BROWN

Industrial Organization

Assignment 0

Diego Gentile Passaro

August 2019

Question 1

Computations were done in STATA using the code in assignment_0.do. The results are shown in the table below.

Table 1: Table IV

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Years of education	0.0802	0.0769	0.0802	0.131	0.0701	0.0669	0.0701	0.101
	(0.0004)	(0.0150)	(0.0004)	(0.0334)	(0.0004)	(0.0151)	(0.0004)	(0.0334)
Adj. R-squared	0.171	0.171	0.171	0.102	0.230	0.229	0.230	0.206
Observations	247199	247199	247199	247199	247199	247199	247199	247199

Standard errors in parentheses

Comments on code: The processing of the raw data (renaming, variable generation, etc) was streamlined to eliminate unnecessary lines of code, and a .csv file was outputted. Also, the regressions were run through abstracted functions to avoid code repetition. Something noteworthy is that when running the IV regressions of models 2 and 4 (the ones that include age and age squared measured in quarters among other controls) there was collinearity with two of the instruments (interactions of quarter indicators and year of birth indicators), as age and age squared measured in quarters are a linear function of these interactions. Therefore, in order to run the exact same regression in MATLAB and avoid STATA dropping automatically different regressors I omitted QTR120 and QTR121 variables on those models.

Question 2

Computations were done in MATLAB using the code in assignment_0.m. I do not tabulate the results as it can be verified that they are exactly the same as the ones in **Question 1**.

Comments on code: The clean .csv file produced in STATA was inputted into MATLAB, and coefficients and standard errors for both the OLS and the IV regressions were computed using basic matrix algebra and the appropriate degrees of freedom corrections. A noteworthy comments is that for computing the correct standard errors in the IV regressions we have to use the right residuals (the ones using the EDUC variable and ot the predict_EDUC variable). Once again, the regressions were run through abstracted functions. QTR120 and QTR121 were omitted from models 2 and 4.

Question 3

Computations were done in STATA using the code in assignment_0.do. The results are shown in the table below.

Table 2: Logit model

	(1)	(2)
	Logit	Average marginal effect
main		
Born in first quarter=1	-0.0478	-0.0116
	(0.0094)	(0.0023)
Observations	247199	247199

Standard errors in parentheses

Comments on code: The logit model and the average marginal effect (AME) of the indicator of being born in the first quarter are straightforward to compute using the logit and margins commands appropriately. The only comment worth doing is that when using margins is key to use factor notation (i.QTR1) to avoid computing the AME as if it was a continuous variable (which will induce using a "small" change instead of an on and off change on the indicator variable QTR1).

Question 4

Computations were done in MATLAB using the code in assignment_0.m. Again, I do not tabulate the results as it can be verified that they are exactly the same as the ones in **Question 3**.

Comments on code: For computing the logit model by hand in MATLAB (using maximum likelihood estimation) I first derived an appropriate FOC for the logit regression given. That is:

$$\sum_{i=1}^{N}(y_i - \Lambda_i(X_i'\beta))X_i = 0$$

Where we have that:

$$X_i = \begin{pmatrix} 1 \\ AGEQ_i \\ AGESQ_i \\ QTR1_i \end{pmatrix}$$

and that:

$$\Lambda_i(X_i'\beta) = \frac{\exp(X_i'\beta)}{1 + \exp(X_i'\beta)}$$

Then, I put that matrix FOC equation into MATLAB and starting from initial β's equal to zero I used fsolve (which is not a MATLAB statistical function) to get the β's that solve it. I did not compute the standard errors in MATLAB. Doing so would involve computing the square-root of the reciprocal of the Fisher-Information matrix evaluated at the maximum likelihood estimates. The Fisher-Information matrix, is the negative Hessian of the log-likelihood which is just the matrix of second derivatives of the log-Likelihood function with respect to the β's.

In order to get the AME I predicted the probability of having 12 or more years of education using the estimated β's and the observed regressors except for the QTR1 variable, which in one case I forced it to be 0 for all observations and in another one I forced it to be 1 for all observations. Then, for each observation I computed the difference in predicted probabilities of the second case minus the first, that is the marginal effect for each observation. Finally, I took the average across observations to obtain the AME, and I obtained the same as in STATA.