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**
Date of exam:	**August 18, 2016**
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 accident on gender (male or female) and economic status (first, second, third class passenger or crew). Consider the model:

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

where *Female* is an indicator for being female; *First*, *Second* and *Third* are indicators for the different classes. *Crew* is the excluded category.

a. 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?
b. Describe other goodness-of-fit measures for this model. What are their comparative advantages?
c. Explain why a linear probability and a probit model will have similar properties if you estimate a fully saturated version of the model above.

2.

a. Explain how a quantile regression model can be used to deal with the problem of censoring of the dependent variable when you estimate a linear regression model. What is the advantage of this model compared to a Tobit model?

b. Suppose you want to estimate the linear relation $Y_i = \alpha + \beta X_i^* + \varepsilon_i$, where you can only observe *X*, which is X_i^* with an error, i.e., $\nu_i = X_i - X_i^*$, where ν_{it} follows an iid distribution with $\sigma_\nu^2 \neq 0$. How would that affect your OLS estimate of *β*? How would this change if you have access to panel data and use a fixed effects estimator?

c. Suppose you now want to estimate a panel data model with fixed effects and a lagged dependent variable, i.e., $Y_{it} = \alpha_i + \gamma Y_{it-1} + \varepsilon_i$. Explain why a conventional fixed effects estimator may give inconsistent estimates in short panels and define the Anderson-Hsiao estimator of this model.

3.

a. In the paper "Workers' Compensation and Injury Duration: Evidence from a Natural Experiment" Bruce Meyer, Kip Viscusi and David Durbin estimate a differences-in-differences model where the dependent variable is the log of the length of time of receipt of workers' compensation. The model is estimated using OLS. Explain why a hazard model is not needed in order to get consistent estimates.

b. Explain the difference between a proportional hazard model based on the exponential distribution and a Cox proportional hazard model. In what context is the exponential model preferred?

c. Explain why individual heterogeneity may appear as duration dependence and describe briefly how you can handle heterogeneity in a proportional hazard model.