Elements of Econometrics. 2022-2023.
Class 15. MLE and Binary Choice Models

Problem 1. (UoL Exam)

a) □ Explain what is the maximum likelihood principle for estimating parameters in statistics and econometrics.

□ What are its advantages and disadvantages in comparison with the method of least squares.

b) An event is hypothesized to occur with probability p. In a sample of n observations, it occurred m times.

□ Demonstrate that the maximum likelihood estimator of p is m/n.

Problem 2. (ICEF Exam) More than a century ago on April 14, 1912, the unthinkable happened when the "unsinkable" Royal Mail Ship (RMS) Titanic crashed into an iceberg and sunk into the Atlantic Ocean. The 20 lifeboats aboard the ship were not enough to save a majority of the passengers, leaving over 1500 passengers and crew members aboard the sinking vessel. A total of 705 passengers escaped onto lifeboats and to safety. But not all passengers had an equal chance of getting onto a lifeboat and surviving the disaster.

A student has found in internet the data on 2201 passengers and crew members of RMS Titanic during his last voyage across the ocean. She is trying now, using regression analysis, to estimate the impact of passenger class (*CLASS*=0 for the crew members, *CLASS*=1 for the first class passengers, *CLASS*=2 for the second class passengers, *CLASS*=3 for the third class passengers), gender (*MALE*=1 for male and 0 for female) and age (*AGE*=1 for adults and 0 for children) on a person's likelihood of surviving the shipwreck *SURV*. In the data *SURV*=1 if the person survived the shipwreck and *SURV*=0 if not. She runs different regressions with the following results (dependent variable *SURV,* regression (i) uses full sample, other regressions use subsample without crew members; (asymptotical) standard errors in parentheses)*.*

	(i) (OLS)	(ii) (OLS)	(iii) (LOGIT)	(iv) (LOGIT)	(v) (PROBIT)
CLASS	−0.0515 (0.0072)	−0.1515 (0.0134)	−0.489 (0.043)	−0.875 (0.085)	−0.502 (0.048)
AGE	−0.165 (0.007)	−0.181 (0.040)	—	−1.056 (0.243)	−0.580 (0.138)
MALE	−0.552 (0.022)	−0.478 (0.023)	—	−2.367 (0.145)	−1.415 (0.084)
Constant	0.985 (0.047)	1.208 (0.053)	0.793 (0.103)	3.895 (0.347)	2.257 (0.193)
N	2,201	1,316	1,316	1,316	1,316
R^2	0.228	0.331			
McFadden R^2			0.0758	0.269	0.267
RSS	371.59	207.31	278.52	204.30	205.21
LogL	−1165.43	−651.45	−807.15	−638.40	−639.93
LR stat,			132.45	469.95	466.90

(a) Give interpretation to the regressions **(i)** and **(ii)**. Are they significant? What are the advantages and the disadvantages of these models? Why in models **(iii)**-**(iv)**-**(v)** these shortcomings are absent or mitigated?

(b) Using regression **(iv)** help the student to evaluate the marginal effect of passenger class, taking as the initial values an adult female person traveling in a first class cabin (use method of derivatives). Compare your result with the corresponding result for regression **(ii)**

(c) What are McFadden R^2 and *LR* statistic? How they can be used for evaluation of the statistical quality of the regression? Comparing regressions **(iii)** and **(iv)** evaluate whether two dummy variables *AGE* and *GENDER* are significant both separately and taken together?

(d) What is the difference between regressions **(iv)** and **(v)?** Help the student to understand whether for female passengers of Titanic the chance to survive is significantly greater. To do this for regression **(v)** evaluate the marginal effect of gender, taking as the initial values an adult male person traveling in a first class cabin (use direct calculation without derivatives). Compare your result with the corresponding result for regression **(ii)**.