

Assignment 6

**Requirements:**

- a. Estimate the model assuming that the probability function is (a) cumulative normal probability distribution function and (b) logistic probability distribution function. Interpret your estimated result (overall test, individual test, pseudo  $R^2$ , counted  $R^2$ ).

Estimate the model assuming that probability function is

(a) cumulative normal pdf

```
1 . probit y x1 x2 x3 x4, nolog
```

Probit regression	Number of obs	=	400
	LR chi2( 4)	=	201.93
	Prob > chi2	=	0.0000
Log likelihood = -147.46881	Pseudo R2	=	0.4064

y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
x1	.3590739	.0371539	9.66	0.000	.2862536 .4318941
x2	-.8525746	.144481	-5.90	0.000	-1.135752 -.569397
x3	-.5735764	.2202882	-2.60	0.009	-1.005333 -.1418195
x4	-1.248569	.226762	-5.51	0.000	-1.693014 -.8041238
_cons	1.45664	.2037279	7.15	0.000	1.057341 1.85594

According to the P-value of the chi square LR test, which is 0.000. Thus, all the independent variables are overall significant. Moreover, looking at each individual z-test — all the variable are individually significant at 5% level. Pseudo  $R^2$  of this test is 0.4064. This  $R^2$  can't be used for interpretation; it is used only for making comparison. In addition, the counted  $R^2$  for the case of  $y=1$  equals to 91.27%,  $y=0$  equals to 60.80% and overall equals to 81.75%. These number are not very far from each other and they are all more than 50%.

```
2 . qui probit y x1 x2 x3 x4, nolog
```

```
3 . estat clas
```

Probit model for y

Classified	True		Total
	D	~D	
+	251	49	300
-	24	76	100
Total	275	125	400

Classified + if predicted Pr(D) >= .5  
True D defined as y != 0

Sensitivity	Pr( +  D)	91.27%
Specificity	Pr( -  ~D)	60.80%
Positive predictive value	Pr( D  +)	83.67%
Negative predictive value	Pr( ~D  -)	76.00%
False + rate for true ~D	Pr( +  ~D)	39.20%
False - rate for true D	Pr( -  D)	8.73%
False + rate for classified +	Pr( ~D  +)	16.33%
False - rate for classified -	Pr( D  -)	24.00%

Correctly classified 81.75%

Therefore, it is quiet good to use this model for prediction or making forecast.



c. From Probit model, show how to compute Overall LR-test.

From Probit, overall LR test is computed by...

$$\begin{aligned} \text{LR-test} &= 2 [\log L_{UR} - \log L_R] \\ &= 2 [-147.46881 + 248.43455] = 201.93 \end{aligned}$$

This LR statistic gives out the P-value of 0.000. Thus, this model is overall significant.

d. From Logit model, compute predicted value of index value and predicted probability of being bad loan by using mean value of all Xs.

7 . mfx, predict(xb)

Marginal effects after logit  
y = Linear prediction (log odds) (predict, xb)  
= 1.32418

same value

The index value ( $\hat{y}$ ) equals to 1.32418. The index value can also be computed as follows...

$$\begin{aligned} \hat{y} &= 2.5165 + 0.63(0.455) - 1.49(0.809) - 0.96(0.557) - 2.16(-0.12) \\ &\approx 1.32 \end{aligned}$$

$$\hat{p} = \frac{1}{1 + e^{-(1.32)}} \approx 0.789$$

∴ Predicted value of index value is 1.32 and predicted prob. of being bad loan is 0.789. #

4 . logit y x1 x2 x3 x4, nolog

Logistic regression

Number of obs = 400  
LR chi2( 4) = 201.05  
Prob > chi2 = 0.0000  
Pseudo R2 = 0.4046

Log likelihood = -147.90869

y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
x1	.6299401	.0708979	8.89	0.000	.4909828	.7688974
x2	-1.488248	.2597744	-5.73	0.000	-1.997396	-.9790992
x3	-.9562902	.3882611	-2.46	0.014	-1.717268	-.1953124
x4	-2.155321	.4055058	-5.32	0.000	-2.950097	-1.360544
_cons	2.5165	.3714373	6.78	0.000	1.788496	3.244503

same value

6 . mfx

Marginal effects after logit

y = Pr(y) (predict)  
= .7898763

variable	dy/dx	Std. Err.	z	P> z	[ 95% C.I. ]		X
x1	.1045522	.01146	9.12	0.000	.082083	.127022	.454973
x2	-.247007	.04388	-5.63	0.000	-.333011	-.161003	.809344
x3	-.1587171	.06397	-2.48	0.013	-.2841	-.033334	.556712
x4	-.3577223	.06679	-5.36	0.000	-.488633	-.226812	-.119684

e. Compute marginal effect at mean and at median for Logit model.

6 . mfx

Marginal effects after logit  
 y = Pr(y) (predict)  
 = .7898763

variable	dy/dx	Std. Err.	z	P> z	[	95% C.I.	]	X
x1	.1045522	.01146	9.12	0.000	.082083	.127022		.454973
x2	-.247007	.04388	-5.63	0.000	-.333011	-.161003		.809344
x3	-.1587171	.06397	-2.48	0.013	-.2841	-.033334		.556712
x4	-.3577223	.06679	-5.36	0.000	-.488633	-.226812		-.119684

8 . mfx, at (median)

Marginal effects after logit  
 y = Pr(y) (predict)  
 = .84127022

variable	dy/dx	Std. Err.	z	P> z	[	95% C.I.	]	X
x1	.0841188	.00961	8.76	0.000	.065292	.102946		.655749
x2	-.1987326	.03349	-5.93	0.000	-.264373	-.133093		.692745
x3	-.1276979	.04944	-2.58	0.010	-.224597	-.030799		.488768
x4	-.28781	.05616	-5.12	0.000	-.397881	-.177739		-.109732

Marginal effect at mean  $\approx 0.79$

" ————— " median  $\approx 0.841$

f. Compute marginal effect at the value of  $X_1=0.5, X_2=1, X_3=0.5, X_4=0$  for the Probit model.

10 . mfx, at (0.5 1 0.5 0)

Marginal effects after probit  
 y = Pr(y) (predict)  
 = .69034

variable	dy/dx	Std. Err.	z	P> z	[	95% C.I.	]	X
x1	.1266183	.01307	9.69	0.000	.101001	.152235		.5
x2	-.3006389	.05452	-5.51	0.000	-.4075	-.193778		1
x3	-.2022572	.07644	-2.65	0.008	-.352085	-.05243		.5
x4	-.4402764	.08419	-5.23	0.000	-.605293	-.27526		0

Marginal effect  $\approx 0.69$

g. Determine counted  $R^2$  using the threshold of predicted value = 0.5 for Logit models.

```
20 . estat clas
```

Logistic model for y

Classified	True		Total
	D	~D	
+	251	49	300
-	24	76	100
Total	275	125	400

Classified + if predicted  $\Pr(D) \geq .5$   
True D defined as  $y \neq 0$

Sensitivity	$\Pr(+ D)$	91.27%
Specificity	$\Pr(- \sim D)$	60.80%
Positive predictive value	$\Pr(D +)$	83.67%
Negative predictive value	$\Pr(\sim D -)$	76.00%
False + rate for true ~D	$\Pr(+ \sim D)$	39.20%
False - rate for true D	$\Pr(- D)$	8.73%
False + rate for classified +	$\Pr(\sim D +)$	16.33%
False - rate for classified -	$\Pr(D -)$	24.00%
Correctly classified		81.75%

Counted  $R^2$  when  $y=1 \rightarrow 91.27\%$   
 " " "  $y=0 \rightarrow 60.80\%$   
 Overall counted  $R^2 \rightarrow 81.75\%$

h. Determine counted  $R^2$  using the threshold of predicted value = 0.7 for Logit models.

```
21 . g yhat1=0 if pr<=0.7
(241 missing values generated)
```

```
22 . replace yhat1=1 if pr>0.7
(241 real changes made)
```

```
23 . tabulate y yhat1
```

y	$\hat{y}$		Total
	0	1	
0	101	24	125
1	58	217	275
Total	159	241	400

Counted  $R^2$  ;

when  $y=1 : \frac{217}{275} = 78.91\%$

when  $y=0 : \frac{101}{125} = 80.8\%$

Overall :  $\frac{101+217}{400} = 79.5\%$





8 . mfx, at (median)

Marginal effects after logit  
 y = Pr(y) (predict)  
 = .84127022

variable	dy/dx	Std. Err.	z	P> z	[	95% C.I.	]	X
x1	.0841188	.00961	8.76	0.000	.065292	.102946		.655749
x2	-.1987326	.03349	-5.93	0.000	-.264373	-.133093		.692745
x3	-.1276979	.04944	-2.58	0.010	-.224597	-.030799		.488768
x4	-.28781	.05616	-5.12	0.000	-.397881	-.177739		-.109732

9 . qui probit y x1 x2 x3 x4

10 . mfx, at (0.5 1 0.5 0)

Marginal effects after probit  
 y = Pr(y) (predict)  
 = .69034

variable	dy/dx	Std. Err.	z	P> z	[	95% C.I.	]	X
x1	.1266183	.01307	9.69	0.000	.101001	.152235		.5
x2	-.3006389	.05452	-5.51	0.000	-.4075	-.193778		1
x3	-.2022572	.07644	-2.65	0.008	-.352085	-.05243		.5
x4	-.4402764	.08419	-5.23	0.000	-.605293	-.27526		0

11 . ssc install fitstat

checking **fitstat** consistency and verifying not already installed...  
 all files already exist and are up to date.

12 . qui logit y x1 x2 x3 x4

13 . fitstat

Measures of Fit for **logit** of **y**

Log-Lik Intercept Only:	-248.435	Log-Lik Full Model:	-147.909
D(395):	295.817	LR(4):	201.052
		Prob > LR:	0.000
McFadden's R2:	0.405	McFadden's Adj R2:	0.385
Maximum Likelihood R2:	0.395	Cragg & Uhler's R2:	0.555
McKelvey and Zavoina's R2:	0.622	Efron's R2:	0.445
Variance of y*:	8.707	Variance of error:	3.290
Count R2:	0.818	Adj Count R2:	0.416
AIC:	0.765	AIC*n:	305.817
BIC:	-2070.811	BIC':	-177.086

14 . predict pr

(option **pr** assumed; Pr(y))

15 . g uhat=0 if pr<=0.5

(300 missing values generated)

16 . drop uhat

17 . g yhat=0 if pr<=0.5

(300 missing values generated)

18 . replace yhat=1 if pr>0.5  
(300 real changes made)

19 . tabulate y yhat

y	yhat		Total
	0	1	
0	<b>76</b>	<b>49</b>	<b>125</b>
1	<b>24</b>	<b>251</b>	<b>275</b>
Total	<b>100</b>	<b>300</b>	<b>400</b>

20 . estat clas

Logistic model for y

Classified	True		Total
	D	~D	
+	<b>251</b>	<b>49</b>	<b>300</b>
-	<b>24</b>	<b>76</b>	<b>100</b>
Total	<b>275</b>	<b>125</b>	<b>400</b>

Classified + if predicted Pr(D) >= .5  
True D defined as y != 0

Sensitivity	Pr( +   D)	<b>91.27%</b>
Specificity	Pr( -   ~D)	<b>60.80%</b>
Positive predictive value	Pr( D   +)	<b>83.67%</b>
Negative predictive value	Pr( ~D   -)	<b>76.00%</b>
False + rate for true ~D	Pr( +   ~D)	<b>39.20%</b>
False - rate for true D	Pr( -   D)	<b>8.73%</b>
False + rate for classified +	Pr( ~D   +)	<b>16.33%</b>
False - rate for classified -	Pr( D   -)	<b>24.00%</b>
Correctly classified		<b>81.75%</b>

21 . g yhat1=0 if pr<=0.7  
(241 missing values generated)

22 . replace yhat1=1 if pr>0.7  
(241 real changes made)

23 . tabulate y yhat1

y	yhat1		Total
	0	1	
0	<b>101</b>	<b>24</b>	<b>125</b>
1	<b>58</b>	<b>217</b>	<b>275</b>
Total	<b>159</b>	<b>241</b>	<b>400</b>

24 . estat clas

Logistic model for y

Classified	True		Total
	D	~D	
+	<b>251</b>	<b>49</b>	<b>300</b>
-	<b>24</b>	<b>76</b>	<b>100</b>
Total	<b>275</b>	<b>125</b>	<b>400</b>

Classified + if predicted Pr(D) >= .5  
 True D defined as y != 0

Sensitivity	Pr( +  D)	<b>91.27%</b>
Specificity	Pr( - ~D)	<b>60.80%</b>
Positive predictive value	Pr( D  +)	<b>83.67%</b>
Negative predictive value	Pr(~D  -)	<b>76.00%</b>
False + rate for true ~D	Pr( + ~D)	<b>39.20%</b>
False - rate for true D	Pr( -  D)	<b>8.73%</b>
False + rate for classified +	Pr(~D  +)	<b>16.33%</b>
False - rate for classified -	Pr( D  -)	<b>24.00%</b>
Correctly classified		<b>81.75%</b>

25 . fitstat

Measures of Fit for **logit** of **y**

Log-Lik Intercept Only:	<b>-248.435</b>	Log-Lik Full Model:	<b>-147.909</b>
D(395):	<b>295.817</b>	LR(4):	<b>201.052</b>
		Prob > LR:	<b>0.000</b>
McFadden's R2:	<b>0.405</b>	McFadden's Adj R2:	<b>0.385</b>
Maximum Likelihood R2:	<b>0.395</b>	Cragg & Uhler's R2:	<b>0.555</b>
McKelvey and Zavoina's R2:	<b>0.622</b>	Efron's R2:	<b>0.445</b>
Variance of y*:	<b>8.707</b>	Variance of error:	<b>3.290</b>
Count R2:	<b>0.818</b>	Adj Count R2:	<b>0.416</b>
AIC:	<b>0.765</b>	AIC*n:	<b>305.817</b>
BIC:	<b>-2070.811</b>	BIC':	<b>-177.086</b>

26 . log close

name: **<unnamed>**  
 log: **C:\Users\user\Documents\BE TU\Year2\EE426\HW6\6.2.smcl**  
 log type: **smcl**  
 closed on: **27 Feb 2021, 16:17:27**