STOCKHOLM UNIVERSITY
Department of Economics

Course name:	**Econometrics 3a: Methods for Analyzing Micro Data**
Course code:	**EC7412**
Examiner:	**Mårten Palme**
Number of credits:	**7,5 credits (hp)**
Date of exam:	**23 August 2015**
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 2 are worth 35 points and question 3 is worth 30 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 accident 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 $\beta_3 = \beta_4 = \beta_5 = 0$.
b. Describe different ways to interpret the magnitude of the β_2 estimate.
c. What would a fully saturated model look like in this context? How can you test between the model suggested above and a fully saturated model and how would you interpret the results?

2. Suppose you want to estimate a linear regression model on cross-section data with one independent variable, i.e., $Y_i = \alpha + \beta X_i + \varepsilon_i$, where you observe X_i with a measurement error, i.e., $X_i^* = X_i + \nu_i$, where ν_i follows an iid distribution with $\sigma_\nu^2 \neq 0$ and X_i is the true work variable.

a. Explain what will happen to your OLS estimate of β in presence of measurement error.

b. Suppose you instead of cross section data you now have a panel. Explain why, depending on the data, the problem with the bias from the measurement error may be better or worse when you use a first difference estimator rather than pooled OLS.
c. Suppose you now want to estimate you panel data model with fixed effects and a lagged dependent variable, i.e., $Y_{it} = \alpha_i + \gamma Y_{it-1} + \beta X_i + \varepsilon_i$. Ignoring any problems with measurement errors, what would happen to the least squares fixed effects estimator when you have a lagged dependent variable in the model?

3.

a. In the context of an exponential proportional hazard model, show that an omitted variable may appear as duration dependence.
b. Describe the consequences of unobserved heterogeneity on the maximum likelihood coefficient estimates in a proportional hazard model.
c. Describe how you can differentiate between different exit routes in the framework of a Cox proportional hazard model. Describe how you can obtain consistent estimates in such model and what assumptions you need?