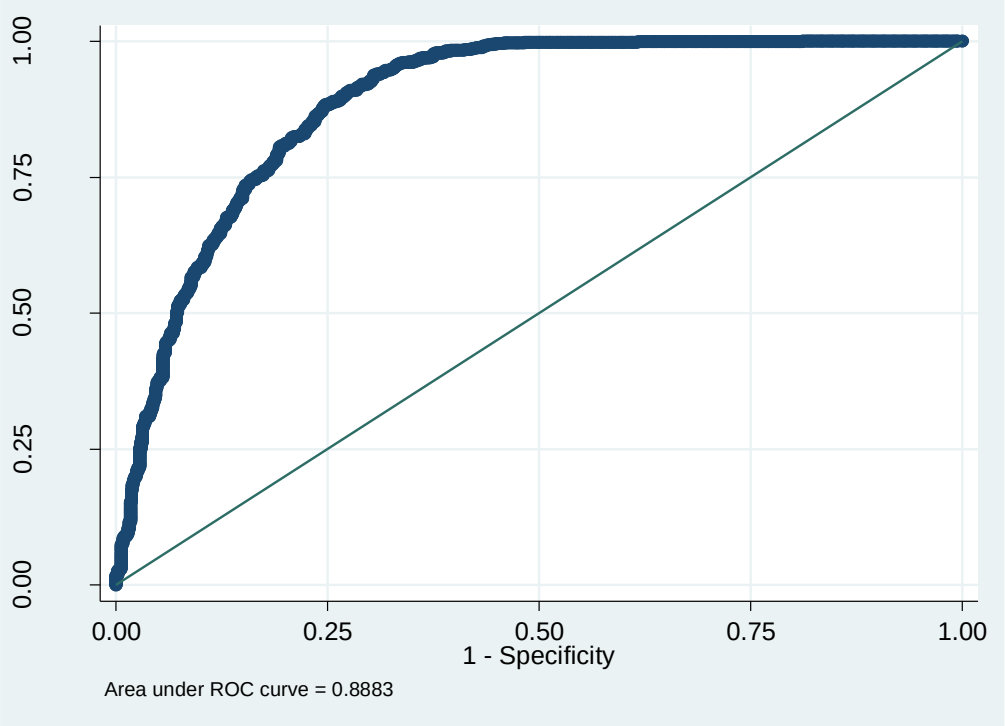
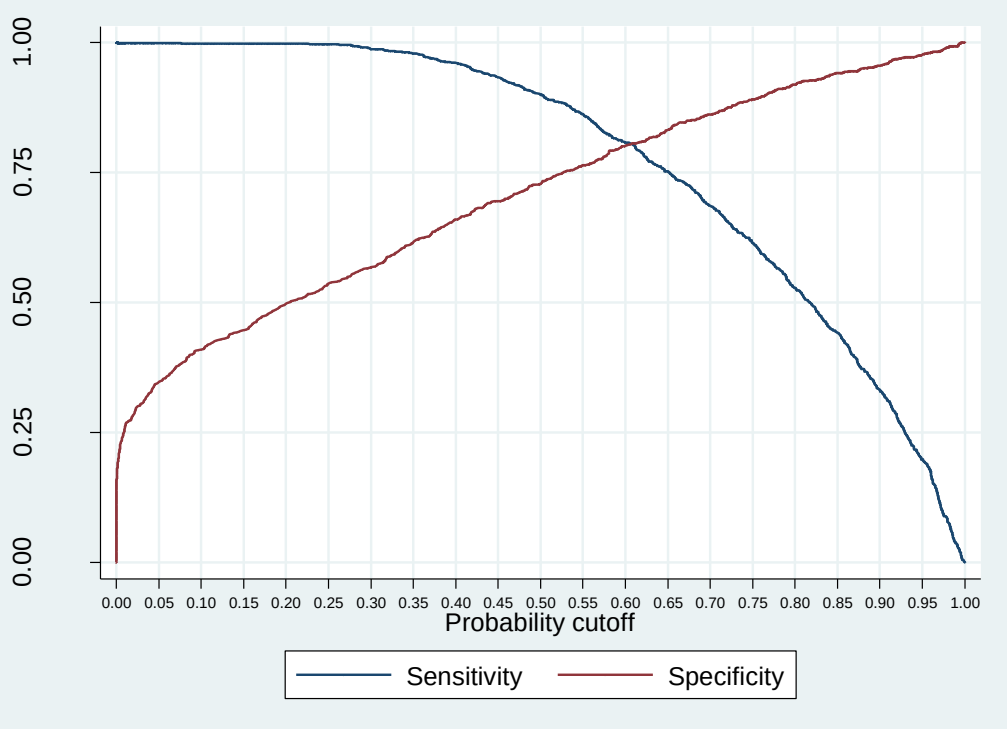
```
. lroc

Logistic model for y

number of observations =      1888
area under ROC curve   =    0.8883
```

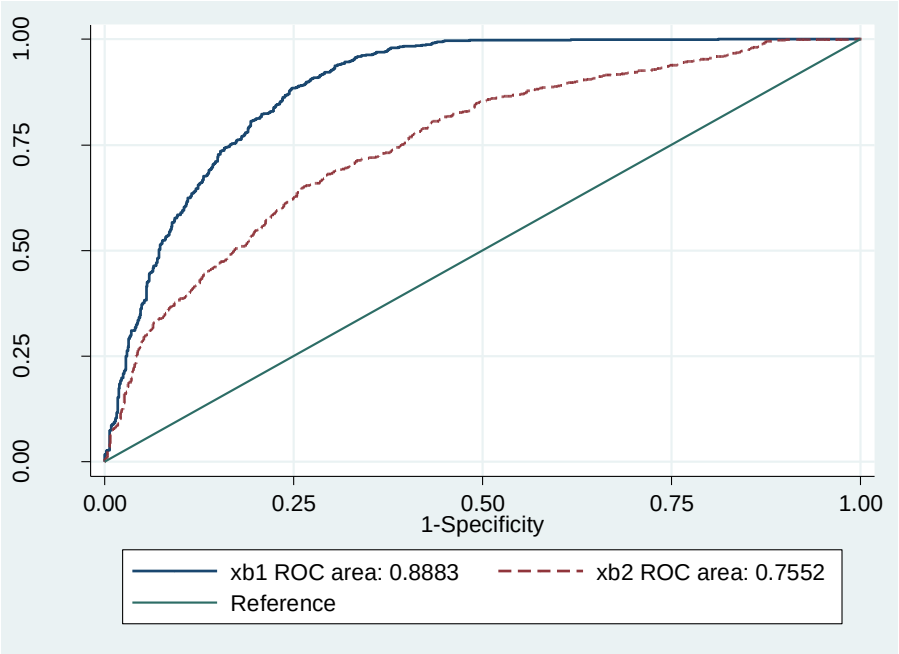



```
. predict xb1, xb
(2 missing values generated)

. lsens
```



```
. estat clas, cut(0.606)

Logistic model for y

              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |       862           159  |       1021
     -     |       208           659  |        867
-----------+--------------------------+-----------
   Total   |      1070           818  |       1888

Classified + if predicted Pr(D) >= .606
True D defined as y != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   80.56%
Specificity                     Pr( -|~D)   80.56%
Positive predictive value       Pr( D| +)   84.43%
Negative predictive value       Pr(~D| -)   76.01%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   19.44%
False - rate for true D         Pr( -| D)   19.44%
False + rate for classified +   Pr(~D| +)   15.57%
False - rate for classified -   Pr( D| -)   23.99%
--------------------------------------------------
Correctly classified                        80.56%
--------------------------------------------------

. logit y x2 x3, nolog

Logistic regression                               Number of obs   =       1888
                                                  LR chi2(2)      =     391.38
                                                  Prob > chi2     =     0.0000
Log likelihood =  -1096.105                       Pseudo R2       =     0.1515

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x2 |   2.857145   .2342892    12.19   0.000     2.397946    3.316343
          x3 |   9.131461   .8380706    10.90   0.000     7.488873    10.77405
       _cons |  -1.472011   .1214552   -12.12   0.000    -1.710059   -1.233963
------------------------------------------------------------------------------

. fitstat, using(m1)

Measures of Fit for logit of y
                               Current             Saved        Difference
Model:                           logit             logit
N:                                1888              1888                 0
Log-Lik Intercept Only       -1291.794         -1291.794             0.000
Log-Lik Full Model           -1096.105          -757.499          -338.606
D                             2192.210(1885)    1514.999(1882)     677.211(3)
LR                             391.378(2)       1068.589(5)        677.211(3)
Prob > LR                        0.000             0.000             0.000
McFadden's R2                    0.151             0.414            -0.262
McFadden's Adj R2                0.149             0.409            -0.260
ML (Cox-Snell) R2                0.187             0.432            -0.245
Cragg-Uhler(Nagelkerke) R2       0.251             0.580            -0.329
McKelvey & Zavoina's R2          0.281             0.999            -0.718
Efron's R2                       0.194             0.490            -0.296
Variance of y*                   4.573          3675.508         -3670.934
Variance of error                3.290             3.290             0.000
Count R2                         0.697             0.827            -0.130
Adj Count R2                     0.301             0.600            -0.300
AIC                              1.164             0.809             0.356
AIC*n                         2198.210          1526.999           671.211
BIC                         -12026.860        -12681.442           654.581
BIC'                           -376.291          -1030.872           654.581
BIC used by Stata              2214.840          1560.259           654.581
AIC used by Stata              2198.210          1526.999           671.211

Difference of  654.581 in BIC' provides very strong support for saved model.
Note: p-value for difference in LR is only valid if models are nested.

. lroc
Logistic model for y
number of observations =     1888
```

```
area under ROC curve   =   0.7552

. predict xb2, xb
(2 missing values generated)

. roccomp y xb1 xb2, graph summary

                                ROC                    -Asymptotic Normal--
                    Obs       Area     Std. Err.      [95% Conf. Interval]
-------------------------------------------------------------------------
xb1                1888     0.8883       0.0080        0.87269     0.90399
xb2                1888     0.7552       0.0111        0.73350     0.77695
-------------------------------------------------------------------------
Ho: area(xb1) = area(xb2)
    chi2(1) =   153.61       Prob>chi2 =   0.0000
```



Probit Model

```
. probit y x1 x2 x3 x4 x5, nolog

Probit regression                                 Number of obs   =       1888
                                                  LR chi2(5)      =    1042.01
                                                  Prob > chi2     =     0.0000
Log likelihood = -770.78852                       Pseudo R2       =     0.4033
------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   2.991683   .1628704    18.37   0.000     2.672462    3.310903
          x2 |   .6882851   .1742238     3.95   0.000     .3468128    1.029757
          x3 |   2.600306   .5990625     4.34   0.000     1.426165    3.774446
          x4 |   .1802702   .0649368     2.78   0.006     .0529964     .307544
          x5 |   .4506521   .1374255     3.28   0.001     .1813031    .7200011
       _cons |  -.9700784   .1261652    -7.69   0.000    -1.217358   -.7227991
------------------------------------------------------------------------------
Note: 95 failures and 0 successes completely determined.

. mfx
Marginal effects after probit
      y  = Pr(y) (predict)
         =   .0151667
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
      x1 |   .1143775      .02894    3.95   0.000   .057651  .171104  -.638752
      x2 |   .0263144      .01081    2.43   0.015   .005118  .047511   .468602
      x3 |   .0994144      .04073    2.44   0.015   .019592  .179236   .050715
      x4 |   .0068921      .00313    2.21   0.027   .000766  .013018   .140109
      x5 |   .0172292      .00746    2.31   0.021   .002605  .031854   .522908
------------------------------------------------------------------------------
```

```
. fitstat

Measures of Fit for probit of y
Log-Lik Intercept Only:      -1291.794   Log-Lik Full Model:           -770.789
D(1882):                      1541.577   LR(5):                       1042.011
                                         Prob > LR:                      0.000
McFadden's R2:                   0.403   McFadden's Adj R2:              0.399
ML (Cox-Snell) R2:               0.424   Cragg-Uhler(Nagelkerke) R2:     0.569
McKelvey & Zavoina's R2:         0.999   Efron's R2:                     0.483
Variance of y*:               1040.916   Variance of error:              1.000
Count R2:                        0.826   Adj Count R2:                   0.598
AIC:                             0.823   AIC*n:                       1553.577
BIC:                        -12654.863   BIC':                       -1004.294
BIC used by Stata:            1586.837   AIC used by Stata:           1553.577

. lstat

Probit model for y
             -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |       963           222  |       1185
     -     |       107           596  |        703
-----------+--------------------------+-----------
   Total   |      1070           818  |       1888

Classified + if predicted Pr(D) >= .5
True D defined as y != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   90.00%
Specificity                     Pr( -|~D)   72.86%
Positive predictive value       Pr( D| +)   81.27%
Negative predictive value       Pr(~D| -)   84.78%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   27.14%
False - rate for true D         Pr( -| D)   10.00%
False + rate for classified +   Pr(~D| +)   18.73%
False - rate for classified -   Pr( D| -)   15.22%
--------------------------------------------------
Correctly classified                        82.57%
--------------------------------------------------
```