Enrolment No. _____

AHMEDABAD UNIVERSITY

AMRUT MODY SCHOOL OF MANAGEMENT

UNDERGRADUATE / GRADUATE PROGRAMMES

2024 WINTER SEMESTER

END SEMESTER EXAMINATION

ECO321 EMPIRICAL RESEARCH METHODS IN ECONOMICS

Date: 25-04-2024 Total Marks: 50

Time: 08.00 am – 10.30 am No. of printed pages: 06

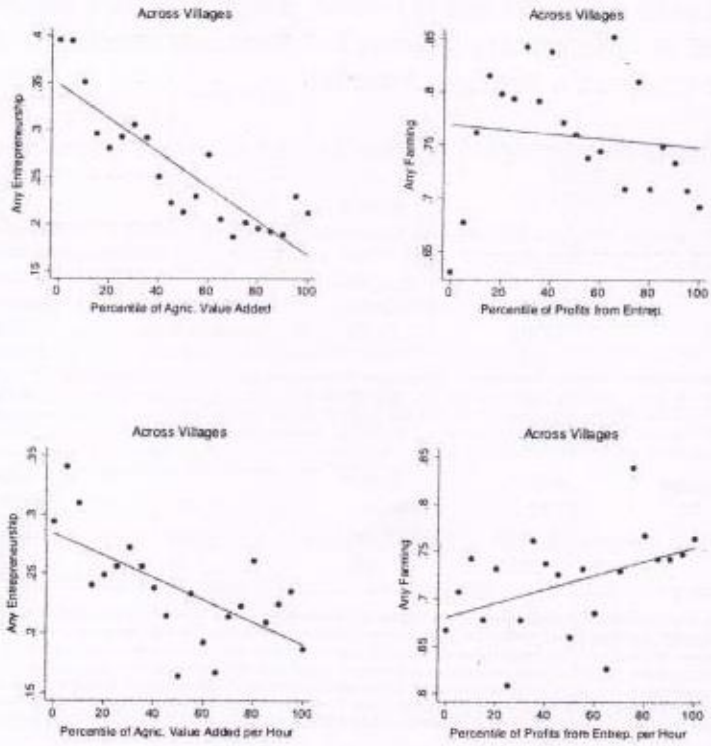
Instruction: The exam has five questions. All questions are compulsory.

Question 1 (10 marks)

1. The following questions are about interpreting scatterplots. For this purpose, I am using the data that belong to the Living Standard Measurement Survey-Integrated Survey on Agriculture (LSMS-ISA) that we have discussed and used in class many times.

We are interested in understanding the relationship between returns from one economic activity and the probability of engaging in another. For instance, we can ask whether farming households that engage in non-farming entrepreneurship (retail, petty trading, charcoal production) have relatively higher or lower agricultural output than those that stick to farming only.

Looking at the graphs below, please answer the following questions



a) Are households that are more productive in agriculture more or less likely to also engage in entrepreneurship? (1 mark)

b) Are households that are more productive in non-farming entrepreneurship more or less likely to also engage in farming? (1 mark)

c) What is the difference between the top and bottom graphs? Does that matter? Why? (2 marks)

2. What is the fundamental assumption of difference-in-differences? Is it testable? Explain why it is needed and how it helps to credibly identify impact. (3 marks)

3. What does LATE mean? Explain what it is in the context of instrumental variable and regression discontinuity design and how it can differ from the ATT. (3 marks)

Question 2 (10 marks)

Answer the following questions based on the materials covered from the paper "Returns to capital in microenterprises: evidence from a field experiment," by De Mel, S., McKenzie, D., & Woodruff, C.

The researchers study the availability and returns from capital among micro-enterprises in Sri Lanka. They estimate the following regression specification

$$Y_{it} = \alpha + \beta K_{it} + \delta_t + \lambda_i + \varepsilon_{it}$$

a) In the paper, the researchers are interested in measuring marginal returns to capital (MRK). Explain what those are in theory. (1 mark)

b) Discuss their identification problem for identifying the coefficient of interest in the above regression. (1 mark)

c) In order to credibly estimate the coefficient of interest, they exploit their experimental design and implement an instrumental variable strategy. Discuss how their instrumental variable strategy satisfies the assumptions for implementing such a strategy. (3 marks)

d) The following Table shows the corresponding results from estimating the above equation.

TABLE IV

INSTRUMENTAL VARIABLE REGRESSIONS MEASURING RETURN TO CAPITAL FROM EXPERIMENT

	Real profits IV-FE (1)	Log real profits IV-FE (2)	Real profits 4 instruments (3)	Real profits adjusted (1) IV-FE (4)	Real profits adjusted (2) IV-FE (5)
Capital stock/log capital stock (excluding land & buildings)	5.85** (2.34)	0.379*** (0.121)	5.16** (2.26)	5.29** (2.28)	4.59** (2.29)
First-stage					
Coefficient on treatment amount	0.91***	0.33***		0.91***	0.91***
F statistic	27.81	49.26	6.79	27.81	27.81
Observations	3,101	3,101	3,101	3,101	3,101
Number of enterprises	384	384	384	384	384

Notes: Data from quarterly surveys conducted by the authors reflecting nine waves of data from March 2005 through March 2007. Capital stock and profits are measured in Sri Lankan rupees, deflated by the Sri Lankan CPI to reflect March 2005 price levels. Profits are measured monthly. The estimated value of the owner's labor is subtracted from profits in columns (4) and (5), as described in the text. In column (4), the owner's time is valued by regression coefficients from a production function using baseline data; in column (5), we use the median hourly earnings in the baseline sample for each of six gender/education groups. A single variable measuring the rupee amount of the treatment is used as the instrument in columns (1) and (2) and (4) and (5). In column (3), we use four separate variables indicating receipt of each treatment type. Except in column (2), the coefficients show the effect of a 100-rupee increase in the capital stock. All regressions include enterprise and period (wave) fixed effects. Standard errors, clustered at the enterprise level, are shown in parentheses. The *F* statistic is the partial *F* statistic in the first-stage regression on the excluded instruments.
***$p < .01$, **$p < .05$, *$p < .1$.

Comment and interpret everything reported in columns (1) and (2). (5 marks)

Question 3 (10 marks)

Answer the following questions based on the materials covered from the paper "Organizational barriers to technology adoption: Evidence from soccer-ball producers in Pakistan," by Atkin, D., Chaudhry, A., Chaudry, S., Khandelwal, A. K., & Verhoogen, E.

a) Explain what is the main hypothesis in their study and the experiment the authors designed to test it. (2 marks)

b) What was the problem they encountered with the first experiment? What seems to be the reason behind the problem? (2 marks)

c) What is the hypothesis they tested using the second experiment? What is the design of the second experiment? (2 marks)

d) The following Table shows the results from the second experiment.

TABLE X

INCENTIVE PAYMENT EXPERIMENT RESULTS (DIE PURCHASE AS OUTCOME)

	Dep. var.: purchased offset die after Sept. 1, 2013			
	First stage (1)	OLS (2)	Reduced form (ITT) (3)	IV (TOT) (4)
Panel A: Short run (as of round 6)				
Received treatment		0.42**		0.40**
		(0.15)		(0.16)
Assigned to Group A	0.68***		0.27**	
	(0.12)		(0.12)	
Stratum dummies	Y	Y	Y	Y
Mean of Group B (control group)		0.00	0.00	0.00
R-squared	0.57	0.40	0.24	0.40
N	31	31	31	31
Panel B: Medium run (as of round 7)				
Received treatment		0.41**		0.38**
		(0.15)		(0.16)
Assigned to Group A	0.72***		0.28**	
	(0.12)		(0.12)	
Stratum dummies	Y	Y	Y	Y
Mean of Group B (control group)		0.00	0.00	0.00
R-squared	0.60	0.40	0.24	0.40
N	29	29	29	29

Notes. Table reports results of incentive payment experiment on adoption rates using additional die purchases (beyond the trade-in offer) after September 1, 2013, as the measure of adoption. Panel A reports short-run results as of round 6 (January–March 2014). Panel B reports medium-run results as of round 7 (October–December 2014). The dependent variable in column 1 is an indicator variable for whether the firm received treatment. Two firms exited between rounds 6 and 7. All regressions include stratum dummies and report robust standard errors. Significance: *.10; **.05; ***.01.

Comment and interpret the results in columns 3 and 4. What is the difference between these two columns? (4 marks)

Question 4 (10 marks)

Answer the following questions based on the materials covered from the paper "Empowerment and Efficiency: Tenancy Reform in West Bengal," by Banerjee, A. V., Gertler, P. J., & Ghatak, M.

a) What is the main research question of the paper? What is the identification problem? (2 marks)

b) What is their identification strategy? What is the associated assumption? (1 mark)

c) Looking at the figure below, please answer the following questions

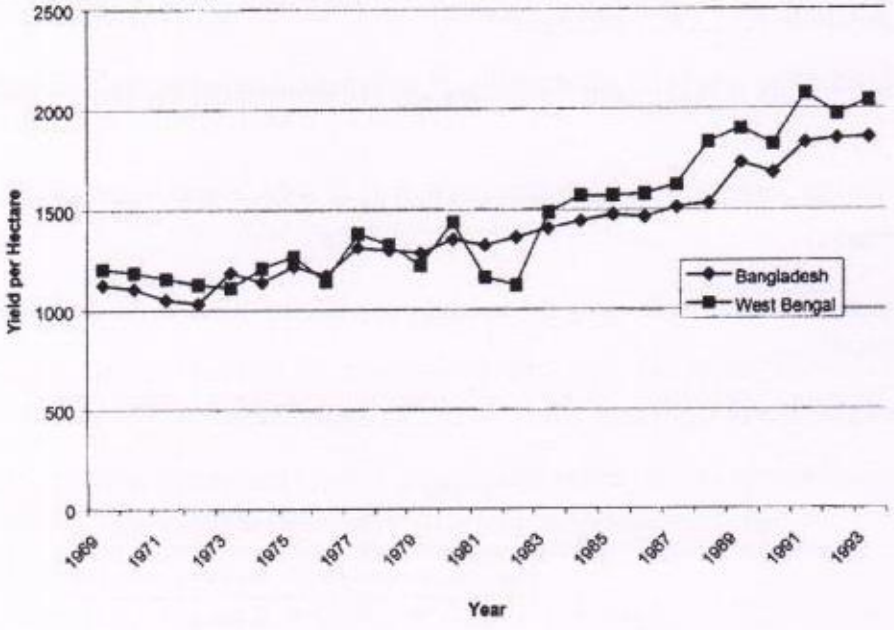


FIG. 4.—Rice yield in West Bengal and Bangladesh, 1969–93

Why is this figure important for their analysis? What important observations can you make from this figure that are relevant to their analysis? (2 marks)

The following table shows some results from their analysis

TABLE 2

DIFFERENCE-IN-DIFFERENCE MODELS OF LOG OF RICE YIELD PER HECTARE (1969–93)

	DIFFERENCE (1969–78) (1)	LEVEL 1969–93 (2)	LEVEL Excluding 1981–82 (3)
West Bengal (=1)	.004 (.17)	...	...
West Bengal × (1979–83)*	...	-.09*** (3.75)	−.01 (.38)
West Bengal × (1984–88)	...	.05** (1.99)	.05** (2.00)
West Bengal × (1988–93)	...	.05* (1.77)	.05* (1.78)
District fixed effects *F*-statistic	...	44.55	42.61
Year fixed effects *F*-statistic	4.26***	29.75***	31.81***
R^2	.12	.80	.81
Sample size	256	717	659

NOTE.—*t*-statistics are in parentheses.

* These variables are obtained by interacting a dummy variable that takes the value one if a district is in West Bengal and zero if it is in Bangladesh with another dummy variable that takes the value one if the observation is in the indicated time period (1979–83 in this case) and zero otherwise.

* Significant at the 10 percent level.

** Significant at the 5 percent level.

*** Significant at the 1 percent level.

d) Interpret the results in columns (2) and (3). What is the difference between these two columns, and why is that important? (3 marks)

e) What is shown in column (1)? Why is that important? (2 marks)

Question 5 (10 marks)

This question is about interpreting STATA output. In what follows, I am using the data that belong to the study "Security, Trade, and Political Violence" by Amodio, Baccini, and Di Maio, published in the Journal of the European Economic Association. We have **NOT** discussed this study in class. However, the question contains all the information necessary to answer the questions below. Please read the descriptions carefully before you start answering the questions.

For each manufacturing sector in the West Bank, the data provide information on inputs used and the value of output for each year in the period from 1999 to 2012. Variables are as follows.

isic4	unique identifier of the manufacturing sector
year	year of observation
lny	log of value of output
lnk	log of value of capital stock
lnl	log of the number of workers

We are interested in identifying returns to capital, defined as the increase in the value of output that follows an increase in capital stock (machinery, land, etc.). In other words, we are interested in the causal effect of increasing capital on output. To do so, we regress the log of value of output *lny* over the log of capital stock *lnk*.

Looking at the output below, please answer the following questions

```
. xtset isic4 year

Panel variable: isic4 (unbalanced)
 Time variable: year, 1999 to 2012, but with gaps
         Delta: 1 unit

. reg lny lnk, cluster(isic4)

Linear regression                               Number of obs     =      1,038
                                                F(1, 107)         =     426.95
                                                Prob > F          =     0.0000
                                                R-squared         =     0.7467
                                                Root MSE          =     1.2411

                                   (Std. err. adjusted for 108 clusters in isic4)
------------------------------------------------------------------------------
             |               Robust
         lny |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnk |   .8045652   .0389379    20.66   0.000     .7273753    .8817551
       _cons |   4.068798   .6058127     6.72   0.000     2.867845    5.269751
------------------------------------------------------------------------------
```

a) What is the first line of command (*xtset* ...), and why is it needed? (1 mark)

b) What does the *cluster* option in the regression command do? Why is it needed? (1 mark)

c) What is the coefficient we are interested in? Is it statistically significant? Can we interpret it as returns to capital? (2 marks)

Consider now the following output, and answer the following questions.

```
. reg lny lnk lnl, cluster(isic4)

Linear regression                               Number of obs     =      1,038
                                                F(2, 107)         =     442.53
                                                Prob > F          =     0.0000
                                                R-squared         =     0.8111
                                                Root MSE          =     1.0723

                                   (Std. err. adjusted for 108 clusters in isic4)
------------------------------------------------------------------------------
             |               Robust
         lny |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnk |   .3715755   .0828134     4.49   0.000     .2074076    .5357434
         lnl |   .6814594   .1053078     6.47   0.000      .472699    .8902199
       _cons |   7.352655   .8060505     9.12   0.000     5.754754    8.950557
------------------------------------------------------------------------------
```

d) What is the difference between the regression equation implemented here and the one before? What would be the reason for this modification? (2 marks)

e) How does the coefficient of interest change? Is this in line with your expectations? Explain. (1 mark)

Consider now the following output and answer the following questions.

```
. xtreg lny lnk lnl, fe cluster(isic4)

Fixed-effects (within) regression               Number of obs     =      1,038
Group variable: isic4                           Number of groups  =        108

R-squared:                                      Obs per group:
     Within  = 0.5757                                         min =          1
     Between = 0.8715                                         avg =        9.6
     Overall = 0.7978                                         max =         13

                                                F(2, 107)         =     157.87
corr(u_i, Xb) = -0.0739                         Prob > F          =     0.0000

                                   (Std. err. adjusted for 108 clusters in isic4)
------------------------------------------------------------------------------
             |               Robust
         lny |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnk |   .1526686   .0537904     2.84   0.005     .0460354    .2593017
         lnl |   .9910766   .0901709    10.99   0.000     .8123234    1.16983
       _cons |   9.172753   .5990215    15.31   0.000     7.985262    10.36024
-------------+----------------------------------------------------------------
     sigma_u |  .82853602
     sigma_e |  .79187591
         rho |  .52261238   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

f) What is the difference between the regression equation implemented here and the one before? What would be the reason for this modification? (2 marks)

g) How does the coefficient of interest change? Is this in line with your expectations? Explain. (1 mark)