Solution Guideline: PS04

Dan N. Olsen, Nikolaj Nielsen

1 PS04

The idea with PS04 is to estimate three different models for binary outcome variables. The models considered are: the linear probability model (LPM), estimated using OLS, and the logit and probit models estimated using MLE.

1.1 Theoretical introduction

For binary outcome data the dependent variable takes on the values,

$$y = \begin{cases} 1 & \text{with prob. } p, \\ 0 & \text{with prob. } 1-p. \end{cases} \tag{1}$$

We can then specify a regression model by parameterizing the probability p to depend on a regressor vector, x, which is $N \times K$ and a parameter vector, β, which is $K \times 1$. The conditional probability is given as,

$$p_i = P\left(y_i = 1 \left| x \right.\right) = G\left(x_i'\beta\right) \tag{2}$$

1.1.1 The Linear Probability Model (LPM)

In the linear probability model the response probability is,

$$p_i = P\left(y_i = 1 \left| x_i \right.\right) = E\left(y_i \left| x_i \right.\right) = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_k x_{i,k} = x_i'\beta \tag{3}$$

The objective function is,

$$Q_N\left(\beta\right) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - x_i'\beta\right)^2. \tag{4}$$

For OLS there exists an analytical solution for the above minimization problem and the OLS estimator, which minimizes the sum of squared residuals, is given as,

$$\hat{\beta} = \left(X'X\right)^{-1}X'y. \tag{5}$$

Note, that $\hat{\beta}$ is an estimate of the population parameter, β.

The heteroscedasticity robust variance-covariance matrix is,

$$\hat{V}\left[\hat{\beta}\right] = \left(X'X\right)^{-1}X'\hat{\Omega}X\left(X'X\right)^{-1} \tag{6}$$

where $\hat{\Omega} = diag_{i=1}^{N}\left[\hat{u}_i\right]$, where $\hat{u}_i = y_i - x_i'\hat{\beta}$ is the estimated OLS residual for individual i. Finally, the standard errors (both robust and non-robust) can be obtained as, $s.e. = \sqrt{diag\left[V\left[\hat{\beta}\right]\right]}$, i.e., the square root of the diagonal elements of the estimated variance-covariance matrix.

If the probability is $p_i = x_i'\beta$ then linear regression is justified (it is, however, preferred to us the logit or probit for the final analysis). The conditional mean and variance of the model are,

$$E\left[y_i\left|x_i\right.\right] = x_i'\beta, \tag{7}$$
$$V\left[y_i\left|x_i\right.\right] = x_i'\beta\left(1 - x_i'\beta\right). \tag{8}$$

As the variance depends upon the regressors the model is heteroscedastic and we need to use heteroscedastic robust standard errors as default.

1.1.2 The Logit Model

For the Logit model the response probability is non-linear,

$$P\left(y_i = 1\left|x\right.\right) = G\left(x'\beta\right) = G\left(x'\beta\right) = \frac{\exp\left(x'\beta\right)}{1 + \exp\left(x'\beta\right)} \tag{9}$$

The outcome in binary models is Bernoulli distributed (the binomial distribution with only one trial). The probability mass function for y_i is,

$$y_i = f\left(y_i\left|x_i\right.\right) = p_i^{y_i}\left(1 - p_i\right)^{1-y_i} \quad y_i \in \{0, 1\} \tag{10}$$

with $p_i = G\left(x_i'\beta\right) = \frac{\exp\left(x_i'\beta\right)}{1+\exp\left(x_i'\beta\right)}$. The log-likelihood contribution for observation i is,

$$L\left(\beta\right) = y_i \log G\left(x_i'\beta\right) + \left(1 - y_i\right)\log\left(1 - G\left(x_i'\beta\right)\right) \tag{11}$$

And the log–likelihood function is,

$$L_N\left(\beta\right) = \sum_{i=1}^{N}\left\{y_i \log G\left(x_i'\beta\right) + \left(1 - y_i\right)\log\left(1 - G\left(x_i'\beta\right)\right)\right\} \tag{12}$$

The robust covariance matrix can be computed using the MLE version of the sandwich formula:

$$\hat{V}(\hat{\theta}_{MLE}) = \frac{1}{N}\hat{B}_0^{-1}\hat{A}_0\hat{B}_0^{-1} \tag{13}$$

Where:

$$\hat{A}_0 = \frac{1}{N}\left.\frac{\partial^2 \log l(\theta)}{\partial\theta\partial\theta'}\right|_{\hat{\theta}} = \frac{1}{N}\hat{H} \qquad \hat{B}_0 = \frac{1}{N}\sum_{i=1}^{N} s_i(\hat{\theta})s_i(\hat{\theta})' = \frac{1}{N}\hat{s}'\hat{s} \tag{14}$$

Where $\hat{H}$ is the Hessian of the log-likelihood function evaluated at $\hat{\theta}$ and $\hat{s}$ is the N-by-K matrix of individual scores.

1.1.3 The Probit Model

For the Probit model the response probability is non-linear,

$$P\left(y_i = 1\left|x\right.\right) = G\left(x'\beta\right) = \Phi\left(x'\beta\right) = \int_{-\infty}^{x'\beta}\phi\left(z\right)dz, \tag{15}$$

where $\Phi(\cdot)$ is the standard normal cdf, with derivative,

$$\phi(z) = \left(\frac{1}{\sqrt{2\pi}}\right)\exp\left(\frac{-z^2}{2}\right) \tag{16}$$

which is the standard normal density function.

The outcome in binary models is Bernoulli distributed (the binomial distribution with only one trial). The probability mass function for y_i is

$$y_i = f(y_i | x_i) = p_i^{y_i}(1-p_i)^{1-y_i} \quad y_i \in \{0,1\}. \tag{17}$$

The log-likelihood contribution for observation i is,

$$L(\beta) = y_i \log G(x_i'\beta) + (1-y_i)\log(1-G(x_i'\beta)) \tag{18}$$

And the log–likelihood function is,

$$L_N(\beta) = \sum_{i=1}^{N}\{y_i \log G(x_i'\beta) + (1-y_i)\log(1-G(x_i'\beta))\} \tag{19}$$

The robust covariance matrix can be computed using the MLE version of the sandwich formula:

$$\hat{V}(\hat{\theta}_{MLE}) = \frac{1}{N}\hat{B}_0^{-1}\hat{A}_0\hat{B}_0^{-1} \tag{20}$$

Where:

$$\hat{A}_0 = \frac{1}{N}\frac{\partial^2 \log l(\theta)}{\partial\theta\partial\theta'}\bigg|_{\hat{\theta}} = \frac{1}{N}\hat{H} \qquad \hat{B}_0 = \frac{1}{N}\sum_{i=1}^{N}s_i(\hat{\theta})s_i(\hat{\theta})' = \frac{1}{N}\hat{s}\hat{s}' \tag{21}$$

Where $\hat{H}$ is the Hessian of the log-likelihood function evaluated at $\hat{\theta}$ and $\hat{s}$ is the N-by-K matrix of individual scores.

1.2 Question 1

This is the table of descriptive statistics

Descriptive statistics

Variable	Mean	Std	Median	Max	Min
labour status	0.4599	0.4987	0.0000	1.0000	0.0000
log non-lab inc	10.6856	0.4125	10.6431	12.3757	7.1869
age	3.9955	1.0552	3.9000	6.2000	2.0000
education	9.3073	3.0363	9.0000	21.0000	1.0000
num young child	0.3119	0.6129	0.0000	3.0000	0.0000
num old child	0.9828	1.0868	1.0000	6.0000	0.0000
foreign dummy	0.2477	0.4319	0.0000	1.0000	0.0000
No. of obs.:	872				

1.3 Questions 2-5

The following table reports the estimates for the three models asked in the assignment: the linear probability model (LPM), probit model and logit model.

Parameter estimates

	LPM		Probit		Logit	
	Param. est.	s.e	Param. est.	s.e	Param. est.	s.e
CST	1.6637	0.3973	3.7491	1.3541	6.1964	2.3033
AGE	0.6825	0.1200	2.0753	0.4096	3.4366	0.6961
AGESQ	-0.0970	0.0145	-0.2943	0.0509	-0.4876	0.0872
EDUC	0.0067	0.0058	0.0192	0.0179	0.0327	0.0300
NYC	-0.2406	0.0301	-0.7145	0.1031	-1.1857	0.1816
NOC	-0.0493	0.0174	-0.1470	0.0514	-0.2409	0.0863
NLINC	-0.2128	0.0355	-0.6669	0.1298	-1.1041	0.2254
FOREIGN	0.2496	0.0402	0.7144	0.1216	1.1683	0.2053

The probit estimates are almost identical to the ones obtained by Gerfin (1996). Note, however, that the variable age is divided by 10 as in Gerfin's paper but the variable age square is not divided by 1000 as he suggests since this would lead to a scaled up coefficient of -29.43. The inconsistency between Gerfin's definition and the coefficient he reports is due to a mistake.

The statistical significance of the parameters can be inferred from the computed t-statistics. For the economic interpretation we are interested in the partial/marginal effects. The coefficients of the LPM are the partial effects and hence their sign and magnitude can be interpreted directly (e.g. one more year of schooling increases the probability (defined between 0 and 1) of being in the labour force by 0.0067, that is 0.67 p.p). The probit and logit coefficients are informative of the direction of the marginal effect but not its size, as we further explain in Question 6.

1.4 Question 6.

The explanation for why the coefficients are not equal for the LPM, the Logit and the Probit model is that these three models use different link functions for the probabilities and, as mentioned by Cameron & Trivedi, it makes more sense to compare the marginal effects across the three models. More specifically, the location and scale are set differently in these models, leading to different parameter estimates. A rule of thumb is that (see Cameron & Trivedi, section 14.3.7),

$$\hat{\beta}_{Logit} \simeq 4\hat{\beta}_{OLS} \tag{22}$$

$$\hat{\beta}_{Probit} \simeq 2.5\hat{\beta}_{OLS} \tag{23}$$

$$\hat{\beta}_{Logit} \simeq 1.6\hat{\beta}_{Probit} \tag{24}$$

The approximation $\hat{\beta}_{Logit} \simeq \left(\pi/\sqrt{3}\right)\hat{\beta}_{Probit}$ is also used. It comes from the fact that the standard normal distribution used as link function for the Probit model has mean 0 and variance 1, while the logistic distribution used for the Logit model has mean 0 and variance $\pi^2/3$.

Your output should look something like the following,

```
Ratio of parameters
-------------------------------------------------------
              Logit/Probit
CST             1.6528
AGE             1.6560
AGESQ           1.6567
EDUC            1.7016
NYC             1.6596
NOC             1.6392
NLINC           1.6555
FOREIGN         1.6355
-------------------------------------------------------
Assumed ratio of sigma:      1.8138
-------------------------------------------------------
```

1.5 Question 7.

The partial (also called marginal) effects in the Logit and Probit models depend upon the regressors, x_k. For continuous variables the partial effects are given as,

$$\frac{\partial P\left(y_i = 1 \left| x_i \right.\right)}{\partial x_{ik}} = \frac{\partial p_i}{\partial x_{ik}} = \frac{\partial G\left(x_i'\beta\right)}{\partial x_{ik}'\beta} \cdot \frac{\partial x_i'\beta}{\partial x_{ik}'} = g\left(x_i'\beta\right)\beta_k \tag{25}$$

where $g\left(z\right) = \frac{\partial G(z)}{\partial z}$ and where $g\left(z\right) = \frac{\exp(z)}{(1+\exp(z))^2}$ for the Logit model and $g\left(z\right) = \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-z^2}{2}\right)$ for the Probit model.

For the LPM the partial effects are constant and are given by the coefficients,

$$\frac{\partial P\left(y_i = 1 \left| x_i \right.\right)}{\partial x_{ik}} = \frac{\partial p_i}{\partial x_{ik}} = \frac{\partial x_i'\beta}{\partial x_{ik}'} = \beta_k \tag{26}$$

In Question 7 we are asked to calculate the partial effects of taking one additional year of education on the probability of participating in the labour market for a woman with some specific characteristics. Denote the vector of characteristics,

$$x_i^{0\prime} = [CST = 1,\ AGE = 2.5,\ AGE2 = 2.5^2,\ EDUC = 10,\ NYC = 1,\ NOC = 0,\\ NLINC = 10,\ FOREIGN = 0]$$

Also, denote the vector of estimated coefficients for the LPM, the Logit and the Probit models, $\hat{\beta}_{OLS}$, $\hat{\beta}_{Logit}$ and $\hat{\beta}_{Probit}$, respectively, which are all $K \times 1$ vectors. Further, denote the coefficient on education (i.e., the fourth element in the corresponding parameter vector) for the three models, $\hat{\beta}_{OLS,4}$, $\hat{\beta}_{Logit,4}$ and $\hat{\beta}_{Probit,4}$. The partial effects can then be calculated as,

$$\frac{\partial P(y_i = 1 \mid x_i)}{\partial x_{i4}} = \frac{\partial p_i}{\partial x_{i4}} = \hat{\beta}_{OLS,4} \tag{27}$$

for the LPM and,

$$\frac{\partial P(y_i = 1 \mid x_i)}{\partial x_{i4}} = \frac{\partial p_i}{\partial x_{i4}} = g\left(x_i^{0\prime}\hat{\beta}_{Logit}\right)\hat{\beta}_{Logit,4} \tag{28}$$

$$= \frac{\exp\left(x_i^{0\prime}\hat{\beta}_{Logit}\right)}{\left(1+\exp\left(x_i^{0\prime}\hat{\beta}_{Logit}\right)\right)^2}\hat{\beta}_{Logit,4} \tag{29}$$

for the Logit model and,

$$\frac{\partial P(y_i = 1 \mid x_i)}{\partial x_{i4}} = \frac{\partial p_i}{\partial x_{i4}} = g\left(x_i^{0\prime}\hat{\beta}_{Probit}\right)\hat{\beta}_{Probit,4} \tag{30}$$

$$= \phi\left(x_i'\hat{\beta}_{Probit}\right)\hat{\beta}_{Probit,4} \tag{31}$$

for the Probit model.[1] Note that the estimated partial effects for the Probit and the Logit models are almost identical while the LPM seems to be biased downwards. Your output should look like the following,

```
Marginal effects
----------------------------------------------------------------
Independent variable           LPM          Probit         Logit
EDUC                        0.0067          0.0076        0.0081
----------------------------------------------------------------
```

The estimated partial effects indicate that when education increases by one year, the probability, for the woman with the given characteristics, to participate in the labour market increases by $1 \cdot 0.0076$ (in the Probit model).

1.6 Question 8.

For discrete variables the partial effects are given as,

$$G(\beta_0 + \beta_1 x_1 + \cdots + \beta_{k-1}x_{k-1} + \beta_k) - G(\beta_0 + \beta_1 x_1 + \cdots + \beta_{k-1}x_{k-1}) \tag{32}$$

[1]Use the command **normpdf** in MATLAB to evaluate expression (**??**).

where $G(x'\beta) = \frac{\exp(x'\beta)}{1+\exp(x'\beta)}$ for the Logit model and $G(x'\beta) = \Phi(x'\beta) = \int_{-\infty}^{x'\beta} \phi(z)\,dz$ and $\phi(z) = \left(\frac{1}{\sqrt{2\pi}}\right)\exp\left(\frac{-z^2}{2}\right)$ for the Probit model.

For the LPM model the partial effects the same as those in the continuous case,

$$\frac{\partial P(y_i = 1 \,|x_i)}{\partial x_{ik}} = \frac{\partial p_i}{\partial x_{ik}} = \frac{\partial x_i'\beta}{\partial x_{ik}'} = \beta_k. \tag{33}$$

In Question 8 we are asked to calculate the partial effects of being a permanent foreign resident on the probability of participating in the labour market for the same woman described in Question 7. Denote the vector of characteristics when the woman is not a foreign resident,

$$\begin{aligned} x_i^{0\prime} = [&CST = 1,\ AGE = 2.5,\ AGE2 = 2.5^2,\ EDUC = 10,\ NYC = 1,\ NOC = 0, \\ &NLINC = 10,\ FOREIGN = 0] \end{aligned}$$

and the vector of characteristics when the woman is a foreign resident,

$$\begin{aligned} x_i^{1\prime} = [&CST = 1,\ AGE = 2.5,\ AGE2 = 2.5^2,\ EDUC = 10,\ NYC = 1,\ NOC = 0, \\ &NLINC = 10,\ FOREIGN = 1] \end{aligned}$$

and the vectors of estimated coefficients for the LPM, the Logit and the Probit models, $\hat{\beta}_{OLS}$, $\hat{\beta}_{Logit}$ and $\hat{\beta}_{Probit}$ respectively. The vectors of estimated coefficients are all $K \times 1$. The partial effects can then be calculated as,

$$\frac{\partial P(y_i = 1 \,|x_i)}{\partial x_{i7}} = \frac{\partial p_i}{\partial x_{i7}} = \hat{\beta}_{OLS,8} \tag{34}$$

for the LPM and,

$$G(\beta_0 + \beta_1 x_1 + \cdots + \beta_{k-1}x_{k-1} + \beta_k) - G(\beta_0 + \beta_1 x_1 + \cdots + \beta_{k-1}x_{k-1}) \tag{35}$$

$$= \frac{\exp\left(x_i^{1\prime}\hat{\beta}_{Logit}\right)}{1+\exp\left(x_i^{1\prime}\hat{\beta}_{Logit}\right)} - \frac{\exp\left(x_i^{0\prime}\hat{\beta}_{Logit}\right)}{1+\exp\left(x_i^{0\prime}\hat{\beta}_{Logit}\right)} \tag{36}$$

for the Logit model and

$$G(\beta_0 + \beta_1 x_1 + \cdots + \beta_{k-1}x_{k-1} + \beta_k) - G(\beta_0 + \beta_1 x_1 + \cdots + \beta_{k-1}x_{k-1}) \tag{37}$$

$$= \Phi\left(x_i^{1\prime}\hat{\beta}_{Logit}\right) - \Phi\left(x_i^{0\prime}\hat{\beta}_{Logit}\right) \tag{38}$$

for the Probit model.[2] Your output should look something like,

```
Marginal effects
----------------------------------------------------------------
Independent variable          LPM          Probit         Logit
FOREIGN                     0.2496         0.2700        0.2726
----------------------------------------------------------------
```

We see that being a permanent foreign resident increases the probability that the woman with the specified characteristics participates in the labour market compared to one with the same characteristics but no foreign residency.

[2] Use the command **normcdf** in MATLAB to evaluate (**??**).

1.7 Question 9.

We will now use the Delta method to compute standard errors for the partial effects. The Delta method can be used to obtain standard errors of a continuous function of other objects which we have standard errors for. In this case the objects are the model parameters and we obtain parameter standard errors using the normal MLE formulas.

In order to calculate the standard errors for the continuous case you need to type equations (8) and (9) from Problem Set 2 into the code. Furthermore, for the discrete case you need to type in equations (10) and (11) from Problem Set 2 into the code. Your output should look like the following,

```
Marginal effects, Probit
-----------------------------------
Variable          dp/dx          s.e.
EDUC             0.0076        0.0071
FOREIGN          0.2700        0.0436
-----------------------------------
```

.