_{0,1} = 1,282$	$\chi^2_{3;0,01} = 11,34$	$\chi^2_{3;0,05} = 7,82$	$\chi^2_{5;0,05} = 11,07$
$\chi^2_{2;0,05} = 5,99$	$\chi^2_{2;0,01} = 9,21$	$\chi^2_{6;0,05} = 12,59$	$\chi^2_{2;0,1} = 4,61$
$\chi^2_{6;0,01} = 16,81$	$\chi^2_{4;0,05} = 9,49$	$\chi^2_{3;0,1} = 6,25$	$\chi^2_{4;0,01} = 13,28$

Z is normal with zero mean and variance equal to one and χ_q^2 is the chi-squared with q degrees of freedom, $\Pr(Z > Z_\alpha) = \alpha$; $\Pr\left(\chi_q^2 > \chi_{q;\alpha}^2\right) = \alpha$.

Remark: Note that the F distribution can be approximated by the χ^2. That is, $\chi_q^2 \sim q \cdot F_{q,n-k-1}$ for $n-k-1$ sufficiently large, , $\Pr\left(\chi_q^2 > \chi_{q;\alpha}^2\right) \simeq \Pr\left(q \cdot F_{q,n-k-1} > \chi_{q;\alpha}^2\right)$.