EXTRAORDINARY FINAL EXAM ECONOMETRICS A

Answer all questions in 1 hour.

(30%) Use data for married women in MROZ dataset of Wooldridge to estimate the hours of work offered by women during a year ($hours$) in terms of $\log(wage)$ controlling for the following variables, which we take as exogenous: years of education ($educ$), age (age), an indicator of having children less than 6 years old ($kidslt6$) and other income in the household ($nwifeinc$).

1. (7.5%) Explain why we consider that $\log(wage)$ is endogenous in the previous regression model (2%). Estimate the labor supply curve by 2SLS using as instruments the years of experience ($exper$) and its squares. Interpret the coefficient of $\log(wage)$ (2%). Provide a 95% confidence interval for the labour supply elasticity at the mean value of worked hours in the sample and interpret it (3.5%).

   .

   In the labor market both hours or work ($hours$) and wages ($wage$) are determined simultaneously, and therefore $\log(wage)$ is an endogenous regressor in the work supply equation of women.

   The labor supply is described by the linear-log equation

$$hours = \beta_0 + \beta_1 \log(wage) + \beta_2 educ + \beta_3 age + \beta_4 kidslt6 + \beta_5 mwifeinc + u,$$

   where β_1 is a semielasticity: $\hat{\beta}_1 = 1639.56$ means that if wages increase by 1%, then the average labor supply of a woman increases approximately by $\hat{\beta}_1/100 = 1639.56/100 = 16.39$ hours a year.

   For the average worked hours of $1302.9 \approx 1303$ in the sample used in the regression as provided in the regression output, the requested elasticity is

$$\xi = \frac{\Delta hours/hours}{\Delta wage/wage} \approx \frac{\Delta hours/hours}{\ln \Delta wage} = \frac{1}{hours}\frac{\Delta hours}{\ln \Delta wage} \approx \frac{1}{hours}\beta_1,$$

   which is estimated by

$$\hat{\xi} = \frac{\hat{\beta}_1}{1303} = \frac{1639.56}{1303} = 1.2583$$

   with standard error equal to

$$SE\left(\hat{\xi}\right) = SE\left(\frac{\hat{\beta}_1}{1303}\right) = \frac{SE\left(\hat{\beta}_1\right)}{1303} = \frac{597.514}{1303} = 0.4586,$$

   so the 95% confidence interval is

$$\hat{\xi} \pm z_{0.025} SE\left(\hat{\xi}\right) \to 1.2583 \pm 1.96 \times 0.4586 \to (0.359\,44, 2.157\,2).$$

   This means that if wages increase in 1%, the average offered hours of labour increase in approximately 1.25% with confidence interval (0.36%, 2.16%) at this level of labor supply ($hours = 1303$).

   If instead we use the average value 740.58 for the whole sample (from the properties of the $wage$ variable) the results would be

$$\hat{\xi} = \frac{\hat{\beta}_1}{740.58} = \frac{1639.56}{740.58} = 2.213\,9$$

   with standard error equal to

$$SE\left(\hat{\xi}\right) = SE\left(\frac{\hat{\beta}_1}{740.58}\right) = \frac{SE\left(\hat{\beta}_1\right)}{740.58} = \frac{597.514}{740.58} = 0.806\,82$$

so the 95% confidence interval is

$$\hat{\xi} \pm z_{0.025} SE\left(\hat{\xi}\right) \rightarrow 2.2139 \pm 1.96 \times 0.80682 \rightarrow (0.63253, 3.7953)\,.$$

This means that if wages increase in 1%, the average offered hours of labour increase in approximately 0.81% with confidence interval (0.63%, 3.80%) at this level of labor supply ($hours$ = 740.58).

.

Model 1: TSLS, using observations 1–428
Dependent variable: hours
Instrumented: lwage
Instruments: const exper expersq age educ kidslt6 nwifeinc
Heteroskedasticity-robust standard errors, variant HC1

	Coefficient	Std. Error	t-ratio	p-value
const	2225.66	607.369	3.664	0.0003
lwage	1639.56	597.514	2.744	0.0063
age	−7.80610	10.5618	−0.7391	0.4603
educ	−183.751	68.2676	−2.692	0.0074
kidslt6	−198.154	209.901	−0.9440	0.3457
nwifeinc	−10.1696	5.32494	−1.910	0.0568
	.	.	.	.

Mean dependent var	1302.930	S.D. dependent var	776.2744
Sum squared resid	7.74e+08	S.E. of regression	1354.205
R^2	0.000543	Adjusted R^2	-0.011299
$F(5, 422)$	2.485229	P-value(F)	0.031022

Hausman test –
Null hypothesis: OLS estimates are consistent
Asymptotic test statistic: $\chi^2(1) = 44.3908$
with p-value = 2.68948e-11

Sargan over-identification test –
Null hypothesis: all instruments are valid
Test statistic: LM = 0.862201
with p-value = $P(\chi^2(1) > 0.862201) = 0.353124$

Weak instrument test –
First-stage $F(2, 421) = 6.17263$

2. (7.5%) Explain how you can test that the two instruments are weak (provide the null and alternative hypothesis, the test statistic and the rejection region) (5%). Implement the test (2.5%).

.

The relevance condition requires that at least one instrument has nonzero coefficient in the reduced form equation of the endogenous regressor, in this case that $\pi_1 \neq 0$ and/or $\pi_2 \neq 0$ in

$$lwage = \pi_0 + \pi_1 exper + \pi_2 exper^2 + \pi_3 age + \pi_4 educ + \pi_5 kidslt6 + \pi_6 nwifeinc + v.$$

With the F statistic provided in the Gretl output, $F = 6.17263$, we reject the null of no relevance

$$\begin{aligned} H_0 &: \pi_1 = \pi_2 = 0 \\ H_1 &: \pi_1 \neq 0 \text{ and/or } \pi_2 \neq 0 \end{aligned}$$

using the critical value $\chi^2_{2,0.05}/2 = 3$ from the $\chi^2_2/2$ null distribution, confirming the relevance of the instruments.

However, for checking whether the instruments are weak we have to see whether $F < 10$ (weak) or $F > 10$ (no weak). As $F < 10$, we conclude that the instruments are weak.

.

Test on Model 2:

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

Test statistic: Robust F(2, 421) = 6.17263, p-value 0.00227937

Model 2: OLS, using observations 1–428
Dependent variable: lwage
Heteroskedasticity-robust standard errors, variant HC1

	Coefficient	Std. Error	t-ratio	p-value
const	−0.447161	0.288901	−1.548	0.1224
age	−0.00255613	0.00591494	−0.4321	0.6659
educ	0.101111	0.0141358	7.153	0.0000
exper	0.0418643	0.0151135	2.770	0.0059
expersq	−0.000762482	0.000406466	−1.876	0.0614
kidslt6	−0.0532185	0.104790	−0.5079	0.6118
nwifeinc	0.00555999	0.00274350	2.027	0.0433
	.	.	.	.

Mean dependent var	1.190173	S.D. dependent var	0.723198
Sum squared resid	186.8577	S.E. of regression	0.666215
R^2	0.163302	Adjusted R^2	0.151377
$F(6, 421)$	14.76233	P-value(F)	2.49e–15
Log-likelihood	−429.9477	Akaike criterion	873.8954
Schwarz criterion	902.3093	Hannan–Quinn	885.1173

3. (7.5%) Explain how you can test that the instrumental variables are exogenous (provide the null and the alternative hypothesis, the statistic and the rejection region) (5%). Implement the test (2.5%).

.

We can test for instrument exogeneity as the model is overidentified, $m = 2$ (two instruments, $exper, exper^2$) while $k = 1$ (only $\log(wage)$ is endogenous). For that we need to compute the J test, which is the robust version of Sargan over-identification test. We need first to store the residuals $\hat{u}$ of the 2SLS estimation and regress them on all the exogenous variables.

$$\hat{u} = \delta_0 + \delta_1 exper + \delta_2 exper^2 + \delta_3 age + \delta_4 educ + \delta_5 kidslt6 + \delta_6 nwifeinc + w$$

and computing the robust F statistic for

$$\begin{aligned} H_0 &: \delta_1 = \delta_2 = 0 \\ H_1 &: \delta_1 \neq 0 \text{ and/or } \delta_2 \neq 0 \end{aligned}$$

so that

$$J = mF = 2 \times 0.599011 = 1.198$$

which is not significant compared to the 5% critical value of a $\chi^2_{m-k} = \chi^2_1$ variable, $\chi^2_{1,0.05} = 1.96^2 = 3.84$, $J < \chi^2_{1,0.05}$, confirming instruments exogeneity, as Sargan's test does in the Gretl output.

.

Test on Model 3:

Null hypothesis: the regression parameters are zero for the variables

exper, expersq

Test statistic: Robust F(2, 421) = 0.599011, p-value 0.549822

Model 3: OLS, using observations 1–428
Dependent variable: uhat1
Heteroskedasticity-robust standard errors, variant HC1

	Coefficient	Std. Error	t-ratio	p-value
const	198.514	556.406	0.3568	0.7214
age	−3.71259	11.7979	−0.3147	0.7532
educ	0.216359	27.1010	0.007983	0.9936
kidslt6	−9.42915	211.797	−0.04452	0.9645
nwifeinc	0.947122	5.49637	0.1723	0.8633
exper	−16.8994	31.2501	−0.5408	0.5889
expersq	0.674344	0.821074	0.8213	0.4119
	.	.	.	.

Mean dependent var	4.00e–13	S.D. dependent var	1346.253
Sum squared resid	7.72e+08	S.E. of regression	1354.446
R^2	0.002014	Adjusted R^2	-0.012209
$F(6, 421)$	0.203165	P-value(F)	0.975733
Log-likelihood	−3690.148	Akaike criterion	7394.296
Schwarz criterion	7422.710	Hannan–Quinn	7405.518

4. (7.5%) Consider now that education (*educ*) is also endogenous because the "ability" of the worker in the job market is omitted from the equation. Why mother education (*motheduc*) and father education ($fatheduc$) can be valid instruments for education? (2%) Estimate the new model with all instruments available and explain how the labor supply of women shifts for different levels of education (2%). Explain how you would perform an exogeneity test of all instruments and perform the test (3.5%).

.

We have to argue the relevance and exogeneity of *motheduc* and $fatheduc$. The relevance would imply that, conditional on other exogenous variables, *motheduc* and $fatheduc$ are jointly correlated with *educ*, which seems likely as education level is persistent across generations. To guarantee exogeneity we need to argue that omitted factors in the error term if the labour supply, like ability, are not (partially) correlated to the parents' level of education.

To estimate the model by 2SLS we have to drop *educ* in the instrument list of in Gretl as is no longer assumed exogenous (but keep it as regressor) and add *motheduc* and *fatheduc* as additional instruments (Model 4). To investigate whether *educ* does affect the labour supply we can test for

$$H_0 \; : \; \beta_{educ} = 0$$
$$H_1 \; : \; \beta_{educ} \neq 0$$

which is not rejected by the corresponding robust t test based on the new 2SLS estimates of Model 4, i.e. $\hat{\beta}_{educ} = -84.5645 < 0$ is not significative different from zero at the 5% level, despite this value of $\hat{\beta}_{educ}$ implies that for every (exogenous) additional year of education, the expected number of hours offered decreases by 84.6 hours.

The J test has to regress the new $2SLS$ residuals on the complete list of exogenous variables, which omits now *educ* and include the two new additional exogenous instruments, *motheduc* and $fatheduc$,

$$\hat{u} = \delta_0 + \delta_1 exper + \delta_2 exper^2 + \delta_3 motheduc + \delta_4 fatheduc + \delta_5 age + \delta_6 kidslt6 + \delta_7 nwifeinc + w$$

compute the robust F statistic for the joint null hypothesis of significance of all instruments

$$H_0 : \delta_1 = \delta_2 = \delta_3 = \delta_4 = 0,$$

where the statistic $J = 4F$ has now 2 degrees of freedom, $m-k = 4-2 = 2$. Then the critical value to compare J is $\chi^2_{2,0.05} = 5.99$ while we obtain from Model 5 $J = 4 \times 0.562853 = 2.2514 < 5.99$, which confirms that the instruments are exogenous as we do not reject H_0 [Sargan test under homoskedasticity in the 2SLS output reaches the same conclusion].

.

Model 4: 2SLS, using observations 1–428
Dependent variable: hours
Instrumented: educ lwage
Instruments: const exper age kidslt6 nwifeinc motheduc fatheduc expersq
Heteroskedasticity-robust standard errors, variant HC1

	Coefficient	Std. Error	t-ratio	p-value
const	1197.92	812.930	1.474	0.1413
lwage	1466.81	521.439	2.813	0.0051
age	−6.01102	9.56379	−0.6285	0.5300
educ	**−84.5645**	**75.5292**	**−1.120**	**0.2635**
kidslt6	−270.326	201.747	−1.340	0.1810
nwifeinc	−14.7889	6.35773	−2.326	0.0205
	.	.	.	.

Mean dependent var	1302.930	S.D. dependent var	776.2744
Sum squared resid	6.82e+08	S.E. of regression	1271.309
R^2	0.000510	Adjusted R^2	-0.011332
$F(5, 422)$	2.656403	P-value(F)	0.022248

Hausman test –

Null hypothesis: OLS estimates are consistent

Asymptotic test statistic: $\chi^2(2) = 42.7892$

with p-value = 5.11034e-10

Sargan over-identification test –

Null hypothesis: all instruments are valid

Test statistic: LM = 1.50242

with p-value = $P(\chi^2(2) > 1.50242) = 0.471795$

.

Model 5: OLS, using observations 1–428

Dependent variable: uhat4 [Residuals Model 4]

Heteroskedasticity-robust standard errors, variant HC1

	Coefficient	Std. Error	t-ratio	p-value
const	137.072	484.971	0.2826	0.7776
exper	−15.4764	29.3204	−0.5278	0.5979
expersq	0.641333	0.771516	0.8313	0.4063
motheduc	13.9167	20.1004	0.6924	0.4891
fatheduc	−12.4611	19.6624	−0.6338	0.5266
age	−3.05688	11.2939	−0.2707	0.7868
kidslt6	−3.22131	195.185	−0.01650	0.9868
nwifeinc	1.18756	4.96809	0.2390	0.8112
	.	.	.	.

Mean dependent var	4.56e–12	S.D. dependent var	1263.844
Sum squared resid	6.80e+08	S.E. of regression	1272.094
R^2	0.003510	Adjusted R^2	-0.013098
$F(7, 420)$	0.323816	P-value(F)	0.943130
Log-likelihood	−3662.791	Akaike criterion	7341.583
Schwarz criterion	7374.056	Hannan–Quinn	7354.408

Test on Model 5:

Null hypothesis: the regression parameters are zero for the variables

exper, expersq, motheduc, fatheduc

Test statistic: Robust F(4, 420) = 0.562853, p-value 0.689752

.

SOME CRITICAL VALUES: $Z_{0.10} = 1.282$, $Z_{0.05} = 1.645$, $Z_{0.025} = 1.96$, $\chi^2_{2,0.05} = 5.99$, $\chi^2_{2,0.025} = 7.378$, $\chi^2_{3,0.05} = 7.81$, $\chi^2_{3,0.025} = 9.3484$, $\chi^2_{4,0.05} = 9.49$, $\chi^2_{4,0.025} = 11.1433$, $\chi^2_{4,0.05} = 9.49$, $\chi^2_{4,0.025} = 11.1433$, $\chi^2_{5,0.05} = 11.07$, $\chi^2_{5,0.025} = 12.8325$, where $\mathbb{P}(Z > Z_\alpha) = \alpha$ and $\mathbb{P}\left(\chi^2_m > \chi^2_{m,\alpha}\right) = \alpha$, Z is distributed as a normal with mean 0 and variance 1, and χ^2_m as a chi-squared with m degrees of freedom.