MINERVA®

Assignment III

Econometrics and Economic Systems

Spring 2019 | SS154 | Taha Bouhoun

Q1) Fair's (1978) *Redbook* survey on extramarital affairs.

a) The regressors of interest are *v1* to *v8*; however, not necessarily all of them belong in your model.

1. Model specification

- Outcome variable: refers to the number of extramarital affairs. For convenience, the variable is coded 1 if cheating occurred once or more, 0 if never been involved in affairs.

A	Freq.	Percent	Cum.
0	4,313	67.75	67.75
1	2,053	32.25	100.00
Total	6,366	100.00	

Fig (0). Tabulating the binary outcome variable

- Explanatory variables:

Marriage rating: as a self-reported metric, this might be subject to overestimation (bias). Given the context of this study, we're not certain about the accuracy of the responses but we still include it in the model as it offers a direct evaluation of the relationship which would have some truth in it when coupled with other metrics.

Age (years): as a personal inherent trait, age could be a powerful explanatory variable. Ranging from 17 to 42, age can highlight important trends in extramarital affairs.

Year married: Intuitively, the length of the relationship plays a crucial role in whether the individual would cheat.

Number of children: Kids being part of the equation can directly affect the way the parents behave. Moving from kid 3 to 4 might not impact the likelihood of cheating so to make it more interesting to interpret, the variable kids was transformed into a binary metric where 0 stands for no kids and 1 otherwise.

Religiosity: one's beliefs and values can impact the way they manage their relationships. In many cases, being religious infers being conservative but this model can put this idea under test so we would like to examine to what extent being devoted to faith can impact the likelihood of cheating.

Education: as one of the personal traits, education might have a role in the likelihood of cheating.

Remark on occupation variable: at first, I thought this would be an interesting addition if it was a binary variable (1 employed 0 if not) as being employed leads to spending time interacting with other people outside of the usual circle. but upon examining the dataset, occupation turns to be just a categorical variable that shows the nature of the job of the husband and wife (v7 and v8). Since all the occupations involve being out there, I thought I would exclude it from the model.

2. Use this data to build a binary choice model for A

Using Probit as a binary choice model for the covariates listed above yielded the results on Table (1). The model specification defines **marriage rating** and **religiosity** as categorical variables. We also have the **children** as a binary variable.

For convenience, table (1) includes the estimates for the Logit and Probit model so that it would be easier to compare. A quick look at the coefficients informs us that both models are similar (the sign and the significance level) however the magnitude can be a bit different which is attributed to the function that each model uses. Probit uses the Cumulative Distribution Function of the normal distribution whereas the Logit model uses the CDF of the logistic distribution.

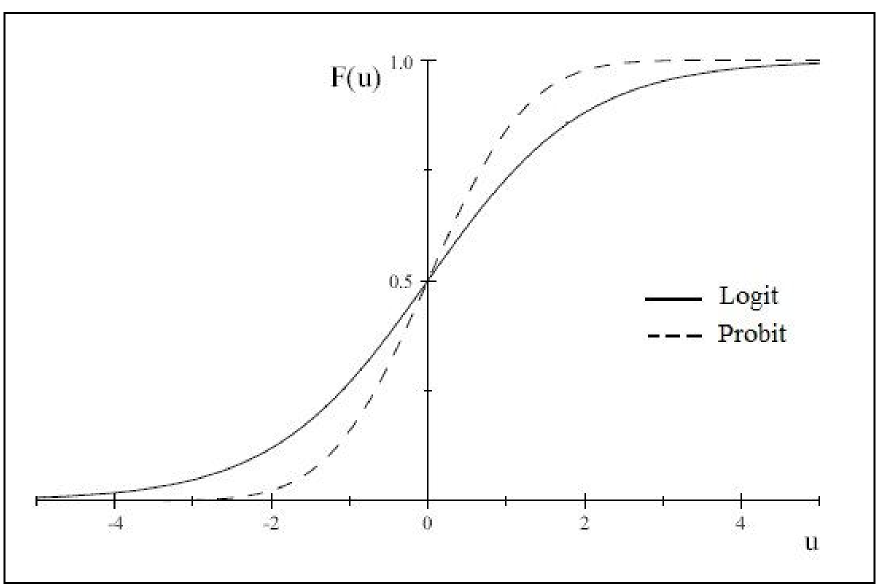

Figure 3: Cumulative Distribution Function of Probit and Logit Models. Source: V. Verardi, Applied Microeconometrics Course, FUNDP.(2008).

We can observe how similar both models can be when plotting the CDF of the distributions they use (normal for Probit and logistic for Logit)

VARIABLES	Probit model	Logit model
Marriage rating = 2	-0.235	-0.391
	(0.156)	(0.263)
Marriage rating = 3	-0.445***	-0.732***
	(0.145)	(0.246)
Marriage rating = 4	-0.982***	-1.604***
	(0.142)	(0.242)
Marriage rating = 5	-1.395***	-2.315***
	(0.142)	(0.243)
Age (years)	-0.032***	-0.055***
	(0.006)	(0.010)
Years married	0.055***	0.094***
	(0.006)	(0.010)
Children	0.211***	0.363***
	(0.046)	(0.079)
Education	-0.003	-0.005
	(0.009)	(0.014)
Religiosity = 2	-0.203***	-0.343***
	(0.051)	(0.085)
Religiosity = 3	-0.386***	-0.653***
	(0.051)	(0.086)
Religiosity = 4	-0.770***	-1.303***
	(0.075)	(0.131)
Constant	1.121***	1.891***
	(0.216)	(0.368)
Observations	6,366	6,366

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table (1). The regression results from a Probit and Logit model

As discussed in class, we can't interpret the estimates as they're embedded in the CDF functional form so we need to compute the marginal effect in order to comment on the indicators.

VARIABLES	Probit model	Logit model
Marriage rating = 2	-0.083	-0.084
	(0.054)	(0.055)
Marriage rating = 3	-0.161***	-0.162***
	(0.049)	(0.051)
Marriage rating = 4	-0.357***	-0.356***
	(0.048)	(0.049)
Marriage rating = 5	-0.481***	-0.482***
	(0.048)	(0.049)
Age (years)	-0.010***	-0.010***
	(0.002)	(0.002)
Years married	0.017***	0.017***
	(0.002)	(0.002)
Children	0.065***	0.066***
	(0.014)	(0.014)
Education	-0.001	-0.001
	(0.003)	(0.003)
Religiosity = 2	-0.067***	-0.068***
	(0.017)	(0.017)
Religiosity = 3	-0.124***	-0.125***
	(0.017)	(0.017)
Religiosity = 4	-0.227***	-0.227***
	(0.020)	(0.021)
Observations	6,366	6,366

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table (2). The marginal effect results of the Probit and Logit model

Marginal effects:
Conforming with Fig. 3 the estimates are almost identical, we notice that **education** has a minor effect (but not significant). Being **religious** brings down the probability of cheating (a rate of 4 brings the likelihood by 22.7% compared to a rate of 1). The shocking part is that having **kids** doesn't lower the likelihood of cheating but raises it by 6.5% ($\alpha = 1\%$) this might indicate that having children can lead parents to invest emotionally on them neglecting their own relationship[1].

[1] #testability: the hypothesis that kids would make it less likely to cheat was tested using the Probit and Logit model, it turns out that having kids raises the likelihood for cheating.

So is the case for **years married** (one extra year married raises the likelihood of extramarital affairs by 1.7%

Marriage rating has a negative relation to the likelihood of cheating which was expected since the more satisfied the people are about their marriage the less likely to seek affairs and betray their partners.

b) The analysis using the self-reported marriage rating and running an ordered choice model:

1. Analyze this variable using an ordered probit model

VARIABLES	Marriage rating
Age	-0.006
	(0.005)
Years married	-0.002
	(0.005)
Children	-0.338***
	(0.036)
Religiosity = 2	-0.020
	(0.041)
Religiosity = 3	0.136***
	(0.041)
Religiosity = 4	0.436***
	(0.057)
Education	0.027***
	(0.007)
Cut1	-2.145
	(0.134)
Cut2	-1.443
	(0.129)
Cut3	-0.700
	(0.128)
Cut4	0.278
	(0.128)
Observations	6,366

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table (3). Ordered probit model coefficients for the marriage rating outcome: As indicated in the table, the explanatory variables are significant to 1% level except for **Age** and **Years married.** The model has 4 different cutoffs which suggest that we have 5 choices to pick from as explained in the metric for marriage rating

2. What variables appear to explain the response to this survey question?

The variable **religiosity** has a strong prediction on how well the marriage is rated. First, the indicator is significant to the 1% level. Second, starting from rating = 2, the coefficients are positive when moving upward in the rating of **religiosity** which indicates that the more religious the person is the more likely to rate their marriage better (*we can't interpret the magnitudes unless we use marginal*)

Children is a significant metric since it's a binary (have or don't have kids), we notice that it impacts the rating negatively and it's 1% level significant

Education is also significant although associated with a milder coefficient.

The following descriptions are based on Table (4) in the appendix

3. Can you obtain the marginal effects for your model? Report them as well

We can obtain the marginal effect of this ordered choice model, the layout of the table would be defined depending on how many outcomes for the dependent variable (Marriage rating). In other words, each variable has a different impact on marriage rating at each level of rating.

4. What do they suggest about the impact of the different independent variables on the reported ratings?

Significance: **Age** and **years married** seem to be insignificant in these results with a p-value that exceeds the 5% threshold.

Estimates: **Religiosity** has a strong impact compared to other covariates, for a religious person it raises their likelihood to rate 5 for marriage status by 17% compared to the non-religious. **Education**, although significant, doesn't affect the rating since it has less than $\pm 1\%$ effect on their rating for their marriage.

Having **children** seems to positively drive the likelihood to have a good rating between 1 and 4 (minor impact), however, the effect is negative between 4 and 5 which signifies that having kids lowers the likelihood of rating a 5 by about 2%

Q2) The German health care data include a count variable ***HospVis*** the number of visits to the hospital. For this application, we will examine this variable. To begin, we treat the full sample (27,326) observations as a cross-section.

a) Begin by fitting a Poisson model to this variable (the exogenous variables are listed in Table F7.1)

1. Determine an appropriate specification for the right-hand side of your model

Explanatory variables:

age: is a decisive metric in health-related studies (intuitively)
female: a personal trait that is associated with subjects, controlling for gender is a good practice to not fall for confoundedness.
handdum: whether a person is handicapped or not is a crucial metric to control for since it implies a significant difference in the frequency of visiting hospitals.
hsat: a self-reported metric, although might be biased, it has some truth to it, hence, it needs to be added to the formula to predict the hospital visits.
hhninc: according to the economic intuition, income has a strong implication on how frequently one would visit the hospital since having medical care implies a cost that one is ought to pay.
educ: another personal trait that needs to be included to highlight any disparities arises from differences in education levels.
public: this ought to be a strong indicator since being insured reflects the frequency of being subjected to medical check-up regularly.

2. Report the regression results and marginal effects:

VARIABLES	(Poisson model) hospvis	(OLS) hospvis
Age (years)	-0.0048***	-0.0008*
	(0.0015)	(0.0005)
Female	0.0630*	0.0087
	(0.0333)	(0.0112)
Handicapped	0.2309***	0.0425***
	(0.0368)	(0.0147)
Health satisfaction	-0.2410***	-0.0400***
	(0.0065)	(0.0042)
Household income	0.0000***	0.0000*
	(0.0000)	(0.0000)
Education	-0.0520***	-0.0062***
	(0.0088)	(0.0023)
Public insurance	-0.0524	-0.0073
	(0.0592)	(0.0199)
Constant	0.0788	0.4913***
	(0.1548)	(0.0673)
Observations	27,286	27,286
R-squared		0.0120

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table (5). The regression results of the Poisson model and the OLS model.

The signs of the coefficients for the Poisson model and the OLS model are consistent, however, the significance levels are different for variables like **Age** and **Female**. We need to run the marginal command to interpret the results.

VARIABLES	Poisson model	OLS model
Age (years)	-0.0007***	-0.0008*
	(0.0002)	(0.0005)
Female	0.0087*	0.0087
	(0.0046)	(0.0112)
Handicapped	0.0320***	0.0425***
	(0.0051)	(0.0147)
Health satisfaction	-0.0334***	-0.0400***
	(0.0010)	(0.0042)
Household income	0.0000***	0.0000*
	(0.0000)	(0.0000)
Education	-0.0072***	-0.0062***
	(0.0012)	(0.0023)
Public insurance	-0.0073	-0.0073
	(0.0082)	(0.0199)
Observations	27,286	27,286

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table (6). The marginal effect for the Poisson and OLS model

The marginal effects for both models are almost identical except for **handicapped** (still too close). The significance level for both models' covariates are similar, but the coefficients are low (0.001 magnitudes) for **Income**, **age**, **education**, and **gender**. On the other hand, intuitively being handicapped raises the number of hospital visits by 0.03/0.042 we can also notice that **health satisfaction** also predicts the number of visits one ought to make to the hospital (the higher the health satisfaction the lower number of visits to the hospital).
Aside from the coefficients, the Poisson model brings down the standard error of all the variables to almost half the SE for the OLS model.

c) Is there evidence of overdispersion in the data? Test for overdispersion.

VARIABLES	hospvis	lnalpha
Age (years)	-0.0055**	
	(0.0022)	
Female	0.0826*	
	(0.0480)	
Handicapped	0.2663***	
	(0.0562)	
Health satisfaction	-0.2262***	
	(0.0099)	
Household	0.0000***	
	(0.0000)	
Education	-0.0530***	
	(0.0119)	
Public insurance	0.0492	
	(0.0825)	
Constant	-0.0792	1.8943***
	(0.2185)	(0.0403)
Observations	27,286	27,286
Likelihood-ratio test of alpha=0:		
chibar2(01) = 5166.92 Prob = chibar2 = 0.000		

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table (7). Negative binomial model to check for overdispersion

As seen before, although the Poisson model mitigates error, it cannot correct for an error in model specification. As we see in Table (7), coefficients for **handicapped** and **health satisfaction** boosted to around $\pm 0.25\%$
According to the test at the bottom of the table which tests for alpha = 0 which is the assumption that the mean and the variance are the same.
(chibar2 = 0.000) means that we reject the hypothesis that mean and variance are the same, hence, there's overdispersion in our data which justifies that we needed to use a negative binomial regression.

Appendix:

1. The marginal effect of the ordered probit model for the Redbook data

```
. margins, dydx(*) atmeans

Conditional marginal effects                    Number of obs     =      6,366
Model VCE    : OIM

dy/dx w.r.t. : Age Years_married Children i.Religiosity Education
1._predict   : Pr(v1==1), predict(pr outcome(1))
2._predict   : Pr(v1==2), predict(pr outcome(2))
3._predict   : Pr(v1==3), predict(pr outcome(3))
4._predict   : Pr(v1==4), predict(pr outcome(4))
5._predict   : Pr(v1==5), predict(pr outcome(5))
at           : Age             =    29.08286 (mean)
               Years_married   =    9.009425 (mean)
               Children        =    1.396874 (mean)
               Religiosity = 2 =    .3561106 (mean)
               Religiosity = 3 =    .3804587 (mean)
               Religiosity = 4 =    .1030474 (mean)
               Education       =    14.20986 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Age          |
    _predict |
          1  |   .0001461   .0001648     0.89   0.375    -.0001769    .0004692
          2  |   .0003882   .0004374     0.89   0.375    -.0004691    .0012455
          3  |    .000705   .0007939     0.89   0.375     -.000851     .002261
          4  |   .0003914    .000441     0.89   0.375    -.0004729    .0012557
          5  |  -.0016307   .0018356    -0.89   0.374    -.0052284    .0019669
-------------+----------------------------------------------------------------
Years_married|
    _predict |
          1  |   .0002794   .0001786     1.56   0.118    -.0000706    .0006294
          2  |   .0007423   .0004711     1.58   0.115    -.0001811    .0016657
          3  |    .001348   .0008547     1.58   0.115    -.0003271    .0030231
          4  |   .0007484   .0004758     1.57   0.116    -.0001841    .0016809
          5  |  -.0031181    .001975    -1.58   0.114     -.006989    .0007528
-------------+----------------------------------------------------------------
Children     |
    _predict |
```

```
           1  |   .0020127   .0005569     3.61   0.000     .0009212    .0031043
           2  |   .0053476   .0014302     3.74   0.000     .0025444    .0081508
           3  |   .0097112   .0025783     3.77   0.000     .0046578    .0147645
           4  |   .0053913   .0014522     3.71   0.000     .0025451    .0082376
           5  |  -.0224628    .005929    -3.79   0.000    -.0340834   -.0108422
--------------+----------------------------------------------------------------
Religiosity 2|
    _predict  |
           1  |    .000869   .0014536     0.60   0.550      -.00198    .0037181
           2  |   .0023089   .0038575     0.60   0.549    -.0052516    .0098694
           3  |    .004193   .0070038     0.60   0.549    -.0095341    .0179201
           4  |   .0023278   .0038901     0.60   0.550    -.0052965    .0099522
           5  |  -.0096988   .0161988    -0.60   0.549    -.0414478    .0220503
--------------+----------------------------------------------------------------
Religiosity 3|
    _predict  |
           1  |  -.0044666   .0014938    -2.99   0.003    -.0073943   -.0015388
           2  |  -.0118671   .0038768    -3.06   0.002    -.0194655   -.0042687
           3  |  -.0215505   .0070099    -3.07   0.002    -.0352897   -.0078113
           4  |  -.0119641    .003929    -3.05   0.002    -.0196649   -.0042634
           5  |   .0498483   .0161485     3.09   0.002     .0181977    .0814988
--------------+----------------------------------------------------------------
Religiosity 4|
    _predict  |
           1  |  -.0154077   .0023985    -6.42   0.000    -.0201086   -.0107068
           2  |  -.0409363   .0056279    -7.27   0.000    -.0519668   -.0299058
           3  |  -.0743397   .0099296    -7.49   0.000    -.0938014   -.0548781
           4  |   -.041271   .0058543    -7.05   0.000    -.0527452   -.0297969
           5  |   .1719547   .0224882     7.65   0.000     .1278787    .2160307
--------------+----------------------------------------------------------------
Education     |
    _predict  |
           1  |  -.0010289    .000252    -4.08   0.000    -.0015229   -.0005349
           2  |  -.0027335   .0006399    -4.27   0.000    -.0039877   -.0014794
           3  |  -.0049641   .0011509    -4.31   0.000    -.0072199   -.0027083
           4  |  -.0027559   .0006516    -4.23   0.000     -.004033   -.0014787
           5  |   .0114824   .0026427     4.34   0.000     .0063027     .016662
-------------------------------------------------------------------------------
```

2. Link to Stata file:

https://github.com/Tahahaha7/Econometrics_and_Economic_Systems/blob/master/Homework%20III.do

3. HC Applications:

#regression: throughout the assignment, I used Stata to run regression models and examine the relationship between the dependent and the explanatory variables. The application also extends to the interpretation of the regression results as well as the comparison between OLS and Poisson model.

#organization: following an easy to read implementation for the results of the regression models used in this assignment as well as an adjacent table with different models estimates to compare and contrast.

#composition: the paper extends to offer a succinct interpretation of the regression results by translating the coefficients in the tables (after defining the significance level) into the context they represent.

#biasidentification: explaining the intuition behind picking the variables for the model specification and highlighting the reasoning behind the choice while highlighting any potential bias (e.g., the self-reporting variables)