Name: _____
ECON3500 Econometrics and Applications

Exam 3 Practice Exam (CH8–10, 12)

There are 50 total points. This practice exam is longer than the actual exam (which is 75 minutes). Show your work to receive full credit. **Regression output for Question 5 appears on the last page.**

1. *(12 points)* A researcher investigating the effect of healthcare spending on health outcomes in the United States has data for 50 states over 20 years (1995-2014). She estimates the following regression using OLS with fixed effects:

$$ln(health)_{it} = \beta_1 ln(spending)_{it} + \beta_2 unemployment_{it} + \alpha_i + \phi_t + u_{it}$$
$$i = 1, ..., 50; \quad t = 1, ..., 20$$

   where *health* is a health outcome measure (e.g., life expectancy), *spending* is per capita healthcare expenditure, *unemployment* is the state unemployment rate, α_i represents state fixed effects, and ϕ_t represents year fixed effects.

   (a) What is the purpose of including state fixed effects (α_i) and year fixed effects (ϕ_t)? What unobserved factors might each type of fixed effect capture? *[4 points]*

(b) The results below show the within-estimates (with fixed effects). Discuss the meaning of the coefficient on healthcare spending and its statistical significance. What is the effect of a 10% increase in healthcare spending on health outcomes?

```
ln(health) = 0.284 * ln(spending) + 0.042 * unemployment
             (0.067)                (0.018)

R^2 = 0.876
N = 1000 (50 states × 20 years)
```

[5 points]

(c) Suppose the regression is re-estimated **without** state and year fixed effects (simple pooled OLS). The coefficient on $ln(spending)$ becomes 0.512 with SE = 0.043. Why might the coefficient differ so much when fixed effects are removed? What does this suggest about the relationship between healthcare spending and unobserved state characteristics? *[3 points]*

2. *(11 points)* A researcher studies the effect of class size on student achievement using data from schools in developing countries. She hypothesizes that smaller classes improve test scores, but student selection is a concern—schools with better teachers may both have smaller classes and higher achieving students.

$$test\_score_i = \beta_0 + \beta_1 class\_size_i + \beta_2 teacher\_experience_i + u_i$$

She proposes using **random assignment of students to classroom** as an instrumental variable for class size. This ensures class size variation is exogenous.

(a) For this instrument to be valid, what assumptions must hold? For each assumption, indicate whether you can test it with data. If you can, explain how; if you cannot, explain why not. *[4 points]*

(b) OLS results show: coefficient on class size = -0.15 (SE = 0.08). IV results show: coefficient on class size = -0.42 (SE = 0.18). Compare these estimates. Why might IV yield a stronger effect? *[4 points]*

(c) Provide one specific threat to the validity of using random assignment as an instrument for class size. *[3 points]*

3. *(13 points)* A policymaker wants to evaluate the effect of a new job training program on employment rates. Some states adopted the program in 2015, while others did not. The researcher has employment data for 20 states over 2010-2018 (before and after 2015).

   (a) Explain why a simple before-after comparison (comparing 2015-2018 to 2010-2014) would be problematic for estimating the causal effect of the training program. *[3 points]*

   (b) Propose a difference-in-differences (DiD) model to estimate the causal effect. Define any new variables. Identify which coefficient represents the program effect. *[5 points]*

   (c) What key assumption must hold for your DiD model to measure a causal effect? Explain in plain English. *[2 points]*

   (d) What concerns do you have about the external validity of these results? Under what conditions would these results generalize to other states or contexts? *[3 points]*

4. *(5 points)* Consider the following causal diagram for the relationship between education and earnings:

   Ability → Education → Earnings
   Ability → Earnings
   Education → Job Type → Earnings

   In this DAG, Education causes Earnings both directly and through Job Type. Ability is an unobserved variable that causes both Education and Earnings.

   (a) List all paths from Education to Earnings. For each path, state whether it is a **causal (front door)** path or a **backdoor** path. *[2 points]*

   (b) What variable(s) should you control for to identify the causal effect of Education on Earnings? What variable(s) should you **not** control for? Explain why for each. *[2 points]*

   (c) Suppose Ability is unobserved and cannot be measured. Name one identification strategy from this course that could help estimate the causal effect of Education on Earnings despite this unobserved confounder. Briefly explain how it would work. *[1 point]*

5. *(9 points)* An education researcher studies factors affecting standardized test scores using data from 620 elementary school students. She is interested in whether the effect of small class sizes differs for boys and girls. **Regression output appears on the last page of this exam.**

Variable	**Description**
$score$	Standardized test score (points)
ln_income	Natural log of family income (dollars)
$small\_class$	$= 1$ if class size ≤ 18 students
$female$	$= 1$ if female
$small\_female$	$small\_class \times female$ (interaction term)

$$score_i = \beta_0 + \beta_1 \, ln\_income_i + \beta_2 \, small\_class_i + \beta_3 \, female_i + \beta_4 \, small\_female_i + u_i$$

(a) What type of model is this with respect to *score* and *income*? Circle one. Interpret the coefficient on ln_income. If family income increases by 10%, what is the predicted effect on test scores? *[3 points]*

Type (circle one): level–level log–level level–log log–log

(b) Using the regression results, calculate the predicted effect of being in a small class for **male** students and for **female** students. Is the small class effect the same for both groups? *[3 points]*

(c) Is the difference in the small class effect between male and female students statistically significant at the 5% level? State which coefficient you are testing and show your work. What does this tell us about whether small classes benefit boys and girls equally? *[3 points]*

QUESTION 5 — REGRESSION OUTPUT

```
. regress score ln_income small_class female small_female

      Source |       SS           df       MS      Number of obs   =       620
-------------+----------------------------------   F(4, 615)       =    107.84
       Model |  193842.60          4   48460.65   Prob > F        =    0.0000
    Residual |  276498.40        615     449.59   R-squared       =    0.4122
-------------+----------------------------------   Adj R-squared   =    0.4083
       Total |  470341.00        619     759.84   Root MSE        =    21.204

------------------------------------------------------------------------------
       score |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   ln_income |     12.840      2.150     5.97   0.000        8.618      17.062
 small_class |      4.520      1.830     2.47   0.014        0.925       8.115
      female |      1.950      1.460     1.34   0.182       -0.918       4.818
small_female |      3.780      2.040     1.85   0.065       -0.226       7.786
       _cons |    430.200     22.100    19.47   0.000      386.800     473.600
------------------------------------------------------------------------------
```