(Revised, 2007)

Part 4

Advanced Cross-Sectional Models

Lecture 9

Binary Choice Models

Read Wooldridge ch 15
Green ch 21
Verbeek ch 7.1

Also, panel choice models

Discrete Choice Models

Dependent variable is a dummy variable.

Eg) Mortgage rate choice $\begin{cases} 1 = \text{fixed rate} \\ 0 = \text{adj. rate} \end{cases}$

Eg) Political Vote $\begin{cases} 1 = \text{Republican} \\ 0 = \text{Democratic} \end{cases}$

Choice Models try to explain determinants of choice.

Eg) $Y_i = \alpha + \beta_1 \text{Income}_i + \beta_2 \text{Edu}_i + \beta_3 \text{Race}_i + \cdots + e_i$

Y_i: $\begin{cases} 1 = \text{Rep} \\ 0 = \text{Demo} \end{cases}$

How do we estimate?

1) Can use OLS

i) $\hat{y}_i = \hat{p}_i = \widehat{P(y_i = 1)}$ "probability"

thus called linear prob. model (LPM).

ii) Marginal effect

$$\hat{\beta}_j = \frac{\Delta \hat{y}}{\Delta X_j} = \frac{\Delta \hat{p}}{\Delta X_j}$$ change (or difference) in prob. induced by 1 unit change in X.

iii) usual tests (t-test, F-test, Wald test) are problematic due to heteroskedasticity $\Rightarrow$ use robust variance for these tests!

problems ① Predicted prob $\hat{P}_i$ can be >1 or <0.

② Heteroskedasticity. $Var(e_i) = p_i(1-p_i) \neq$ constant

can do GLS, though.

$\boxed{\begin{array}{l} Var(y|x) \\ = x\beta(1-x\beta) \\ \therefore \text{depends on } X\text{'s} \end{array}}$

Note these problems may not be serious.

Note Advantages of the CPM

① Good approximation for common values of the covariates
(Problem occurs only at extreme cases; $\hat{p}_i > 1$ or < 0)
- Good estimates of constant partial effects near the center of x.

② Can use robust std error or WLS (weighted LS : GLS) to correct for heteroskedasticity

WLS

$y_i/\hat{\sigma}_i$ on $x_i/\hat{\sigma}_i$
where $\hat{\sigma}_i = \sqrt{\hat{y}_i(1-\hat{y}_i)}$

Then usual tests are valid.

Do this transformation even on the constant and dummy variables!

③ More useful if most X's take on only a few values or saturated.
; No reason to worry about the predicted probability lying outside the bound.

④ Easier to use in the panel data models with the LPM.
cf)
- Dynamic panel choice model
(no clear solutions available for the FE or RE logit or probit models)
- (Nonlinear) Panel choice models often hard to estimate.

To guarantee $0 \le \hat{p}_i \le 1$, we consider two different transformations using cdf of

i) std. normal dist $\rightarrow$ probit model

ii) logistic dist $\rightarrow$ logit model

2) LOGIT Model

Note other distributions can be used as well.

$$p_i = G(\beta' X_i) + e_i$$

where $\beta' X_i = \alpha + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}$

G is the cdf of the logistic dist.

s.t. $G(\beta' X_i) = \boxed{\dfrac{1}{1 + e^{-\beta' X_i}} \overset{\text{let}}{=} p_i}$

Rewriting gives:

$$\log \frac{p_i}{1-p_i} = \beta' X_i + e_i \quad \text{or} \quad \text{logit } p_i = \beta' X_i + e_i$$

log of "odds"

"$\log \frac{p_i}{1-p_i}$"

Estimation : MLE

$$L = \prod_{i=1}^{n} p_i^{y_i} (1-p_i)^{1-y_i}$$

$$\log_\beta L = \sum_{i=1}^{n} (y_i \log p_i + (1-y_i) \log(1-p_i)]$$

where $p_i = \dfrac{1}{1+e^{-\beta' X_i}}$

Interpretation

i) sign $(\hat{\beta}_j)$: direction of the effect of X on P

ii) Marginal effect (Note it's not $\hat{\beta}_j$)

$$\frac{d\hat{P}_i}{dX_{ij}} = g(\hat{\beta}'X_i)\cdot\hat{\beta}_j \qquad \text{of course } \underline{\text{not}} \text{ constant.}$$

Report at "means" of X's

If X^2 is used, use $(\bar{X})^2$, rather than $(\overline{X^2})$.

.. depends on ① the coef. of X_j (ie $\hat{\beta}_j$) as well as ② all other coefficients and ③ all (other) indep. var. ($\hat{\beta}'X_i$)

where g is the pdf of G st. $g = G'$

$$g(\hat{\beta}'X_i) = \frac{e^{-\hat{\beta}'X_i}}{(1+e^{-\hat{\beta}'X_i})^2}$$

which on to report? $\hat{\beta}$ or marginal effects?

iii) prediction

$$\hat{P}_i = \frac{1}{1+e^{-\hat{\beta}'X_i}}$$

iv) Testing hypothesis

R^2 is hardly defined → thus No F-test

log L is well defined → thus LR test is used.

Note t-test can be used on each coef. (Fine.)

Note Standard error of marginal effect can be computed, by the delta method.
.. "partial derivatives of probabilities" at means of indep. variables.

Distinguish!

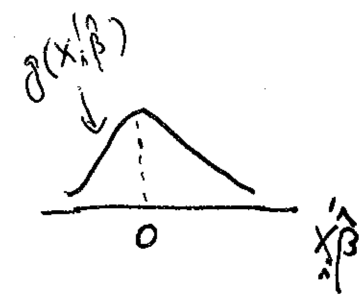
$\hat{\beta} \rightarrow s(\hat{\beta})$

$\frac{d\hat{P}_i}{dX_{ij}} \rightarrow s\left(\frac{d\hat{P}_i}{dX_{ij}}\right)$

(marginal effect) $(g(\hat{\beta}'X_i)\hat{\beta}_j)$ delta method (next page)

(More on marginal effect)

i) $\dfrac{d\hat{P}_i}{dX_{ij}} = g(\hat{\beta}' X_i)\cdot\hat{\beta}_j$



: Maximized when $X_i'\hat{\beta} = 0$

and diminishes as $X_i'\hat{\beta}$ increases (or decreases)

Eg)
$P = G(c + \beta_1 X_1 + \beta_2 X_2^2 + \beta_3 \lg X_3)$
$dP/dX_1 = \beta_1 \cdot g(X_i'\beta)$
$dP/dX_2 = (2\beta_2 X_2)\cdot g(X_i'\beta)$
$dP/d\log X_3 = \beta_3 \cdot g(X_i'\beta)$
= increase in P when X_3 increases by 100%

ii) In the logit models, the relative effects do not depend on Xs.

$$\left(\frac{d\hat{P}}{dX_j}\right) \Big/ \left(\frac{d\hat{P}}{dX_k}\right) = \hat{\beta}_j / \hat{\beta}_k$$

iii) If X_K is binary (dummy or discrete), $\hat{\beta}_K$ is NEVER a partial effect. Instead,

partial effect is to be evaluated as

$$= G(\underbrace{X_{K-1}\hat{\beta}_{K-1} + \hat{\beta}_K \cdot 1}_{(X_K=1)}) - G(\underbrace{X_{K-1}\hat{\beta}_{K-1} + \underbrace{\hat{\beta}_K \cdot 0}_{0}}_{(X_K=0)})$$

thus it also depends on all other $\hat{\beta}$'s

: Actually, this method is general, and can be applied to cases of continuous X's

: This is obvious, but it might be one common mistake, when X_j is discrete.

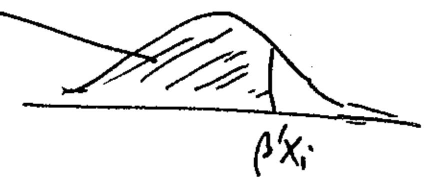
* iv) Std. error of $\dfrac{d\hat{P}_i}{dX_{ij}}$ can be obtained by the delta method; $var\left(\dfrac{d\hat{P}}{dX_i}\right) = \dfrac{dg(\cdot)}{d\beta'}\cdot var(\hat{\beta})\cdot\dfrac{dg(\cdot)}{d\hat{\beta}}$

3) PROBIT Model

Let

$$p_i = \Phi(\beta' X_i)$$

where $\Phi(\beta' X_i)$ is the cdf of the std normal dist.



then

$$\Phi^{-1}(p_i) = \beta' X_i \qquad \text{i.e.}$$

$$= \alpha + \beta' X_i + e_i$$

what is this: Inverse function of $\Phi(\cdot)$
there is no closed form solution,
thus hard to express it
(People use Table, instead).

Interpretation

i) sign of $\hat{\beta}_j$

ii) Marginal effect

$$\frac{d\hat{p}_i}{dx_{ij}} = \phi(\hat{\beta}' X_i)\,\hat{\beta}_j$$

where $\phi(\hat{\beta}' X_i)$ is the pdf. of the std. normal dist

$$\phi(\hat{\beta}' X_i) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(\hat{\beta}' X_i)^2\right]$$

iii) prediction

$\hat{p}_i = \Phi(\hat{\beta}' X_i)$ using the table of the std. normal dist.

iv) Testing hypothesis

No F-test. Use LR test.

t-test can be used for each coefficient.

(same as logit models)

(Exercise)

<u>Examples</u> labor participation of women.

see: choice-binary.log output.

use each of probit & logit model

i) Interpret the estimated coefficient.

(sign)

ii) Marginal effect.

what is the effect of 2 additional years of education?

a) at means

b) of the individual who is a college graduate 25 years old female.

iii) Predicted probability

of a women who is a college graduate and 25 years old.

iv) Test on significance of each coeff.

v) Test on the joint significance of AGE & GENDER.

<u>Solution</u> Here, 25 yrs old, 16 yrs of ed.

Gender $= 0$

First, <u>Logit</u>

i) <u>pred</u> $\hat{P}_i = \dfrac{1}{1+\exp(-\hat{\beta}' X_i)}$

where $\hat{\beta}' X_i = -5.266 - .0098\,Age + .674\,Edu - 2.608\,Gender$

$= -5.266 - .0098(25) + .674(16) - 2.608(0)$

$= 4.599$

$$\hat{P}_i = \frac{1}{1+\exp(-4.599)} = .990$$

ii) Two more yrs of edu.

$$= 2 \cdot \frac{\Delta \hat{P}_i}{\Delta Edu} = 2 \cdot g(\underbrace{\hat{\beta}' X_i}_{4.599}) \cdot \underbrace{\hat{\beta}_j}_{.674}$$

where $g(\hat{\beta}' X_i) = \dfrac{e^{-4.599}}{(1+e^{-4.599})^2} = .0098$

$= 2 \times (.0098) \times .674 = .013$

.. hard to improve prob., as the prob is already close to 1.0

Second, probit

i) Prediction

$$\hat{\beta}' X_i = -3.048 - .00613\,Age + .3869\,Edu - 1.454\,Gender$$

$$= -3.048 - .00613(25) + .3869(16) - 1.454(0)$$

$$= 2.603$$

$\hat{P}_x = .9953$ using z-table

ii) Two more yrs of edu

$$= 2 \cdot \frac{\Delta \hat{P}_i}{\Delta Edu} = 2 \cdot \phi(\underbrace{\hat{\beta}' X_i}_{2.603}) \cdot \underbrace{\hat{\beta}_j}_{.3869}$$

where $\phi(2.603) = \frac{1}{\sqrt{2\pi}} \exp[-\frac{1}{2}(2.603)^2]$

$$= .0054$$

$$= 2 \times (.0054) \times .3869 = .00415$$

Exercise, Repeat for a 34 yrs old, 14 yrs of educated man.

Ans) $\hat{P}$ = .7733 or .7602 (logit or probit)

2xmarg.effect = .2363 or .24 (")

```
log using choice_binary_revised.log, replace
use choice_binary.dta
set more off

list  job age school gender

regress  job age school gender
predict p_lpm, xb

regress  job age school gender, robust

logit  job age school gender
predict p_logit, p
mfx compute
*dlogit2  job age school gender

logistic  job age school gender
predict p_logistic, p
mfx compute

probit  job age school gender
predict p_probit, p
mfx compute
*dprobit  job age school gender

list  p_lpm p_probit p_logit p_logistic

* WLS

gen sigi = sqrt(p_lpm*(1-p_lpm))
gen job1 = job/sigi
gen age1 = age/sigi
gen school1 = school/sigi
gen gender1 = gender/sigi
gen sigi_inv = 1/sigi

regress  job1 age1 school1 gender1 sigi_inv, noconst

regress  job age school gender [aw = 1/sigi]
wls0 job age school gender, wvar(sigi) type(abse) noconst graph
wls0 job age school gender, wvar(sigi) type(e2) noconst graph


/* WLS: example of Greene chapter 12 */

use http://www.ats.ucla.edu/stat/stata/examples/greene/TBL5-1, clear

rename x1 age
rename x2 income
rename x3 exp
rename x4 ownrent
rename x5 selfemp

generate incomesq = income^2
drop if exp==0
save chapter12, replace

/* summary check */

gen age1     = age  / income^0.5
gen income1  = income / income^0.5
gen exp1     = exp / income^0.5
gen ownrent1 = ownrent / income^0.5
gen selfemp1 = selfemp / income^0.5
```

```
gen incomesq1 = incomesq / income^0.5
gen const1 = 1 / income^0.5

wls0 exp age ownrent income incomesq , wvar(income) type(abse) noconst
/* 12.3a */
regress exp age ownrent income incomesq [aw = 1/income]
regress exp1 age1 ownrent1 income1 incomesq1 const1, noconstant
regress exp1 age1 ownrent1 income1 incomesq1

/* end of checkup */

log close
```

12

```
-----------------------------------------------------------------------------
      log:  choice_binary_revised.log
 log type:  text
opened on:  26 Sep 2007, 01:21:59

. use choice_binary.dta

. set more off

.
. list  job age school gender

     +-------------------------------+
     | job   age   school   gender |
     |-------------------------------|
  1. |   1    31       16        0 |
  2. |   1    34       14        1 |
  3. |   1    41       16        1 |
  4. |   0    67        9        0 |
  5. |   1    25       12        0 |
     |-------------------------------|
  6. |   0    58       12        1 |
  7. |   1    45       14        0 |
  8. |   1    55       10        0 |
  9. |   0    43       12        0 |
 10. |   1    55        8        0 |
     |-------------------------------|
 11. |   1    25       11        0 |
 12. |   1    41       14        0 |
 13. |   0    62       12        1 |
 14. |   1    51       13        1 |
 15. |   0    39        9        1 |
     |-------------------------------|
 16. |   1    35       10        0 |
 17. |   1    40       14        1 |
 18. |   0    43       10        1 |
 19. |   0    37       12        1 |
 20. |   1    27       13        0 |
     |-------------------------------|
 21. |   1    28       14        0 |
 22. |   1    48       12        1 |
 23. |   0    66        7        1 |
 24. |   0    44       11        1 |
 25. |   0    21       12        1 |
     |-------------------------------|
 26. |   1    40       10        1 |
 27. |   1    41       15        0 |
 28. |   0    23       10        1 |
 29. |   0    31       11        1 |
 30. |   1    44       12        1 |
     +-------------------------------+

.
. regress  job age school gender

      Source |       SS       df       MS              Number of obs =      30
-------------+------------------------------           F(  3,    26) =    4.96
       Model |  2.62055889     3  .87351963            Prob > F      =  0.0075
    Residual |  4.57944111    26  .17613235            R-squared     =  0.3640
-------------+------------------------------           Adj R-squared =  0.2906
       Total |         7.2    29  .248275862           Root MSE      =  .41968

------------------------------------------------------------------------------
         job |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
```

13

```
-------------+----------------------------------------------------------------
         age |  -.0009682   .0066986    -0.14   0.886  -.0147373     .012801
      school |   .0911182   .0375996     2.42   0.023   .0138312    .1684052
      gender |  -.3803107   .1562374    -2.43   0.022  -.7014612   -.0591602
       _cons |  -.2227047   .6153338    -0.36   0.720  -1.487541    1.042132
------------------------------------------------------------------------------
-

. predict p_lpm, xb

.
. regress  job age school gender, robust

Regression with robust standard errors                 Number of obs =      30
                                                       F(  3,    26) =   10.49
                                                       Prob > F      =  0.0001
                                                       R-squared     =  0.3640
                                                       Root MSE      =  .41968

------------------------------------------------------------------------------
             |               Robust
         job |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0009682   .0064178    -0.15   0.881  -.0141601    .0122237
      school |   .0911182    .030241     3.01   0.006    .028957    .1532795
      gender |  -.3803107   .1616239    -2.35   0.026  -.7125334   -.0480881
       _cons |  -.2227047   .5294223    -0.42   0.677  -1.310948    .8655384
------------------------------------------------------------------------------

.
. logit  job age school gender

Iteration 0:   log likelihood =  -20.19035
Iteration 1:   log likelihood =  -14.08727
Iteration 2:   log likelihood = -13.333675
Iteration 3:   log likelihood = -13.246511
Iteration 4:   log likelihood = -13.244612
Iteration 5:   log likelihood = -13.244611

Logit estimates                                   Number of obs   =         30
                                                  LR chi2(3)      =      13.89
                                                  Prob > chi2     =     0.0031
Log likelihood = -13.244611                       Pseudo R2       =     0.3440

------------------------------------------------------------------------------
         job |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0098882   .0389641    -0.25   0.800  -.0862564      .06648
      school |   .6741644   .3215752     2.10   0.036   .0438886     1.30444
      gender |  -2.608477   1.199745    -2.17   0.030  -4.959935   -.2570198
       _cons |  -5.266331   4.120738    -1.28   0.201  -13.34283    2.810168
------------------------------------------------------------------------------

. predict p_logit, p

. mfx compute

Marginal effects after logit
      y  = Pr(job) (predict)
         =  .69517648
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|          X
---------+--------------------------------------------------------------------
     age |  -.0020954      .00821   -0.26   0.799     41.3333
```

14

```
  school |   .1428596        .06298     2.27    0.023       11.8333
  gender*|  -.4849757        .16352    -2.97    0.003       .566667
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1

. *dlogit2  job age school gender
.
. logistic  job age school gender

Logistic regression                               Number of obs   =         30
                                                  LR chi2(3)      =      13.89
                                                  Prob > chi2     =     0.0031
Log likelihood = -13.244611                       Pseudo R2       =     0.3440

------------------------------------------------------------------------------
         job | Odds Ratio   Std. Err.      z    P>|z|  [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .9901606   .0385807    -0.25   0.800   .917359    1.06874
      school |   1.962392   .6310567     2.10   0.036  1.044866   3.685625
      gender |   .0736466   .0883572    -2.17   0.030  .0070134   .7733529
------------------------------------------------------------------------------

. predict p_logistic, p

. mfx compute

Marginal effects after logistic
      y  = Pr(job) (predict)
         =  .69517648
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|        X
---------+--------------------------------------------------------------------
     age |  -.0020954      .00821   -0.26   0.799     41.3333
  school |   .1428596      .06298    2.27   0.023     11.8333
  gender*|  -.4849757     1.19975   -0.40   0.686     .566667
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1

.
. probit  job age school gender

Iteration 0:   log likelihood =  -20.19035
Iteration 1:   log likelihood = -13.937494
Iteration 2:   log likelihood = -13.305501
Iteration 3:   log likelihood = -13.267639
Iteration 4:   log likelihood =  -13.26742

Probit estimates                                  Number of obs   =         30
                                                  LR chi2(3)      =      13.85
                                                  Prob > chi2     =     0.0031
Log likelihood =  -13.26742                       Pseudo R2       =     0.3429

------------------------------------------------------------------------------
         job |      Coef.   Std. Err.      z    P>|z|  [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0061331   .0237865    -0.26   0.797 -.0527538    .0404876
      school |   .3869652   .1759277     2.20   0.028  .0421532    .7317773
      gender |   -1.45382   .6299747    -2.31   0.021 -2.688547   -.2190919
       _cons |  -3.047916   2.399211    -1.27   0.204 -7.750285    1.654452
------------------------------------------------------------------------------

. predict p_probit, p

. mfx compute
```

14-1

```
Marginal effects after probit
      y  = Pr(job) (predict)
         =  .67502799
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|         X
---------+--------------------------------------------------------------------
     age |  -.0022073       .00854   -0.26   0.796    41.3333
  school |   .1392695       .06012    2.32   0.021    11.8333
 gender*|  -.4692287       .16282   -2.88   0.004    .566667
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1

. *dprobit  job age school gender

.
. list  p_lpm p_probit p_logit p_logistic

     +-------------------------------------------+
     |    p_lpm   p_probit     p_logit   p_logi~c |
     |-------------------------------------------|
  1. | 1.205173   .9984285    .9945883   .9945883 |
  2. | .6397215   .7602951    .7733448   .7733448 |
  3. | .8151807   .9248186    .9245898   .9245898 |
  4. |  .532491   .5095153    .5346048   .5346048 |
  5. | .8465095   .9253966    .9293296   .9293296 |
     |-------------------------------------------|
  6. | .4342487   .4153232    .4113591   .4113591 |
  7. | 1.009382   .9818525    .9765006   .9765006 |
  8. | .6352275   .6859546    .7173721   .7173721 |
  9. | .8290821   .9085606    .9167091   .9167091 |
 10. |  .452991   .3860938    .3972671   .3972671 |
     |-------------------------------------------|
 11. | .7553912   .8543728    .8701486   .8701486 |
 12. | 1.013255   .9829184    .9773914   .9773914 |
 13. | .4303759   .4057835    .4018164   .4018164 |
 14. | .5321442   .5855156    .5950862   .5950862 |
 15. | .1792894   .1041526    .1003838   .1003838 |
     |-------------------------------------------|
 16. | .6545911   .7281004    .7556962   .7556962 |
 17. | .6339124   .7487157    .7627772   .7627772 |
 18. |  .266535    .185178    .1738827   .1738827 |
 19. | .4545805   .4660999    .4623967   .4623967 |
 20. | .9356914   .9653944     .961978    .961978 |
     |-------------------------------------------|
 21. | 1.025841   .9860208    .9800642   .9800642 |
 22. | .4439305   .4393799    .4354945   .4354945 |
 23. | -.029088    .013983    .0217048   .0217048 |
 24. |  .356685   .3032854    .2902687   .2902687 |
 25. | .4700715   .5052068    .5018782   .5018782 |
     |-------------------------------------------|
 26. | .2694395   .1901327    .1781852   .1781852 |
 27. | 1.104373   .9938793    .9883499   .9883499 |
 28. | .2858986   .2197184     .204144    .204144 |
 29. | .3692714   .3316926    .3174461   .3174461 |
 30. | .4478032   .4490708    .4452417   .4452417 |
     +-------------------------------------------+

.
. * WLS
.
. gen sigi = sqrt(p_lpm*(1-p_lpm))
(6 missing values generated)

. gen job1 = job/sigi
```

14-2

```
(6 missing values generated)

. gen age1 = age/sigi
(6 missing values generated)

. gen school1 = school/sigi
(6 missing values generated)

. gen gender1 = gender/sigi
(6 missing values generated)

. gen sigi_inv = 1/sigi
(6 missing values generated)

.
. regress  job1 age1 school1 gender1 sigi_inv, noconst

      Source |       SS       df       MS              Number of obs =      24
-------------+------------------------------           F(  4,    20) =   12.50
       Model |   53.56152     4    13.39038            Prob > F      =  0.0000
    Residual | 21.4168641    20   1.0708432            R-squared     =  0.7144
-------------+------------------------------           Adj R-squared =  0.6572
       Total | 74.9783841    24  3.12409934            Root MSE      =  1.0348

------------------------------------------------------------------------------
        job1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        age1 |  -.0021052   .0080463    -0.26   0.796   -.0188894     .014679
     school1 |   .1128579   .0529995     2.13   0.046    .0023028     .223413
     gender1 |   -.401973   .1913488    -2.10   0.049   -.8011197   -.0028263
    sigi_inv |  -.4112543   .7492448    -0.55   0.589   -1.974151    1.151643
------------------------------------------------------------------------------

.
. regress  job age school gender [aw = 1/sigi]
(sum of wgt is   5.4335e+01)

      Source |       SS       df       MS              Number of obs =      24
-------------+------------------------------           F(  3,    20) =    2.51
       Model | 1.62003772     3  .540012574            Prob > F      =  0.0876
    Residual | 4.29587769    20  .214793884            R-squared     =  0.2738
-------------+------------------------------           Adj R-squared =  0.1649
       Total | 5.91591541    23  .257213713            Root MSE      =  .46346

------------------------------------------------------------------------------
         job |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0014612   .0080317    -0.18   0.857    -.018215    .0152926
      school |   .1148828   .0557078     2.06   0.052   -.0013215    .2310871
      gender |  -.4295052   .2035091    -2.11   0.048   -.8540177   -.0049927
       _cons |  -.4355718   .7657082    -0.57   0.576   -2.032811    1.161667
------------------------------------------------------------------------------

. wls0 job age school gender, wvar(sigi) type(abse) noconst graph
(6 missing values generated)

WLS regression -  type: proportional to abs(e)

(sum of wgt is   6.1313e+01)

      Source |       SS       df       MS              Number of obs =      24
-------------+------------------------------           F(  3,    20) =    2.51
       Model | 1.62003773     3  .540012577            Prob > F      =  0.0876
    Residual | 4.29587767    20  .214793883            R-squared     =  0.2738
```

14-3

```
-------------+------------------------------           Adj R-squared =  0.1649
       Total |  5.9159154    23  .257213713            Root MSE      =  .46346

------------------------------------------------------------------------------
         job |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0014612   .0080317    -0.18   0.857    -.018215    .0152926
      school |   .1148828   .0557078     2.06   0.052   -.0013215    .2310871
      gender |  -.4295052   .2035091    -2.11   0.048   -.8540177   -.0049927
       _cons |  -.4355718   .7657082    -0.57   0.576   -2.032811    1.161667
------------------------------------------------------------------------------

. wls0 job age school gender, wvar(sigi) type(e2) noconst graph
(6 missing values generated)

WLS regression -  type: proportional to e^2

(sum of wgt is   1.2977e+02)

      Source |       SS       df       MS              Number of obs =      24
-------------+------------------------------           F(  3,    20) =    2.51
       Model |  1.6200377     3  .540012566            Prob > F      =  0.0876
    Residual |  4.2958777    20  .214793885            R-squared     =  0.2738
-------------+------------------------------           Adj R-squared =  0.1649
       Total |  5.9159154    23  .257213713            Root MSE      =  .46346

------------------------------------------------------------------------------
         job |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0014612   .0080317    -0.18   0.857    -.018215    .0152926
      school |   .1148828   .0557078     2.06   0.052   -.0013216    .2310871
      gender |  -.4295052   .2035091    -2.11   0.048   -.8540177   -.0049927
       _cons |  -.4355718   .7657082    -0.57   0.576   -2.032811    1.161667
------------------------------------------------------------------------------

.
.
. /* WLS: example of Greene chapter 12 */
.
. use http://www.ats.ucla.edu/stat/stata/examples/greene/TBL5-1, clear

.
. rename x1 age

. rename x2 income

. rename x3 exp

. rename x4 ownrent

. rename x5 selfemp

.
. generate incomesq = income^2

. drop if exp==0
(28 observations deleted)

. save chapter12, replace
file chapter12.dta saved

.
. /* summary check */
.
```

14-4

```
. gen age1     = age   / income^0.5

. gen income1  = income / income^0.5

. gen exp1     = exp / income^0.5

. gen ownrent1 = ownrent / income^0.5

. gen selfemp1 = selfemp / income^0.5

. gen incomesq1 = incomesq / income^0.5

. gen const1 = 1 / income^0.5

.
. wls0 exp age ownrent income incomesq , wvar(income) type(abse) noconst
/* 12.3a */

WLS regression -  type: proportional to abs(e)

(sum of wgt is   5.7161e-01)

      Source |       SS       df       MS              Number of obs =      72
-------------+------------------------------           F(  4,    67) =    5.73
       Model |  1266234.75     4  316558.686           Prob > F      =  0.0005
    Residual |   3703808.1    67  55280.7179           R-squared     =  0.2548
-------------+------------------------------           Adj R-squared =  0.2103
       Total |  4970042.85    71  70000.6035           Root MSE      =  235.12

------------------------------------------------------------------------------
         exp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -2.935011   4.603331    -0.64   0.526    -12.1233    6.253276
     ownrent |   50.49364   69.87914     0.72   0.472    -88.9857     189.973
      income |   202.1694   76.78152     2.63   0.010    48.91285     355.426
    incomesq |  -12.11364    8.27314    -1.46   0.148   -28.62689     4.39962
       _cons |  -181.8706   165.5191    -1.10   0.276   -512.2481    148.5068
------------------------------------------------------------------------------

. regress exp age ownrent income incomesq [aw = 1/income]
(sum of wgt is   2.4956e+01)

      Source |       SS       df       MS              Number of obs =      72
-------------+------------------------------           F(  4,    67) =    5.73
       Model |  1266234.79     4  316558.697           Prob > F      =  0.0005
    Residual |  3703808.18    67   55280.719           R-squared     =  0.2548
-------------+------------------------------           Adj R-squared =  0.2103
       Total |  4970042.96    71  70000.6051           Root MSE      =  235.12

------------------------------------------------------------------------------
         exp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -2.935011   4.603331    -0.64   0.526    -12.1233    6.253276
     ownrent |   50.49364   69.87914     0.72   0.472    -88.9857     189.973
      income |   202.1694   76.78152     2.63   0.010    48.91285     355.426
    incomesq |  -12.11364    8.27314    -1.46   0.148   -28.62689     4.39962
       _cons |  -181.8706   165.5191    -1.10   0.276   -512.2481    148.5068
------------------------------------------------------------------------------

. regress exp1 age1 ownrent1 income1 incomesq1 const1, noconstant

      Source |       SS       df       MS              Number of obs =      72
-------------+------------------------------           F(  5,    67) =   15.84
       Model |  1518004.77     5  303600.953           Prob > F      =  0.0000
```

14-5

```
    Residual |  1283774.35    67  19160.8111           R-squared     =  0.5418
-------------+------------------------------           Adj R-squared =  0.5076
       Total |  2801779.11    72  38913.5988           Root MSE      =  138.42

------------------------------------------------------------------------------
        exp1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        age1 |  -2.935011   4.603331    -0.64   0.526    -12.1233    6.253275
    ownrent1 |   50.49365   69.87914     0.72   0.472    -88.9857     189.973
     income1 |   202.1694   76.78152     2.63   0.010    48.91284     355.426
   incomesq1 |  -12.11364    8.27314    -1.46   0.148   -28.62689    4.399621
      const1 |  -181.8706   165.5191    -1.10   0.276   -512.2481    148.5068
------------------------------------------------------------------------------

. regress exp1 age1 ownrent1 income1 incomesq1

      Source |       SS       df       MS              Number of obs =      72
-------------+------------------------------           F(  4,    67) =    2.83
       Model |  216096.606     4  54024.1516           Prob > F      =  0.0313
    Residual |   1279473.4    67  19096.6179           R-squared     =  0.1445
-------------+------------------------------           Adj R-squared =  0.0934
       Total |  1495570.01    71  21064.3663           Root MSE      =  138.19

------------------------------------------------------------------------------
        exp1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        age1 |  -2.647726   4.616128    -0.57   0.568   -11.86156    6.566103
    ownrent1 |    48.1849   69.89543     0.69   0.493   -91.32696    187.6968
     income1 |   311.1198   148.1804     2.10   0.040    15.35038    606.8892
   incomesq1 |  -16.40469   10.50467    -1.56   0.123    -37.3721    4.562719
       _cons |  -277.2822   231.3404    -1.20   0.235   -739.0396    184.4752
------------------------------------------------------------------------------

.
. /* end of checkup */
.
. log close
       log:  choice_binary_revised.log
  log type:  text
 closed on:  26 Sep 2007, 01:22:02
-------------------------------------------------------------------------------
-------------------------
```

Choice Model : Motivation

$$\begin{cases} y_i^* = \beta' X_i + e_i \\ \text{where } y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* \le 0 \end{cases} \end{cases}$$

"latent regression"
"random utility models (see Greene p670)
[Also, see Wooldridge Ex 15.2, p.510

Note y_i^* = benefit − cost ; latent variable

- We do not know what are the factors separately for benefit or cost. We do not observe y_i^*, either
- We only observe y_i and X_i.

eg) I observe $y_i = 1$ if a student takes this class. I do not observe his (her) benefit or cost. Simply I believe that $y_i^* > 0$.
I also observe X_i (gender, age, GPA, ..) which will determine y_i^*.

Note Common mistake in explaining the choice model

$$y_i = \beta' X_i + e_i \,, \quad y_i = \begin{cases} 1 & \text{if yes} \\ 0 & \text{if no} \end{cases}$$

this should be y_i^*.

Note $P(y_i^* > 0) = \Phi(\beta' X_i)$ or $G(\beta' X_i)$ or $\alpha + \beta X_i$
probit — logit — LPM

$P(y_i = 1) = P(y_i^* > 0)$

MLE (more)

Let $G(X_i'\beta)$ be the cdf of the std normal or logistic dist.

$$\mathcal{L} = \prod_{i=1}^{n} [G(X_i'\beta)]^{y_i} [1 - G(X_i'\beta)]^{1-y_i}$$

$$\log \mathcal{L}_i = y_i \log [G(X_i'\beta)] + (1-y_i) \log [1 - G(X_i'\beta)]$$

$$\log \mathcal{L} = \sum_{i=1}^{n} \log \mathcal{L}_i$$

- foc (Score vector)

$$S_i(\beta) = \left[\frac{g(X_i'\beta) \cdot [y_i - G(X_i'\beta)]}{G(X_i'\beta)[1 - G(X_i'\beta)]} \right] \cdot X_i = \begin{pmatrix} \text{generalized} \\ \text{residual} \end{pmatrix} \cdot X_i$$

- soc (Hessian, information matrix)

$$-E[H_i(\beta) | X_i] = \frac{[g(X_i'\beta)]^2 X_i X_i'}{G(X_i'\beta)[1 - G(X_i'\beta)]} \equiv A(X_i, \beta)$$

$$Var(\hat{\beta}) = \left[\sum_{i=1}^{n} A(X_i, \hat{\beta}) \right]^{-1} \equiv \hat{V}$$

<u>Note</u> Q-MLE and Huber-White sandwich estimator

$$Var(\hat{\beta}) = \left(\sum_i^n \hat{A}_i \right)^{-1} \left(\sum_i^n \hat{S}_i \hat{S}_i' \right) \left(\sum_i^n \hat{A}_i \right)^{-1}$$

* conceptual problem
Even when the model is mis-specified, we just correct for the std. error?

See p.40 of this lecture note.

This can be used when the model is suspected to be <u>mis-specified</u>.

Then, it's QMLE.

<u>point</u>: We're <u>not</u> using this estimator for the reason of heteroskedasticity.

Exercise (a) show that in the logit model, the generalized residual is given as 17

$$\left[y_i - \frac{\exp(X_i'\beta)}{1+\exp(X_i'\beta)} \right]$$

(b) let $q = 2y-1$. show that

$$\ln L = \sum_i \ln G(q_i X_i \beta), \quad S_i(\hat{\beta}) = \frac{q_i \cdot g(q_i X_i \beta)}{G(q_i X_i'\beta)} \cdot X_i$$

Note we can write

$$S(\hat{\beta}) = \sum_i^n S_i(\hat{\beta}) = \sum_{i=1}^n \left[y_i - \frac{\exp(X_i'\hat{\beta})}{1+\exp(X_i'\hat{\beta})} \right] \cdot X_i$$

$$= \sum_{i=1}^n (y_i - \hat{p}_i) X_i = 0$$

this implies

$$\sum \hat{p}_i X_i = \sum y_i X_i$$

Testing hypothesis in binary choice models

$H_0: \gamma = 0$ (subset of parameters) : exclusion restrictions

i) LR test

$$LR = 2(\log L_u - \log L_R) \sim \chi^2_m$$

m = # of restrictions

$\log L_R$ = restricted $\log L$ imposing $\gamma = 0$

ii) LM test

let $y_i^* = X_i'\beta + Z_i'\gamma + e_i$, and test $H_0: \underset{(m\times 1)}{\gamma} = 0$.

let $\tilde{\beta}$ be the probit estimator by imposing $\gamma = 0$

$$y_i^* = X_i'\beta + u_i \Rightarrow \tilde{\beta}, \tilde{u}_i$$

Then, obtain $\tilde{u}_i$ from the restricted model

$$\tilde{u}_i = y_i - G(x_i'\tilde{\beta})$$

Regress

$$\frac{\tilde{u}_i}{\sqrt{\hat{G}_i(1-\hat{G}_i)}} \quad \text{on} \quad \frac{\hat{g}_i}{\sqrt{\hat{G}_i(1-\hat{G}_i)}} x_i \ , \quad \frac{\hat{g}_i}{\sqrt{\hat{G}_i(1-\hat{G}_i)}} z_i$$

$LM = nR^2$ from this regression

$\sim \chi^2_m$

iii) Wald test

$H_0: R\beta = r$ or $\gamma = 0$

Wald $= \hat{\beta}' \hat{V}^{-1} \hat{\beta}$ where $\hat{V}$ is given in page 11.

or $(R\hat{\beta} - r)'(R\hat{V}R')^{-1}(R\hat{\beta} - r)$

or $(R\hat{\beta} - r)'[g(\hat{\beta})\ \hat{V}\ g(\hat{\beta})']^{-1}(R\hat{\beta} - r)$ if $R\beta = r$ is of a non linear form.

with $g(\hat{\beta}) = \dfrac{d R\hat{\beta}}{d\hat{\beta}'}$

Testing for heteroskedasticity, Solution of the issue

point : the probit (logit) estimator becomes inconsistent in the presence of heteroskedasticity in e_i.

If $Var(e_i) = (\exp(z_i'\gamma))^2$, say, we can modify

$$\ln L = \sum \left[y_i \ln G\left(\frac{x_i'\beta}{\exp(z_i'\gamma)}\right) + (1-y_i)\ln\left(1 - G\left(\frac{x_i'\beta}{\exp(z_i'\gamma)}\right)\right)\right]$$

... Stata "HETPROB" procedure, eg.

Reason: Heteroskedasticity in $Var(e_i|x_i)$ changes the functional form of $P(Y|X)$, thus $\log L_i$

Also, one can use LM test for testing heteroskedasticity

$$y_i^* = x_i'\beta + \left[-(x_i'\hat{\beta}) z_i\right]\gamma + e_i$$

$$H_0: \gamma = 0$$

and do the LM test ; see Green, p.681.

Solution if it exists.

Note Wooldridge (book, p.479) provides some insights on heteroskedasticity & non-normality.

Example
If $y^* = c + \beta_1 X_1 + \beta_2 X_2 + e$,
$Var(e) = \sigma^2 X_1^2$,
then do the probit (or logit) of Y on $1/X_1$, 1, X_2/X_1

$$\begin{cases} P(y_i = 1 \mid x_i) = G(x_i'\beta) & * \\ y_i^* = x_i'\beta + e_i, \ \underbrace{Var(e_i) \neq \sigma^2}_{\text{heteroskedasticity}} \text{ or } \underbrace{e_i \text{ is non-normal}}_{\text{non-normality}} \end{cases}$$

- We're more interested in the 1st equation. If this is invalid, we lose the whole point. But QMLE is useful!
- We're not much interested in $E(y_i^* \mid x_i)$, where the heteroskedasticity issue arises.
- If $e_i \sim$ logistic, but probit (std. normal) is used. Even in this case, the marginal effects of the probit model are reasonable. Thus, non-normality may not be a big issue.

* Choice-based Sampling : Green, p673

eg) 'loan default' Our sample has more observations with default than the true population.

population :	w_1	w_0	
sample :	p_1	p_0	$(p_1 \gg w_1)$
	$(y_i=1)$	$(y_i=0)$	

Let $w_i = y_i\left(\frac{w_1}{p_1}\right) + (1-y_i)\left(\frac{w_0}{p_0}\right)$

$$\log L = \Sigma\, w_i \ln G(q_i \beta' x_i) \quad : \text{weighted sampling MLE}$$

Goodness of fit

- there are pseudo R^2 and McFadden R^2

$$1 - \frac{1}{1+2(\log L_1 - \log L_0)/n}$$

$1 - \log L_1 / \log L_0$ where $\log L_0$ is from the model with a constant only.

Frequencies of actual & predicted outcomes

		Actual 1	Actual 0
predicted	1	14	3
	0	4	9

threshold = 0.5

7 obs mis-predicted.

stata: "lstat"

<u>Note</u> β and σ are not separately identified since y^* does not have a well defined unit of measurement. Thus, we obtain $\hat{\beta}/\hat{\sigma}$ from usual softwares.

<u>Note</u> $\hat{\beta}$'s from the logit model & probit models are different. But the marginal effects $d\hat{p}/dx_j$ appear similar.

* Proportions Data

grouped data p_i = # of 1's out of n_i observation of group i

$0 \le p_i \le 1$

$$\log L = \sum_{i=1}^{n} n_i \{ p_i \ln G(x_i'\beta) + (1-p_i) \ln [1 - G(x_i'\beta)] \}$$

or

$$\log L = \sum_{i=1}^{n} \{ y_i \ln G(x_i'\beta) + (1-y_i) \ln [1 - G(x_i'\beta)] \}$$

where y_i = proportions (<u>not</u> 1 or 0), "Wooldridge fractional logit."

Other estimators

- semi-parametric estimators of the slope parameters 2/
  (non-parametric kernel estimate)

  ... we do not need the function $G(\cdot)$.

  and can estimate $\hat{\beta}$.

  But, we cannot predict probabilities.

  : Stoker (1986) & others

- Median estimator : Minimum score estimator by Manski (1975)

  $$\text{Min} \sum_{i=1}^{n} |y_i - I(X_i\beta > 0)| \qquad \text{"absolute value"}$$

More on $\hat{\beta}$s

- In probit (also in logit), we actually obtain $\hat{\beta}/\hat{\sigma}$ since $\hat{\beta}$ & $\hat{\sigma}$ are not separately identified; standardized coefficients.
- Neglected heterogeneity (Omitted variables indep. of X's)

  $P(y_i=1 \mid x_i, \alpha_i) = \Phi(X_i\beta + \beta_2\alpha_i)$ where α_i = omitted variables independent of X_i

  In a linear model, if omitted variables are independent of other regressors, they do not affect their coefficients.
  But it is not the case in the probit model.

  $\hat{\beta}$ will be biased (attenuation bias; $\hat{\beta} = \hat{\beta}_0/\sqrt{Var(\beta_2\alpha + e)}$)
  (smaller)

  However, the marginal effects, dP/dX, will Not be affected.
  Also, $\widehat{P(Y|x)} = \Phi(X\hat{\beta}_0/\sqrt{Var(\beta_2\alpha + e)})$ is consistently obtained.
- Probit vs logit coefficients

  $\phi(0) = \frac{1}{\sqrt{2\pi}} = 0.4$, $g(0) = \frac{1}{(1+1)^2} = 0.25$

  Thus, logit estimates are $0.4/0.25 = 1.6$ times larger than probit est.
  The probit estimates $* 0.4 \simeq$ LPM estimates. logit est $* 0.25 \approx$ LPM est.

Endogeneity problem & 2SLS

$$y_1^* = z_1\delta_1 + \gamma y_2 + u_1$$

endogenous (continuous)

We simply ASSUME that a reduced form for y_2 exists.

(if y_1 appears in the (structural) y_2 equation, there is no reduced form equation for y_2; due to non-linearity)

$\Rightarrow$ Maddala: it exists if y_1^* is included in y_2 equation

Note LPM usual IV in linear models will be used!

$$y_2 = z_1 c_1 + z_2 c_2 + v_2 = zc + v_2 \qquad z = (z_1, z_2)$$

(u_1 & v_2 are correlated)

1) Testing for endogeneity in probit or logit models.

write $u_1 = \theta_1 v_2 + e_1$ then $\theta_1 = \dfrac{Cov(u_1, v_2)}{Var(v_2)}$

$$y_1^* = z_1\delta_1 + \gamma y_2 + (\theta_1 v_2 + e_1)$$

where $Var(e_1) = Var(u_1) - CORR(u_1, v_2)^2$

The test is valid even if normality or homoskedasity of v_2 is not satisfied (under H_0)

Thus, we can set up the test

① Run OLS on the reduced form of y_2

$$y_2 = z\hat{c} + \hat{v}_2$$

and obtain the residual, $\hat{v}_2$.

② Run Probit y_1 on z_1, y_2 and $\hat{v}_2$

$$y_1^* = z_1\delta_1 + \gamma y_2 + \theta_1\hat{v}_2 + e_1$$

Test $H_0: \theta_1 = 0$ implying $CORR(u_1, v_2) = 0$

(t-test)

2) Estimation

i) 2-step estimation (the same as the previous testing procedure)

"The estimates of the Marginal (partial) effects are still valid, even if H_0 is rejected."

... do not use $\hat{y}_2$. Just add $\hat{V}_2$ in the probit or logit estimation.

If $\theta_1 = 0$ is rejected (t-test), then reported $s(\hat{\theta}_1)$ is not valid. $\Rightarrow$ correction for 2-step estimators.

Corrections

1. use $g(w_i, \theta)$ not $s(w_i, \theta)$ of M-estimation
2. Robust variance is not a solution

... when $s(\hat{\theta}_1)$ is obtained, the correlation between (u_1, V_2) is not accounted for.

.. see Rivers & Vuong (1988); textbook p474 p361 (correction)

ii) full MLE

see textbook, p.476

$$f(y_1, y_2) = \underset{\substack{\downarrow \\ \Phi(w)}}{f(y_1 | y_2)}\ \underset{\substack{\downarrow \\ \text{normal}}}{f(y_2)}$$

where

$$w_i = \left[z_i \delta_1 + \alpha_1 y_{2i} + (\rho_1/\tau_2)(y_{2i} - z_i \delta)\right] / (1-\rho_1^2)^{\frac{1}{2}}$$

$$= (\Phi(w))^{y_1} (1 - \Phi(w))^{1-y_1} \cdot f(y_2)$$

$$\mathcal{L} = \prod_{i=1}^{n} f(y_{1i}, y_{2i})$$

$H_0: \rho = 0$ is a test for endogeneity

the MLE accounts for $\mathrm{CORR}(u_1, V_2) \equiv \rho$ but this parameter is often troublesome. ($\rho \to 1$ or -1) not converging, if R.O.

Binary Endogeneous Regressor

$$y_1^* = z_1\delta_1 + \gamma y_2 + u_1 \quad : y_2 \text{ is } \underline{\text{binary}}, \text{ too.}$$

$$y_2^* = zc + v_2$$

$$z = (z_1, z_2)$$

Eg) $y_1 = \begin{cases} 1 & \text{dividend pay} \\ 0 & \text{o/w} \end{cases}$ $\quad y_2 = \begin{cases} 1 & \text{bank ownership} \\ 0 & \text{o/w} \end{cases}$

Eg) $y_1 = \begin{cases} 1 & \text{Catholic school} \\ 0 & \text{o/w} \end{cases}$ $\quad y_2 = \begin{cases} 1 & \text{Catholic} \\ 0 & \text{o/w} \end{cases}$

One <u>tempting</u> but <u>incorrect</u> approach:

<u>1st</u> Do probit of y_2^*, and obtain $\Phi(z\hat{c})$

<u>2nd</u> Replace y_2 with $\Phi(z\hat{c})$

... This is an exaple of "forbidden regression" (see book, p. 478 and p 230 – 235)

"the expectation cannot pass through non-linear indicator functions"

MLE is easier (really?) and more efficient.

Combine 4 possible outcomes of (y_1, y_2) and construct $\log L$ using $P(y_1 = i \mid y_2 = j)$ $i, j = 1, 0$; see textbook p. 478.

Panel choice Models

1) pooling probit or logit

$$\max_{\beta} \sum_{i=1}^{N}\sum_{t=1}^{T}\left[y_{it}\log G(X_{it}\beta) + (1-y_{it})\log G(X_{it}\beta)\right]$$

$$y_{it}^{*} = X_{it}'\beta + u_{it}$$

<u>Note</u> Testing for "dynamic" models
(possible lagged dep. variables)

① Do pooling probit or logit, and obtain residuals,

$$y_{it} - \Phi(X_{it}'\hat{\beta}) \overset{\text{let}}{\equiv} \hat{u}_{it}$$

② Do pooling probit or logit with $\hat{u}_{i,t-1}$ added.

$$P(y_{it}=1) = \Phi(X_{it}\beta + \gamma_1 \hat{u}_{i,t-1})$$

Test $H_0: \gamma_1 = 0$ (no dynamics)

If H_0 is rejected, consider a dynamic model; say

$$y_{it}^{*} = X_{it}'\beta + \alpha y_{i,t-1} + u_{it}$$

.. still, can do dynamic LPM.

2) Panel choice models

$$P(y_{it}=1) = \Phi(X_{it}'\beta + \underbrace{c_i}_{\text{unobserved heterogeneity}})$$

Assume: $\boxed{\text{strict exogeneity}}$

example:
employed = $\Phi(\cdot\cdot, \#\text{ of arrests}, \cdot\cdot)$
$\|$
1/0
If y_{it} affects future arrests ($w_{i,t+s}$) the strict exog. assumption fails

Test: Add $w_{i,t+s}$ and test its significance

① FE probit (seldom)

c_i is the parameter to estimate.

– 'demeaning' is not possible; $y_{it}^* = y_{it} - \bar{y}_i$ works only for linear models.

– incidental parameter problems
(too many parameters to estimate)

⇒ FE probit is "rarely" used.

② RE probit (assume $\text{cov}(X_{it}, c_i) = 0$)

$$f(y_1, \dots, y_T) = \int_{-\infty}^{\infty} f(y_1, \cdots y_T, c)\, dc$$

∵ like $f(y) = \int f(x,y)\,dx$

$$= \int_{-\infty}^{\infty} \left[\prod_{t=1}^{T} f(y_t \mid c)\right] f(c)$$

… like $f(x,y) = f(y|x)\,f(x)$

This is not trivial; Butler & Moffit (1982) provides an approximation of the above integration.

Note "Integrating out" method for RE models
(Butler & Moffit's Gaussian-Hermite quadrature method)   Greene, p690

$$\varepsilon_{it} = v_{it} + \alpha_i$$

$$Var(\varepsilon_{it}) = \sigma_v^2 + \sigma_\alpha^2 \qquad (\text{assume } \sigma_v^2 = 1)$$

$$\mathcal{L}_i = P(y_{i1}, \dots, y_{iT_i}) = \int_{L_{iT_i}}^{U_{iT_i}} \cdots \int_{L_{i1}}^{U_{i1}} f(\varepsilon_{i1}, \dots \varepsilon_{iT_i}) \, d\varepsilon_{i1} \cdots d\varepsilon_{iT}$$

where $L_i \to U_i$ implies $\begin{cases} (-\infty, -x_i'\beta) & \text{if } y_i = 0 \\ (-x_i'\beta, \infty) & \text{if } y_i = 1 \end{cases}$

and

$$f(\varepsilon_{i1}, \dots, \varepsilon_{iT_i}, \alpha_i)$$

$$= f(\varepsilon_{i1}, \dots, \varepsilon_{iT_i} \mid \alpha_i) f(\alpha_i)$$

$$f(\varepsilon_{i1}, \dots, \varepsilon_{iT_i}) = \int_{-\infty}^{\infty} f(\varepsilon_{i1}, \dots, \varepsilon_{iT_i} \mid \alpha_i) f(\alpha_i) \, d\alpha_i$$

$$= \int_{-\infty}^{\infty} \prod_{t=1}^{T_i} f(\varepsilon_{it} \mid \alpha_i) f(\alpha_i) \, d\alpha_i$$

Thus,

added to evaluate $P(y_{i1}, \dots y_{iT} \mid \alpha_i)$

$$\mathcal{L}_i = \int_{-\infty}^{\infty} \left[\prod_{t=1}^{T_i} \left(\int_{L_{it}}^{U_{it}} f(\varepsilon_{it} \mid \alpha_i) \right) \right] f(\alpha_i) \, d\alpha_i \qquad (*)$$

$$= \int_{-\infty}^{\infty} \left[\prod_{t=1}^{T_i} Prob(Y_{it} = y_{it} \mid x_{it}'\beta + \alpha_i) \right] \frac{1}{\sigma_\alpha \sqrt{2\pi}} e^{-\frac{\alpha_i^2}{2\sigma_\alpha^2}} \, d\alpha_i$$

Let $\frac{\alpha_i^2}{2\sigma_\alpha^2} = r_i^2 \Rightarrow \begin{cases} \alpha_i = (\sigma_\alpha \sqrt{2}) r_i \\ d\alpha_i = (\sigma_\alpha \sqrt{2}) dr_i \end{cases}$

$$= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-r_i^2} \left[\prod_{t=1}^{T_i} \Phi(q_{it}(x_{it}'\beta + (\sigma_\alpha \sqrt{2}) r_i) \right] dr_i$$

$\Rightarrow$ Transform the integral problem into a summation problem via Gaussian quadrature.

"Gaussian quadrature"

$$\int_L^U w(x) f(x)\, dx \cong \sum_{j=1}^{M} w_j f(a_j)$$

where $w(x)$: weighting function

w_j : quadrature weight

a_j : quadrature abscissa

when $w(x) = e^{-x^2}$, it's Gaussian-Hermite quadrature.

$$\mathcal{L} = \sum_{i=1}^{n} \left\{ \ln \left[\frac{1}{\sqrt{\pi}} \sum_{h=1}^{H} w_h \prod_{t=1}^{T_i} \Phi\left(q_{it}(X_{it}'\beta + \theta z_h)\right) \right] \right\}$$

where $q_{it} = 2y_{it} - 1$

$\theta = \sigma_\alpha \sqrt{2}$

$z_h = \sqrt{\frac{2\rho}{1-\rho}}$, $\rho = \frac{\sigma_\alpha^2}{\sigma_v^2 + \sigma_\alpha^2}$

Here, H needs to be given, and the result may be sensitive to its choice.

(STATA : "quadchk" checks sensitivity, default $H = 12$)

Alternative method : maximum simulated likelihood (MSL)

(*) equation (last page)

$$\mathcal{L}_i = \int_{-\infty}^{\infty} \left[\prod_{t=1}^{T_i} \mathrm{Prob}(Y_{it} = y_{it} \mid X_{it}'\beta + \alpha_i) \right] f(u_i)\, du_i$$

$$= E_{u_i}\left[\quad " \quad \right]$$

call inside $h(u_i)$

$$= \mathrm{plim}\ \frac{1}{R} \sum_{r=1}^{R} h(u_{ir})$$

using a sample of observations

(simulated) $u_{i1}, \ldots, u_{iR}$

$$\ln L = \sum_{i=1}^{n} \mathcal{L}_i$$

* ③ FE logit (Common) see WI, p490-492

Find the joint dist of $(y_{i1} \dots y_{iT_i})$, conditional on c_i and $n_i = \sum_{t=1}^{T_i} y_{it}$

then the conditional ML does not depend on c_i. (c_i cancelled out.)

$\Rightarrow$ CMLE (Chamberlain's method) (next page)

- we can obtain log-odd expressions, but cannot obtain partial effects since c_i is not estimated.

- Requires conditional independence. thus FE logit cannot allow for dynamic models.

$P(y_{it}=1)$

$$= \frac{e^{\alpha_i + \beta' X_{it}}}{1+e^{\alpha_i+\beta' X_{it}}}$$

Not possible to sweep out α_i by taking differences or deviation from group means

$\boxed{\text{FE logit models do NOT require the condition } Cov(X_{it}, \alpha_i) = 0}$

④ RE logit

- no feasible simple form
; Integrating the logit cdf wrt normal density has no simple form solution

⑤ FE probit using the conditional likelihood function (next page) is not possible.

Remark : Conditional Likelihood Function (FE logit) Chamberlain's method.

A FE logit model is

$$P(y_{it}=1 \mid x_{it}) = \frac{e^{\alpha_i + x_{it}'\beta}}{1+e^{\alpha_i + x_{it}'\beta}}$$

unconditional likelihood

$$L = \prod_i \prod_t (F_{it})^{y_{it}}(1-F_{it})^{1-y_{it}}$$

conditional likelihood

$$L^c = \prod_{i=1}^{n} P\left(Y_{i1}=y_{i1}, \ldots, Y_{iT}=y_{iT_i} \,\Big|\, \sum_{t=1}^{T_i} y_{it}\right)$$

(∵ this is free of α_i)

$$= \prod_{i=1}^{n} \frac{\exp\left(\sum_{t=1}^{T_i} y_{it} x_{it}'\beta\right)}{\Sigma_{\text{all set}} \exp\left(\sum_{t=1}^{T_i} d_{it} x_{it}'\beta\right)}$$

where the denominator is summed over the set of all ${}_{T_i}C_{S_i}$ sequences with $S_i = \sum_{t=1}^{T_i} y_{it}$.

ex) $T_i = 2$

i) $y_{i1}=0,\ y_{i2}=0 \Rightarrow P(0,0 \mid \text{sum}=0)=1$

ii) $y_{i1}=1,\ y_{i2}=1 \Rightarrow P(1,1 \mid \text{sum}=2)=1$

iii) $y_{i1}=0,\ y_{i2}=1 \Rightarrow P(0,1 \mid \text{sum}=1)$

$$= \frac{P(0,1)}{P(0,1)+P(1,0)}$$

This is given as

$$\frac{\frac{1}{1+\exp(\alpha_i+x_{i1}'\beta)} \cdot \frac{\exp(\alpha_i+x_{i2}'\beta)}{1+\exp(\alpha_i+x_{i2}'\beta)}}{(\text{top}) + \frac{\exp(\alpha_i+x_{i1}'\beta)}{1+\exp(\alpha_i+x_{i1}'\beta)} \cdot \frac{1}{1+\exp(\alpha_i+x_{i2}'\beta)}} = \frac{e^{x_{i2}'\beta}}{e^{x_{i1}'\beta}+e^{x_{i2}'\beta}}$$

... free of α_i

Note

pair id	y_i	x_{1i}	x_{2i}
1	0	:	.
1	1	:	.
2	0	:	.
2	1	:	.
3	0	:	.
3	1	:	!
⋮			

eg) T=2 "Matched sample pairs" (not necessarily panel)

"conditional Logit" uses the information on the matching.

- this model is different from the usual logit model. "pair" vs "1 or 0" decision (treated vs control)

Thus, by conditioning on the sum of two observations, we can remove α_i.

Note i) One can test CML vs ML

Hausman test

$$(\hat{\beta}_{CML} - \hat{\beta}_{ML})' [Var(CML) - Var(ML)]^{-1} (\hat{\beta}_{CML} - \hat{\beta}_{ML})$$

ii) We cannot estimate the marginal effects on the predicted probabilities unless we can plug α_i (which was cancelled out!)

Dynamic choice Models (Panel)

$$P(y_{it} = 1 \mid y_{i,t-1}, \ldots, y_{i0}, X_{it}, \alpha_i)$$

$$= G(X_{it}'\beta + \rho y_{i,t-1} + \alpha_i)$$

$H_0: \rho = 0$ (state dependence is absent)

- FE cannot be used
- RE version is possible (integrating out α_i)

$$P(y_{it} = 1 \mid \cdot\cdot) = G(\phi + X_{it}'\beta + \rho y_{i,t-1} + \underbrace{\varphi_0 y_{i0} + X_i'\xi}_{\text{added}} + \underbrace{\alpha_i}_{\text{integrate out}} + e_{it})$$

Exercise Wooldridge 15.19 (p.515) : dynamic model

(a) ~ (e).

Other panel estimates

(1) FIML using multivariate-normal dist

$$y_{it}^* = x_{it}\beta + \alpha_i + e_{it}, \quad y_{it} = 1(y_{it}^* > 0)$$

where e_i ~ multivariate normal

$$Var(e_i) = \Omega_{N \times N}$$

.. Hard to estimate ..

(2) Mundlak's method (or chamberlain's RE)

Add $\bar{X}_{i.}$ to regressors

$$\alpha_i | X_i \sim N(\psi + \bar{X}_{i.}\xi, \sigma_a^2)$$

where $\alpha_i = \psi + \bar{X}_{i.}\xi + a_i$

Thus

$$P(y_{it} = 1 | X_{it}) = \Phi[c_0 + X_{it}\beta + \bar{X}_{i.}\xi]$$

Note Time invariant variables, time dummies cannot be included in $\bar{X}_{i.}$

Then DO RE

(3) Semi-parametric estimator

no dist. assumption about α_i

$\Rightarrow$ not efficient (not $\sqrt{N}$-consistent)

(4) Paired-sample $\Rightarrow$ use FE logit (conditional likelihood)

(not panel, but cluster)

Logit

logit fits maximum-likelihood logit models. depvar==0 indicates a negative outcome; depvar!=0 & depvar!=. (typically depvar==1) indicates a positive outcome.

Also see help logistic; many users prefer the logistic command to logit. Results are the same regardless of which you use, but logistic reports odds ratios rather than coefficients by default and some people simply prefer the name logistic to logit. In Stata, both are the maximum-likelihood estimator. A number of commands are documented under help logistic that may be run after logit or logistic.

If estimating on grouped data, see help glogit.

Logistic

logistic fits maximum-likelihood logistic regression models. depvar==0 indicates a negative outcome; depvar!=0 & depvar!=. (typically depvar==1) indicates a positive outcome.

logistic reports odds ratios. You can type "logit" after logistic estimation to obtain the coefficients. In addition, there are a number of other commands that can be used after logistic to explore the nature of the fit:

help lfit Perform goodness-of-fit tests
help lstat Report summary statistics including classification table
help lroc Graph the ROC curve
help lsens Graph sensitivity and specificity vs. P cutoff

probit

probit fits maximum-likelihood probit models.

dprobit also fits maximum-likelihood probit models. Rather than reporting coefficients, dprobit reports the change in the probability for an infinitesimal change in each independent, continuous variable and, by default, the discrete change in the probability for dummy variables. probit may be typed without arguments after dprobit estimation to see the model in coefficient form.

depvar==0 indicates a negative outcome; depvar!=0 & depvar!=. (typically depvar==1) indicates a positive outcome.

If you are estimating on grouped data, see help glogit (sic).

blogit, bprobit, glogit and gprobit

blogit and bprobit produce maximum-likelihood logit and probit estimates on grouped ("blocked") data; glogit and gprobit produce weighted least-squares estimates. In the syntax diagrams, pos_var and pop_var refer to variables containing the total number of positive responses and the total population.

See help logit and help probit for obtaining maximum-likelihood estimates on ungrouped (individual or micro) data.

xtprobit

xtprobit fits random-effects (re) and population-averaged (pa) probit models for cross-sectional time-series datasets.

For the random-effects model, the likelihood (for an independent unit i) is expressed as an integral which is computed using Gauss-Hermite quadrature. After fitting your final model, you may want to run quadchk to check the numerical soundness of the Gauss-Hermite quadrature approximation; see help quadchk and [XT] quadchk for details.

By default, the population-averaged model is an equal-correlation model; that is, xtprobit, pa assumes corr(exchangeable). See help xtgee for details on how to fit other population-averaged models.

Xtlogit

xtlogit fits a fixed-effects (fe), an random-effects (re), or a population-averaged (pa) logit model for cross-sectional time-series datasets. Whenever we refer to a fixed-effects model, we mean the conditional fixed-effects model.

For the random-effects model, the likelihood (for an independent unit i) is expressed as an integral which is computed using Gauss-Hermite quadrature. After fitting your final model, you may want to run quadchk to check the numerical soundness of the Gauss-Hermite quadrature approximation; see help quadchk and [XT] quadchk for details.

By default, the population-averaged model is an equal-correlation model; that is, xtlogit, pa assumes corr(exchangeable). See help xtgee for details on how to fit other population-averaged models.

Review Questions on Binary Choice Models

1. What are possible limitations of using the OLS estimator when the dependent variable is a dummy variable? Why is it called the LPM?

2. Discuss about how we use each of the LPM, Probit and Logit models, for the following.

   (i) prediction equation
   (ii) marginal effect (Is it constant over different individuals? If not, what is the term for this?)
   (iii) testing hypothesis (Which tests can be used?)

3. Using the estimation result of the example on labor participation, for each of LPM, Probit and logit models,

   (a) What is the predicted probability of labor participation for a 34 years old man with 14 years of education?

   (b) If he is not employed, what will be the marginal effect on the increased probability of labor participation when he receives two more years of education?

Exercise on Choice Models

Objectives: Analyze Qualitative Choice Models using the aacsb.xls file (web site).

Data Background:

The data file aacsb.txt contains data for 404 graduate business school programs on six variables, the names of which are contained in the first row of the file. The variable **AACSB** is a 1,0 dummy variable that indicates whether or not the American Assembly of Collegiate Schools of Business (AACSB) accredits the business school program. Three other variables indicate the number of students enrolled (**SIZE**), the average **GMAT** score of enrollees, and the proportion of faculty with doctorates (**FACPHD**). Two additional 1,0 dummy variables indicate whether the program granted a doctoral business degree (**PHD**), and whether the program was a public university (**STATE**). The data for this file come from a study of accrediting practices published by Jantzen and Pendleton ("Preferences of the American Assembly of Collegiate Schools of Business," *Journal of Education for Business,* Vol. 70, Sept/Oct 1994, pp. 6-11). About two-fifths of all graduate business programs were accredited at the time the data were collected (1992).

Assignment:

(a) Read the data in. The appropriate STATA command is:

insheet using aacsb.txt (saved as a text file at the right folder)

(b) To examine how the probability of being accredited is likely to be affected by a graduate business program's size, student GMAT scores, faculty doctorates, the presence of a doctoral business program, or being a publicly funded program, run the following ordinary least squares (OLS) regression:

regress aacsb size gmat facphd phd state

Interpret the regression coefficients, and conduct appropriate t-tests.

(c) For how many observations, will the predicted probability lie outside the bound or 1 and 1?

redict pred_lpm, xb
list pred_lpm

(d) Estimate a logit regression model explaining which schools became accredited by the AACSB and which were not. Then, conduct appropriate t-tests on the logit regression coefficients. Repeat for probit models.

logit aacsb size gmat facphd phd state
probit aacsb size gmat facphd phd state

(e) Conduct a likelihood ratio test on whether or not the coefficients on the FACPHD and the PHD variables are both zeros each of the logit and probit models.

logit aacsb size gmat state
probit aacsb size gmat state

(f) Using the marginal effects at means, determine which variables significantly affect the probability of being AACSB accredited in each of the LPM, logit and probit models.

dlogit2 aacsb size gmat facphd phd state
dprobit aacsb size gmat facphd phd state

[Note: dlogit2 needs to be installed. Help-search-all-dlogit2-install]

(g) Find the partial effect of PHD, at means of regressors, each of the LPM, logit and probit models.

dlogit2 aacsb size gmat facphd phd state
dprobit aacsb size gmat facphd phd state

Suppose that one has collected the following data for Greene University:

SIZE	GMAT	FACPHD	PHD	STATE
300.	540.	90	0	1

(h) Using each of the LPM, Logit and Probit models, find the probability that AACSB will accredit the business school program of the Greene University.

(i) Using each of (i) LPM and (ii) Logit and (iii) Probit models, determine how much the probability will be changed if the Greene University implements a Ph.D. program.

Exercises

1. Wooldridge Ex 15.3 (p.510) "probit partial effect"
2. " Ex 15.7 (p.512) "probit, crime.RAW"
3. " Ex 15.19 (p.515) "dynamic probit"
4. " Ex 15.20 (p.516)

Problem Set (FE logit)

Q. A FE logit model is based on the following probability.

$$P(y_{it} = 1|x_{it}) = \frac{\exp(\alpha_i+X_{it}\beta)}{1+\exp(\alpha_i+X_{it}\beta)}.$$

(a) Consider the following sample with N=4 observations, and T=2.

	y_{it}	
	t=1	t=2
i=1	1	0
i=2	0	0
i=3	0	1
i=4	1	1

Find the conditional likelihood function of this FE panel logit model.

(b) Suppose that $T_i = 3$. Find the contribution of the i-th observation to the conditional likelihood function L_i, when

		y_{it}	
	t=1	t=2	t=3
...			
ith	1	0	0

40

What are the advantages of using the robust variance estimator over the standard maximum-likelihood variance estimator in logistic regression?

Title	Advantages of the robust variance estimator
Author	Bill Sribney, StataCorp
Date	January 1998

I once overheard a famous statistician say that the robust variance estimator for (unclustered) logistic regression is stupid. His reason was that if the outcome variable is binary then it's got to be a Bernoulli distribution. It's not like linear regression with data that stands a good chance of being nonnormal and heteroscedastic.

Well, it's not as simple as this; there's a bunch of subtle nuisances here. Let me lay them out here. I'm sure that the famous statistician is aware of them, but they don't necessarily lead to his conclusion.

It's true that it's got to be a Bernoulli distribution. That is, if $\mathbf{Y_i}$ is a random variable for the outcome of the **i-th** unit, then

$$P(Y_{i=1}) = p_i$$

or equivalently,

$$E(Y_i) = p_i$$

This *has* to be true. Note how I indexed the RHS by **i**. The term $\mathbf{p_i}$ is dependent on **i**. It's certainly not true in general that $\mathbf{P(Y_{i=1}) = p}$, where **p** is a constant independent of **i**.

In logistic regression, we model $\mathbf{p_i}$ with a likelihood that assumes

$$\text{logit}(p_i) = x_i{*}b$$

So these are our assumptions:

1. That the **logit** link function is correct.
2. That $\mathbf{logit(p_i) = x_i{*}b}$; i.e., that the relation is linear and all necessary predictors are in the model; i.e., that the model is correctly specified.
3. That the **i=1**,..., N observations are independent.

41

The robust variance estimator is robust to the assumptions (1) and (2). It does require (3), but you can specify clusters and just assume independence of the clusters if you wish.

The MLE is also quite robust to (1) being wrong. If the link function is really probit and you estimate a logit, everything's almost always fine.

Now if (2) is wrong, strictly speaking, you are in trouble with the interpretation of the point estimates of your model, never mind the variance estimates. Imagine that **logit(p_i)** is truly quadratic in $\mathbf{x}_i$, but you fit it linear in $\mathbf{x}_i$. What's the interpretation of the coefficient? It's the best fit of a straight line to something that's not straight! It's got some meaning, but it's somewhat problematic.

Well, the robust variance estimator will do a good job of giving you variance estimates and confidence intervals for this problematic case of a misspecified model. That is, if one imagines resampling the data and each time fitting the same misspecified model, then you get good coverage probabilities w.r.t. the "true" population parameters of the misspecified model, i.e., the besting fitting straight line in the population to something that's not straight.

On the other hand, if you have confidence that your model is not misspecified, then the ML variance estimator is theoretically more efficient.

Advice

In summary, my personal advice (and I have respect for conflicting opinions) is

- I never worry about whether (1) is true. I assume the **logit** link is OK.
- If I think that the model is reasonably specified, I use the ML variance estimator for logistic regression.
- Only if I have good reason to believe that the model is poorly specified would I use the robust variance estimator. That is, if the model fails goodness-of-fit tests, etc. Sometimes one just has to live with missing predictors and badly fitting models because data was only collected for a few predictors. In this case, I'd use the robust variance estimator.

And, obviously, I'd use the robust variance estimator if I had clustered data.

This is in contrast to the advice I'd give for linear regression when I'd say *always* use the robust variance estimator.

The robust variance estimator is only approximate for ML models

Note that there are also other theoretical reasons to be keener on the robust variance estimator for linear regression than for general ML models. The robust variance estimator uses a one-term Taylor series approximation. In linear regression, the coefficient estimates **b** are a linear function of **y**; namely,

$$b = (X'X)^{-1} X'y$$

Thus, the one-term Taylor series is exact and not an approximation.

For logistic regression and other MLEs, the ignored higher-order terms in the Taylor series are nonzero. So it's truly an approximation in these cases.

42

One can come up with a robust variance estimator that uses a second order correction. It's been worked out for logistic regression, and it will likely be implemented in Stata at some point in the future.

Follow-up question

What are the "true" population parameters for which the robust variance estimator gives good coverage properties?

Linear regression model

First let me assume a linear regression model, then later I'll discuss MLEs.

Consider the entire population from which the sample is drawn. Let **X** be the matrix of independent variables and **Y** the vector of dependent variables for the *entire population*. Then consider

```
B = (X'X)^-1 X'Y
```

The parameter **B** is the coefficient vector for the linear model for the entire population. **Y** may be linear in **X** or it may not. **B** is simply the best least-squares coefficients for the entire population. **B** is what I was referring to when I said "the 'true' population parameters" in my above explanation.

The parameter **B** is what the robust variance estimator considers you to be estimating. The sample estimate is

```
b = (x'x)^-1 x'y
```

where **x** is the matrix of the sample values of the independent variables and **y** is the vector of sample dependent variables. If there were sampling weights, the above equation would have weights added in the appropriate places.

Anyhow, **b** is an estimate of **B**. The robust variance estimator estimates **V(b)** such that nominal (1-alpha) confidence intervals constructed from it have **B** in the interval about (1-alpha) of the time if one was to repeatedly resample from this population.

If **Y** is not linear in **X** due to incorrect functional form or missing predictors, then the interpretation of **B** is problematic. **B** can be considered to be the best least-squares linear fit for this set of predictors. **b** and **V(b)** are "robust to misspecification" in that **b** estimates **B** and that **V(b)** is a valid estimate of the variance of **b** even though misspecification is present.

Note that this theory requires no assumptions about the distribution of **Y**.

Contrast this to the case of OLS estimates, which do not give valid variance estimates, in general, under misspecification, and which do require distributional assumptions on **Y** — i.e., normality and homoscedasticity.

ML Models

For ML models, consider **L(B; Y, X)**, an arbitrary likelihood function with data **Y, X** for the entire population. Let **B*** give the maximum of **L(B; Y, X)**. The sample estimate **b*** is the maximum of **L(b; y, x)**. If there are weights, we add weights to the likelihood function so that **L(b; y, x)** estimates **L(B; Y, X)**. Because **L(b; y, x)** estimates **L(B; Y, X)**, **b*** estimates **B***. The robust variance estimator produces correct variance estimates **V(b*)** for **b*** in the same sense discussed above: nominal (1-alpha) confidence intervals constructed from it have **B*** in the interval about (1-alpha) of the time if one was to repeatedly resample from this population.

43

Note that **L(B; Y, X)** is not necessarily the true likelihood for the population; i.e., it is not necessarily the correct distribution of **Y|X**. The theory doesn't require it; it can be any function.

Standard MLE theory, on the other hand, requires **L(b; y, x)** to be the true distribution function for the sample.

Note that when there are sampling weights or clustering, **L(b; y, x)** is in no sense a valid likelihood; it's clearly not the distribution of the sample when there are weights or cluster sampling. **L(b; y, x)** merely has to estimate the arbitrary **L(B; Y, X)** for our theory to hold. This is why the survey theorists call **L(b; y, x)** a pseudo-likelihood, and it's also why you can't do standard likelihood ratio tests with it.

However, if **L(B; Y, X)** is not close to the true distribution, its interpretation is problematic, just as in the case of a misspecified linear regression.
