

Econometrics, Spring 2022

Final Exam

The time limit is 90 minutes and the exam is worth a total of 30 points. You are NOT allowed to use any books, lecture notes or electronic devices except electronic calculators. You cannot cooperate with your classmates. In case of any such cooperation being detected, both parties will receive zero points in the final. Any violation of academic honesty will be punished to the fullest extent possible. The necessary formulas and statistical tables are included below this setup.

Multiple choice (max 5)	True/False (max 5)	Problem 1 (max 6)	Problem 2 (max 14)	Total (max 30)

Multiple choice questions (5 points max)

1. The general approach to obtaining fully robust standard errors and test statistics in the context of panel data is known as _____.

- a. confounding
- b. differencing
- c. clustering**
- d. attenuating

2. If a process is said to be integrated of order one, or $I(1)$, _____.

- a. the first difference of the process is weakly dependent**
- b. it is stationary at level
- c. averages of such processes already satisfy the standard limit theorems
- d. it does not have a unit root

3. Which of the following is used to test whether a time series follows a unit root process?

- a. Wald test
- b. White test
- c. Johansen test
- d. Augmented Dickey-Fuller test**

4) Consider the following regression model: $\log(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_3 + u$. This model will suffer from functional form misspecification if _____.

- a. β_0 is omitted from the model
- b. u is heteroskedastic
- c. x_1^2 is omitted from the model**
- d. x_3 is a binary variable

5. The significance level of a test is:

- a. the probability of rejecting the null hypothesis when it is false.
- b. the probability of rejecting the null hypothesis when it is true.**
- c. one minus the probability of rejecting the null hypothesis when it is false.
- d. one minus the probability of rejecting the null hypothesis when it is true.

True/False questions (5 points max)

1. The two stage least squares estimator is less efficient than the ordinary least squares estimator when the explanatory variables are exogenous. **(True)**

2. One of the disadvantages of the logistic regression function is that predicted probabilities can be outside of (0,1) interval. **(False)**

3. A data set is called an unbalanced panel if it has missing years for at least some cross-sectional units in the sample. **(True)**
4. Weakly dependent processes are said to be integrated of order zero. **(True)**
5. The multiple linear regression model with a binary dependent variable is called the linear probability model. **(True)**

Problem 1 (6 points max)

- a) What approach would you use to estimate the causal effect if you have dataset with two group of observational units and two-period panel data?

The correct answer: difference-in-difference approach is used for estimation of causal effect in case of two-period panel data and two observation groups (treated and control). However, I also gave full points for fixed effects or first-difference model.

- b) What are the conditions that instrumental variable should satisfy to be valid?

Instrumental variable (or instrument) should be a variable z such that

1. z is uncorrelated with the error term: $Cov(z, \epsilon) = 0$
2. z is correlated with the explanatory variable x : $Cov(x, z) \neq 0$

Problem 2 (14 points max)

Suppose you want to assess the impact of a race of an individual on the likelihood of approving a mortgage loan. In the example below the key explanatory variable is white, a dummy variable equal to one if the applicant was white. The other applicants in the data set are black and Hispanic. To test for the discrimination in the mortgage loan market, a linear probability model (LPM) can be used:

$$approve = \alpha_0 + \alpha_1 white + u$$

- a) **(2pt)** Suppose you obtain the following output from the regression above. Interpret the coefficient on *white*.

VARIABLES	(1) approve
white	0.201*** (0.0198)
Constant	0.708*** (0.0182)
Observations	1,989
R-squared	0.049
Standard errors in parentheses	
*** p<0.01, ** p<0.05, * p<0.1	

White people are 20% more likely to obtain approved mortgage than blacks and Hispanics.

- b) (3pt) Name at least one pro and one con of using an LPM. **It is easy to interpret, predicted values goes out of [0,1] interval.**
- c) (3pt) Suppose now that you run probit and logit models as well, interpret the coefficients on *white* for probit and logit models and compare them with the LPM model.

VARIABLES	(LPM) approve	(Probit) approve	(Logit) approve
white	0.201*** (0.0198)	0.784*** (0.0867)	1.409*** (0.151)
Constant	0.708*** (0.0182)	0.547*** (0.0754)	0.885*** (0.125)
Observations	1,989	1,989	1,989
R-squared	0.049		

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In probit and logit models you cannot directly interpret the coefficients, what matters is only signs – if a person is white, he is more likely to get mortgage approved.

- d) (3pt) By how much is it more likely for white people to obtain mortgage loan in comparison to minorities according to probit model? How different is this result from LPM result?

$$P(Y = 1|X) = \Phi(0.547 + 0.784 * 1) = \Phi(1.331) = 0.908$$

$$P(Y = 0 | X) = \Phi(0.547) = 0.705$$

Comparison: 0.908-0.705=0.203 -> by 20% more likely just like in the LPM.

- e) (3pt) By how much is it more likely for white people to obtain mortgage loan in comparison to minorities according to logit model? Note that the functional form of the logit model is $\Lambda(\cdot) = \frac{\exp(\cdot)}{1+\exp(\cdot)}$.

$$\Lambda(2.294) = \frac{\exp(2.294)}{1+\exp(2.294)} = 0.908.$$

$$\Lambda(0.885) = \frac{\exp(0.885)}{1+\exp(0.885)} = 0.708.$$

Difference is about 20% meaning that white people are 20% more likely to get mortgage approved.