线

序号

姓名：

封

学号：

专业、班级：

密

浙江财经大学 2022～2023 学年第二学期

《计量经济学(双语)》课程期末考试试卷上机部分

考核方式：　上机考试　　　　　　　　考试日期：2023 年 6 月 14 日

适用专业、班级：20 应用统计

题　号	一	二	三	四	五	六	七	总分
得　分								
评卷人								

（共　六　大题）

1: [16 marks] Use the data in **SLEEP75.dta** for this exercise. The equation of interest is

$$sleep = \beta_0 + \beta_1 totwrk + \beta_2 educ + \beta_3 age + \beta_4 age^2 + \beta_5 ynkid + \mu$$

(i) Estimate this equation **separately for men and women** and report the results in the usual form. Are there **notable differences** in the two estimated equations ?

(ii) Compute the **Chow test** for equality of the parameters in the sleep equation for men and women. Use the form of the test that adds ***male*** and the interaction terms ***maletotwrk*, …, *male yngkid*** and uses the full set of observations. What are the relevant ***df*** for the test? Should you reject the null at the 5% level?

(iii) Now allow for a **different intercept** for males and females and determine **whether the interaction terms involving *male* are jointly significant.**

(iv) Given the results from parts (ii) and (iii), what would be **your final model**?

2: [18 marks] Use the data in **JTRAIN98.dta** to answer this question. The variable **unem98** is a binary variable indicating whether a worker was unemployed in 1998. It can be used to measure the effectiveness of the job training program in reducing the probability of being unemployed.

(i) What percentage of workers was unemployed in 1998, after the job training program? How does this compare with the unemployment rate in 1996?

(ii) Run the simple regression **unem98** on **train.** How do you interpret the coefficient on train? Is it statistically significant?

(iii) Add to the regression in part (ii) the explanatory variables **earn96**, **educ**, **age**, and **married**. Now interpret the estimated training effect. Why does it differ so much from that in part (ii)?

(iv) Now perform **full regression adjustment** by running a regression with a full set of interactions, where all variables (except the training indicator) are centered around their sample means:

$$unem98_i \text{ on } train_i,\ earn96_i,\ educ_i,\ age_i,\ married,\ train_i \cdot (earn96_i - \overline{earn96}),$$
$$train_i \cdot (educ_i - \overline{educ}),\ train_i \cdot (age_i - \overline{age}),\ train_i \cdot (married_i - \overline{married}).$$

This regression uses all of the data. What happens to the estimated **average treatment effect** of train compared with part (iii).

(v) Are the **interaction terms** in part (iv) jointly significant?

(vi) Verify that you obtain exactly the same average treatment effect if you run two separate regressions. That is, run two separate regressions for the control and treated groups, obtain the fitted values $unem98_i^0$ and $unem98_i^1$ for everyone in the sample, and then compute $\tau_{ura} = \sum_{n=1}^{\infty} \frac{1}{n}[unem98_i^1 - unem98_i^0]$. Check this with the coefficient on train in part (iv). Which approach is more convenient for obtaining a standard error?

3: [21 marks]Use the data in **LABSUP.dta** to answer the following questions. These are data on almost 32,000 black or Hispanic women. **Every woman in the sample is married**. It is a subset of the data used in Angrist and Evans (1998). Our interest here is in determining how weekly hours worked, ***hours***, changes with number of children (***kids***). **All women in the sample have at least two children**. The two potential **instrumental variables** for ***kids***, which is suspected as being endogenous, work to generate exogenous variation starting with two children.

(i) Estimate the equation

$$hours = \beta_0 + \beta_1 kids + \beta_2 nonmomi + \beta_3 educ + \beta_4 age + \beta_5 age^2 + \beta_6 black + \beta_7 hispan + u$$

by OLS and obtain the **heteroskedasticity-robust standard errors**. Interpret the coefficient on ***kids***. Discuss its statistical significance.

(ii) A variable that Angrist and Evans propose as an instrument is ***samesex***, a binary variable equal to one if the first two children are the same biological sex. What do you think is the argument for why it is a relevant instrument for kids?

(iii) Run the regression $\boldsymbol{kids_i}$ on $\boldsymbol{samesex_i}$, $\boldsymbol{nonmomi_i}$, $\boldsymbol{educ_i}$, $\boldsymbol{age_i}$ ， $\boldsymbol{age^2{}_i}$, $\boldsymbol{black_i}$ $\boldsymbol{hispan_i}$ and see if the story from part (ii) holds up. In particular, interepret the coefficient on ***samesex***. How statistically significant is ***samesex*** ?

(iv) Using ***samesex*** as an **IV** for ***kids***, obtain the IV estimates of the equation in part (i). How does the ***kids*** coefficient compare with the OLS estimate? Is the IV estimate precise?

(v) Now add ***multi2nd*** as an instrument. Obtain the F statistic from the first stage regression and determining whether ***samesex*** and ***multi2nd*** are sufficiently strong.

(vi) Using ***samesex*** and ***multi2nd*** both as instruments for ***kids***, What is the IV estimate coefficient on ***kids*** ?

(vii) In part (vi), how many **overidentification restrictions** are there? Does the **overidentification test** pass?

4: **[18 marks]** Use the data **CPS91.dta** for this exercise. These data are for married women, where we also have information on each husband's income and demographics.

(i) What **fraction** of the women report being **in the labor force**?

(ii) Using only the data for **working women**--you have no choice--estimate the wage equation

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + \beta_4 black + \beta_5 hispanic + u$$

by ordinary least squares（**OLS**）. Do there appear to be significant **wage differences** by **race and ethnicity**?

(iii) Estimate a **probit model** for ***inlf*** that includes the explanatory variables in the wage equation from part (ii) as well as ***nwifeinc*** and ***kidlt6***. Do these last two variables have coefficients of the **expected sign**? **Are they statistically significant**?

(iv) Test hypothesis $\beta_2 = \beta_3 = 0$. Report the value of **likelihood ratio (LR) statistic.**

(v) Compute the average partial effect (**APE**) for ***educ*** from t probit model in part (iii).

(vi) Compute the **inverse Mills ratio** (for each observation) and add it as additional regressor to the wage equation from part (ii). What is its two-sided **p-value**?

5: [11 marks]Use the data in JTRAIN98 to answer the following questions. Here you will use a **Tobit model** because the outcome, ***earn98***, sometimes is zero.

(i) How many observations (men) in the sample have earn98 = 0? Is it a large percentage of the sample?

(ii) Estimate a Tobit model for ***earn98***, using train, earn96, educ, and married as the explanatory variables. Report the β_{train} and its standard error. Is the sign what you expect? How statistically significant is it?

(iii) In part (ii), obtain the average partial effect—which is the average treatment effect—of train, and obtain its standard error.

6: [16 marks] Use the data in **wagepan.dta** for this exercise.

(i) Estimate the model

$$lwage_{it} = \beta_0 + \beta_1 \text{exper}_{it} + \beta_2 \text{expersq}_{it} + \beta_3 educ_{it} + \beta_4 black_{it} + \beta_5 hisp_{it} + v_{it}, v_{it} = a_i + u_{it}$$

by **pooled OLS**, and report the estimates and standard errors in the usual form.

(ii) Estimate the **random effects model** (thinking that $v_{it} = a_i + u_{it}$), and then carry out the **Lagrange multiplier test** of the hypothesis that the classical model without the unobserved effect applies.

(iii) Estimate the **fixed effects model** and then test the hypothesis that the constant term(a_i) is the same for all i.

(iv) Carry out **Hausman's test** for the random versus the fixed effect model. [5marks]