Example Logit Probit 2

The model

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

$$Prob(Y=1|X) = f(X_1, X_2, X_3, X_4)$$

Dependent variable $Yi = 1$ 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

Requirements:

1. 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).
2. Make comparison of the goodness of fit of the two models.
3. From Probit model, show how to compute Overall LR-test.
4. From Logit model, compute predicted value of index value and predicted probability of being bad loan by using mean value of all *Xs*.
5. Compute marginal effect at mean and at median for Logit model.
6. 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.
7. Determine counted R^2 using the threshold of predicted value = 0.5 for Logit models.
8. Determine counted R^2 using the threshold of predicted value = 0.7 for Logit models.

1 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).

```
. probit y x1 x2 x3 x4

Iteration 0:   log likelihood = -265.87901
Iteration 1:   log likelihood = -152.30041
Iteration 2:   log likelihood = -147.75466
Iteration 3:   log likelihood =  -147.7109
Iteration 4:   log likelihood = -147.71089

Probit regression                                 Number of obs   =        450
                                                  LR chi2(4)      =     236.34
                                                  Prob > chi2     =     0.0000
Log likelihood = -147.71089                       Pseudo R2       =     0.4444

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .3636552    .036416     9.99   0.000     .2922811    .4350292
          x2 |   -.862947   .1436001    -6.01   0.000    -1.144398   -.5814958
          x3 |  -.5817432   .2191799    -2.65   0.008    -1.011328   -.1521586
          x4 |  -1.262129   .2253211    -5.60   0.000     -1.70375    -.820508
       _cons |   1.474976   .2012109     7.33   0.000      1.08061    1.869342
------------------------------------------------------------------------------

. fitstat

Measures of Fit for probit of y

Log-Lik Intercept Only:       -265.879   Log-Lik Full Model:           -147.711
D(445):                        295.422   LR(4):                         236.336
                                         Prob > LR:                       0.000
McFadden's R2:                   0.444   McFadden's Adj R2:               0.426
ML (Cox-Snell) R2:               0.409   Cragg-Uhler(Nagelkerke) R2:      0.589
McKelvey & Zavoina's R2:         0.669   Efron's R2:                      0.472
Variance of y*:                  3.021   Variance of error:               1.000
Count R2:                        0.838   Adj Count R2:                    0.416
AIC:                             0.679   AIC*n:                         305.422
BIC:                         -2423.193   BIC':                         -211.899
BIC used by Stata:             325.968   AIC used by Stata:             305.422

. logit y x1 x2 x3 x4

Iteration 0:   log likelihood = -265.87901
Iteration 1:   log likelihood = -159.64911
Iteration 2:   log likelihood = -148.73788
Iteration 3:   log likelihood = -148.40325
Iteration 4:   log likelihood =   -148.403
Iteration 5:   log likelihood =   -148.403

Logistic regression                               Number of obs   =        450
                                                  LR chi2(4)      =     234.95
                                                  Prob > chi2     =     0.0000
Log likelihood =   -148.403                       Pseudo R2       =     0.4418

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .6431189   .0697691     9.22   0.000      .506374    .7798637
          x2 |  -1.518271   .2591056    -5.86   0.000    -2.026108   -1.010433
          x3 |  -.9789811   .3883357    -2.52   0.012    -1.740105   -.2178571
          x4 |  -2.194103   .4045806    -5.42   0.000    -2.987067    -1.40114
       _cons |   2.568447   .3682716     6.97   0.000     1.846648    3.290246
------------------------------------------------------------------------------

. fitstat

Measures of Fit for logit of y

Log-Lik Intercept Only:       -265.879   Log-Lik Full Model:           -148.403
D(445):                        296.806   LR(4):                         234.952
                                         Prob > LR:                       0.000
McFadden's R2:                   0.442   McFadden's Adj R2:               0.423
ML (Cox-Snell) R2:               0.407   Cragg-Uhler(Nagelkerke) R2:      0.587
McKelvey & Zavoina's R2:         0.656   Efron's R2:                      0.471
Variance of y*:                  9.559   Variance of error:               3.290
Count R2:                        0.838   Adj Count R2:                    0.416
AIC:                             0.682   AIC*n:                         306.806
BIC:                         -2421.809   BIC':                         -210.515
BIC used by Stata:             327.352   AIC used by Stata:             306.806
```

2 Make comparison of the goodness of fit of the two models.
According to the above results, it seems that logit model provide a little bit better fitting the data compare to probit model since McFadden's R2 from logit is a little bit higher than from probit (0.444>0.442).

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

From the above result LR-test = 2 ((-147.711) - (-265.879)) = 236.336

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

```
. mfx, predict(xb)

Marginal effects after logit
      y  = Linear prediction (log odds) (predict, xb)
         =  1.7377664
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
      x1 |   .6431189      .06977    9.22   0.000   .506374  .779864   .897658
      x2 |  -1.518271      .25911   -5.86   0.000  -2.02611 -1.01043   .765019
      x3 |  -.9789811      .38834   -2.52   0.012  -1.74011 -.217857   .543069
      x4 |  -2.194103      .40458   -5.42   0.000  -2.98707 -1.40114  -.129975
------------------------------------------------------------------------------
```

From the estimated result,

$$\hat{I} = -2.19 + 0.643(0.898) - 1.518(0.765) - 0.979(0.543) - 2.194(-0.13) = 1.738$$

$$\hat{P} = \frac{1}{1+e^{-(-1.738)}} = 0.8504$$

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

```
. mfx

Marginal effects after logit
      y  = Pr(y) (predict)
         =  .85040313
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
      x1 |   .0818161        .009    9.09   0.000   .064172  .099461   .897658
      x2 |  -.1931508      .03431   -5.63   0.000  -.260389 -.125912   .765019
      x3 |  -.1245437      .04894   -2.54   0.011   -.22046 -.028627   .543069
      x4 |  -.2791286      .05252   -5.31   0.000  -.382062 -.176195  -.129975
------------------------------------------------------------------------------

. mfx, at(median)

Marginal effects after logit
      y  = Pr(y) (predict)
         =  .89138055
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
      x1 |   .0622676       .0079    7.88   0.000   .046777  .077758   1.04907
      x2 |  -.1470009      .02571   -5.72   0.000    -.1974 -.096602   .649129
      x3 |  -.0947862      .03605   -2.63   0.009  -.165443 -.024129   .482146
      x4 |  -.2124358      .04212   -5.04   0.000  -.294989 -.129883  -.145555
------------------------------------------------------------------------------
```

6 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.

```
. mfx, at(0.5 1 0.5 0)

Marginal effects after probit
      y  = Pr(y) (predict)
         =  .69251249
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
      x1 |   .1278389      .01294    9.88   0.000   .102481  .153196        .5
      x2 |  -.3033593      .05437   -5.58   0.000  -.409922 -.196796         1
      x3 |  -.2045053      .07588   -2.70   0.007  -.353228 -.055783        .5
      x4 |  -.4436873      .08375   -5.30   0.000   -.60783 -.279544         0
------------------------------------------------------------------------------
```

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

```
. estat clas

Logistic model for y

              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |       301            49  |        350
     -     |        24            76  |        100
-----------+--------------------------+-----------
   Total   |       325           125  |        450

Classified + if predicted Pr(D) >= .5
True D defined as y != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   92.62%
Specificity                     Pr( -|~D)   60.80%
Positive predictive value       Pr( D| +)   86.00%
Negative predictive value       Pr(~D| -)   76.00%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   39.20%
False - rate for true D         Pr( -| D)    7.38%
False + rate for classified +   Pr(~D| +)   14.00%
False - rate for classified -   Pr( D| -)   24.00%
--------------------------------------------------
Correctly classified                        83.78%
--------------------------------------------------
```

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

```
. estat clas, cut(0.7)

Logistic model for y

              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |       267            24  |        291
     -     |        58           101  |        159
-----------+--------------------------+-----------
   Total   |       325           125  |        450

Classified + if predicted Pr(D) >= .7
True D defined as y != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   82.15%
Specificity                     Pr( -|~D)   80.80%
Positive predictive value       Pr( D| +)   91.75%
Negative predictive value       Pr(~D| -)   63.52%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   19.20%
False - rate for true D         Pr( -| D)   17.85%
False + rate for classified +   Pr(~D| +)    8.25%
False - rate for classified -   Pr( D| -)   36.48%
--------------------------------------------------
Correctly classified                        81.78%
--------------------------------------------------
```