2. Limited Dependent Variables Models

[1] Binary choice models (Review)

(1) 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 $y_t^* < 0$;

  and 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).
  - $f(y_t) = \left(\Phi(x_{t\bullet}'\beta)\right)^{y_t}\left(1-\Phi(x_{t\bullet}'\beta)\right)^{1-y_t}$.

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

- 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:

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

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

Theorem: Under suitable regularity conditions,

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

Implication:

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

  $Cov(\hat{\theta})$ 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.

- 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:

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

- Testing Hypothesis:

  1. Wald test:

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

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

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

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

  (Overall significance test)

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

(2) Logit Models

- Model:

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

  $\varepsilon_t \sim$ 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

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

    $\to \beta_j$ can be interpreted as the effect of x_{jt} on "log odds".

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

(3) 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.′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.

(4) Probit/Logit Panel Models

1) Model:

$$y_{it}^* = x_{it}\beta + z_i\gamma + \alpha_i + \varepsilon_{it},$$

where ε_{it} are iid N(0,1) and y_{it} = 1 if $y_{it}^* > 0$; = 0 otherwise.

2) Fixed effects model

- Treat the α_i as parameters to be estimated.
- MLE

  [For probit]

$$l_T(\beta,\gamma,\alpha_1,...,\alpha_N) = \Sigma_i\Sigma_t\begin{bmatrix} y_{it}\ln\Phi(x_{it}\beta + z_i\gamma + \alpha_i) \\ +(1-y_{it})\ln\left(1-\Phi(x_{it}\beta + z_i\gamma + \alpha_i)\right)\end{bmatrix}.$$

[For logit]

$$l_T(\beta,\gamma,\alpha_1,...,\alpha_N) = \Sigma_i\Sigma_t \begin{bmatrix} y_{it}(x_{it}\beta + z_i\gamma + \alpha_i) \\ -\ln\left(1+\exp(x_{it}\beta + z_i\gamma + \alpha_i)\right) \end{bmatrix}.$$

- Facts:
  - If N is large, probit (logit) ML estimators are computationally burdensome.
  - If T is small, probit (logit) ML estimators are severely biased: Chamberlain (1980, RES) derives the asymptotic bias of ML estimator for a simple logit model (scalar β, no time invariant regressor, T =2). He found that $p\lim_{N\to\infty} \hat{\beta}_{ML} = 2\beta$!
  - Some Monte Carlo experiments (e.g., Heckman, 1981) show that ML estimators behave relatively well if T is large (T = 10 or more).

3) Random Effects Model I

- Assume that regressors (x_{it} and z_i) are uncorrelated with α_i.
- α_i iid $N(0,\sigma_\alpha^2)$. Let $\alpha_i = \sigma_\alpha g_i$ where $g_i \sim N(0,1)$.
- See Butler and Moffitt (ECON, 1982) and Hsiao (Econometrics Reviews, 1984).

- The joint pdf of y_{i1} , ... , y_{iT} is given by:

$$f(y_{i1},...,y_{iT}) = E_{g_i}[r_i(\beta,\gamma,\sigma_\alpha g_i)] = \int r_i(\beta,\gamma,\sigma_\alpha g_i) f(g_i)dg_i, \qquad (1)$$

where

$$r_i(\beta,\gamma,\sigma_\alpha g_i) = \prod_{t=1}^{T} \begin{pmatrix} \Phi(x_{it}\beta + z_i\gamma + \sigma_\alpha g_i)^{y_{it}} \\ \times\left(1-\Phi(x_{it}\beta + z_i\gamma + \sigma_\alpha g_i)\right)^{1-y_{it}} \end{pmatrix}.$$

- Log-likelihood function:

$$l_N(\beta,\gamma,\sigma_\alpha) = \Sigma_i \ln\left[f(y_{i1},...,y_{iT})\right].$$

- MLE requires integrations. LIMDEP can do integrations using an approximation procedure. [LIMDEP computes β, γ and ρ = $\sigma_\alpha^2/(1+\sigma_\alpha^2)$. Data do not have to be balanced.

- Simulated ML (SML) method:
  - Gourieroux and Monfort (1993, Journal of Econometrics).
  - Generate random numbers, $g_i^{(1)}$, ... , $g_i^{(H)}$ for each i (all are N(0,1)).
  - If H is large,

$$r_{iH}(\beta,\gamma,\sigma_\alpha) = \frac{1}{H}\Sigma_{h=1}^{H} r_i(\beta,\gamma,\sigma_\alpha g_i^{(h)}) \approx E_{g_i}[r_i(\beta,\gamma,\sigma_\alpha g_i)]. \qquad (2)$$

  - Do MLE using (2) instead of (1). This alternative MLE is called Simulated ML (SML). SML is an efficient as MLE, and is computationally easier.

4) Random Effects Model II

- Regressors (x_{it} and z_i) are correlated with α_i.
  - See Chamberlain (1984, Handbook of Econometrics).
  - $\alpha_i = x_{i1}\lambda_1 + ... + x_{iT}\lambda_T + z_i\pi + \eta_i$, where η_i are iid $N(0,\sigma_\eta^2)$.
  - $y_{it}^* = x_{it}\beta + z_i\gamma + \alpha_i + \varepsilon_{it} = x_{it}\beta + x_i^o\lambda + z_i(\gamma+\pi) + \eta_i + \varepsilon_{it}$,

    where $x_i^o = (x_{i1},...,x_{iT})$ and $\lambda' = (\lambda_1',...,\lambda_T')$.
  - Do MLE as in the case I, and estimate β, λ and $(\gamma+\pi)$.

5) Logit model with fixed effects.

- Use conditional MLE (Chamberlain, 1980, ReStud).
- Logistic distribution:

  pdf: $f(h) = \exp(h)/[1+\exp(h)]^2$;

  cdf: $F(h) = \exp(h)/[1+\exp(h)]$.
- Case in which T = 2. The results obtained below can apply to more general cases. [LIMDEP can do this.]
- Possible outcomes for (y_{i1},y_{i2}):

  $(y_{i1}, y_{i2}) \in \{(1,1),(1,0),(0,1),(0,0)\}$.

- Choose the observations with (1,0) and (0,1) only.

$$\Pr[(y_{i1}, y_{i2}) = (1,0) \; or \; (0,1)]$$

$$= \Pr(y_{i1}=1)\Pr(y_{i2}=0) + \Pr(y_{i1}=0)\Pr(y_{i2}=1)$$

$$= \frac{\exp(x_{i1}\beta + z_i\gamma + \alpha_i)}{1+\exp(x_{i1}\beta + z_i\gamma + \alpha_i)} \frac{1}{1+\exp(x_{i2}\beta + z_i\gamma + \alpha_i)}$$

$$+ \frac{1}{1+\exp(x_{i1}\beta + z_i\gamma + \alpha_i)} \frac{\exp(x_{i2}\beta + z_i\gamma + \alpha_i)}{1+\exp(x_{i2}\beta + z_i\gamma + \alpha_i)}$$

$$= \frac{\exp(x_{i1}\beta + z_i\gamma + \alpha_i) + \exp(x_{i2}\beta + z_i\gamma + \alpha_i)]}{[1+\exp(x_{i1}\beta + z_i\gamma + \alpha_i)][1+\exp(x_{i2}\beta + z_i\gamma + \alpha_i)]}.$$

$$\Pr[(y_{i1},y_{i2}) = (1,0)] = \Pr(y_{i1} = 1)\Pr(y_{i2} = 0).$$

$$\Pr[(y_{i1},y_{i2}) = (1,0) | (y_{i1},y_{i2}) = (1,0) \text{ or } (0,1)]$$

$$= \Pr[(y_{i1},y_{i2}) = (1,0)]/\Pr[(y_{i1},y_{i2}) = (1,0) \text{ or } (0,1)]$$

$$= \frac{\dfrac{\exp(x_{i1}\beta + z_i\gamma + \alpha_i)}{[1+\exp(x_{i1}\beta + z_i\gamma + \alpha_i)][1+\exp(x_{i2}\beta + z_i\gamma + \alpha_i)]}}{\dfrac{\exp(x_{i1}\beta + z_i\gamma + \alpha_i) + \exp(x_{i2}\beta + z_i\gamma + \alpha_i)]}{[1+\exp(x_{i1}\beta + z_i\gamma + \alpha_i)][1+\exp(x_{i2}\beta + z_i\gamma + \alpha_i)]}}$$

$$= \frac{\exp(x_{i1}\beta + z_i\gamma + \alpha_i)}{\exp(x_{i1}\beta + z_i\gamma + \alpha_i) + \exp(x_{i2}\beta + z_i\gamma + \alpha_i)]}$$

$$= \frac{\exp(x_{i1}\beta)}{\exp(x_{i1}\beta) + \exp(x_{i2}\beta)]} \equiv \Lambda(x_{i1},x_{i2},\beta).$$

- Then:

$$f(y_{i1}, y_{i2} \mid (y_{i1}, y_{i2}) = (1,0)\ or\ (0,1))$$
$$= \left[\Lambda(x_{i1}, x_{i2}, \beta)\right]^{y_i} \left[1 - \Lambda(x_{i1}, x_{i2}, \beta)\right]^{1-y_i}$$

where $y_i = 1$ if $(y_{i1},y_{i2}) = (1,0)$; $= 0$ if $(y_{i1},y_{i2}) = (0,1)$.

- The log-likelihood function:

$$l_N(\beta) = \Sigma_{i=1}^{N} \ln f(y_{i1}, y_{i2} \mid (y_{i1}, y_{i2}) = (1,0)\ or\ (0,1)).$$

- A drawback is that it can't estimate γ.

[2] Tobit Panel Models

(1) Model:

$$y_{it}^{*} = x_{it}\beta + z_i\gamma + \alpha_i + \varepsilon_{it},$$

where the ε_{it} are iid $N(0,\sigma_\varepsilon^2)$ and only the $y_{it} = \max(0, y_{it}^{*})$ are observed.

Let $\alpha_i = \sigma_\alpha g_i$ where $g_i \sim N(0,1)$.

(2) Random Effects Model I

- Regressors (x_{it} and z_i) are uncorrelated with α_i.
- The joint pdf of $y_{i1}, \ldots, y_{iT}$ is given by:

$$f(y_{i1},\ldots,y_{iT}) = E_{g_i}\left[r_i(\beta,\gamma,\sigma_\varepsilon,\sigma_\alpha g_i)\right] = \int r_i(\beta,\gamma,\sigma_\varepsilon,\sigma_\alpha g_i)dg_i,$$

where,

$$r_i(\beta,\gamma,\sigma_\varepsilon,\sigma_\alpha g_i) = \prod_{y_{it}>0}\frac{1}{\sigma_\varepsilon}\phi\left\{\frac{y_{it} - x_{it}\beta - z_i\gamma - \sigma_\alpha g_i}{\sigma_\varepsilon}\right\} \times \prod_{y_{it}=0}\left(1 - \Phi\left[\frac{x_{it}\beta + z_i\gamma + \sigma_\alpha g_i}{\sigma_\varepsilon}\right]\right).$$

- Log-likelihood function:

$$l_N(\beta,\gamma,\sigma_\alpha) = \Sigma_i \ln\left[f(y_{i1},\ldots,y_{iT})\right].$$

- MLE requires integrations. LIMDEP can do integrations using an approximation procedure. [LIMDEP computes β, γ and $\rho = \sigma_\alpha^2/(1+\sigma_\alpha^2)$. Data do not have to be balanced.

- Simulated ML (SML) method:
  - Gourieroux and Monfort (1993, Journal of Econometrics).
  - Generate random numbers, $g_i^{(1)}, \ldots, g_i^{(H)}$ for each i (all are N(0,1)).
  - If H is large,

$$r_{iH}(\beta,\gamma,\sigma_\varepsilon,\sigma_\alpha) = \frac{1}{H}\Sigma_{h=1}^{H} r_i(\beta,\gamma,\sigma_\varepsilon,\sigma_\alpha g_i^{(h)})$$
$$\approx E_{g_i}[r_i(\beta,\gamma,\sigma_\varepsilon,\sigma_\alpha g_i)]$$

- Do MLE using r_{iH}. This alternative MLE is called Simulated ML (SML). SML is an efficient as MLE, and is computationally easier.

(3) Random Effects Model II

- Regressors (x_{it} and z_i) are correlated with α_i.
  - $\alpha_i = x_{i1}\lambda_1 + \ldots + x_{iT}\lambda_T + z_i\pi + \eta_i$, where η_i are iid $N(0,\sigma_\eta^2)$.
  - $y_{it}^* = x_{it}\beta + z_i\gamma + \alpha_i + \varepsilon_{it} = x_{it}\beta + x_i^o\lambda + z_i(\gamma+\pi) + \eta_i + \varepsilon_{it}$,

    where $x_i^o = (x_{i1},\ldots,x_{iT})$ and $\lambda' = (\lambda_1',\ldots,\lambda_T')$.
  - Do MLE as in the case I, and estimate β, λ, (γ+π), σ_η^2 and σ_ε^2.

(4) Fixed Effects Models

- See Honore (1992, ECON).
  - Proposes a GMM type of estimator (complicated).
  - Based on the assumption that the ε_{it} are iid and symmetric around zero mean. (The ε_{it} do not have to be normal.)
  - Can't estimate γ.
- Honore (1993, Journal of Econometrics).
  - Extending to a dynamic model.

[4] Panel Selection Model

- Model:

$$y_{it} = x_{it}\beta + z_i\gamma + \alpha_i + \varepsilon_{it};$$

$$h_{it}^* = w_{it}\theta + q_i\xi + \eta_i + v_{it}.$$

  - Observe h_{it} ($h_{it} = 1$ if $h_{it}^* > 0$ and $h_{it} = 0$ if $h_{it}^* < 0$).
  - Observe y_{it} only if $h_{it} = 1$.

- Can use the random effect assumptions to estimate the model.
- For the fixed effects treatments, see Kyriazidou (1997, ECON).

  → See Honore and Kyriazidou (2000, ECON).

[5] Ordered probit model

(1) Basic Model

- $y_t^* = x_t\beta + \varepsilon_t$, $\varepsilon_t \sim N(0,1)$.
- Observe y_t, where

$$
\begin{aligned}
y_t &= 0, \text{ if } y_t^* < \mu_0 \\
&= 1, \text{ if } \mu_0 < y_t^* < \mu_1 \\
&= 2, \text{ if } \mu_1 < y_t^* < \mu_2 \\
&\vdots \\
&= J, \text{ if } y_t^* > \mu_{J-1}.
\end{aligned}
$$

Note:

- Need a restriction, $\mu_0 = 0$, for identification.

  → OK for MLE of β (except the overall intercept term).

Example:

- Survey data (hate, so-so, like): y_t^* = degree of preference.

Unknown Parameters:

β, μ_1, ... , μ_{J-1}.

Note:

If J = 1, the model becomes the usual probit model.

Probabilities:

- p_{0t} $\equiv \Pr(y_t = 0) = \Pr(y_t^* < \mu_0 = 0) = \Pr(\varepsilon_t < 0 - x_t\beta) = \Phi(-x_t\beta)$;

  p_{1t} $\equiv \Pr(y_t = 1) = \Pr(0 = \mu_0 < y_t^* < \mu_1)$

  $= \Pr(y_t^* < \mu_1) - \Pr(y_t^* < \mu_0) = \Phi(\mu_1 - x_t\beta) - \Phi(-x_t\beta)$;

  p_{2t} $\equiv P(y_t = 2) = \Phi(\mu_2 - x_t\beta) - \Phi(\mu_1 - x_t\beta)$;

  :

  p_{Jt} $\equiv \Pr(y_t = J) = 1 - \Phi(\mu_{J-1} - x_t\beta)$.

Log-likelihood function:

- Define $d_{jt} = 1$ if $y_t = j$; $= 0$ otherwise.
- Then, the pdf of y_t is given by $(p_{0t})^{d_{0t}}(p_{1t})^{d_{1t}}...(p_{Jt})^{d_{Jt}}$.
- $l_T(\beta, \mu_1, ..., \mu_{J-1}) = \Sigma_{t=1}^T \{d_{0t}\ln(p_{0t}) + ...d_{Jt}\ln(p_{Jt})\}$.

(2) Model with Heteroskedasticity

- $y_t^{**} = x_t\beta + \varepsilon_t$, $\varepsilon_t \sim N\left(0, \left[\exp(z_t\gamma)\right]^2\right)$:
  - z_t may include some variables in x_t.
  - z_t should not include overall intercept term.

- Observe y_t, where

$$
\begin{aligned}
y_t &= 0, \text{ if } y_t^{**} < \mu_0 = 0 \\
&= 1, \text{ if } \mu_0 < y_t^{**} < \mu_1 \\
&= 2, \text{ if } \mu_1 < y_t^{**} < \mu_2 \\
&\quad \vdots \\
&= J, \text{ if } y_t^{**} > \mu_{J-1}.
\end{aligned}
$$

Unknown Parameters:

$\beta, \gamma, \mu_1, \dots, \mu_{J-1}$.

Redefine the model:

$$y_t^* = y_t^{**} / \exp(z_t\gamma)$$

$$\rightarrow y_t^* = x_t\beta / \exp(z_t\gamma) + v_t, \text{ where } v_t \sim N(0,1).$$

Probabilities:

- $p_{0t} \equiv \Pr(y_t = 0) = \Pr(y_t^{**} < \mu_0 = 0) = \Pr(y_t^* < 0/\exp(z_t\gamma))$

$$= \Pr(x_t\beta/\exp(z_t\gamma) + v_t < 0) = \Pr(v_t < -x_t\beta/\exp(z_t\gamma))$$

$$= \Phi(-x_t\beta/\exp(z_t\gamma));$$

$p_{1t} \equiv \Pr(y_t = 1) = \Phi((\mu_1 - x_t\beta)/\exp(z_t\gamma)) - \Phi(-x_t\beta/\exp(z_t\gamma));$

$p_{2t} \equiv P(y_t = 2)$

$$= \Phi((\mu_2 - x_t\beta)/\exp(z_t\gamma)) - \Phi((\mu_1 - x_t\beta)/\exp(z_t\gamma));$$

:

$p_{Jt} \equiv \Pr(y_t = J) = 1 - \Phi((\mu_{J-1} - x_t\beta)/\exp(z_t\gamma)).$

Log-likelihood function:

- Define d_{jt} = 1 if y_t = j; = 0 otherwise.
- $l_T(\beta, \mu_1, ..., \mu_{J-1}) = \Sigma_{t=1}^{T} \{d_{0t} \ln(p_{0t}) + ... d_{Jt} \ln(p_{Jt})\}$.

(3) Application [Hausman, Lo and Mackinlay (1992, JF)]

Situation:

- Changes in prices of stock are quoted in discrete units (ticks).

  $\rightarrow$ 1 tick for equities = $0.125 (1/8);

  $\rightarrow$ 1 ticks for US treasury bond = $ 0.03125 (1/32).

- For NYSE, most of transactions occurs with zero or one-tick price changes. And, price changes greater than 4 ticks are greatly rare.

- Let y_t = change in transaction prices in ticks = -4,-3, ..., 0,...,3, 4.

- Let y_t^{**} = changes in actual continuous prices.

- Wish to estimate the effects of some exogenous variables x_t on y_t^{**}.
- Wish to estimate the effects of x_t on $\Pr(y_t = s)$, s = -4, ... , 4.

Solution:

- Let $y_t^{**} = x_t\beta + \varepsilon_t$, $\varepsilon_t \sim N\left(0, [\exp(z_t\gamma)]^2\right)$.
- Redefine:

  $y_t = 0$ if actual $y_t = -4$ or more,

  $y_t = 1$ if actual $y_t = -3$, ...
- Do MLE for the ordered probit with heteroskedasticity.

[6] Unordered choice models

Example:

The dependent variable y may take many different values, 1, 2, ... , n.

For example, $y_t = 1$ if drive, $y_t = 2$ if bus; and $y_t = 3$ if taxi.

(1) Multinomial Logit Models (Theil)

- Assume:

  $\ln[\Pr(y_t = j)/\Pr(y_t = i)] = x_t(\beta_j - \beta_i)$, i, j = 1, 2, ... , n,

  where x_t contains individual characteristics and β_j for choice j.

- This assumption (with the fact that the sum of probs = 1) implies:

$$\Pr(y_t = j) = \frac{\exp(x_t\beta_j)}{\Sigma_{i=1}^{n} \exp(x_t\beta_i)}.$$

- Need to normalize β's (See Greene, 0. 721). Usually, $\beta_1 = 0$.
  - Under this normalization,

    $\Pr(y_t = 1) = 1/\xi_t$, $\xi_t = 1 + \Sigma_{i=2}^{n} \exp(x_t\beta_i)$;

    $\Pr(y_t = j) = \exp(x_t\beta)/\xi_t$, j = 2, ... , n.

- Interpretation of estimates:
    - sgn(β_{jh}) (sign of β_{jh}) indicates the direction of the effects of x_{th} on $Pr(y_t = j)/Pr(y_t = 1)$.

(2) Conditional Logit Model (McFadden)

- $\ln[Pr(y_t = j)/Pr(y_t = i)] = (x_{tj} - x_{ti})\theta$, j, i = 1, 2, ... , n,

  where x_{jt} includes the characteristics of choice.

- Example: (Boskin, JPE, 1982)

  y_t = occupation;

  x_{tj} = variables such as the present value for the jth occupation, training cost/net worth of the jth occupation, and the present value of time unemployed for the jth occupation.

- $Pr(y_t = j) = \dfrac{\exp(x_{tj}\theta)}{\Sigma_{i=1}^{n} \exp(x_{ti}\theta)}$.

(3) Unordered Multiple Probit (Hausman and Wise, ECON, '78)

- Problems in multinomial logit models (MLM):
  - In MLM, any possible correlation among choices is not allowed.
    - Called IIA (Independence of Irrelevant Alternatives).
- IIA: In MLM, $Pr(y_t = j)/Pr(y_t = i)$ does not depend on the number or nature of other alternatives.

- Red bus-blue bus problem:
  - Suppose you have two alternative choices: blue bus and red buses. These choices must be highly correlated. However, MLM does not allow this.
  - You initially have two choices: red bus and drive. Assume:

    Pr(red bus)=Pr(drive) = 0.5 $\rightarrow$ Pr(red bus)/Pr(drive) = 1.
    - Now, let's add blue bus to the choice set.
    - Intuitively, Pr(red bus)/Pr(blue bus) = 1.
    - In MLM, Pr(red bus)/Pr(drive) = 1:
      - Pr(red bus) = Pr(blue bus) = Pr(drive) = 1/3.
      - Quite unreasonable.
    - Correct probabilities must be:

      Pr(red bus) = Pr(blue bus) = 1/4, and Pr(drive) = 1/2.
  - To avoid IIA, need to use multivariate normal distributions. But this alternative is very messy.

[7] Bivariate Probit Models

(1) Bivariate Normal Distribution

- $\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$.

- $f(\varepsilon_1, \varepsilon_2 \mid \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{\varepsilon_1^2 - 2\rho\varepsilon_1\varepsilon_2 + \varepsilon_2^2}{2(1-\rho^2)} \right]$, where $-1 \le \rho \le 1$.

- The cdf is denoted by:

$$F(h,k \mid \rho) = \Pr(\varepsilon_1 < h, \varepsilon_2 < k) = \Pr(\varepsilon_1 > -h, \varepsilon_2 > -k)$$

$$= \int_{-h}^{\infty} \int_{-k}^{\infty} f(\varepsilon_1, \varepsilon_2 \mid \rho) d\varepsilon_2 d\varepsilon_1.$$

- Facts:
  - $\Pr(\varepsilon_1 > -h, \varepsilon_2 > -k) = F(h,k|\rho)$.
  - $\Pr(\varepsilon_1 > -h, \varepsilon_2 < -k) = \Pr(\varepsilon_1 > -h) - \Pr(\varepsilon_1 > -h, \varepsilon_2 > -k)$
    $= \Phi(h) - F(h,k \mid \rho)$
  - $\Pr(\varepsilon_1 < -h, \varepsilon_2 > -k) = \Phi(k) - F(h,k \mid \rho)$
  - $\Pr(\varepsilon_1 < -h, \varepsilon_2 < -k) = 1 - \Phi(h) - \Phi(k) + F(h,k \mid \rho)$.
  - $\frac{\partial F(h,k \mid \rho)}{\partial h} = \phi(h)\Phi\left[\frac{k - \rho h}{\sqrt{1-\rho^2}}\right]$; $\frac{\partial F(h,k \mid \rho)}{\partial k} = \phi(k)\Phi\left[\frac{h - \rho k}{\sqrt{1-\rho^2}}\right]$;

- $\dfrac{\partial F(h,k\mid\rho)}{\partial\rho} = f(h,k\mid\rho)$.

(2) Full Observability Model

Model:

$y_{1t}^* = x_{1t}\beta_1 + \varepsilon_{1t}$;

$y_{2t}^* = x_{2t}\beta_2 + \varepsilon_{2t}$.

- Observe: $y_{1t} = 1$ if $y_{1t}^* > 0$; $y_{1t} = 0$ if $y_{1t}^* < 0$

  $y_{2t} = 1$ if $y_{2t}^* > 0$; $y_{2t} = 0$ if $y_{2t}^* < 0$

Example: AMEX card.

y_{1t}: buy a good from CostCo or not.

y_{2t}: use AMEX card or not.

Four possible outcomes:

$$p_{11,t} \equiv \Pr(y_{1t}=1, y_{2t}=1) = \Pr(\varepsilon_{1t} > -x_{1t}\beta_1, \varepsilon_{2t} > -x_{2t}\beta_2)$$
$$= F(x_{1t}\beta_1, x_{2t}\beta_2 \mid \rho) \equiv F_t$$

$$p_{10,t} \equiv \Pr(y_{1t}=1, y_{2t}=0) = \Phi(x_{1t}\beta_1) - F_t \equiv \Phi_{1t} - F_t$$

$$p_{01,t} \equiv \Pr(y_{1t}=0, y_{2t}=1) = \Phi(x_{2t}\beta_2) - F_t \equiv \Phi_{2t} - F_t$$

$$p_{00,t} \equiv \Pr(y_{1t}=0, y_{2t}=0) = 1 - \Phi_{1t} - \Phi_{2t} + F_t$$

PDF of y_{1t} and y_{2t}:

$$\left(p_{11,t}\right)^{y_{1t}y_{2t}}\left(p_{10,t}\right)^{y_{1t}(1-y_{2t})}\left(p_{01,t}\right)^{(1-y_{1t})y_{2t}}\left(p_{00,t}\right)^{(1-y_{1t})(1-y_{2t})}.$$

Log-likelihood function:

$$l_T(\beta_1,\beta_2,\rho)=\Sigma_{t=1}^{T}\begin{Bmatrix} y_{1t}y_{2t}\ln(p_{11,t})+y_{1t}(1-y_{2t})\ln(p_{10,t}) \\ +(1-y_{1t})y_{2t}\ln(p_{01,t})+(1-y_{1t})(1-y_{2t})\ln(p_{00,t}) \end{Bmatrix}.$$

Note:

- Suppose that $\rho = 0$. Then, $F(x_{1t}\beta_1, x_{2t}\beta_2 \mid \rho) = \Phi(x_{1t}\beta_1)\Phi(x_{2t}\beta_2)$.
  - $l_T(\beta_1,\beta_2)$ = probit $l_T(\beta_1)$ for y_{1t} + probit $l_T(\beta_2)$ for y_{2t}.
  - β_1 and β_2 can be estimated separately by separate probits.
- Even if $\rho \neq 0$, separate estimators are consistent, but not efficient. The bivariate probit ML is more efficient.

(3) Censored Probit (Bivariate Probit with Selection)

Model:

- We always observe $y_{1t} = 1$ if $y_{1t}^* > 0$ and $y_{1t} = 0$ if $y_{1t}^* > 0$.
- We observe y_{2t} iff $y_{1t} = 1$,

  and $y_{2t} = 1$ if $y_{2t}^* > 0$ and $y_{2t} = 0$ if $y_{2t}^* > 0$.

Example: Farber (1983, Research in Labor Economics)

y_{1t} = whether a worker wants to join union or not.

y_{2t} = whether union wants the worker or not.

Three cases:

$$\Pr(y_{1t}=1, y_{2t}=1) = F(x_{1t}\beta_1, x_{2t}\beta_2 \mid \rho) \equiv F_t$$

$$\Pr(y_{1t}=1, y_{2t}=0) = \Phi(x_{1t}\beta_1) - F_t \equiv \Phi_{1t} - F_t$$

$$\Pr(y_{1t}=0) = 1 - \Phi_{1t}.$$

PDF of y_{1t} and y_{2t}:

$$\left(F_t\right)^{y_{1t}y_{2t}} \left(\Phi_{1t} - F_t\right)^{y_{1t}(1-y_{2t})} \left(1-\Phi_{1t}\right)^{(1-y_{1t})}.$$

Log-likelihood function:

$$l_T(\beta_1, \beta_2, \rho) = \Sigma_{t=1}^{T} \begin{Bmatrix} y_{1t}y_{2t}\ln(F_t) + y_{1t}(1-y_{2t})\ln(\Phi_{1t} - F_t) \\ +(1-y_{1t})\ln(1-\Phi_{1t}) \end{Bmatrix}$$

Note:

- If $\rho = 0$,

  $l_T(\beta_1, \beta_2, \rho)$ = probit for y_{1t} with all observations + probit for y_{2t} with the observations with $y_{1t} = 1$.

  - β_1 and β_2 can be estimated by separate probits.
  - Notice that the probit for β_2 uses observations with $y_{1t} = 1$ only, not all observations.

- If $\rho \neq 0$,
  - the probit ML estimator of β_1 is still consistent, but probit of β_2 is inconsistent.

- Very often, you may fail to obtain the censored MLE.
  - May need to restrict $\rho = 0$.
  - If we do, have to interpret the y_{1t}^* equation as a conditional one defined given $y_{1t} = 1$. (It describes $\Pr(y_{2t} = 1 | y_{1t} = 1)$.)
- For the censored probit with unrestricted ρ, the second equation is interpreted as unconditional one.

(4) Poirier Probit (Journal of Econometrics, 1980)

Model:

- Observe only $y_t = y_{1t}y_{2t}$: $y_t = 1$ if $y_{1t}^* > 0$ and $y_{2t}^* > 0$; $= 0$, otherwise.

Example:

- Two member committee with unanimity rule.

Two cases:

$$\Pr(y_t = 1) = \Pr(y_{1t}^* > 0, y_{2t}^* > 0) = F_t$$

$$\Pr(y_t = 0) = 1 - F_t$$

PDF of y_{1t} and y_{2t}:

$$\left(F_t\right)^{y_t}\left(1 - F_t\right)^{1-y_t}.$$

Log-likelihood function:

$$l_T(\beta_1, \beta_2, \rho) = \Sigma_{t=1}^T \left\{ y_t \ln(F_t) + (1 - y_{1t})\ln(1 - F_t) \right\}$$

Note:

- Separate probits are impossible even if ρ = 0.
- If ρ = 0, $l_T(\beta_1, \beta_2) = \Sigma_{t=1}^T \left\{ y_t \ln(\Phi_{1t}\Phi_{2t}) + (1 - y_{1t})\ln(1 - \Phi_{1t}\Phi_{2t}) \right\}$.
    - MLE of Abowd and Farber (1982, ILRR).
    - When ρ is restricted at zero, the second equation should be interpreted as conditional one.
- If ρ ≠ 0, the A-F MLE in inconsistent for the estimation of unconditional equations for y_{1t} and y_{2t}.
- If $x_{1t} = x_{2t}$, can't distinguish which estimates are for which equations.

[8] Double Selection Model

Basic Model:

1) $y_{1t}^{*} = x_{1t}\beta_1 + \varepsilon_{1t}$

2) $y_{2t}^{*} = x_{2t}\beta_2 + \varepsilon_{2t}$

3) $y_{3t} = x_{3t}\beta_3 + \varepsilon_{3t}$.

Assumptions:

- 1) and 2): a bivariate probit model.
- Let y_{1t} and y_{3t} be the dummy variables for 1) and 2).
- Observe y_{3t} only if $y_{1t} = y_{2t} = 1$.
- $\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho & \sigma_{13} \\ \rho & 1 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix} \right)$.

Example:

y_{1t} = LF; y_{2t} = EMP_t → consored probit.

y_{3t} = $LRATE_t$ (log of wage rate).

Two-Stage Estimation:

- $$\begin{aligned} &E(y_{3t} \mid y_{1t}^* > 0, y_{2t}^* > 0) \\ &\quad = x_{3t}\beta_3 + E(\varepsilon_{3t} \mid \varepsilon_{1t} > -x_{1t}\beta_1, \varepsilon_{2t} > -x_{2t}\beta_2) \\ &\quad = x_{3t}\beta_3 + \sigma_{13}\lambda_{1t} + \sigma_{23}\lambda_{2t}, \end{aligned}$$

  where,

  $$\lambda_{1t} = \frac{\phi(x_{1t}\beta_1)\Phi\left[\dfrac{x_{2t}\beta_2 - \rho x_{1t}\beta_1}{\sqrt{1-\rho^2}}\right]}{F(x_{1t}\beta_1, x_{2t}\beta_2 \mid \rho)};$$

  $$\lambda_{2t} = \frac{\phi(x_{2t}\beta_2)\Phi\left[\dfrac{x_{1t}\beta_1 - \rho x_{2t}\beta_2}{\sqrt{1-\rho^2}}\right]}{F(x_{1t}\beta_1, x_{2t}\beta_2 \mid \rho)}.$$

Note:

- If ρ = 0, we have:

  $$\lambda_{1t} = \frac{\phi(x_{1t}\beta_1)\Phi(x_{2t}\beta_2)}{\Phi(x_{1t}\beta_1)\Phi(x_{2t}\beta_2)} = \frac{\phi(x_{1t}\beta_1)}{\Phi(x_{1t}\beta_1)};\ \lambda_{2t} = \frac{\phi(x_{2t}\beta_2)}{\Phi(x_{2t}\beta_2)}.$$

  → inverse Mill's ratios.

Note:

- For observed y_{3t},

  $y_{3t} = x_{3t}\beta_3 + \sigma_{13}\lambda_{1t} + \sigma_{23}\lambda_{2t} + v_t$,

  where,

  $E(v_t \mid y_{2t}^* > 0, y_{2t}^* > 0) = 0$;

  $\text{var}(v_t \mid y_{1t}^* > 0, y_{2t}^* > 0) = \pi_t = \sigma_{33} - \xi_t$;

  $$\xi_t = \sigma_{13}^2[(x_{1t}\beta_1)\lambda_{1t} + \lambda_{2t}^2 + \rho\lambda_{3t}] + \sigma_{23}^2[(x_{2t}\beta_2)\lambda_{2t} + \lambda_{1t}^2 + \rho\lambda_{3t}] \\ -2\sigma_{13}\sigma_{23}(\lambda_{3t} - \lambda_{1t}\lambda_{2t});$$

  $\lambda_{3t} = f(x_{1t}\beta_1, x_{2t}\beta_2 \mid \rho) / F(x_{1t}\beta_1, x_{2t}\beta_2 \mid \rho)$.

Two-step estimation

- Do bivariate probit and get $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\rho}$, $\hat{\lambda}_{1t}$ and $\hat{\lambda}_{2t}$.
- Do OLS on $y_{3t} = x_{3t}\beta_3 + \sigma_{13}\hat{\lambda}_{1t} + \sigma_{23}\hat{\lambda}_{2t} + error$.

Facts on the two-step estimator:

- Consistent.
- F or Wald tests for $\sigma_{13} = \sigma_{23} = 0$ (no selection) using usual OLS covariance matrix ≈ LM test, while individual t tests for $\sigma_{13} = 0$ and $\sigma_{23} = 0$ are wrong [Ahn (Economic Letters, 1992)].
- All other t or F tests based on usual OLS covariance matrix are all wrong.

Details on two-step estimation:

- The model we wish to estimate:

  $$y_{3t} = x_{3t}\beta_3 + \sigma_{13}\lambda_{1t} + \sigma_{23}\lambda_{2t} + v_t .$$

  But need to use $\hat{\lambda}_{1t}$ and $\hat{\lambda}_{2t}$.

- Let's consider the consequence of this substitution:

  $$y_{3t} = x_{3t}\beta_3 + \sigma_{13}\hat{\lambda}_{1t} + \sigma_{23}\hat{\lambda}_{2t} + [\sigma_{13}(\lambda_{1t} - \hat{\lambda}_{1t}) + \sigma_{23}(\lambda_{2t} - \hat{\lambda}_{2t}) + v_t],$$

  where [•] is the error term in the model we estimate.

- As we discussed above, the error term v_t is heteroskedastic unless $\sigma_{13} = \sigma_{23} = 0$.

- The error component $\sigma_{13}(\lambda_{1t} - \hat{\lambda}_{1t}) + \sigma_{23}(\lambda_{2t} - \hat{\lambda}_{2t})$ are autocorrelated because they are the functions of estimated β_1, β_2, and ρ.

Derivation of the Corrected Covariance Matrix of the Two-Step Estimator [Ham, ReSTUD, 1982]

- Some notation:

  $\theta = (\beta_1, \beta_2, \rho)'$;

  $\hat{\theta}$ = bivariate probit ML estimator with $\hat{\Omega}$ = estimated $Cov(\hat{\theta})$

  $z_t = (x_{3t}, \hat{\lambda}_{1t}, \hat{\lambda}_{2t})$; $\gamma = (\beta_3', \sigma_{13}, \sigma_{23})'$;

  $\hat{v}_t$ = OLS residual from the second stage OLS (only for observed y_{3t}).

- In order to create F_t, use BVN command in LIMDEP and CDFBVN in GAUSS.

- By Taylor expansion,

$$
\begin{aligned}
y_{3t} &= x_{3t}\beta_3 + \sigma_{13}\hat{\lambda}_{1t} + \sigma_{23}\hat{\lambda}_{2t} + [\sigma_{13}(\lambda_{1t} - \hat{\lambda}_{1t}) + \sigma_{23}(\lambda_{2t} - \hat{\lambda}_{2t}) + v_t] \\
&= x_{3t}\beta_3 + \sigma_{13}\hat{\lambda}_{1t} + \sigma_{23}\hat{\lambda}_{2t} \\
&\quad + \left(-\frac{\partial \lambda_{1t}}{\partial \theta'}(\hat{\theta} - \theta) - \frac{\partial \lambda_{2t}}{\partial \theta'}(\hat{\theta} - \theta) + v_t \right) \\
&= x_{3t}\beta_3 + \sigma_{13}\hat{\lambda}_{1t} + \sigma_{23}\hat{\lambda}_{2t} \\
&\quad + \left(\left(-\frac{\partial \lambda_{1t}}{\partial \theta'} - \frac{\partial \lambda_{2t}}{\partial \theta'} \right)(\hat{\theta} - \theta) + v_t \right)
\end{aligned}
$$

- Important terms:

$\lambda_{3t} = f(x_{1t}\beta_1, x_{2t}\beta_2 \mid \rho) / F(x_{1t}\beta_1, x_{2t}\beta_2 \mid \rho)$;

$$
\begin{aligned}
\xi_t &= \sigma_{13}^2[(x_{1t}\beta_1)\lambda_{1t} + \lambda_{2t}^2 + \rho\lambda_{3t}] + \sigma_{23}^2[(x_{2t}\beta_2)\lambda_{2t} + \lambda_{1t}^2 + \rho\lambda_{3t}] \\
&\quad -2\sigma_{13}\sigma_{23}(\lambda_{3t} - \lambda_{1t}\lambda_{2t});
\end{aligned}
$$

$\pi_t = \sigma_{33} - \xi_t$;

$w_{2t} = \sigma_{13}(\lambda_{3t} - \lambda_{1t}\lambda_{2t}) + \sigma_{23}[-(x_{2t}\beta_2)\lambda_{2t} - \lambda_{2t}^2 - \rho\lambda_{3t}]$;

$w_{1t} = \sigma_{23}(\lambda_{3t} - \lambda_{1t}\lambda_{2t}) + \sigma_{13}[-(x_{1t}\beta_1)\lambda_{1t} - \lambda_{1t}^2 - \rho\lambda_{3t}]$;

$$
\begin{aligned}
w_{3t} &= \sigma_{13}[-\{(x_{1t}\beta_1 - \rho x_{2t}\beta_2)/(1-\rho^2)\}\lambda_{3t} - \lambda_{1t}\lambda_{3t}] \\
&\quad +\sigma_{23}[-\{(x_{2t}\beta_2 - \rho x_{1t}\beta_1)/(1-\rho^2)\}\lambda_{3t} - \lambda_{2t}\lambda_{3t}];
\end{aligned}
$$

$$\hat{\sigma}_{33} = \frac{1}{T_3}\Sigma_{t=1}^{T_3}\hat{v}_t + \frac{1}{T_3}\Sigma_{t=1}^{T_3}\hat{\xi}_t .$$

- In LIMDEP, XBR computes sample mean. Remember to use data with observed y_{3t} only.

- $M_1 = \Sigma_{t=1}^{T_3} \begin{pmatrix} w_{2t} x_{2t}' z_t \\ w_{1t} x_{1t}' z_t \\ w_{3t} z_t \end{pmatrix}; M_2 = \Sigma_{t=1}^{T_3} \pi_t z_t' z_t; M_3 = \Sigma_{t=1}^{T_3} z_t' z_t$.

- Covariance matrix:

$$Cov(\hat{\gamma}) = M_3^{-1} \left(M_1' \hat{\Omega} M_1 + M_2 \right) M_3^{-1}.$$

- Procedure:
  - Get $\hat{\theta}$ and $\hat{\Omega}$.
  - Do OLS on $y_{3t} = z_t \gamma + err$ (using data with observed y_{3t}), and get $\hat{\gamma}$.
  - For $Cov(\hat{\gamma})$, use observations with observed y_{3t}.
    - Estimate M_1, M_2 using $\hat{v}_t^2$ instead of π_t.