Course name:	**Econometrics 3a: Methods for Analyzing Micro Data**
Course code:	**EC7412**
Examiner:	**Mårten Palme**
Number of credits:	**7,5 credits**
Date of exam:	**May 31, 2017**
Examination time:	**3 hours**

Write your identification number on each paper and cover sheet (the number stated in the upper right hand corner on your exam cover).

Do not write answers to more than one question in the same cover sheet. Explain notions/concepts and symbols. If you think that a question is vaguely formulated, specify the conditions used for solving it. Only legible exams will be marked. No aids are allowed.

The exam consists of 3 questions. Questions 1 and 3 are worth 30 points and question 2 is worth 40 points, 100 points in total. For the grade E 45 points are required, for D 50 points, C 60 points, B 75 points and A 90 points.

Results will be posted on mitt.su.se three weeks after the exam, at the latest

----

Good luck!

1.

Suppose you have data survival from the Titanic ship catastrophe on survival and categorical variables on age (adult or child), gender (male or female) and economic status (first, second, third class passenger or crew). You use logit to estimate the following model:

$$P(Y_i = 1) = \beta_0 + \beta_1 Child_i + \beta_2 Female_i + \beta_3 First_i + \beta_4 Second_i + \beta_5 Third_i,$$

where *Child* is an indicator for being a child, *Female* is an indicator for being female; *First*, *Second* and *Third* are indicators for the different classes.

    a. Describe how you can test the hypothesis $\beta_3 = 0$ using a likelihood ratio and a Wald test, respectively.

    b. Describe three different ways to measure goodness-of-fit in this model.

    c. Explain why a Linear probability model (LPM) may give inconsistent estimates and explain why this can be avoided by changing the specification of the model while maintaining the LPM framework.

2.

a. Suppose you want to study the average wage differentials between two different demographic groups. One of the groups has a much higher rate of labor force participation than the other. Explain why the underlying distribution of wage offers may be the same in the two groups, although you observe different average wages for those who work. Explain also under what assumptions the observed means consistently estimates the underlying means in the wage offer distribution.

b. By using the method proposed by Blundell et al. (2007), show how you could estimate "worse case bounds" for the median of the underlying distribution of wage offers. What is required for these bounds to be identified?

c. Describe how you could estimate the mean difference by using Heckman's two stage model. Set up the model and state the underlying assumptions. What is required for it to be identified?

3.

    a. Suppose you want study differences in unemployment durations between two demographic groups in a sample of unemployed workers. What assumptions are needed in order to get consistent estimates if you simply use an OLS model with log duration of unemployment spells as a dependent variable to do this?

    b. Suppose you study unemployment duration. In the context of an exponential proportional hazard model, show that unobserved heterogeneity may appear as duration dependence. How will it affect the estimates of the model?

    c. Suppose that you want to study the relation between education and mortality in circulatory diseases. Describe how you can estimate a competing risk model for this problem. What assumptions are required? Describe also how you can construct Peterson bounds for comparing mortality in circulatory diseases between two education groups. Give an example for how these bounds may look like at a particular age.