

Assignment 6

The model

In the study of default probability of the loan, determination factors include:

$$\text{Prob}(Y=1|X) = f(X_1, X_2, X_3, X_4)$$

Dependent variable $Y_i = 1$ if the firm is bad loan, and $= 0$ for good loan.

Independent variables

X_1 is debt coverage ratio.

X_2 is liquidity ratio represented by current assets to current liabilities

X_3 is profitability ratio represented by sales to total assets

X_4 is solidity ratio represented by retained earnings to total assets

From Data `assign6.dta`:

Requirements:

- 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).
- Make comparison of the goodness of fit of the two models.
- From Probit model, show how to compute Overall LR-test.
- From Logit model, compute predicted value of index value and predicted probability of being bad loan by using mean value of all X s.
- Compute marginal effect at mean and at median for Logit model.
- 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.
- Determine counted R^2 using the threshold of predicted value = 0.5 for Logit models.
- Determine counted R^2 using the threshold of predicted value = 0.7 for Logit models.

- 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).

6204641218

(1) probit

overall test: LR $\chi^2(4) = 201.93$, Prob $\chi^2 = 0.0000$

individual test:	z	P-value
X_1	9.66	0.000
X_2	-5.90	0.000
X_3	-2.60	0.000
X_4	-5.57	0.000

Pseudo $R^2 = 0.4064$

counted $R^2 = 0.818$

Adjusted Pseudo $R^2 = 0.386$

Adjusted counted $R^2 = 0.416$

(2) Logit

overall test: LR $\chi^2(4) = 201.05$, Prob $\chi^2 = 0.0000$

individual test:	z	P-value
X_1	8.89	0.000
X_2	-5.73	0.000
X_3	-2.46	0.014
X_4	-5.32	0.000

Pseudo $R^2 = 0.4046$

counted $R^2 = 0.818$

Adjusted Pseudo $R^2 = 0.385$

Adjusted counted $R^2 = 0.416$

(a). Estimated results under both logit and probit probability distribution function are significant in overall test and also for individual test which all independent variables are significant under both cases. Pseudo R-squared and adjusted Pseudo R-squared from both cases are having slightly different values; almost the same value if rounded to a 2 decimal places. Counted R-squared and adjusted counted R-squared are not changed, implying that both models have the same level of predictability.

b. Make comparison of the goodness of fit of the two models.

$$\text{log-likelihood for probit} = -147.469, \quad \text{logit} = -147.909$$

$$\text{From Pseudo } R^2 = 1 - \frac{\ln L_{UR}}{\ln L_R} \quad \text{and Pseudo } \bar{R}^2 = 1 - \frac{\ln L_{UR-k}}{\ln L_R}$$

can be seen that higher log-likelihood value ($\ln L_{UR}$) results in higher R^2 . As we use mle method to estimate, we're trying to maximize the likelihood function, higher R^2 (and adjusted one) represents better goodness of fit. In this case, probit is slightly higher.

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

$$\begin{aligned} \text{LR Test} &= 2[\log_{UR} - \log_R] \sim \chi^2_{(4-1)} \\ &= 2(-147.469 + 248.435) = 201.932 \end{aligned}$$

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

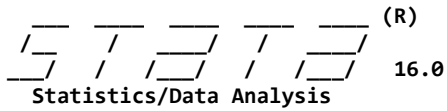
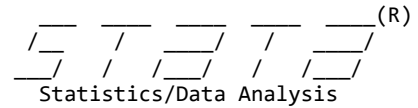
$$\hat{I} = 1.32418 \quad \hat{p} = .7898763 \quad | \quad \text{est.}$$

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

$$\text{counted } R^2 = 0.8175$$

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

$$\text{counted } R^2 = \frac{101 + 217}{400} = 0.795$$



MP - Parallel Edition

16.0 Copyright 1985-2019 StataCorp LLC
 StataCorp
 4905 Lakeway Drive
 College Station, Texas 77845 USA
 800-STATA-PC <http://www.stata.com>
 979-696-4600 stata@stata.com
 979-696-4601 (fax)

20-student 2-core Stata lab perpetual license:

Serial number: 501606222284
 Licensed to: Faculty of economics
 Thammasat University

Notes:

1. Unicode is supported; see [help unicode advice](#).
2. More than 2 billion observations are allowed; see [help obs advice](#).
3. Maximum number of variables is set to 5000; see [help set maxvar](#).

```
1 . use "C:\Users\A\Downloads\assign6.dta"
2 . log using "C:\Users\A\Desktop\426.a6.smcl"
```

```
name: <unnamed>
log: C:\Users\A\Desktop\426.a6.smcl
log type: smcl
opened on: 3 Mar 2021, 18:50:10
```

```
3 . 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

```
4 . fitstat
```

		probit
Log-likelihood		
Model		-147.469
Intercept-only		-248.435
Chi-square		
Deviance(df=395)		294.938
LR(df=4)		201.931
p-value		0.000
R2		
McFadden		0.406
McFadden(adjusted)		0.386
McKelvey & Zavoina		0.640
Cox-Snell/ML		0.396

AIC divided by N	0.765
BIC(df=5)	325.775
<hr/>	
Variance of e	3.290
y-star	8.707

7 . sum x1 x2 x3 x4

Variable	Obs	Mean	Std. Dev.	Min	Max
x1	400	.4549731	3.100887	-8.155268	9.74565
x2	400	.8093444	.6100628	.0004192	2.852366
x3	400	.5567118	.3871706	.0016088	1.874493
x4	400	-.119684	.387141	-1.213392	.9788648

8 . mfx, predict(xb)

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

variable	dy/dx	Std. Err.	z	P> z	[95% C.I.]	X
x1	.6299401	.0709	8.89	0.000	.490983 .768897	.454973
x2	-1.488248	.25977	-5.73	0.000	-1.9974 -.979099	.809344
x3	-.9562902	.38826	-2.46	0.014	-1.71727 -.195312	.556712
x4	-2.155321	.40551	-5.32	0.000	-2.9501 -1.36054	-.119684

9 . predict pr
(option pr assumed; Pr(y))

10 . 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).

11 . 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

(f).

12 . mfx, at(0.5 1 0.5 0)

Marginal effects after logit
 $y = \text{Pr}(y)$ (predict)
 = .70372027

variable	dy/dx	Std. Err.	z	P> z	[95% C.I.]	X
x1	.1313413	.01481	8.87	0.000	.102307	.160376		.5
x2	-.3102967	.0596	-5.21	0.000	-.427116	-.193478		1
x3	-.1993846	.07892	-2.53	0.012	-.354074	-.044695		.5
x4	-.4493802	.09126	-4.92	0.000	-.628243	-.270518		0

13 . g yhat=0 if pr<=0.5
 (300 missing values generated)

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

15 . tabulate yy hat
variable yy not found
r(111);

16 . tabulate y yhat

y	yhat		Total
	0	1	
0	76	49	125
1	24	251	275
Total	100	300	400

17 . estat clas

Logistic model for y

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

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

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

7 . g yhat=0 if pr<=0.7
(241 missing values generated)

8 . replace yhat=1 if pr>0.7
(241 real changes made)

9 . 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 $\sim D$	$\Pr(+ \sim D)$	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%

10 . tabulate y yhat

y	yhat		Total
	0	1	
0	101	24	125
1	58	217	275
Total	159	241	400

(h).