7. QUALITATIVE DEPENDENT VARIABLES

[1] Binary choice models

- Motivation:
  - Dependent variable (y_t) is a yes/no variable (e.g., unionism, migration, labor force participation, or dealth ...).

(1) Linear Model (Somewhat Defective)

Digression to Bernoulli's Distribution:

- Y is a random variable with pdf; p = Pr(Y=1) and (1-p) = Pr(Y=0).
- $f(y) = p^y(1-p)^{1-y}$.
- $E(y) = \Sigma_y yf(y) = 1\bullet p + 0\bullet (1-p) = p$;
- $var(y) = \Sigma_y y^2 f(y) - [E(y)]^2 = p - p^2 = p(1-p)$.

End of Digression

- Linear Model:

  $y_t = x_{t\bullet}'\beta + \varepsilon_t$, where $y_t = 1$ if yes and $y_t = 0$ if no.

  - Assume that the $x_{t\bullet}$ are nonstochastic and $E(\varepsilon_t) = 0$.

    [Or assume that $E(y_t|x_{t\bullet}) = 0$]

  - $E(y_t) = E(x_{t\bullet}'\beta+\varepsilon_t) = x_{t\bullet}'\beta \equiv p_t$ (= $Pr(y_t = 1)$).
  - $\frac{\partial p_t}{\partial x_{tj}} = \beta_j$: So, the coefficients measure effects of x_{tj} on p_t.

- Problems in the linear model:

1) The ε_t are nonnormal and heteroskedastic.

  - Note that $y_t = 1$ or 0.

    $\rightarrow \varepsilon_t = 1 - x_{t\bullet}'\beta$ with prob $= p_t = x_{t\bullet}'\beta$

    $= -x_{t\bullet}'\beta$ with prob $= 1 - p_t = 1 - x_{t\bullet}'\beta$.

  - $E(\varepsilon_t) = (1-x_{t\bullet}'\beta)x_{t\bullet}'\beta + (-x_{t\bullet}'\beta)(1-x_{t\bullet}'\beta) = 0$.

  - $var(\varepsilon_t) = E(\varepsilon_t^2) = (1-x_{t\bullet}'\beta)^2 x_{t\bullet}'\beta + (-x_{t\bullet}'\beta)^2(1-x_{t\bullet}'\beta) = x_{t\bullet}'\beta(1-x_{t\bullet}'\beta)$.

    $\rightarrow$ Not constant over t.

  - OLS is unbiased but not efficient.

  - GLS using $\hat{\sigma}_t^2 = \left(x_{t\bullet}'\hat{\beta}\right)\left(1 - x_{t\bullet}'\hat{\beta}\right)$ is more efficient than OLS.

2) Suppose that we wish to predict $p_o = P(y_o = 1)$ at $x_{o,\bullet}$. The natural predictor of p_o is $x_{o,\bullet}'\hat{\beta}$ where $\hat{\beta}$ is OLS or GLS. But $x_{o,\bullet}'\hat{\beta}$ would be outside of the range (0,1).

(2) Probit Model

- Model:

$y_t^* = x_{t\bullet}'\beta + \varepsilon_t$, t = 1, ... , T,

where y_t^* is a unobservable latent variable (e.g., level of utility);

$y_t = 1$ if $y_t^* > 0$

$= 0$ if $y_t^* < 0$;

the $(-\varepsilon_t)$ are i.i.d. N(0,1).

Digression to normal pdf and cdf

- $X \sim N(\mu,\sigma^2)$: $f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$, $-\infty < x < \infty$.
- $Z \sim N(0,1)$: $\phi(z) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)$; $\Phi(z) = \Pr(Z < z) = \int_{-\infty}^{z}\phi(v)dv$.
- In LIMDEP, $\phi(z)$ = N01(z) and $\Phi(z)$ = PHI(z).

  In GAUSS, $\phi(z)$ = pdfn(z) and $\Phi(z)$ = cdfn(z).
- Some useful facts:

  $d\Phi(z)/dz = \phi(z)$; $d\phi/dz = -z\phi(z)$; $\Phi(-z) = 1 - \Phi(z)$; $\phi(z) = \phi(-z)$.

End of digression

- Return to the Probit model

- PDF of the y_t:
  - $\Pr(y_t = 1) = \Pr(y_t^* > 0) = \Pr(x_{t\bullet}'\beta + \varepsilon_t > 0) = \Pr(x_{t\bullet}'\beta > -\varepsilon_t)$

    $= \Pr(-\varepsilon_t < x_{t\bullet}'\beta) = \Phi(x_{t\bullet}'\beta)$.

    → This gaurantees $p_t \equiv \Pr(y_t = 1)$ being in the range (0,1).
  - $f(y_t) = \left(\Phi(x_{t\bullet}'\beta)\right)^{y_t}\left(1-\Phi(x_{t\bullet}'\beta)\right)^{1-y_t}$.

Short Digression

$y_t^* = x_{t\bullet}'\beta + \varepsilon_t$, t = 1, ... , T,

$(-\varepsilon_t)$ are i.i.d. U(0,1).

Then, $\Pr(y_t = 1) = x_{t\bullet}'\beta$ (linear) (Heckman and Snyder, Rand, 1997).

End of Digression

- Log-likelihood Function of the Probit model
  - $L_T(\beta) = \Pi_{t=1}^{T} f(y_t)$.
  - $l_T(\beta) = \Sigma_t \ln(f(y_t)) = \Sigma_t\left\{y_t \ln\Phi(x_{t\bullet}'\beta) + (1-y_t)\ln\left(1-\Phi(x_{t\bullet}'\beta)\right)\right\}$

- Some useful facts:
  - $E(y_t) = \Phi(x_{t\bullet}'\beta)$.
  - $\dfrac{\partial\Phi(x_{t\bullet}'\beta)}{\partial\beta} = \dfrac{\partial\Phi(x_{t\bullet}'\beta)}{\partial x_{t\bullet}'\beta}\dfrac{\partial x_{t\bullet}'\beta}{\partial\beta} = \phi(x_{t\bullet}'\beta)x_{t\bullet}$.

- $\frac{\partial^2\Phi(x_{t\bullet}'\beta)}{\partial\beta\partial\beta'} = \left(\frac{\partial^2\Phi(x_{t\bullet}'\beta)}{\partial\beta_i\partial\beta_j}\right)_{k\times k} = -(x_{t\bullet}'\beta)\phi(x_{t\bullet}'\beta)x_{t\bullet}x_{t\bullet}'$.

- $\frac{\partial l_T(\beta)}{\partial\beta} = \Sigma_t\left\{y_t\frac{\partial\ln\Phi(x_{t\bullet}'\beta)}{\partial\beta} + (1-y_t)\frac{\partial\ln(1-\Phi(x_{t\bullet}'\beta))}{\partial\beta}\right\}$

$$= \Sigma_t\left\{y_t\frac{\phi(x_{t\bullet}'\beta)}{\Phi(x_{t\bullet}'\beta)}x_{t\bullet} + (1-y_t)\frac{-\phi(x_{t\bullet}'\beta)}{1-\Phi(x_{t\bullet}'\beta)}x_{t\bullet}\right\}$$

$$= \Sigma_t\frac{(y_t-\Phi(x_{t\bullet}'\beta))\phi(x_{t\bullet}'\beta)}{\Phi(x_{t\bullet}'\beta)(1-\Phi(x_{t\bullet}'\beta))}x_{t\bullet}.$$

- Numerical Property of the MLE of β ($\hat{\beta}$)

  - $\frac{\partial l_T(\hat{\beta})}{\partial\beta} = \Sigma_t\frac{(y_t-\Phi(x_{t\bullet}'\hat{\beta}))\phi(x_{t\bullet}'\hat{\beta})}{\Phi(x_{t\bullet}'\hat{\beta})(1-\Phi(x_{t\bullet}'\hat{\beta}))}x_{t\bullet} = 0_{k\times 1}$.

  - $H_T(\hat{\beta}) = \frac{\partial l_T(\hat{\beta})}{\partial\beta\partial\beta'}$ should be negative definite.

    [See Judge, et al for the exact form of H_T]

  - $l_T(\beta)$ is globally concave with respect to β; that is, $H_T(\beta)$ is negative definite. [Amemiya (1985, Advanced Econometrics)].

  - Use $[-H_T(\hat{\beta})]^{-1}$ as $Cov(\hat{\beta})$.

- Consistency of MLE
  - Almost Theorem:

    For any model with a unknown p×1 vector of unknown parameters θ,

    MLE is consistent if $E\left(\frac{\partial l_T(\theta)}{\partial \theta}\right) = 0_{p\times 1}$.

  - For Probit;

$$E\left(\frac{\partial l_T(\theta)}{\partial \theta}\right) = E\left(\Sigma_t \frac{(y_t - \Phi(x_{t\bullet}'\beta))\phi(x_{t\bullet}'\beta)}{\Phi(x_{t\bullet}'\beta)(1-\Phi(x_{t\bullet}'\beta))} x_{t\bullet}\right)$$

$$= \Sigma_t \frac{(E(y_t) - \Phi(x_{t\bullet}'\beta))\phi(x_{t\bullet}'\beta)}{\Phi(x_{t\bullet}'\beta)(1-\Phi(x_{t\bullet}'\beta))} x_{t\bullet}$$

$$= \Sigma_t \frac{(\Phi(x_{t\bullet}'\beta) - \Phi(x_{t\bullet}'\beta))\phi(x_{t\bullet}'\beta)}{\Phi(x_{t\bullet}'\beta)(1-\Phi(x_{t\bullet}'\beta))} x_{t\bullet} = 0_{k\times 1}.$$

$\rightarrow \hat{\beta}$ is consistent.

- How to find MLE (See Greene Ch. 5 or Hamilton, Ch. 5)

1. Newton-Raphson's algorithm:

   STEP 1: Choose an initial $\hat{\theta}_o$. Then compute

   (*) $\hat{\theta}_1 = \hat{\theta}_o + [-H_T(\hat{\theta}_o)]^{-1} s_T(\hat{\theta}_o)$.

   STEP 2: Using $\hat{\theta}_1$, compute $\hat{\theta}_2$ by (*).

   STEP 3: Continue until $\hat{\theta}_{q+1} \cong \hat{\theta}_q$.

Note: N-R method is the best if $l_T(\theta)$ is globally concave (i.e., the Hessian matrix is always negative definite for any θ). N-R may not work, if $l_T(\theta)$ is not globally concave.

2. BHHH [Berndt, Hall, Hall, Hausman]

- $l_T(\theta) = \Sigma_t \ln[f_t(\theta)]$.
- Define:

$$g_t(\theta) = \frac{\partial \ln[f_t(\theta)]}{\partial \theta} \text{ [p×1] } (s_T(\theta) = \Sigma_t g_t(\theta).)$$

$$B_T(\theta) = \Sigma_t g_t(\theta) g_t(\theta)' \text{ [cross product of first derivatives].}$$

Theorem: Under suitable regularity conditions,

$$\frac{1}{T} B_T(\hat{\theta}) \rightarrow_p \lim_{T\rightarrow\infty} E\left(-\frac{1}{T} H_T(\theta_o) \right).$$

Implication:

- $B_T(\hat{\theta}) \approx -H_T(\hat{\theta})$, as $T \rightarrow \infty$.

  $Cov(\hat{\theta})$ can be estimated by $[B_T(\hat{\theta})]^{-1}$ or $[-H_T(\hat{\theta})]^{-1}$.

- BHHH algorithm uses

$$\hat{\theta}_1 = \hat{\theta}_o + \lambda_o \left(B_T(\hat{\theta}_o) \right)^{-1} s_T(\hat{\theta}_o),$$

  where λ is called step length.

- When BHHH is used, no need to compute second derivatives.
- Other available algorithms: BFGS, BFGS-SC, DFP.

BHHH for Probit:

Can show $g_t(\beta) = \xi_t x_{t\bullet}$, where,

$$\xi_t = \frac{(y_t - \Phi_t)\phi_t}{\Phi_t(1-\Phi_t)}; \phi_t = \phi(x_{t\bullet}'\beta); \Phi_t = \Phi(x_{t\bullet}'\beta)$$

$$\rightarrow B_T(\hat{\beta}) = \Sigma_t \hat{g}_t \hat{g}_t' = \Sigma_t \hat{\xi}_t^2 x_{t\bullet} x_{t\bullet}'.$$

$[B_T(\hat{\beta})]^{-1}$ is $Cov(\hat{\beta})$ by BHHH.

- Interpretation of β

1) β_j shows direction of influence of x_{tj} on $\Pr(y_t = 1) = \Phi(x_{t\bullet}'\beta)$.

$\rightarrow \beta_j > 0$ means that $\Pr(y_t=1)$ increases with x_{tj}

2) Rate of change:

$$\frac{\partial \Pr(y_t = 1)}{\partial x_{tj}} = \frac{\partial \Phi(x_{t\bullet}'\beta)}{\partial x_{tj}} = \phi(x_{t\bullet}'\beta)\beta_j.$$

- Estimation of probabilities and rates of changes
  - Estimation of $p_t = \Pr(y_t=1)$ at mean of $x_{t\bullet}$
    - Use $\hat{p} = \Phi(\bar{x}'\hat{\beta})$.
    - $\text{var}(\hat{p}) = \left(\phi(\bar{x}'\hat{\beta})\right)^2 \bar{x}'\hat{\Omega}\bar{x}$ where $\hat{\Omega} = Cov(\hat{\beta})$ [by delta-method].

- Estimation of rates of change
    - Use $\hat{p}^j = \frac{\partial \Phi(\bar{x}'\hat{\beta})}{\partial x_{tj}} = \phi(\bar{x}'\hat{\beta})\hat{\beta}_j$.
    - $\text{var}(\hat{p}^j) = \frac{\partial p^j(\hat{\beta})}{\partial \beta'} \hat{\Omega} \frac{\partial p^j(\hat{\beta})}{\partial \beta}$. [by delta-method].
    - Note that:

$$\frac{\partial p^j(\beta)}{\partial \beta'} = -(\bar{x}'\beta)\phi(\bar{x}'\beta)\beta_j \bar{x}' + \phi(\bar{x}'\beta)J_j,$$

where J_j = 1×k vector of zeros except that the j'th element = 1.

- Note on normalization:
    - Model: $y_t^* = x_{t\bullet}'\beta + \varepsilon_t$, $-\varepsilon_t \sim N(0,\sigma^2)$.

      $y_t = 1$ iff $y_t^* > 0$.
    - $p_t = \Pr(y_t = 1) = \Pr(y_t^* > 0) = \Pr(x_{t\bullet}'\beta + \varepsilon_t > 0) = \Pr(-\varepsilon_t < x_{t\bullet}'\beta)$

      $= \Pr(-\varepsilon_t/\sigma < x_{t\bullet}'(\beta/\sigma)) = \Phi[x_{t\bullet}'(\beta/\sigma)]$.
    - Can estimate β/σ, but not β and σ separately.

- Testing Hypothesis:

1. Wald test:
    - H_o: $w(\beta) = 0$.
    - $W_T = w(\hat{\beta})'\left[W(\hat{\beta})\hat{\Omega}W(\hat{\beta})'\right]^{-1} w(\hat{\beta}) \to_d \chi^2$(df = # of restrictions),

    where $\hat{\beta}$ = probit MLE and $W(\beta) = \frac{\partial w(\beta)}{\partial \beta'}$.

2. LR test:

- Easy for equality or zero restrictions (i.e., H_o: $\beta_2 = \beta_3$, or H_o: $\beta_2 = \beta_3 = 0$).
- EX 1: Suppose you wish to test H_o: $\beta_4 = \beta_5 = 0$.

  STEP 1: Do Probit without restriction and get $l_{T,UR} = \ln(L_{T,UR})$.

  STEP 2: Do Probit with the restrictrions and get $l_{T,R} = \ln(L_{T,R})$.

  $\rightarrow$ Probit without x_{t4} and x_{t5}.

  STEP 3: $LR_T = 2[l_{T,UR} - l_{T,R})] \Rightarrow \chi^2(df = 2)$.

- EX 2: Suppose you wish to test H_o: $\beta_2 = ... = \beta_k = 0$.

  (Overall significance test)

  - Let $n = \Sigma_t y_t$.
  - $l_T^* = n \ln(n/T) + (T-n) \ln[(T-n)/T]$.
  - $LR_T = 2[l_{T,UR} - l_T^*] \rightarrow_p \chi^2(k-1)$.

- Pseudo-R^2 (McFadden, 1974)
  - $\rho^2 = 1 - l_{T,UR} / l_T^*$.

    $\rightarrow 0 \leq \rho^2 \leq 1$.
  - If $\Phi(x_{t\bullet}'\hat{\beta}) = 1$ whenever $y_t = 1$, and if $\Phi(x_{t\bullet}'\hat{\beta}) = 0$ whenever $y_t = 0$, $\rho^2 = 1$.
  - If $0 < \rho^2 < 1$, no clear meaning.

LIMDEP CORNER:

YOU CAN ACCESS LIMDEP 7.0 FOR WINDOWS IN THE ECONOMICS DEPARTMENT COMPUTER ROOM. OR YOU CAN ACCESS LIMDEP THROUGH MY INSTRUCTOR□S VOLUMN FOR ECN527.

INSTRUCTION FOR ACCESSING AN INSTRUCTOR VOLUME

Special Note:

Before you can use the computers at ASU, you must first obtain an ASURITE ID. You may obtain the ASURITE ID at BAC, BA, Goldwater and Computer Commons computing sites (see the support staff for assistance). Once you receive your ASURITE ID and have confirmed that it works, please follow these steps to access my instructor volume.

Problem Tips:

A. If you have difficulty signing on, push the restart button (or turn the computer off and then on again) and start over at Step 1 below.

B. DO NOT enter your ASURITE ID anywhere EXCEPT on the screen display over the ASU logo and photograph.

The computer should already be on. If the last student did not log out and the desktop screen still shows a set of icons, click on the **Log Out** icon and then click on **Log Me Out**.

Accessing the Instructor's Volume

1. At the ASU PC Network logon you will get a message: "Click OK for the next two requests."

   Click on the **OK** button here.
   Click **OK** and wait 1-2 minutes while the logon scripts execute.
   Click **OK** to get to the sign on screen with the ASU logo displayed over an ASU photograph back ground.

2. At the sign on screen enter your **ASURITE ID** and **password**. Enter both items in <u>lower case</u>.
   Click on **OK**.
   Wait during the message "Mounting AFS volumes." Soon the Window 95 desktop will be displayed.

3. Double click on the **Applications** folder icon on the desktop.

4. Double click on the **Instructor Volumes** folder icon on the desktop.

5. Find the icon named **ECN527**, and double click on it.

6. The U: drive instructor volume is now mounted but you cannot see it until the current window is closed. Close the instructor volume window by clicking on **X** in the upper right corner of the window.

7. Double click on the **U: drive** icon on the desktop.

8. Go to the directory Limdep/Program. Click on the **LIMDEP** icon.

9. Now you entered the LIMDEP program (Version 7.0 for windows).

When you have finished using the instructors volume, be sure to LOG OUT so that the next computer user does not have access to your files.

10. Double click on the **Log Out** icon on the desktop. Click on **Log Me Out**. DO NOT turn the computer off.

HOW TO READ DATA

Basic Format:

```
READ ; NOBS = ...
     ; NVAR = ...
     ; NAMES = ... (THE NAMES OF VARIABLES)
     ; FILE = ... (THE FILE CONTAINING RAW DATA)
     ; FORMAT = ... (SEE LIMDEP MANUAL)     $
```

Example: Using MWPSID82.DB (MW_READ.LIM)

```
READ ; NOBS=1962; NVAR=25
     ; FILE=MWPSID82.DB
     ; NAMES=  NLF,    EMP,    WRATE,  LRATE,  ED,
               URB,    MINOR,  AGE,    TENURE, EXP,
               REGS,   OCCW,   OCCB,   INDUMG, INDUMN,
               UNION,  UNEMPR, OFINC,  LOFINC, KIDS,
               WUNE80, HWORK,  USPELL, SEARCH, KIDS5 $

CREATE; LF = 1 - NLF $
```

(1) This problem is available in MW_READ.LIM. Run the program to read the data, MWPSID82.DB

(2) To save the data, click **File/Project save as**. Type **mw.lpj**.

INFORMATION ON MWPSID82.DB

This is the data set of married women in 1981 sampled from PSID. Total number of observations are 1962, and 25 variables are observed.

VARIABLES	DEFINITION
NLF	NLF=1 IF NON-LABOR-FORCE (HOUSEWIFE)
EMP	EMP=1 IF EMPLOYED
WRATE	HOURLY WAGE RATE ($)
LRATE	Log of WRATE = LOG(WRATE+1)
ED	YEARS OF EDUCATION
URB	URB=1 IF RESIDENT IN SMSA
MINOR	MINOR=1 IF BLACK AND HISPANIC
AGE	YEARS OF AGE
TENURE	MONTHS UNDER THE CURRENT EMPLOYER
EXP	NUMBER OF YEARS WORKED SINCE AGE 18
REGS	REGS=1 IF LIVES IN THE SOUTH OF U.S.
OCCW	OCCW=1 IF WHITE COLOR
OCCB	OCCB=1 IF BLUE COLOR
INDUMG	INDUMG=1 IF IN THE MANUFACTURING INDUSTRY
INDUMN	INDUMN=1 IF NOT IN MANUFACTURING SECTOR
UNION	UNION=1 IF UNION MEMBER
UNEMPR	% UNEMPLOYMENT RATE IN THE RESIDENT'S COUNTY, 1980
OFINC	OTHER FAMILY MEMBER'S INCOME IN 1980 ($)
LOFINC	LOG OF (OFINC+1)
KIDS	NUMBER OF CHILDREN # 17 YEARS OF AGE
HWORK	HOURS OF HOMEWORK PER WEEK
USPELL	UNEMPLOYED WEEKS FOR EMPLOYED WIFE
SEARCH	WEEKS LOOKING FOR JOB IN 1980
WUNE80	[UNKNOWN]
KIDS5	NUMBER OF CHILDREN # 5 YEARS OF AGE

PROBIT ESTIMATION

Example 1: (mw_prob1.lim)

```
/* Basic Program */

probit ; lhs = emp
       ; rhs = one,ed,urb,minor,lofinc
       ; maxit = 1000
       ; start = 0,0,0,0,0
       ; tlf = 0.00001 ; tlb = 0.00001 ; tbg = 0.00001
       ; alg = newton   ? Can choose bhhh, bfgs, dfp, stedes
       ; margin $       ? Estimate dPr(y=1)/dx at sample mean
```

OUTPUT

```
+-----------------------------------------------------------------------+
| Dependent variable is binary, y=0 or y not equal 0                    |
| Ordinary    least squares regression    Weighting variable = none     |
| Dep. var. = EMP      Mean=   .4704383282    , S.D.=   .4992525893     |
| Model size: Observations =    1962, Parameters =   5, Deg.Fr.=   1957 |
| Residuals:  Sum of squares= 474.9962035    , Std.Dev.=         .49266 |
| Fit:        R-squared=  .028211, Adjusted R-squared =          .02622 |
| Model test: F[  4,   1957] =   14.20,    Prob value =          .00000 |
| Diagnostic: Log-L =  -1392.4945, Restricted(b=0) Log-L =   -1420.5675 |
|             LogAmemiyaPrCrt.=   -1.413, Akaike Info. Crt.=      1.425 |
+-----------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant  .9170578290      .17362375        5.282   .0000
 ED        .2915849316E-01  .51513883E-02    5.660   .0000  12.205403
 URB       .6459244774E-01  .24441441E-01    2.643   .0082  .68654434
 MINOR     .2566128748E-01  .26742220E-01     .960   .3373  .27166157
 LOFINC   -.8619967756E-01  .17367289E-01   -4.963   .0000  9.9052275

Normal exit from iterations. Exit status=0.
```

```
+---------------------------------------------+
| Binomial Probit Model                       |
| Maximum Likelihood Estimates                |
| Dependent variable                EMP       |
| Weighting variable                ONE       |
| Number of observations           1962       |
| Iterations completed                4       |
| Log likelihood function     -1327.781       |
| Restricted log likelihood   -1356.524       |
| Chi-squared                  57.48581       |
| Degrees of freedom                  4       |
| Significance level           .0000000       |
+---------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
	Index function for probability				
Constant	1.228085172	.47320614	2.595	.0095	
ED	.7629799947E-01	.13446644E-01	5.674	.0000	12.205403
URB	.1708719953	.63111585E-01	2.707	.0068	.68654434
MINOR	.5968648933E-01	.68899573E-01	.866	.3863	.27166157
LOFINC	-.2390755442	.48122027E-01	-4.968	.0000	9.9052275

```
+-------------------------------------------+
| Partial derivatives of E[y] = F[*]   with |
| respect to the vector of characteristics. |
| They are computed at the means of the Xs. |
| Observations used for means are All Obs.  |
+-------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
	Index function for probability				
Constant	.4885503224	.18831825	2.594	.0095	
ED	.3035246504E-01	.53493381E-02	5.674	.0000	12.205403
URB	.6797538994E-01	.25106007E-01	2.708	.0068	.68654434
MINOR	.2374416228E-01	.27409137E-01	.866	.3863	.27166157
LOFINC	-.9510776362E-01	.19144903E-01	-4.968	.0000	9.9052275

Frequencies of actual & predicted outcomes
Predicted outcome has maximum probability.

	Predicted		
Actual	0	1	Total
0	740	299	1039
1	527	396	923
Total	1267	695	1962

Example 2: (mw_prob2.lim)

```
/* Testing hypotheses */

/*
   In LIMDEP7, Wald tests can be computed for any restriction
               LR tests are handy only for certain restrictions:
                    equality between parameters or zero restriction
  For example,

        Hypo 1: b(1) = 0, b(2) = b(3)
        Hypo 2: b(1) = 0, b(2) = b(3), b(4) = b(2)+ b(3) + b(5)
        Hypo 3: b(1) = 0, b(2) = b(3), b(4)^2 = b(5)

  Wald can be used for any of these hypotheses.
  LR is easy to use for Hypo 1.
*/

/* Testing hypo 1 */

? Unrestricted Model

namelist ; x   = one,ed,urb,minor,lofinc $
probit   ; lhs = emp
         ; rhs = x
         ; maxit = 1000         $

matrix ; uprb = b ; uprc = varb $
calc   ; ulogl = logl           $

? Restricted Model

probit ; lhs = emp
       ; rhs = x
       ; maxit = 1000
       ; rst = 0,b2,b2,b4,b5 $

matrix ; rprb = b     $
calc   ; rlogl = logl $
```

```
? [Wald Test]

title; Wald test for b1 = 0 and b2 = b3 $
wald ; labels = b1,b2,b3,b4,b5
     ; start  = uprb
     ; var    = uprc
     ; fn1    = b1
     ; fn2    = b2 - b3      $

? [LR test]

title; LR test for b1 = 0 and b2 = b3 $
calc ; list
     ; lrt  = 2*(ulogl - rlogl)
     ; pval = 1 - chi(lrt,2) $

/* Wald for hypo 3 */

title; Wald test for b1 = 0, b2 = b3 and b4^4 = b5  $
wald ; labels = b1,b2,b3,b4,b5
     ; start  = uprb
     ; var    = uprc
     ; fn1    = b1
     ; fn2    = b2 - b3
     ; fn3    = b4^2 - b5  $
```

[Output]

Unrestricted Model:

```
+---------------------------------------------+
| Binomial Probit Model                       |
| Maximum Likelihood Estimates                |
| Dependent variable                 EMP      |
| Weighting variable                 ONE      |
| Number of observations            1962      |
| Iterations completed                 4      |
| Log likelihood function      -1327.781      |
| Restricted log likelihood    -1356.524      |
| Chi-squared                   57.48581      |
| Degrees of freedom                   4      |
| Significance level            .0000000      |
+---------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
	Index function for probability				
Constant	1.228085188	.47320614	2.595	.0095	
ED	.7629799963E-01	.13446644E-01	5.674	.0000	12.205403
URB	.1708719958	.63111585E-01	2.707	.0068	.68654434
MINOR	.5968648873E-01	.68899573E-01	.866	.3863	.27166157
LOFINC	-.2390755461	.48122027E-01	-4.968	.0000	9.9052275

Frequencies of actual & predicted outcomes
Predicted outcome has maximum probability.

Actual	Predicted 0	Predicted 1	Total
0	740	299	1039
1	527	396	923
Total	1267	695	1962

Restricted Model:

```
               +---------------------------------------------+
               | Binomial Probit Model                       |
               | Maximum Likelihood Estimates                |
               | Dependent variable                  EMP     |
               | Weighting variable                  ONE     |
               | Number of observations             1962     |
               | Iterations completed                  4     |
               | Log likelihood function       -1331.862     |
               | Restricted log likelihood     -1356.524     |
               | Chi-squared                    49.32278     |
               | Degrees of freedom                    2     |
               | Significance level             .0000000     |
               +---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Index function for probability
 Constant  .0000000000    ........(Fixed Parameter)........
 ED        .8579347233E-01  .12778130E-01    6.714   .0000  12.205403
 URB       .8579347233E-01  .12778130E-01    6.714   .0000  .68654434
 MINOR     .1336072788      .63756953E-01    2.096   .0361  .27166157
 LOFINC   -.1233762049      .17163930E-01   -7.188   .0000  9.9052275
```

```
Frequencies of actual & predicted outcomes
Predicted outcome has maximum probability.

            Predicted
------  ----------  +  -----
Actual      0    1  |  Total
------  ----------  +  -----
  0       756  283  |   1039
  1       571  352  |    923
------  ----------  +  -----
Total    1327  635  |   1962
```

Wald Test for Hypo 1:

```
               +-----------------------------------------------+
               | WALD procedure. Estimates and standard errors |
               | for nonlinear functions and joint test of     |
               | nonlinear restrictions.                       |
               | Wald Statistic             =      8.06527     |
               | Prob. from Chi-squared[ 2] =       .01773     |
               +-----------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Fncn( 1)  1.228085188      .47320614        2.595   .0095
 Fncn( 2) -.9457399616E-01  .65465798E-01   -1.445   .1486
```

LR test for Hypo 1:

```
--> title; LR test for b1 = 0 and b2 = b3 $
--> calc ; list
    ; lrt  = 2*(ulogl - rlogl)
    ; pval = 1 - chi(lrt,2) $
    LRT     =  .81630295544032380D+01
    PVAL    =  .16881874007718900D-01
```

Wald test for Hypo 3:

```
               +------------------------------------------------+
               | WALD procedure. Estimates and standard errors |
               | for nonlinear functions and joint test of     |
               | nonlinear restrictions.                        |
               | Wald Statistic              =     52.38105     |
               | Prob. from Chi-squared[ 3] =       .00000     |
               +------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Fncn( 1)  1.228085188      .47320614        2.595   .0095
 Fncn( 2) -.9457399616E-01  .65465798E-01   -1.445   .1486
 Fncn( 3)  .2426380230      .46728575E-01    5.192   .0000
```

[EXERCISE]Use MWPSID82.DB. Do probit for:

LHS=EMP,
RHS=ONE, ED, URB, MINOR, AGE, REGS, UNEMPR, LOFINC, KIDS5, EXP.

Construct Wald, LR and LM statistics for H_o: no effect of other family income, and the effect of AGE = the effect of EXP.

[EXERCISE]Use MWPSID82.DB. Do PROBIT as you did above. Estimate the same model by BHHH. Compare your new result with the result from the above program.

(3) Logit Models

- Model:

  $y_t^* = x_{t\bullet}'\beta + \varepsilon_t$,

  ε_t ~ logistic with $g(\varepsilon) = e^{\varepsilon}/(1+e^{\varepsilon})^2$ and $G(\varepsilon) = e^{\varepsilon}/(1+e^{\varepsilon})$.

- Use $\Pr(y_t = 1) \equiv p_t = G(x_{t\bullet}'\beta)$ (instead of $\Phi(x_{t\bullet}'\beta)$).

- Logit MLE $\hat{\beta}_{\log it}$ max.

$$\ln(L_T) = \Sigma_t \left\{ y_t \ln\left(G(x_{t\bullet}'\beta)\right) + (1-y_t)\ln\left(1-G(x_{t\bullet}'\beta)\right) \right\}.$$

  Use $[-H_T(\hat{\beta}_{\log it})]^{-1}$ or $[B_T(\hat{\beta}_{\log it})]^{-1}$ as $Cov(\hat{\beta}_{\log it})$.

- Interpretation of β

  - $p_t = \dfrac{e^{x_{t\bullet}'\beta}}{1+e^{x_{t\bullet}'\beta}} \quad \rightarrow \quad \ln\left(\dfrac{p_t}{1-p_t}\right) = x_{t\bullet}'\beta$.

    → β_j can be interpreted as the effect of x_{jt} on "log odds".

  - $\dfrac{\partial p_t}{\partial x_{jt}} = g(x_{t\bullet}'\beta)\beta_j$.

- Facts:
  - The logistic dist. is quite similar to standard normal dist. except that the logistic dist. has thicker tails (similarly to t(7)).
  - If data contain few obs. with y = 1 or y = 0, then probit and logit may be quite different. Other than that, probit and logit yield very similar predictions. Especially, marginal effects are quite similar.
  - Roughly, $\hat{\beta}_{\log it} = 1.6\hat{\beta}_{probit}$.

Empirical Example:

[Program (mw_log.lim)]

```
title  ; Employment Probability $
logit  ; lhs = emp
       ; rhs = one,ed,urb,minor,lofinc
       ; margin $
```

[Output]

```
              +---------------------------------------------+
              | Multinomial Logit Model                     |
              | Maximum Likelihood Estimates                |
              | Dependent variable                 EMP      |
              | Weighting variable                 ONE      |
              | Number of observations            1962      |
              | Iterations completed                 4      |
              | Log likelihood function      -1327.765      |
              | Restricted log likelihood    -1356.524      |
              | Chi-squared                   57.51681      |
              | Degrees of freedom                   4      |
              | Significance level            .0000000      |
              +---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Characteristics in numerator of Prob[Y = 1]
 Constant  1.966854554      .76093734        2.585   .0097
 ED         .1232783526     .21886605E-01    5.633   .0000  12.205403
 URB        .2743110439     .10148559        2.703   .0069  .68654434
 MINOR      .9461106393E-01 .11073756         .854   .3929  .27166157
 LOFINC   -.3842593851      .77640232E-01   -4.949   .0000  9.9052275


+--------------------------------------------+
| Partial derivatives of probabilities with |
| respect to the vector of characteristics. |
| They are computed at the means of the Xs. |
| Observations used for means are All Obs.  |
+--------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Marginal effects on Prob[Y = 0]
 Constant  .4899291263      .18963601        2.584   .0098
 ED        .3070773865E-01  .54513459E-02    5.633   .0000  12.205403
 URB       .6832888064E-01  .25278167E-01    2.703   .0069  .68654434
 MINOR     .2356692608E-01  .27583842E-01     .854   .3929  .27166157
 LOFINC   -.9571621064E-01  .19341033E-01   -4.949   .0000  9.9052275


Frequencies of actual & predicted outcomes
Predicted outcome has maximum probability.

            Predicted
------  ----------  +  -----
Actual      0    1  |  Total
------  ----------  +  -----
  0       740  299  |   1039
  1       527  396  |    923
------  ----------  +  -----
Total    1267  695  |   1962
```

(4) Nonparametric estimation of binary choice model

1) Cosslett (Econometrica, 1983)

- See also Amemiya (1985, book)
- $\Pr(y_t = 1) = F(x_{t\bullet}'\beta)$, where F is a unknown cdf.
- Joint estimation of β and F is feasible, although it is not easy.
- Asymptotic distribution of the estimator is not known.

2) Nonparametric Estimation of $F(x_{t\bullet}'\beta)$

- For binary choice models,

  $E(y_t|x_{t\bullet}) = F(x_{t\bullet})$ ($F(\bullet)$ = pdf of ε)

  $\rightarrow$ For example, $F(x_{t\bullet}) = \Phi(x_{t\bullet}'\beta)$ for probit.

  $\rightarrow$ The functional form of $F(\bullet)$ is not known in general.

- Possible to estimate $F(x_{t\bullet}'\beta)$ [but not F and β] for any t

  by Kernel Smoothing.

  $\rightarrow$ See Härdle (1990, Applied Nonparametric Regression.)

- LIMDEP can do this.

3) Nonparametric Estimation of β:

See Powell, Stock and Stoker (1989, Econ, 1403-30).

4) Manski (Journal of Econometrics, 1975)

- "Maximum Score Estimator." (MSE)
- Motivation: The distribution of ε_t not known.
- Assumptions:
  - $Med(\varepsilon_t) = 0 \rightarrow Pr(\varepsilon_t < 0) = 1/2$.
  - The $x_{t\bullet}$ are iid over t.
- The model:

  $y_t^* = x_{t\bullet}'\beta + \varepsilon_t$; $y_t = 1$ iff $y_t^* > 0$.
- Define:

  $z_t = sgn(y_t^*) = 1$ if $y_t^* > 0$, and $= -1$, if $y_t^* < 0$.
- Define $b = \beta/(\beta'\beta)^{1/2}$. [Note that $b'b = 1$.]

  [Need it for identification.]
- The MSE estimator, $\hat{b}$, maximizes

  $S(b) = (1/N)\, \Sigma_t[z_t sgn(x_{t\bullet}'b)]$.
- Intuition:
  - $sgn(x_{t\bullet}'\hat{b})$ = predicted z_t.
  - If the prediction is correct, $z_t\, sgn(x_{t\bullet}'\hat{b}) = 1$.
  - If the prediction is incorrect, $z_t\, sgn(x_{t\bullet}'\hat{b}) = -1$.
  - max. S(b)

    = max. # of correct predictions with penalty !!!

- Maximizing S(b) is equivalent to:

  min $\Sigma_t|y_t$ - max(0,sgn($x_{t\bullet}'b$))|. (*)

- LIMDEP uses (*). [you don't have to define z_t.]
- Properties of MSE:
  - Consistent.
  - It does not have a standard asymptotic distribution.
  - LIMDEP computes covariance matrix of $\hat{b}$ using bootstrapping. But the method is not based on clean theories.

[2] Censoring vs. Truncation

(Greene, ch. 20)

(1) Classical distinction

- Consider shots on target.

  Truncation: cases where you have data on "hole" only.

  Censoring: cases where you know how many shots missed.

(2) Censoring

- $y^* \sim$ pdf: $f(y^*)$.
- Observe $y = y^*$ if $A < y^* < B$; A if $y^* \leq A$; B if $y^* \geq B$.

  (For obs. with $y = A$ or $y = B$, y^* is unknown.)
- Log-likelihood function:

  $OB = \{t|y_t^* \text{ observed}\}$; $NOB = \{t|y_t^* \text{ unobserved}\}$,

$$l_T = \Sigma_{t\in OB} \ln\left[f(y_t^* | t\in OB)\times \Pr(t\in OB)\right] + \Sigma_{t\in NOB} \ln\left(\Pr(t\in NOB)\right).$$

- Note:

$f(y_t^*|t\in OB)\Pr(t\in OB)$

$$= f(y_t^*|A < y_t^* < B)\Pr(A < y_t^* < B)$$

$$= [f(y_t^*)/\Pr(A < y_t^* < B)]\Pr(A < y_t^* < B) = f(y_t^*).$$

$$\to l_T = \sum_{A<y_t<B} \ln\left(f(y_t^*)\right) + \sum_{y_t=A} \ln\left(\Pr(y_t^* \leq A)\right) + \sum_{y_t=B} \ln\left(\Pr(y_t^* \geq B)\right)$$

$$\to l_T = \sum_{A<y_t<B} \ln\left(f(y_t)\right) + \sum_{y_t=A} \ln\left(\Pr(y_t^* \leq A)\right) + \sum_{y_t=B} \ln\left(\Pr(y_t^* \geq B)\right)$$

(3) Truncation

- Observe $y = y^*$ iff $A \le y^* \le B$
- Log-likelihood function:

pdf of y_t: $g(y_t) = f(y_t^* | A \le y_t^* \le B)$

$$= \frac{f(y_t^*)}{\Pr(A \le y_t^* \le B)} = \frac{f(y_t)}{\Pr(A \le y_t^* \le B)}.$$

$$l_T = \Sigma_t \left\{ \ln(f(y_t)) - \ln[\Pr(A \le y_t^* \le B) \right\}.$$

(4) Tobit (A censored model)

1) Latent model: $y_t^* = x_t.'\beta + \varepsilon_t$, ε_t iid $N(0,\sigma^2)$. $[y_t^* \sim N(x_t.'\beta, \sigma^2)]$

2) 3 possible cases:

A. Observe $y_t = y_t^*$ if $y_t^* > 0$; = 0, otherwise. $\rightarrow y_t = \max(0, y_t^*)$

B. Observe $y_t = y_t^*$ if $y_t^* < 0$; = 0 otherwise. $\rightarrow y_t = \min(0, y_t^*)$.

C. Observe $y_t = y_t^*$ if $y_t^* < L_t$; = L_t otherwise.

3) Log-likelihood for A

- $\Pr(y_t^* \le 0) = \Pr(x_t.'\beta + \varepsilon_t \le 0) = \Pr(\varepsilon_t \le -x_t.'\beta)$

$$= \Pr(\varepsilon_t/\sigma \le -x_t.'(\beta/\sigma)) = 1 - \Phi[x_t.'(\beta/\sigma)].$$

- $$f(y_t^*) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_t^* - x_t.'\beta)^2}{2\sigma^2} \right).$$

- Therefore,

$$l_T(\beta,\sigma) = \sum_{y_t>0} \ln f(y_t^*) + \sum_{y_t=0} \ln\left(1-\Phi\left(\frac{x_{t\bullet}'\beta}{\sigma}\right)\right)$$

$$= \sum_{y_t>0} \ln f(y_t) + \sum_{y_t=0} \ln\left(1-\Phi\left(\frac{x_{t\bullet}'\beta}{\sigma}\right)\right)$$

$$= \sum_{y_t>0}\left\{-\frac{1}{2}\ln(2\pi) - \ln(\sigma) - \frac{1}{2\sigma^2}(y_t - x_{t\bullet}'\beta)^2\right\}$$

$$+\sum_{y_t=0} \ln\left(1-\Phi\left(\frac{x_{t\bullet}'\beta}{\sigma}\right)\right)$$

4) Estimation procedure

- Let $\beta_o = \beta/\sigma$ and $h = 1/\sigma$.
- l_T is globally concave in terms of β_o and h.
- MLE for β_o and h by N-R.
- Convert them to β and σ.

5) Interpretation:

(i) $E(y_t^*)$ = E[latent var. (e.g., desired consumption)] = $x_{t\bullet}'\beta$.

$$\rightarrow \beta_j = \frac{\partial E(y_t^*)}{\partial x_{tj}}.$$

(ii) $E(y_t)$ = E[observed variable (e.g., actual expenditure)]

$= \Pr(y_t^* \geq 0)E(y_t^*|y_t^* \geq 0) + \Pr(y_t^* < 0)E(0|y_t^* < 0)$

$= \Phi(x_{t\bullet}'\beta/\sigma)E(x_{t\bullet}'\beta + \varepsilon_t|\varepsilon_t \geq -x_{t\bullet}'\beta)$

$= \Phi(x_{t\bullet}'\beta/\sigma)[x_{t\bullet}'\beta + \sigma\lambda(x_{t\bullet}'\beta/\sigma)]$

[where $\lambda(x_{t\bullet}'\beta/\sigma) = \phi(x_{t\bullet}'\beta/\sigma)/\Phi(x_{t\bullet}'\beta/\sigma)$ (inverse Mill's ratio)]

$= \Phi(x_{t\bullet}'\beta/\sigma)x_{t\bullet}'\beta + \sigma\phi(x_{t\bullet}'\beta/\sigma)$

Note:

$$\frac{\partial E(y_t)}{\partial x_{tj}} = \phi\left(\frac{x_{t\bullet}'\beta}{\sigma}\right)\left(\frac{\beta_j}{\sigma}\right)(x_{t\bullet}'\beta) + \Phi\left(\frac{x_{t\bullet}'\beta}{\sigma}\right)\beta_j$$

$$+\sigma\left(-x_{t\bullet}'\left(\frac{\beta}{\sigma}\right)\right)\phi\left(\frac{x_{t\bullet}'\beta}{\sigma}\right)\frac{\beta_j}{\sigma}$$

$$= \Phi\left(\frac{x_{t\bullet}'\beta}{\sigma}\right)\beta_j$$

6) Estimation of $E(y^*)$ and E(y)

- Let $g_1(\beta) = \bar{x}'\beta$.
  - Estimated $E(y_t^*)$ at sample mean = $g_1(\hat{\beta})$.
  - $se = \sqrt{\hat{G}_1\hat{\Omega}\hat{G}_1'}$, where $\hat{\Omega} = Cov(\hat{\beta})$.

- Let $g_2(\beta,\sigma) = \Phi\left(\frac{\bar{x}'\beta}{\sigma}\right)\bar{x}'\beta + \sigma\phi\left(\frac{\bar{x}'\beta}{\sigma}\right)$.
  - Estimated $E(y_t)$ at sample mean = $g_2(\hat{\beta},\hat{\sigma})$.
  - $G_2(\beta,\sigma) = \frac{\partial g_2}{\partial(\beta',\sigma)} = \left[\Phi\left(\frac{\bar{x}'\beta}{\sigma}\right)\bar{x}', \phi\left(\frac{\bar{x}'\beta}{\sigma}\right)\right]$.
  - se = $\sqrt{\hat{G}_2\ddot{\Omega}\hat{G}_2'}$, where $\ddot{\Omega} = Cov\begin{pmatrix}\hat{\beta}\\ \hat{\sigma}\end{pmatrix}$.

[Empirical Example]

The model is given: $y_i^* = x_{i\bullet}'\beta + \varepsilon_i$; ε_i iid $N(0,\sigma^2)$.

Basic Command:

(1) $y_i = \max(0,y_i^*)$:

```
TOBIT; LHS = Y ; RHS = ONE,X1,X2,... ; MARGIN ; PAR $
```

(2) $y_i = \min(0,y_i^*)$:

```
TOBIT; LHS = Y ; RHS = ONE,X1,X2,... ; UPPER ; MARGIN ; PAR $
```

(3) $y_i = \min(L_i,y_i^*)$

```
TOBIT; LHS = Y ; RHS = ONE,X1,X2,... ; UPPER ; LIMIT = L ; MARGIN
     ; PAR $
```

NOTE: For truncation model, replace TOBIT by TRUNC.
NOTE: MARGIN computes $\partial E(y_i|x_i)/\partial x_i$.
NOTE: The estimate of σ is stored in S. Can retrieve it using the CALC command.
NOTE: If you want to use a subsample, you can use REJECT command. For example,

```
reject; emp # 1 $
```

If you execute this command, LIMDEP will choose observations for employed people.

NOTE: If you want to return to the whole sample, type:

```
sample; all $
```

Program (mw_tob.lim)

```
reject ; emp # 1 $
dstats ; rhs =uspell,kids5, kids, ed, age, exp $
tobit  ; lhs = uspell
       ; rhs = one,kids5,kids,ed,lofinc, age, exp, wrate,
            occw, occb
       ; maxit = 1000
       ; par
       ; margin  $
```

[OUTPUT]

```
                             Descriptive Statistics
               All results based on nonmissing observations.
Variable        Mean          Std.Dev.         Minimum          Maximum      Cases
--------------------------------------------------------------------------------
USPELL    2.65438787      7.95095799      .000000000       52.0000000        923
KIDS5     .322860238      .656840899      .000000000       4.00000000        923
KIDS      1.21993499      1.33139677      .000000000       8.00000000        923
ED        12.4572048      2.26427973      3.00000000       17.0000000        923
AGE       37.5904659      10.8058791      19.0000000       60.0000000        923
EXP       10.2481040      7.37794275      1.00000000       40.0000000        923


              +---------------------------------------------+
              | Limited Dependent Variable Model - CENSORED |
              | Maximum Likelihood Estimates                |
              | Dependent variable              USPELL      |
              | Weighting variable                 ONE      |
              | Number of observations             923      |
              | Iterations completed                 8      |
              | Log likelihood function      -933.4950      |
              | Threshold values for the model:             |
              | Lower=     .0000     Upper=+infinity        |
              +---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Primary Index Equation for Model
 Constant -4.833973094      22.303662        -.217   .8284
 KIDS5     3.582724387      2.4633637        1.454   .1458  .32286024
 KIDS     -.1792458053      1.2565693        -.143   .8866  1.2199350
 ED       -.2718397231      .81269621        -.334   .7380  12.457205
 LOFINC    .8886534139      2.2497026         .395   .6928  9.8432227
 AGE      -.5739202502      .21947205       -2.615   .0089  37.590466

 EXP      -.5503976319E-01  .30189369        -.182   .8553  10.248104
 WRATE    -1.864677423      .69733993       -2.674   .0075  5.7422752
 OCCW     -2.101651414      3.9873969        -.527   .5981  .63163597
 OCCB      16.70138334      4.4150633        3.783   .0002  .16468039
         Disturbance standard deviation
 Sigma     28.05515780      1.9431904       14.438   .0000


+--------------------------------------------+
| Partial derivatives of expected val. with |
| respect to the vector of characteristics. |
| They are computed at the means of the Xs. |
| Observations used for means are All Obs.  |
| Conditional Mean at Sample Point   2.0508 |
| Scale Factor for Marginal Effects   .1428 |
```

```
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant -.6903042430      3.1845764        -.217   .8284
 KIDS5     .5116225923      .35313787        1.449   .1474   .32286024
 KIDS     -.2559677879E-01  .17950044        -.143   .8866  1.2199350
 ED       -.3881943706E-01  .11607992        -.334   .7381  12.457205
 LOFINC    .1269020763      .32130726         .395   .6929  9.8432227
 AGE      -.8195734154E-01  .30992979E-01  -2.644   .0082  37.590466
 EXP      -.7859824894E-02  .43116344E-01   -.182   .8554  10.248104
 WRATE    -.2662809063      .98744284E-01  -2.697   .0070  5.7422752
 OCCW     -.3001214239      .56901180        -.527   .5979   .63163597
 OCCB      2.385002060      .62686441        3.805   .0001   .16468039
```

[EXERCISE]Use MWPSID82.DB. Consider the housework supply of employed married women:

$$HWORK_i = x_{i\bullet}'\beta + \varepsilon_i,$$

where $x_{i\bullet}$ includes ONE, ED, URB, AGE, UNEMPR, OFINC, EXP, OCCW, OCCB, TENURE, WRATE. Estimate this model by Tobit. (Some women reported zero hours on homework.)

(5) Truncation (Maddala, Ch. 6)

1) Example 1:

- Earnings function from a sample of poor people (Hausman and Wise, ECON. 1979):

  $y_t^* = x_{t\bullet}'\beta + \varepsilon_t$, $\varepsilon_t \sim N(0,\sigma^2)$.

  Observe $y_t = y_t^*$ iff $y_t^* < L_t$ ($L_t = 1.5\times$"poverty line" dep. on family size)

- Log-likelihood function:

  - pdf of y_t: $g(y_t) = f(y_t^*|y_t^* \leq L_t) = \frac{f(y_t^*)}{\Pr(y_t^* \leq L_t)}$.

  $$\Pr(y_t^* \leq L_t) = \Pr\left(\frac{\varepsilon_t}{\sigma} \leq \frac{L_t - x_{t\bullet}'\beta}{\sigma}\right) = \Phi\left(\frac{L_t - x_{t\bullet}'\beta}{\sigma}\right)$$

  $$\rightarrow \ln L = \Sigma_t \left\{\ln(f(y_t)) - \ln\Phi\left(\frac{L_t - x_{t\bullet}'\beta}{\sigma}\right)\right\}.$$

- $E(y_t) = E(y_t^*|y_t^* \leq L_t) = x_t'\beta - \sigma\lambda\left(\frac{L_t - x_{t\bullet}'\beta}{\sigma}\right)$.

2) Example 2:

- Observe $y_t = y_t^*$ iff $y_t^* > L_t$
- $f(y_t^*|y_t^* \geq L_t) = f(y_t^*)/\Pr(y_t^* \geq L_t)$
- $\Pr(y_t^* \geq L_t) = 1 - \Phi\left[\frac{L_t - x_{t\bullet}'\beta}{\sigma}\right]$

$$\rightarrow l_T = \Sigma_t \left\{ \ln(f(y_t)) - \ln\left[1 - \Phi\left(\frac{L_t - x_{t\bullet}'\beta}{\sigma}\right)\right]\right\}.$$

- $E(y_t) = E(y_t^*|y_t^* \geq L_t) = x_t'\beta + \sigma\lambda\left(-\frac{L_t - x_{t\bullet}'\beta}{\sigma}\right)$
- If $L_t = 0$ for all t, $1 - \Phi\left(\frac{L_t - x_{t\bullet}'\beta}{\sigma}\right) = 1 - \Phi\left(\frac{-x_{t\bullet}'\beta}{\sigma}\right) = \Phi\left(\frac{x_{t\bullet}'\beta}{\sigma}\right)$.
- Consider tobit. Choose observations with $y_t > 0$ and do truncation. This is the case where we observe $y_t = y_t^*$ iff $y_t^* > 0$. The truncation MLE using the truncated data is consistent even if it is inefficient.

[Empirical Example]

Program: (mw_trun.lim)

```
reject  ; emp # 1 $
reject  ; uspell=0 $
dstats  ; rhs =uspell,kids5, kids, ed, age, exp $
trunc   ; lhs = uspell
        ; rhs = one,kids5,kids,ed,lofinc, age, exp, wrate,
          occw, occb
        ; maxit = 1000
        ; par
        ; margin  $
sample; all $
```

[Output]

```
+--------------------------------------------+
| Limited Dependent Variable Model - TRUNCATE |
| Maximum Likelihood Estimates                |
| Dependent variable              USPELL      |
| Weighting variable                 ONE      |
| Number of observations             149      |
| Iterations completed                 6      |
| Log likelihood function      -540.9255      |
| Threshold values for the model:             |
| Lower=     .0000     Upper=+infinity        |
| Observations after truncation      149      |
+--------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Primary Index Equation for Model
 Constant -73.88277922      45.833561       -1.612   .1070
 KIDS5     9.715640830      3.5815789        2.713   .0067  .44295302
 KIDS     -2.615500990      1.9749742       -1.324   .1854  1.3892617
 ED       -3.070879632      1.4049426       -2.186   .0288  12.040268
 LOFINC    14.49645829      4.8695164        2.977   .0029  9.6509025
 AGE      -.1635388695      .36410090       -.449   .6533  34.308725
 EXP      -.6647435311      .53775597       -1.236   .2164  8.7382550
 WRATE    -2.766990567      1.3434528       -2.060   .0394  4.9815436
 OCCW      2.291854685      5.7187514         .401   .6886  .46308725
 OCCB     -3.940289956      6.1203299        -.644   .5197  .35570470
         Disturbance standard deviation
 Sigma     16.66202239      2.0837068        7.996   .0000
```

```
+--------------------------------------------+
| Partial derivatives of expected val. with  |
| respect to the vector of characteristics.  |
| They are computed at the means of the Xs.  |
| Observations used for means are All Obs.   |
| Conditional Mean at Sample Point  14.9297  |
| Scale Factor for Marginal Effects   .4215  |
+--------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Constant	-31.14496814	19.320941	-1.612	.1070	
KIDS5	4.095586648	1.5097992	2.713	.0067	.44295302
KIDS	-1.102553205	.83254190	-1.324	.1854	1.3892617
ED	-1.294516115	.59224753	-2.186	.0288	12.040268
LOFINC	6.110919709	2.0527237	2.977	.0029	9.6509025
AGE	-.6893910779E-01	.15348517	-.449	.6533	34.308725
EXP	-.2802197794	.22668872	-1.236	.2164	8.7382550
WRATE	-1.166412985	.56632677	-2.060	.0394	4.9815436
OCCW	.9661214953	2.4107151	.401	.6886	.46308725
OCCB	-1.661012302	2.5799988	-.644	.5197	.35570470

(6) Two-part Model

Cragg (ECON, 1971), Lin and Schmidt (Review of Economics and Statistics (RESTAT), 1984)

1) Model:

- $y_t = x_{t\bullet}'\beta + \varepsilon_t$, where $f(\varepsilon_t) = \dfrac{\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2\sigma^2}\varepsilon_t\right)}{\Phi\left(\frac{x_{t\bullet}'\beta}{\sigma}\right)}$ and $\varepsilon_t > - x_{t\bullet}'\beta$;
- $h_t^* = z_{t\bullet}'\gamma + v_t$ with $v_t \sim N(0,1)$;
- $h_t = 1$ iff $h_t^* > 0$; $= 0$, otherwise.

2) Assumptions

i) Observe y_t iff $h_t = 1$.

ii) ε_t and v_t are stochastically independent.

3) Example:

y_t: desired spending on clothing; h_t: timing to buy.

4) Distribution of y_t:

$$g(y_t) = \frac{\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2\sigma^2}\varepsilon_t\right)}{\Phi\left(\frac{x_{t\bullet}'\beta}{\sigma}\right)}.$$

$\Pr(h_t^* > 0) = \Phi(z_{t\bullet}'\gamma)$

$$\rightarrow \sum_{h_t=1} \ln[g(y_t \mid h_t^* > 0)\Pr(h_t^* > 0)] + \sum_{h_t=0} \ln[\Pr(h_t^* < 0)].$$

Note:

$g(y_t|h_t^* > 0) = g(y_t)$, because ε_t and v_t are sto. indep.

$$l_T = \sum_{h_t=1} \left\{ \begin{array}{l} -\frac{1}{2}\ln(2\pi) - \ln(\sigma) - \frac{1}{2\sigma^2}(y_t - x_{t\bullet}'\beta)^2 \\ + \ln\Phi(z_{t\bullet}'\gamma) - \ln\Phi\left(\frac{x_{t\bullet}'\beta}{\sigma}\right) \end{array} \right\}$$

$$+ \sum_{h_t=0} \ln\left[1 - \Phi(z_{t\bullet}'\gamma)\right]$$

$$= \sum_{h_t=1} \left\{ \begin{array}{l} -\frac{1}{2}\ln(2\pi) - \ln(\sigma) - \frac{1}{2\sigma^2}(y_t - x_{t\bullet}'\beta)^2 \\ - \ln\Phi\left(\frac{x_{t\bullet}'\beta}{\sigma}\right) \end{array} \right\}$$

$$+ \textit{probit} \log \textit{likelihood function for } h_t^* = z_{t\bullet}'\gamma + v_t.$$

$\rightarrow$ trunc. for $y_t > 0$ + probit for all obs.

$\rightarrow$ Estimate (β,σ) by trunc. and γ by probit.

$\rightarrow$ $l_{Cragg} = l_{trunc} + l_{probit}$.

Note:

Let $z_{t\bullet} = x_{t\bullet}$. If $\gamma = \beta/\sigma$, Cragg becomes tobit!!!

5) LR test for tobit specification

STEP 1:Do tobit and get l_{tobit}

STEP 2:Do trunc using observations with $y_t > 0$ and get l_{trunc}.

STEP 3:Do probit using all observations, and get l_{probit}

STEP 4:$l_{cragg} = l_{trunc} + l_{probit}$.

STEP 5:$LR = 2[l_{cragg} - l_{tobit}] \rightarrow_d \chi^2(k)$.

[3] Selection Model

- Heckman, ECON, 1979.

Motivation:

- Model of interest:

  $y_{1t} = x_{1t,\bullet}'\beta_1 + \varepsilon_{1t}$.

- Observe y_{1t} (or $x_{1t,\bullet}$) under a certain condition ("selection rule").

Example:

- Observe a woman's market wage if she works.

Complete Model:

$$y_{1t} = x_{1t,\bullet}'\beta_1 + \varepsilon_{1t} ,$$

$$y_{2t}^* = x_{2t,\bullet}'\beta_2 + \varepsilon_{2t} .$$

$$y_{2t} = 1 \text{ if } y_{2t}^* > 0; \quad = 0 \text{ if } y_{2t}^* < 0.$$

We observe y_{1t} iff $y_{2t} = 1$ (x_{2t} must be observable for any t.)

Assumptions:

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right].$$

Note:

- In LIMDEP, σ_{12} is called theta (θ).

Theorem:

Suppose:

$$\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right].$$

Then, $E(h_1|h_2 > -a) = \sigma_{12}\dfrac{\phi(a)}{\Phi(a)}$.

Facts:

- $E(\varepsilon_{1t}|y_{2t}^* > 0) = E(\varepsilon_{1t}|\varepsilon_{2t} > -x_{2t,\bullet}'\beta_2) = \sigma_{12}\lambda(x_{2t,\bullet}'\beta_2)$,

  where $\lambda(x_{2t,\bullet}'\beta_2) = \dfrac{\phi(x_{2t,\bullet}'\beta_2)}{\Phi(x_{2t,\bullet}'\beta_2)} \equiv \lambda_t$ [inverse Mill's ratio]

- $E(y_{1t}|y_{2t}^* > 0) = x_{1t}'\beta_1 + E(\varepsilon_{1t}|\varepsilon_{2t} > -x_{2t,\bullet}'\beta_2) = x_{1t,\bullet}'\beta_1 + \sigma_{12}\lambda_t$

- $y_{1t} = x_{1t,\bullet}'\beta_1 + \sigma_{12}\lambda_t + v_t$,

  where $E(v_t|\varepsilon_{2t} > -x_{2t,\bullet}'\beta_2) = 0$;

  $\mathrm{var}(v_t|\varepsilon_{2t} > -x_{2t,\bullet}'\beta_2) \equiv \pi_t = \sigma_1^2 - \xi_t$, $\xi_t = \sigma_{12}^2[(x_{2t,\bullet}'\beta_2)\lambda_t + \lambda_t^2]$

Two-Step Estimation:

STEP 1: Do probit for all t, and get $\hat{\beta}_2$, and $\hat{\lambda}_t = \dfrac{\phi(x_{2t,\bullet}'\hat{\beta}_2)}{\Phi(x_{2t,\bullet}'\hat{\beta}_2)}$.

STEP 2: Do OLS on $y_{1t} = x_{1t,\bullet}'\beta_1 + \sigma_{12}\hat{\lambda}_t + \eta_t$, and get $\hat{\beta}_1$ and $\hat{\sigma}_{12}$.

Facts on the Two-Step Estimator:

- Consistent.
- t-test for H_o: $\sigma_{12} = 0$ (no selection) in STEP 2 is the LM test (Melino, Review of Economic Studies, 1982)
- But all other t-tests are wrong!!! $\rightarrow s^2(X'X)^{-1}$ is inconsistent.
- So, have to compute corrected covariance matrix

  [See, Heckman (1979, Econ.), Greene (1981, Econ.).]

  - Sometimes, corrected covariance matrix is not computable (Greene, Econ, 1981).

Covariance Matrix of the Two-Step Estimator:

- Let $\hat{\Omega} = Cov(\hat{\beta}_2)$.
- $y_{1t} = x_{1t,\bullet}'\beta_1 + \sigma_{12}\lambda_t + v_t$

  $\rightarrow \quad y_{1t} = x_{1t,\bullet}'\beta_1 + \sigma_{12}\hat{\lambda}_t + [-\sigma_{12}(\hat{\lambda}_t - \lambda_t) + v_t]$.

Short Digression:

By Taylor expansion around the true value of β_2,

$$\hat{\lambda}_t = \lambda(x_{2t,\bullet}'\hat{\beta}_2) \approx \lambda(x_{2t,\bullet}'\beta_2) + \frac{\partial\lambda(x_{2t,\bullet}'\beta_2)}{\partial\beta_2}(\hat{\beta}_2 - \beta_2).$$

End of Digression

$\rightarrow \quad y_{1t} = \begin{pmatrix} x_{1t,\bullet}' & \hat{\lambda}_t \end{pmatrix} \begin{pmatrix} \beta_1 \\ \sigma_{12} \end{pmatrix} + h_t'(\hat{\beta}_2 - \beta_2) + v_t = z_{t\bullet}'\gamma + [h_t'(\hat{\beta}_2 - \beta_2) + v_t],$

where $h_t = \sigma_{12}[(x_{2t,\bullet}'\beta_2)\lambda_t + \lambda_t^2]x_{2t,\bullet}$.

$\rightarrow$ In matrix notation,

$y_1 = Z\gamma + [H(\hat{\beta}_2 - \beta_2) + v]$.

$$\hat{\gamma}_{TS} = (Z'Z)^{-1}Z'y_1$$

$$\rightarrow \quad = (Z'Z)^{-1}Z'(Z\gamma + H(\hat{\beta}_2 - \beta_2) + v) \quad .$$

$$= \gamma + (Z'Z)^{-1}Z'H(\hat{\beta}_2 - \beta_2) + (Z'Z)^{-1}Z'v$$

$\rightarrow$ Can show that $(\hat{\beta}_2 - \beta_2)$ and v are uncorrelated. Then, intuitively,

$$Cov(\hat{\gamma}_{TS}) = Cov[(Z'Z)^{-1}Z'H(\hat{\beta}_2 - \beta_2) + (Z'Z)^{-1}Z'v]$$

$$= Cov[(Z'Z)^{-1}Z'H(\hat{\beta}_2 - \beta_2)] + Cov[(Z'Z)^{-1}Z'v]$$

$$= (Z'Z)^{-1}Z'HCov(\hat{\beta}_2 - \beta_2)H'Z(Z'Z)^{-1}$$

$$+ (Z'Z)^{-1}Z'Cov(v)Z(Z'Z)^{-1}$$

$$= (Z'Z)^{-1}Z'H\Omega H'Z(Z'Z)^{-1} + (Z'Z)^{-1}Z'\Pi Z(Z'Z)^{-1},$$

where $\Pi = diag(\pi_1, ..., \pi_T)$.

$$\approx (Z'Z)^{-1}Z'\hat{H}\hat{\Omega}\hat{H}'Z(Z'Z)^{-1} + (Z'Z)^{-1}Z'\ddot{\Pi}Z(Z'Z)^{-1},$$

where $\ddot{\Pi} = diag(\hat{v}_1^2, ..., \hat{v}_T^2)$ and $\hat{H}$ is an estimated H.

- MLE (which is more efficient than two-step estimator)
    - $\Pr(y_{1t}$ is not observed) = $\Pr(y_{2t}^* < 0) = 1 - \Phi(x_{2t,\bullet}'\beta_2)$.
    - $\Pr(y_{1t}$ is observed) = $\Phi(x_{2t,\bullet}'\beta_2)$.
    - $f(y_{1t}, y_{1t}$ is observed)

$= f(y_{1t}|y_{2t}^* \geq 0)$

$$= \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left[-\frac{(y_{1t} - x_{1t,\bullet}'\beta_1)^2}{2\sigma_1^2}\right] \times\Phi\left[\frac{\sigma_1 x_{2t,\bullet}'\beta_2 + (\sigma_{12}/\sigma_1)(y_{1t} - x_{1t,\bullet}'\beta_1)}{\sqrt{\sigma_1^2 - \sigma_{12}}}\right] \div\Phi(x_{2t,\bullet}'\beta_2)$$

$\div\Phi(x_{2t,\bullet}'\beta_2)$.

- $l_T = \sum_{y_{1t}\ observed} \ln[f(y_{1t} | y_{1t}\ is\ observed)\Pr(y_{1t}\ is\ observed)$
$$+ \sum_{y_{1t}\ is\ not\ observed} \ln \Pr(y_{1t}\ is\ not\ observed).$$

Note:

In LIMDEP, rho = σ_{12}/σ_1

(Correlation coefficient between ε_{1t} and ε_{2t})

[Empirical Example]

Consider the model:

$y_{1t} = x_{1t,\cdot}'\beta_1 + \varepsilon_{1t}$; $y_{2t}^* = x_{2t,\cdot}'\beta_2 + \varepsilon_{2t}$.
$y_{2t} = 1$ if $y_{2t}^* > 0$; $= 0$, otherwise.
y_{1t} are observed only if $y_{2t} = 1$.

The model can be estimated by following commands:

PROBIT; LHS = Y2 ; RHS = list for x_{2t} ; HOLD (IMR = LAM) $
SELECT; LHS = Y1 ; RHS = list for x_{1t} $

This program returns result for the two-stage estimation. B contains β_1 and θ where $\theta = \sigma_{12}$; VARB contains the corrected covariance matrix. LIMDEP also calculates $\rho = \sigma_{12}/\sigma_1$ and store it in the name of RHO.

If you want to do MLE, then, use:

SELECT; LHS = Y1 ; RHS = list for x_{1i} ; MLE $

[EXERCISE]Consider the market wage equation of married women. The model is given:

$LRATE_t = x_{1t}'\beta_1 + \varepsilon_{1t}$
$EMP_t^* = x_{2t}'\beta_2 + \varepsilon_{2t}$,

where EMP_t^* denotes the latent variable which represents the t'th woman's willingness to work. We assume that the woman can get a job immediately if she wants. W_t is observed only if $EMP_t = 1$. Both of x_{1t} and x_{2t} contain ONE, ED, URB, MINOR, AGE, REGS, UNEMPR, LOFINC, KIDS5, EXP. Estimate this model by two-stage and MLE. Can you find some efficiency gains by using MLE?

Program: (mw_heck1.lim)

```
OPEN    ; OUTPUT = MW_HECK1.OUT $
NAMELIST; X  = ONE,   ED,     URB,   MINOR,   AGE,
               REGS,  UNEMPR, EXP
        ; X1 = X, OCCW, OCCB, UNION, TENURE
        ; X2 = X, LOFINC, KIDS5       $
PROBIT  ; LHS = EMP ; RHS = X2 ; HOLD   $
SELEC   ; LHS = LRATE ; RHS = X1 ; MLE ; MAXIT = 1000  $
```

[Output]

```
              +----------------------------------------------+
              | Binomial Probit Model                        |
              | Maximum Likelihood Estimates                 |
              | Dependent variable                 EMP       |
              | Weighting variable                 ONE       |
              | Number of observations            1962       |
              | Iterations completed                 5       |
              | Log likelihood function      -1164.046       |
              | Restricted log likelihood    -1356.524       |
              | Chi-squared                   384.9554       |
              | Degrees of freedom                   9       |
              | Significance level            .0000000       |
              | Results retained for SELECTION model.        |
              +----------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Index function for probability
 Constant  2.266181446      .54202728        4.181   .0000
 ED         .7879400741E-01  .15165875E-01    5.195   .0000  12.205403
 URB        .1758888543      .68203409E-01    2.579   .0099  .68654434
 MINOR      .4110693370E-01  .77625279E-01     .530   .5964  .27166157
 AGE       -.3396916054E-01  .39855933E-02   -8.523   .0000  37.484709
 REGS       .1166792983      .72182531E-01    1.616   .1060  .34352701
 UNEMPR   -3.690413913      1.2812595        -2.880   .0040  .74255861E-01
 EXP        .7349731553E-01  .58759795E-02   12.508   .0000  8.3577982
 LOFINC    -.2315954611      .55699140E-01   -4.158   .0000  9.9052275
 KIDS5     -.5068819912      .45726877E-01  -11.085   .0000  .50764526

Frequencies of actual & predicted outcomes
Predicted outcome has maximum probability.

            Predicted
------  ----------  +  -----
Actual      0    1  |  Total
------  ----------  +  -----
  0       776  263  |   1039
  1       344  579  |    923
------  ----------  +  -----
Total    1120  842  |   1962
```

```
+------------------------------------------------------------+
| Sample Selection Model                                     |
| Probit selection equation based on EMP                     |
| Selection rule is: Observations with EMP      =  1         |
| Results of selection:                                      |
|                   Data points     Sum of weights           |
| Data set              1962            1962.0               |
| Selected sample        923             923.0               |
+------------------------------------------------------------+

+-----------------------------------------------------------------------+
| Sample Selection Model                                                |
| Two stage   least squares regression    Weighting variable = none     |
| Dep. var. = LRATE    Mean=   1.662758599    , S.D.=   .3999430413     |
| Model size: Observations =     923, Parameters =  13, Deg.Fr.=    910 |
| Residuals:  Sum of squares= 88.81466508    , Std.Dev.=         .31241 |
| Fit:        R-squared=  .389174, Adjusted R-squared =          .38112 |
|             (Note:  Not using OLS.  R-squared is not bounded in [0,1] |
| Model test: F[ 12,    910] =   48.32,    Prob value =          .00000 |
| Diagnostic: Log-L =   -229.2730, Restricted(b=0) Log-L =    -463.3122 |
|             LogAmemiyaPrCrt.=   -2.313, Akaike Info. Crt.=       .525 |
| Standard error corrected for selection.....    .33121                 |
| Correlation of disturbance in regression                              |
| and Selection Criterion (Rho)..............    .43718                 |
+-----------------------------------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Constant	.5169131630	.11437465	4.519	.0000	
ED	.6649307325E-01	.60206920E-02	11.044	.0000	12.457205
URB	.1773759094	.25080466E-01	7.072	.0000	.71614301
MINOR	-.9162049057E-01	.27555154E-01	-3.325	.0009	.28494041
AGE	-.4052069464E-02	.16570669E-02	-2.445	.0145	37.590466
REGS	-.3921158142E-01	.26063159E-01	-1.504	.1325	.35861322
UNEMPR	-.8525455562	.47233534	-1.805	.0711	.72383532E-01
EXP	.1114265941E-01	.31600426E-02	3.526	.0004	10.248104
OCCW	.1619122454	.29327728E-01	5.521	.0000	.63163597
OCCB	.1268119881	.35417953E-01	3.580	.0003	.16468039
UNION	.1509058350	.28282270E-01	5.336	.0000	.18526544
TENURE	.1712234936E-01	.28990529E-02	5.906	.0000	4.4117010
LAMBDA	.1448006954	.48487172E-01	2.986	.0028	.71724931

```
+----------------------------------------------+
| ML Estimates of Selection Model              |
| Maximum Likelihood Estimates                 |
| Dependent variable                LRATE      |
| Weighting variable                  ONE      |
| Number of observations             1962      |
| Iterations completed                 34      |
| Log likelihood function       -1394.007      |
| FIRST 10 estimates are probit equation.      |
+----------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Selection (probit) equation for EMP
 Constant  3.115102731      .53290613        5.845   .0000
 ED         .9309490438E-01  .14900731E-01    6.248   .0000
 URB        .1926803829      .68356876E-01    2.819   .0048
 MINOR     -.5249932085E-01  .78718530E-01    -.667   .5048
 AGE       -.2783301692E-01  .39870895E-02   -6.981   .0000
 REGS       .8956208114E-01  .72652319E-01    1.233   .2177
 UNEMPR    -3.688661523     1.3155590        -2.804   .0050
 EXP        .7008950745E-01  .59336828E-02   11.812   .0000
 LOFINC    -.3575482529      .54066061E-01   -6.613   .0000
 KIDS5     -.4130739462      .35401212E-01  -11.668   .0000
          Corrected regression, Regime 1
 Constant   .4044289991      .10183058        3.972   .0001
 ED         .7103225068E-01  .50385133E-02   14.098   .0000
 URB        .1871581170      .26897106E-01    6.958   .0000
 MINOR     -.8436757242E-01  .31160709E-01   -2.707   .0068
 AGE       -.6049562510E-02  .15511175E-02   -3.900   .0001
 REGS      -.3210445926E-01  .28081038E-01   -1.143   .2529
 UNEMPR    -1.088359891      .47861865       -2.274   .0230
 EXP        .1670674352E-01  .28320181E-02    5.899   .0000
 OCCW       .1573414608      .29381105E-01    5.355   .0000
 OCCB       .1320411808      .36071358E-01    3.661   .0003
 UNION      .1501912661      .28208061E-01    5.324   .0000
 TENURE     .1666189277E-01  .28726903E-02    5.800   .0000
 SIGMA(1)   .3668855166      .15393033E-01   23.835   .0000
 RHO(1,2)   .7079833477      .62997072E-01   11.238   .0000
```

Dependent Variable: LRATE
Method: Least Squares
Date: 11/21/02 Time: 08:41
Sample: 1 923
Included observations: 923

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.662917	0.102027	6.497494	0.0000
ED	0.060361	0.005598	10.78351	0.0000
URB	0.165747	0.024398	6.793562	0.0000
MINOR	-0.093992	0.027109	-3.467156	0.0006
AGE	-0.001527	0.001419	-1.076222	0.2821
REGS	-0.044511	0.025576	-1.740373	0.0821
UNEMPR	-0.598670	0.457466	-1.308664	0.1910
EXPP	0.004728	0.002291	2.063932	0.0393
OCCW	0.157935	0.029718	5.314402	0.0000
OCCB	0.117249	0.035754	3.279365	0.0011
UNION	0.154210	0.028571	5.397502	0.0000
TENURE	0.017368	0.002903	5.983214	0.0000

R-squared	0.382966	Mean dependent var	1.662759
Adjusted R-squared	0.375516	S.D. dependent var	0.399943
S.E. of regression	0.316052	Akaike info criterion	0.547097
Sum squared resid	90.99889	Schwarz criterion	0.609862
Log likelihood	-240.4854	F-statistic	51.40169
Durbin-Watson stat	1.804551	Prob(F-statistic)	0.000000

[5] Switching Model (Mover/Stayer)

(Lee, International Economic Review, 1978)

Model:

- Three equations:

  $y_{1t} = x_{1t,\bullet}'\beta_1 + \varepsilon_{1t}$

  $y_{0t} = x_{0t,\bullet}'\beta_0 + \varepsilon_{0t}$

  $I_t^* = z_{t\bullet}'\gamma + u_t$.

  where $\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{0t} \\ u_t \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{1,0} & \sigma_{1,u} \\ \sigma_{1,0} & \sigma_0^2 & \sigma_{0,u} \\ \sigma_{1,u} & \sigma_{0,u} & 1 \end{pmatrix}\right]$.

- Assume $\sigma_{1,0} = 0$ (in LIMDEP).
- Observe:

  $I_t = 1$ if $I_t^* > 0$; $= 0$, otherwise

  $y_t = y_{1t}$ if $I_t = 1$; $y_t = y_{0t}$ if $I_t = 0$.

Two Step Estimation:

- $E(y_t|I_t^* > 0) = E(y_{1t}|I_t^* > 0) = x_{1t,\bullet}'\beta_1 + E(\varepsilon_{1t}|u_t > -z_{t\bullet}'\gamma)$

  $= x_{1t,\bullet}'\beta_1 + \sigma_{1,u}\lambda(z_{t\bullet}'\gamma)$

  - $y_t = x_{1t,\bullet}'\beta_1 + \sigma_{1,u}[\phi(z_{t\bullet}'\gamma)/\Phi(z_{t\bullet}'\gamma)] + v_{1t}$. (*)

Note:

- $cov(\varepsilon_{0t},-u_t) = -cov(\varepsilon_{0t},u_t) = -\sigma_{0,u}$.
- $E(y_t|I_t^* < 0) = E(y_{0t}|I_t^* < 0) = x_{0t,\bullet}'\beta_0 + E(\varepsilon_{0t}|u_t < -z_{t\bullet}'\gamma)$

$$= x_{0t,\bullet}'\beta_0 + E(\varepsilon_{2t}|-u_t > -(-z_{t\bullet}'\gamma))$$

$$= x_{0t,\bullet}'\beta_0 - \sigma_{0,u}\lambda(-z_{t\bullet}'\gamma).$$

$$= x_{0t,\bullet}'\beta_0 - \sigma_{0,u}\lambda(-z_{t\bullet}'\gamma).$$

- $y_t = x_{2t,\bullet}'\beta_2 - \sigma_{0,u}\lambda(-z_{t\bullet}'\gamma) + v_{2t}$. (**)
- Can do the two-step estimation for (*) and (**), separately.

MLE: For l_T, see Maddala (book).

Example:

$\ln(w_{ut}) = x_{ut,\bullet}'\beta_u + \varepsilon_{ut}$ [union wage]

$\ln(w_{nt}) = x_{nt,\bullet}'\beta_n + \varepsilon_{nt}$ [nonunion wage]

$I_t^* = x_{pt,\bullet}'\beta_p + \varepsilon_{pt}$ [preference index]

$I_t = 1$ if $I_t^* > 0$; $= 0$ if $I_t^* < 0$

Note:

- $x_{pt,\bullet}$ should be observed for all t. It includes variables related with wages, cost of moving to union jobs, and preference.

Some interesting hypotheses on union:

- Unions tend to equalize wages over race, experience and age.
- Unions tend to narrow dispersion of wages, other things being equal.

[Empirical Example]

MOVER/STAYER MODEL

$y_{1t} = x_{1t,\bullet}'\beta_1 + \varepsilon_{1t}$, $\varepsilon_{1t} \sim N(0,\sigma_1^2)$
$y_{0t} = x_{0t,\bullet}'\beta_0 + \varepsilon_{0t}$, $\varepsilon_{0t} \sim N(0,\sigma_0^2)$
$I_t^* = z_{t\bullet}'\gamma + u_t$. $u_t \sim N(0,1)$

$\sigma_{u1} = cov(u,\varepsilon_1)$; $\sigma_{u0} = cov(u,\varepsilon_0)$
$\rho_{u1} = corr(\varepsilon_1,u) = \sigma_{u1}/\sigma_1$; $\rho_{u0} = corr(\varepsilon_0,u)$.

$I_t = 1$ if $I_t^*>0$; 0, otherwise.
y_{1t} are observed if $I_t = 1$.
y_{0t} is observed if $I_t = 0$.

This model is estimated by following commands:

```
NAMELIST; Z = ... ; X1 = ... ; X0= ... $

PROBIT   ; LHS = I ; RHS = Z ; HOLD $
SELECT   ; LHS = Y ; RHS = X1 $
MATRIX   ; BETA1 = BSR1 $ ? BSR1 = (b1,s1,rho_{1,u})
SELECT   ; LHS = Y ; RHS = X0 ; LIMITS = 1 $
MATRIX   ; BETA0 = BSR0 $ ? BSR0 = (b0,s0,rho_{0,u})
SELECT   ; LHS = Y ; RH1 = X1 ; RH2 = X0
         ; MLE ; ALL ; START = BETA1,BETA0
         ; TLF = 0.0000001 ; TLB = 0.0000001 ; TLG = 0.0000001
         ; MAXIT = 1000 $
```

If you run the program, you get following results:

(1) probit γ.
(2) two-stage β_1, σ_{u1} (LIMDEP denotes this as coefficient of LAMBDA.)
(3) two-stage β_0, σ_{u0} (LIMDEP denotes this as coefficient of LAMBDA.)
(4) MLE results: γ, β_1, β_0, σ_0, ρ_{u0}, σ_1, ρ_{u1}.

[EXERCISE] Use MWPSID82.DB. Choose the employed only. You want to estimate union and nonunion wages by LEE's method. In this case, Y is LRATE; all of x_1, x_0, and z contain ONE, ED, URB, MINOR, AGE, EXP, REGS, UNEMPR, OCCW, OCCB. Estimate the model by the MOVER/STAYER model.

Program:(mw_lee.lim)

```
? UNION-NONUNION WAGE

NAMELIST  ; X = ONE, ED, URB, MINOR, AGE, EXP, REGS, UNEMPR,
                OCCW, OCCB $

? CHOOSING EMPLOYED ONLY

REJECT    ; EMP # 1 $

? TWO-STAGE AND MLE

PROBIT    ; LHS = UNION ; RHS = X ; HOLD $
SELECT    ; LHS = LRATE ; RHS = X $
MATRIX    ; BETA1 = BSR1 $ ? BSR1 = (b1,s1,rho_{u1})
SELECT    ; LHS = LRATE ; RHS = X ; LIMITS = 1 $
MATRIX    ; BETA0 = BSR0 $ ? BASR0 = (b0,s0,rho{u2})
SELECT    ; LHS = LRATE ; RH1 = X ; RH2 = X
          ; MLE ; ALL ; START = PB, BETA1,BETA0
          ; TLF = 0.0000001; TLB = 0.0000001 ; TLG = 0.0000001
          ; MAXIT = 1000 $
```

[Output]

PROBIT RESULTS:

```
+---------------------------------------------+
| Binomial Probit Model                       |
| Maximum Likelihood Estimates                |
| Dependent variable                UNION     |
| Weighting variable                  ONE     |
| Number of observations              923     |
| Iterations completed                  6     |
| Log likelihood function       -386.1421     |
| Restricted log likelihood     -442.3796     |
| Chi-squared                    112.4750     |
| Degrees of freedom                    9     |
| Significance level             .0000000     |
| Results retained for SELECTION model.       |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Index function for probability
 Constant -2.617015862      .51023856       -5.129   .0000
 ED         .8574023976E-01  .27671484E-01    3.099   .0019  12.457205
 URB        .7211858131E-01  .12557978         .574   .5658  .71614301
 MINOR      .7066238648     .13340295        5.297   .0000  .28494041
 AGE      -.4853860407E-02  .72094613E-02    -.673   .5008  37.590466
 EXP        .2051808803E-01  .10054021E-01    2.041   .0413  10.248104
 REGS     -1.040157346      .14426730       -7.210   .0000  .35861322
 UNEMPR    5.984353663     2.1650843         2.764   .0057  .72383532E-01
 OCCW       .9235575096E-01  .15303386         .603   .5462  .63163597
 OCCB       .6670184806     .17576278        3.795   .0001  .16468039


Frequencies of actual & predicted outcomes
Predicted outcome has maximum probability.

            Predicted
------  ----------  +  -----
Actual      0    1  |  Total
------  ----------  +  -----
  0       741   11  |    752
  1       157   14  |    171
------  ----------  +  -----
Total     898   25  |    923
```

TWO-STEP RESULTS FOR UNION WORKERS:

```
+----------------------------------------------------------+
| Sample Selection Model                                   |
| Probit selection equation based on UNION                 |
| Selection rule is: Observations with UNION    =  1       |
| Results of selection:                                    |
|                    Data points     Sum of weights        |
| Data set                923              923.0           |
| Selected sample         171              171.0           |
+----------------------------------------------------------+
```

```
+-----------------------------------------------------------------------+
| Sample Selection Model                                                |
| Two stage   least squares regression    Weighting variable = none     |
| Dep. var. = LRATE    Mean=   1.855300971    , S.D.=   .3845529020     |
| Model size: Observations =     171, Parameters =  11, Deg.Fr.=    160 |
| Residuals:  Sum of squares= 15.20892926    , Std.Dev.=         .30831 |
| Fit:        R-squared=  .353433, Adjusted R-squared =          .31302 |
|             (Note:  Not using OLS.  R-squared is not bounded in [0,1] |
| Model test: F[ 10,    160] =    8.75,    Prob value =          .00000 |
| Diagnostic: Log-L =    -35.7472, Restricted(b=0) Log-L =     -78.7168 |
|             LogAmemiyaPrCrt.=   -2.291, Akaike Info. Crt.=       .547 |
| Standard error corrected for selection.....    .32745                 |
| Correlation of disturbance in regression                              |
| and Selection Criterion (Rho).............    .39146                  |
+-----------------------------------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Constant	.1989702391	2.7164770	.073	.9416	
ED	.8559907217E-01	.64371321E-01	1.330	.1836	12.783626
URB	.1529249252	.81125963E-01	1.885	.0594	.78362573
MINOR	.1898086360E-01	.51152404	.037	.9704	.35672515
AGE	-.1457889180E-02	.51539188E-02	-.283	.7773	37.877193
EXP	.1180394543E-01	.15753760E-01	.749	.4537	11.327485
REGS	-.2464753117	.79047052	-.312	.7552	.14619883
UNEMPR	1.246237354	4.5980742	.271	.7864	.78479532E-01
OCCW	.1374318348	.11162874	1.231	.2183	.62573099
OCCB	.2051807506	.49555645	.414	.6788	.23391813
LAMBDA	.1281841681	.99030465	.129	.8970	1.2576321

TWO-STEP RESULTS FOR NON-UNION WORKERS:

```
+----------------------------------------------------------+
| Sample Selection Model                                   |
| Probit selection equation based on UNION                 |
| Selection rule is: Observations with UNION    =  0       |
| Results of selection:                                    |
|                  Data points     Sum of weights          |
| Data set              923              923.0             |
| Selected sample       752              752.0             |
+----------------------------------------------------------+
```

```
+-----------------------------------------------------------------------+
| Sample Selection Model                                                |
| Two stage   least squares regression    Weighting variable = none     |
| Dep. var. = LRATE    Mean=   1.618975693    , S.D.=   .3905632551     |
| Model size: Observations =     752, Parameters =  11, Deg.Fr.=    741 |
| Residuals:  Sum of squares= 76.52245253    , Std.Dev.=         .32136 |
| Fit:        R-squared=  .322100, Adjusted R-squared =          .31295 |
|             (Note:  Not using OLS.  R-squared is not bounded in [0,1] |
| Model test: F[ 10,    741] =   35.21,    Prob value =          .00000 |
| Diagnostic: Log-L =   -207.8246, Restricted(b=0) Log-L =    -359.5371 |
|             LogAmemiyaPrCrt.=   -2.256, Akaike Info. Crt.=       .582 |
| Standard error corrected for selection.....    .32957                 |
| Correlation of disturbance in regression                              |
| and Selection Criterion (Rho).............   -.38101                  |
+-----------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant  .6870010481      .12547033        5.475   .0000
 ED         .5525611013E-01  .86533202E-02    6.386   .0000  12.382979
 URB        .1697129228      .28025204E-01    6.056   .0000  .70079787
 MINOR    -.1315714426      .56715460E-01   -2.320   .0203  .26861702
 AGE      -.4908581666E-03  .16396287E-02    -.299   .7647  37.525266
 EXP       .1067372216E-01  .26605199E-02    4.012   .0001  10.002660
 REGS      .8051619073E-02  .71147410E-01     .113   .9099  .40691489
 UNEMPR   -1.215456362      .67381632       -1.804   .0713  .70997340E-01
 OCCW      .1699618459      .34467701E-01    4.931   .0000  .63297872
 OCCB      .1029094990      .61714854E-01    1.667   .0954  .14893617
 LAMBDA   -.1255690455      .20709644        -.606   .5443 -.28597751
```

MLE

```
+---------------------------------------------+
| ML Estimates of Selection Model             |
| Maximum Likelihood Estimates                |
| Dependent variable                LRATE     |
| Weighting variable                  ONE     |
| Number of observations              923     |
| Iterations completed                 44     |
| Log likelihood function       -641.0801     |
| MOVER/STAYER model (MLE). LHS= LRATE        |
| FIRST 10 estimates are probit equation.     |
| next  10 slopes are for the Y=1 equation    |
+---------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
	Selection (probit) equation for UNION				
Constant	-2.645990376	.51373267	-5.151	.0000	
ED	.8681051837E-01	.29267538E-01	2.966	.0030	
URB	.6778735412E-01	.13318452	.509	.6108	
MINOR	.7034367729	.14398344	4.886	.0000	
AGE	-.4356677231E-02	.78974340E-02	-.552	.5812	
EXP	.1970606495E-01	.11909515E-01	1.655	.0980	
REGS	-1.041862767	.14835841	-7.023	.0000	
UNEMPR	6.078392656	2.1363848	2.845	.0044	
OCCW	.9594217936E-01	.15392204	.623	.5331	
OCCB	.6709876141	.18244826	3.678	.0002	
	Corrected regression, Regime 0 (UNION)				
Constant	.1355470774	1.2229363	.111	.9117	
ED	.8700546409E-01	.30284069E-01	2.873	.0041	
URB	.1538360672	.76138920E-01	2.020	.0433	
MINOR	.3196077722E-01	.22800768	.140	.8885	
AGE	-.1693435375E-02	.46276619E-02	-.366	.7144	
EXP	.1247116063E-01	.10223242E-01	1.220	.2225	
REGS	-.2646562264	.35006795	-.756	.4496	
UNEMPR	1.357888546	2.4790218	.548	.5839	
OCCW	.1401773694	.85041030E-01	1.648	.0993	
OCCB	.2171028920	.22973254	.945	.3446	
	(NON-UNION)				
Constant	.6672868219	.11712790	5.697	.0000	
ED	.5801080113E-01	.83830854E-02	6.920	.0000	
URB	.1720733461	.28206307E-01	6.101	.0000	
MINOR	-.1089672386	.58067188E-01	-1.877	.0606	
AGE	-.6830625434E-03	.15614740E-02	-.437	.6618	
EXP	.1136329807E-01	.25239037E-02	4.502	.0000	
REGS	-.2331002590E-01	.70028085E-01	-.333	.7392	
UNEMPR	-1.021918636	.66968541	-1.526	.1270	
OCCW	.1739410670	.34376508E-01	5.060	.0000	
OCCB	.1250030435	.63825851E-01	1.959	.0502	
	Variance parameters				
SIGMA(0)	.3217756467	.96729648E-02	33.265	.0000	
RHO(0,u)	.7947841292E-01	.63641564	.125	.9006	
SIGMA(1)	.3351599662	.14730995	2.275	.0229	
RHO(1,u)	.4556733832	1.1116953	.410	.6819	