HSE and University of London Double Degree Programme in Data Science and Business Analytics

Elements of Econometrics, 2023-2024

Classteacher: Ksenia Kasianova.

Class 20: SEM. Bianary choice models

Problem 1

Consider an example of a system (a simplified version of the accelerator multiplier model) in which there are linear constraints of the non-excluding conditions:

$$\begin{cases} C_t = a_0 + a_1 Y_t + a_2 C_{t-1} + u_{t1} \\ I_t = b_0 + b_1 (Y_t - Y_{t-1}) + u_{t2} \\ Y_t = C_t + I_t, \end{cases}$$

where C_t is consumption;
I_t - investments;
Y_t - income.
(a) Is system identified?

Problem 2

(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.

Let $y_1, \ldots, y_n$ be i.i.d. random variables from Exp(θ), where θ is scale parameter.
(b) Find the maximum likelihood estimator.
(c) Find Fisher information $i(\theta)$.

Problem 3

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 yessel. 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 ($\boldsymbol{CLASS} = 0$ for the crew members, $\boldsymbol{CLASS} = 1$ for the first class passengers, $\boldsymbol{CLASS} = 2$ for the second class passengers, $\boldsymbol{CLASS} = 3$ for the third class passengers), gender ($\boldsymbol{MALE} = 1$ for male and 0 for female) and age ($\boldsymbol{AGE} = 1$ 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 $\boldsymbol{SURV}$, 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	-0.1515	-0.489	-0.875	-0.502
	(0.0072)	(0.0134)	(0.043)	(0.085)	(0.048)
AGE	-0.165	-0.181		-1.056	-0.580
	(0.007)	(0.040)	-	(0.243)	(0.138)
MALE	-0.552	-0.478		-2.367	-1.415
	(0.022)	(0.023)		(0.145)	(0.084)
Constant	0.985	1.208	0.793	3.895	2.257
	(0.047)	(0.053)	(0.103)	(0.347)	(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).

Problem 4

A student during non-study time works in a record company. It was not a banner year but there appeared potential clients - two musical bands. In case of successful release of their albums in December company turns a profit, which allows to make bonus payment for its employees and happily celebrate the New Year. There are data on 200 albums on the base of which the student wants to analyze and predict the success of future albums.

Success - dependent binary variable which takes value 1 if number of sales of an album in the week after release is greater than 200 thousands.

Budget - the amount (in thousands of US dollars) spent promoting the album before release.

Airplay - number of times songs from an album were played on radio in the week before release.

Rank - people's estimate of the attractiveness and stylishness of the performing musicians out of 10

The student estimates different model specifications (standard errors in parentheses):

	(i) OLS	(ii) Logit	(iii) Logit
Constant	-0.688	-9.9075	-10.911
	(0.1286)	(1.783)	(2.2275)
Budget	0.00047	0.00419	0.0042
	(0.00005)	(0.0007)	(0.0007)
Airplay	0.0185	0.1483	0.2222
	(0.0021)	(0.0245)	(0.0958)
$Airplay^2$	-	-	-0.0012
			(0.0015)
Rank	0.0589	0.5047	0.5082
	(0.0181)	(0.2081)	(0.2074)
R^2	0.522312		
McFadden R^2		0.5163	0.5186
RSS	23.874	20.520	20.309
LogLikelihood	-71.238	-67.027	-66.703

One band **A** plans to spend \$10000 on advertising and order 55 plays on radio in the week before release, the majority of people in focus-group found them "gorgeous" giving rank 7. The other band **B** is the band of young students who can afford to spend only \$2000 on advertising and order 25 plays on radio but people liked the way they look even much more and ranked them 10.

(a) Give interpretations to the coefficients of model (i).

What's the difference between models (ii) and (i)?

What are advantages and disadvantages of each type of models?

(b) According to model (ii) what are the chances for each band to issue a successful album?

Investigate the marginal effects of budget, airplay and ranking. Compare the results with ones obtained in model (i).

What are the maximum possible marginal effects of budget, airplay and ranking according to model (ii)?