BMECV1314D

COPENHAGEN BUSINESS SCHOOL

DEPARTMENT OF ECONOMICS

Winter Semester, 2024

Final Ordinary Resit Exam, 22.1.2025

Econometrics

Time allowed: TWO hours

Students must answer all parts of the question.

General guidance:
You should not devote more than one hour to answer each of the questions.

DO NOT TURN THIS PAGE OVER UNTIL INSTRUCTED TO DO SO BY THE INVIGILATOR

QUESTION

This question considers the estimation of the determinants of participation in private pension plans. The data are a random sample of 9,275 employees. The data comprise of the following variables:

inc	= annual income in 1,000s of €
marr	= 1 if married (0 otherwise)
male	= 1 if male (0 otherwise)
age	= age in years
fsize	= number of family members
pira	= 1 if participating in a private pension insurance plan (0 otherwise)
incsq	= *inc* squared
agesq	= *age* squared

The descriptive statistics are given in Table 1A.

TABLE 1A

Variable	Obs	Mean	Std. Dev.	Min	Max
inc	9,275	39.25464	24.09	10.008	199.041
marr	9,275	.6285714	.4832128	0	1
male	9,275	.2044205	.4032993	0	1
age	9,275	41.08022	10.29952	25	64
fsize	9,275	2.885067	1.525835	1	13
pira	9,275	.2543396	.4355128	0	1
incsq	9,275	2121.192	3001.469	100.1601	39617.32
agesq	9,275	1793.653	895.6488	625	4096

The dependent variable of the analysis is *pira*. The empirical analysis is conducted with R.

CONTINUED

Suppose that $y \in \{0,1\}$ has the probability density function

f(y|$\boldsymbol{x_i}, \boldsymbol{\beta}) = [G(\boldsymbol{x_i}\boldsymbol{\beta})]^y[\mathbf{1} - G(\boldsymbol{x_i}\boldsymbol{\beta})]^{1-y}$*,*

where *E(y|*$\boldsymbol{x}_i$*)=*$p(\boldsymbol{x}_i)$ and *Var(y|*$\boldsymbol{x}_i$*)=*$p(\boldsymbol{x}_i)[1 - p(\boldsymbol{x}_i)]$ *with* $p(\boldsymbol{x}_i) = P(y = 1|\boldsymbol{x}_i)$ *.*$\boldsymbol{x}_i$ *is 1xk. G is a known function, while* $\boldsymbol{\beta}$ is *kx1* and unknown*.* There are i=1,…,n observations *(*y_i, x_i*).*

(a) Explain why it would be suitable to apply conditional maximum likelihood estimation (CMLE).

(b) What is the sample log-likelihood of CMLE?

(c) What is the necessary condition for the log-likelihood to be maximised?

(d) The following Table 1B presents regression output for the data described above.

CONTINUED

TABLE 1B

```
> reg1<-glm(pira ~ inc+incsq+age+agesq+marr+male+fsize, family = binomial(link=logit), data=pension)
> summary(reg1)

Call:
glm(formula = pira ~ inc + incsq + age + agesq + marr + male + 
    fsize, family = binomial(link = logit), data = pension)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)    
(Intercept) -5.866e+00  4.942e-01 -11.869  < 2e-16 ***
inc          6.310e-02  3.470e-03  18.183  < 2e-16 ***
incsq       -2.422e-04  2.525e-05  -9.595  < 2e-16 ***
age          9.694e-02  2.344e-02   4.136 3.53e-05 ***
agesq       -5.085e-04  2.629e-04  -1.934   0.0531 .  
marr         1.432e-01  7.830e-02   1.829   0.0674 .  
male        -8.300e-02  7.465e-02  -1.112   0.2662    
fsize       -1.985e-01  2.430e-02  -8.167 3.17e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 10518.8  on 9274  degrees of freedom
Residual deviance:  8757.7  on 9267  degrees of freedom
AIC: 8773.7

Number of Fisher Scoring iterations: 5
```

Explain the econometric model that has been estimated in Table 1B. Use a latent variable model approach in your explanations.

(e) What is the partial relationship of annual income and the dependent variable in the model in Table 1B? Explain.

(f) More code is being executed and the output is shown in Output 1C.

OUTPUT 1C

```
> library(margins)
> lev<-data.frame(mean(pension$inc),mean(pension$inc)^2,mean(pension$age),mean(pension$age)^2,mean(pension$marr),mean(pension$male),mean(pens
ion$fsize))
> names(lev)<-c("inc","incsq","age","agesq","marr","male","fsize")
> margins(reg1, at=lev)
Average marginal effects at specified values
glm(formula = pira ~ inc + incsq + age + agesq + marr + male +     fsize, family = binomial(link = logit), data = pension)

 at(inc) at(incsq) at(age) at(agesq) at(marr) at(male) at(fsize)    inc      incsq     age      agesq    marr     male    fsize
   39.25      1541   41.08      1688   0.6286   0.2044     2.885 0.0116 -4.454e-05 0.01782 -9.348e-05 0.02633 -0.01526 -0.03649
```

Explain in detail what has been done in Output 1C? What is the role of income when it is 50,000$?

(g) Additional code gives the Output in Output 1D.

CONTINUED

OUTPUT 1D

```
> l_ur<-logLik(reg1)
> reg2<-glm(pira ~ inc+incsq+age+agesq+fsize, family = binomial(link=logit), data=pension)
> l_r <-logLik(reg2)
> lr=2*(l_ur-l_r)
> lr
'log Lik.' 6.057492 (df=8)
> 1-pchisq(lr, 2)
'log Lik.' 0.04837625 (df=8)
```

Explain in detail what has been done in Output 1D What do you conclude?

(h) Define the pseudo-R-squared for maximum likelihood estimation. Why is this used instead of the conventional R-squared after maximum likelihood estimation?

END OF PAPER