(i) अधिकतम संभावना अनुमान (maximum likelihood estimation) क्या है? रैखिक संभाव्यता मॉडल (linear probability model) के बजाय लॉगिट (logit) और प्रोबिट (probit) जैसे अधिकतम संभावना अनुमानकों (maximum likelihood estimators) का उपयोग करने के क्या लाभ हैं? आप प्रोबिट और लॉगिट के बीच किस प्रकार चयन करेंगे?

(ii) प्राकृतिक प्रयोग (natural experiments) क्या हैं? आप नीतिगत बदलाव के मामले में औसत व्यवहार प्रभाव (average treatment effect) किस प्रकार प्राप्त करेंगे, चर्चा कीजिए?

[This question paper contains 16 printed pages.]

Your Roll No.....

Sr. No. of Question Paper : 4769 **I**

Unique Paper Code : 2273100008

Name of the Paper : Advanced Econometrics (NEP)

Name of the Course : **Pool of Discipline Specific Elective (DSE)**

Semester : V

Duration : 3 Hours Maximum Marks : 90

Instructions for Candidates

1. Write your Roll No. on the top immediately on receipt of this question paper.
2. The question paper consists of **six** questions.
3. **Q1** is compulsory. It carries **10** marks.
4. Do any **four** out of **five** questions from **Q2** to **Q6**. Each question carries **20** marks.
5. Use of simple non programmable calculator is allowed.
6. Statistical tables are attached for your reference.

7. Answers may be written either in English or Hindi; but the same medium should be used throughout the paper.

<u>छात्रों के लिए निर्देश</u>

1. इस प्रश्न-पत्र के मिलते ही ऊपर दिए गए निर्धारित स्थान पर अपना अनुक्रमांक लिखिए ।

2. प्रश्न पत्र में छह प्रश्न हैं ।

3. प्रश्न 1 अनिवार्य है । इसके 10 अंक हैं ।

4. प्रश्न 2 से प्रश्न 6 तक **पाँच** में से कोई **चार** प्रश्न हल करें । प्रत्येक प्रश्न के 20 अंक हैं ।

5. सामान्य गैर-प्रोग्रामीय कैलकुलेटर का उपयोग करने की अनुमति है ।

6. आपके संदर्भ के लिए सांख्यिकीय तालिकाएँ संलग्न हैं ।

7. इस प्रश्न-पत्र का उत्तर अंग्रेजी या हिंदी किसी एक भाषा में दीजिए, लेकिन सभी उत्तरों का माध्यम एक ही होना चाहिए ।

1. The following equation describes the median housing price in a community in terms of amount of pollution (nox for nitrous oxide) and the average number of rooms in houses in the community (rooms) :

(ii) निश्चित प्रभाव परिवर्तन (fixed effects transformation) किस प्रकार फर्म-विशिष्ट अप्रत्याशित विषमता (unobserved heterogeneity) को समाप्त करता है?

(iii) इस मॉडल में, आप गुणांकों (coefficients) की व्याख्या किस प्रकार करेंगे?

(iv) निश्चित प्रभाव अनुमान के अपने ज्ञान के आधार पर, यदि विशिष्ट त्रुटियों (idiosyncratic errors) u_{it} में क्रमिक सहसंबंध हो, तो क्या संभावित समस्याएँ उत्पन्न हो सकती हैं?

(v) आप इसका परीक्षण किस प्रकार करेंगे? निश्चित प्रभाव अनुमानक (fixed effects estimator) के लिए स्वतंत्रता की डिग्रियां (degrees of freedom) क्या होगी?

6. (i) What is maximum likelihood estimation? What are the advantages of using maximum likelihood estimators such as logit and probit, instead of linear probability model? How would you choose between the probit and the logit?

(ii) What are natural experiments? Discuss how you would obtain the average treatment effect in case of a policy change? (10×2=20)

एक पैनल डेटा सेट पर विचार कीजिए जहाँ आश्रित चर Y_{it} वर्ष t में फर्म i के लिए फर्म उत्पादकता के लॉग का प्रतिनिधित्व करता है। यह डेटा 54 फर्मों के संबंध में, जिन्होंने प्रत्येक वर्ष स्क्रैप दरों की सूचना दी, तीन वर्षों, 1987, 1988 और 1989 के लिए हैं। आपकी रुचि फर्म उत्पादकता के संबंध में नौकरी प्रशिक्षण कार्यक्रम (बाइनरी चर $grant_{it}$ द्वारा इंगित है) के प्रभाव का आकलन करने में है। मॉडल में समय-नियत प्रभावों को नियंत्रित करने के लिए डमी वर्ष (d88, d89) शामिल हैं। अप्रत्याशित फर्म-विशिष्ट प्रभाव (unobserved firm-specific effects) (a_i) व्याख्यात्मक चर (explanatory variables) के साथ सहसंबंधित हैं।

मॉडल निम्नानुसार विनिर्दिष्ट है :

$$\log(scrap_{it}) = \beta_0 + \beta_1\, grant_{it} + \beta_2\, grant_{it-} + \beta_3\, \mathrm{d88} + \beta_4\, \mathrm{d89} + a_i + u_{it}$$

(i) इस संदर्भ में निश्चित प्रभाव अनुमानक (fixed effects estimator) क्यों उपयुक्त है, स्पष्ट कीजिए।

$$\log(\text{price}) = \beta_0 + \beta_1 \log(\text{nox}) + \beta_2\, \text{rooms} + u$$

(i) What are the probable signs of β_1 and β_2? What is the interpretation of β_1? Explain.

(ii) Why might log(nox) and rooms be negatively correlated? If this is the case, does the simple regression of log(price) on log (nox) produce an upward or a downward biased estimator of β_1?

(iii) Using the data, the following equations were estimated :

$$\widehat{\log(price)} = 11.71 - 1.043 \log(nox), \quad N = 506, \quad R^2 = .264$$

$$\widehat{\log(price)} = 9.23 - .718 \log(nox) + .306\, rooms, \quad N = 506, \quad R^2 = .514$$

Is the relationship between the simple and multiple regression estimates of the elasticity of price with respect to nox what you have predicted given your answer in part (ii)? Does this mean that -.718 is definitely close to the true elasticity than -1.043?

(iv) State whether the following statement is true or false. Also provide justification for your answer: 'Multicollinearity and micro-numerosity are identical concepts.' (2.5×4=10)

प्रदूषण की मात्रा (नाइट्रस ऑक्साइड के लिए nox) और समुदाय में घरों में कमरों की औसत संख्या (rooms) के संदर्भ में उस समुदाय में औसत आवास मूल्य का वर्णन निम्नलिखित समीकरण में किया गया है :

$$\log(\text{price}) = \beta_0 + \beta_1 \log(\text{nox}) + \beta_2 \text{ rooms} + u$$

(i) β_1 और β_2 के संभावित संकेत (probable signs) क्या हैं? β_1 का क्या अर्थ (interpretation) है? स्पष्ट कीजिए।

(ii) क्यों log(nox) और rooms नकारात्मक रूप से सहसंबंधित (negatively correlated) हो सकते हैं? यदि नकारात्मक रूप से सहसंबंधित हैं, तो क्या log(price) on log (nox) का सरल प्रतिगमन (simple regression) β_1 का उर्ध्वगामी (upward) या अधोगामी (downward) पक्षपातपूर्ण अनुमानक (biased estimator) उत्पन्न करता है?

(iii) डेटा का उपयोग करते हुए निम्नलिखित समीकरणों का अनुमान लगाया गया :

$$\widehat{\log(price)} = 11.71 - 1.043 \log(nox), \quad N = 506, \quad R^2 = .264$$

$$\widehat{\log(price)} = 9.23 - .718 \log(nox) + .306\, rooms, \quad N = 506, \quad R^2 = .514$$

nox के संबंध में चतपबम की लोच (elasticity) के सरल और बहु प्रतिगमन अनुमानों (simple and multiple regression

The model is specified as :

$$\log(scrap_{it}) = \beta_0 + \beta_1\, grant_{it} + \beta_2\, grant_{it-} + \beta_3\, \text{d88} + \beta_4\, \text{d89} + a_i + u_{it}$$

(i) Explain why the fixed effects estimator is appropriate in this context.

(ii) How does the fixed effects transformation eliminate the firm-specific unobserved heterogeneity?

(iii) In this model, how would you interpret the coefficients?

(iv) Based on your knowledge of fixed effects estimation, what potential problems could arise if there is serial correlation in the idiosyncratic errors u_{it}?

(v) How would you test for this? What will be the degrees of freedom for the fixed effects estimator? (4×5=20)

$$\log(earnings) = \beta_0 + \beta_1\, alcohol + \beta_2\, educ + u_1$$

$$alcohol = \gamma_0 + \gamma_1 \log(earnings) + \gamma_2 educ + \gamma_3 \log(\text{price}) + u_2$$

जहाँ price शराब का स्थानीय मूल्य सूचकांक है, जिसमें राज्य का और स्थानीय कर शामिल हैं। मान लीजिए कि educ और price बहिर्जात (exogeneous) हैं।

(i) यदि β_1, β_2, γ_1, γ_2 और γ_3 ये सभी शून्य से भिन्न हों, तो कौन सा समीकरण चिह्नित (identified) किया जाता है?

(ii) आप उस समीकरण का अनुमान किस प्रकार लगाएंगे?

5. Consider a panel data set where the dependent variable Y_{it} represents the log of firm productivity for firm i in year t. The data is for three years; 1987, 1988 and 1989, on the 54 firms that reported scrap rates in each year. You are interested in estimating the impact of a job training program (indicated by a binary variable $grant_{it}$) on firm productivity. The model includes year dummies (d88, d89) to control for time-fixed effects. The unobserved firm-specific effects (a_i) are correlated with the explanatory variables.

estimates) के बीच संबंध वही है जो आपने भाग (ii) में अपने उत्तर में पूर्वानुमान लगाया है? क्या इसका तात्पर्य यह है कि −.718 निश्चित रूप से −1.043 की तुलना में वास्तविक लोच (true elasticity) के करीब है?

(iv) बताइए कि निम्नलिखित कथन सत्य है अथवा असत्य। अपने उत्तर का औचित्य भी बताइए: 'बहुसंरेखता और सूक्ष्मसंख्या (Multicollinearity and micro-numerosity) समान अवधारणाएँ हैं।'

2. (a) Four hundred driver's license applicants were randomly selected and asked whether they passed their driving test ($Pass_i = 1$) or failed their test ($Pass_i = 0$); data were also collected on their gender ($Male_i = 1$ if male and = 0 if female). The following table summarizes the results from the estimated models :

Dependent Variable: *Pass*			
Explanatory variable	(1) Probit	(2) Logit	(3) LPM
Male	-0.333 (0.161)	0.622 (0.303)	-0.071 (0.034)
Constant	1.282 (0.124)	2.197 (0.242)	0.900 (0.022)

Note: The figures in column (1) indicate estimated coefficients and the figures below in parenthesis indicate standard errors.

Using results from Column (1) to (3),

(i) Compute the estimated probability of passing the test for men and for women for each of three models.

(ii) Are the models in (1) – (3) different? Why or why not? Explain your answer.

(b) What measures of fit are typically used to assess binary dependent variable regression models? Discuss. (12+4+4)

(क) ड्राइविंग लाइसेंस के चार सौ आवेदकों को यादृच्छिक रूप से चुना गया और उनसे पूछा गया कि वे अपने ड्राइविंग टेस्ट में उत्तीर्ण हुए हैं ($Pass_i = 1$) या अनुत्तीर्ण ($Pass_i = 0$); हुए हैं। उनके लैंगिकता के संबंध में, यदि पुरुष हों तो ($Male_i = 1$) और यदि महिला हों तो ($Male_i = 0$), भी डेटा संगृहीत किया गया। निम्नलिखित तालिका में अनुमानित मॉडलों से परिणामों का सार दिया गया है :

(i) If β_1, β_2, γ_1, γ_2 and γ_3 are all different from zero, which equation is identified?

(ii) How would you estimate that equation? (4×5=20)

(क) बाईं ओर दिखाई दे रहे समान चर y_1 (typically "quantity") के साथ "आपूर्ति और मांग के रूप" में द्वि-समीकरण प्रणाली पर विचार कीजिए :

$$y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1$$

$$y_1 = \alpha_2 y_2 + \beta_2 z_2 + u_2$$

(i) यदि $\alpha_1 = 0$ या $\alpha_1 = 0$ है, तो समझाइए कि y_1 के लिए कम किया गया रूप (reduced form) क्यों विद्यमान है।

(ii) यदि $\alpha_1 \neq 0$ और $\alpha_2 = 0$ है, तो y_2 के लिए कम किया गया रूप ज्ञात कीजिए।

(iii) यदि $\alpha_1 \neq 0$, $\alpha_2 \neq 0$ और $\alpha_1 \neq \alpha_2$ है, तो y_1 और y_2 के लिए कम किया गया रूप (reduced form) ज्ञात कीजिए।

(ख) मान लीजिए कि वार्षिक आय और शराब की खपत को निम्नलिखित युगपत समीकरण मॉडल (simultaneous equation model) द्वारा निर्धारित किया जाता है :

4. (a) Consider a two-equation system in "supply and demand form" with the same variable y_1 (typically "quantity") appearing on the left-hand side :

$$y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1$$

$$y_1 = \alpha_2 y_2 + \beta_2 z_2 + u_2$$

(i) If $\alpha_1 = 0$ or $\alpha_2 = 0$, explain why a reduced form exists for y_1.

(ii) If $\alpha_1 \neq 0$ and $\alpha_2 = 0$, find the reduced form for y_2.

(iii) If $\alpha_1 \neq 0$, $\alpha_2 \neq 0$ and $\alpha_1 \neq \alpha_2$, find the reduced form for y_1 and y_2.

(b) Suppose that annual earnings and alcohol consumption are determined by the simultaneous equation model:

$$\log(earnings) = \beta_0 + \beta_1\, alcohol + \beta_2\, educ + u_1$$

$$alcohol = \gamma_0 + \gamma_1 \log(earnings) + \gamma_2 educ + \gamma_3 \log(\text{price}) + u_2$$

where price is a local price index for alcohol, which includes state and local taxes. Assume that educ and price are exogeneous.

आश्रित चर: *पास*			
व्याख्यात्मक चर	(1) प्रोबिट	(2) लॉगिट	(3) एलपीएम
पुरुष	-0.333 (0.161)	0.622 (0.303)	-0.071 (0.034)
स्थिर	1.282 (0.124)	2.197 (0.242)	0.900 (0.022)

नोट: कॉलम (1) में आंकड़े अनुमानित गुणांक दर्शाते हैं और नीचे कोष्ठक में दिये गये आंकड़े मानक त्रुटि को दर्शाते हैं।

कॉलम (1) से (3) में दिये गये परिणामों का उपयोग करते हुए निम्नलिखित ज्ञात कीजिए,

(i) तीनों मॉडलों में से प्रत्येक के लिए पुरुषों और महिलाओं के लिए परीक्षण में उत्तीर्ण होने की अनुमानित संभाव्यता (estimated probability) की गणना कीजिए।

(ii) क्या (1) - (3) में मॉडल अलग-अलग हैं? यदि हैं तो क्यों अथवा नहिं हैं तो क्यों नहीं है? अपना उत्तर स्पष्ट कीजिए।

(ख) बाइनरी आश्रित चर प्रतिगमन मॉडल (binary dependent variable regression models) का आकलन करने के लिए आमतौर पर फिट के कौन से पैमाने (measures) का उपयोग किया जाता है? चर्चा कीजिए।

3. Consider the following model of earnings :

$\ln(\text{earnings}_i) = \beta_0 + \beta_1 \text{ age}_i + \beta_2 \text{ education}_i + u_i$

where earnings is hourly earnings, age is in years, education is highest qualification obtained, u is the unobservable component of the model, i indexes the observation and the β's are the unknown parameters.

(i) What could be the problem in using OLS to estimate the above model?

(ii) Suppose you have data on two additional variables, distance to nearest school (dschool) and distance to nearest college (dcollege) at 16 years of age. Discuss whether these are likely to be good instruments.

(iii) Suppose you use dschool and dcollege as instruments in the estimation of the above model. Write down the reduced form for education.

(iv) State the conditions under which the parameters of the model in part (iii) above are identified.

(5×4=20)

आय के निम्नलिखित मॉडल पर विचार कीजिए :

$\ln(\text{earnings}_i) = \beta_0 + \beta_1 \text{ age}_i + \beta_2 \text{ education}_i + u_i$

जहाँ earnings प्रति घंटे आय है, age वर्षों में है, education प्राप्त उच्चतम योग्यता है, u मॉडल का अप्राप्य घटक (unobservable component) है, i अवलोकन (observation) को अनुक्रमित (indexes) करता है और β's अज्ञात पैरामीटर हैं।

(i) उपरोक्त मॉडल का आकलन (estimate) के लिए OLS का उपयोग करने में क्या समस्या हो सकती है?

(ii) मान लीजिए कि आपके पास दो अतिरिक्त चरों, 16 वर्ष की आयु में निकटतम स्कूल की दूरी (dschool) और निकटतम कॉलेज की दूरी (dcollege) से संबंधित डेटा है। चर्चा कीजिए कि क्या ये अच्छे साधन (instruments) हो सकते हैं।

(iii) मान लीजिए कि आप उपर्युक्त मॉडल के आकलन (instruments) में dschool और dcollege का साधन के रूप में उपयोग करते हैं। शिक्षा के लिए छोटा किया गया रूप (reduced form) लिखिए।

(iv) वे स्थितियां बताइए जिनके तहत उपर्युक्त भाग (iii) में मॉडल के मापदंडों की पहचान की जाती है।