# 1. 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:

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

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

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

  → $\beta_j > 0$ 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:

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

       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 $l_{T,R} = \ln(L_{T,R})$.

       → Probit without $x_{t4}$ and $x_{t5}$.

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

- 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:

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

  $\varepsilon_t$ ~ 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 $\beta$ 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 $\varepsilon$)

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

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

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

  by Kernel Smoothing.

  → See Härdle (1990, Applied Nonparametric Regression.)

- LIMDEP can do this.

3) Nonparametric Estimation of $\beta$:

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

4) Manski (Journal of Econometrics, 1975)

- "Maximum Score Estimator." (MSE)
- Motivation: The distribution of $\varepsilon_t$ 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] Probit/Logit Panel Models

(1) Model:

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

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

(2) Fixed effects model

- Treat the $\alpha_i$ 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 $\alpha_i$.
- $\alpha_i$ 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:

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

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 $\beta$, $\gamma$ 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,

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

  - 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 $\alpha_i$.
  - See Chamberlain (1984, Handbook of Econometrics).
  - $\alpha_i = x_{i1}\lambda_1 + \ldots + x_{iT}\lambda_T + z_i\pi + \eta_i$, where $\eta_i$ 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 and (γ+π).

(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.

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

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

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

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

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

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

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

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

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

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

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

- Then:

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

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

- The log-likelihood function:

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

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

## [3] Tobit Panel Models

(1) Model:

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

where the $\varepsilon_{it}$ 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 $\alpha_i$.
- The joint pdf of $y_{i1}$, ... , $y_{iT}$ is given by:

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

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:

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

- MLE requires integrations. LIMDEP can do integrations using an approximation procedure. [LIMDEP computes $\beta$, $\gamma$ 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 $\alpha_i$.
  - $\alpha_i = x_{i1}\lambda_1 + \ldots + x_{iT}\lambda_T + z_i\pi + \eta_i$, where $\eta_i$ 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), $\sigma_\eta^2$ and $\sigma_\varepsilon^2$.

(4) Fixed Effects Models

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

## [4] Panel Selection Models

- 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

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

**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:**

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

**Note:**

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

**Probabilities:**

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

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

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

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

  :

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

**Log-likelihood function:**

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

(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

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

**Unknown Parameters:**

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

**Redefine the model:**

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

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

**Probabilities:**

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

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

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

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

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

$\vdots$

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

**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 $\beta_j$ 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:
  - $\text{sgn}(\beta_{jh})$ (sign of $\beta_{jh}$) 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 $-1 \leq \rho \leq 1$.

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

## (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:

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

Note:

- Suppose that ρ = 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 $y_{2t}$.
  - $\beta_1$ and $\beta_2$ can be estimated separately by separate probits.
- Even if ρ ≠ 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 $y_{1t} = 1$.

  - $\beta_1$ and $\beta_2$ can be estimated by separate probits.
  - Notice that the probit for $\beta_2$ uses observations with $y_{1t} = 1$ only, not all observations.

- If $\rho \neq 0$,
  - the probit ML estimator of $\beta_1$ is still consistent, but probit of $\beta_2$ 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 $\rho$, 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 $\rho = 0$.
- If $\rho = 0$, $l_T(\beta_1, \beta_2) = \Sigma_{t=1}^T \left\{ y_t \ln(\Phi_{1t}\Phi_{2t}) + (1 - y_{1t}) \ln(1 - \Phi_{1t}\Phi_{2t}) \right\}$ .
  - MLE of Abowd and Farber (1982, ILRR).
  - When $\rho$ is restricted at zero, the second equation should be interpreted as conditional one.
- If $\rho \neq 0$, 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 $y_{2t}$ 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 components $\sigma_{13}(\lambda_{1t} - \hat{\lambda}_{1t}) + \sigma_{23}(\lambda_{2t} - \hat{\lambda}_{2t})$ are autocorrelated because they are the functions of estimated $\beta_1$, $\beta_2$, and $\rho$.

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 $\pi_t$.

## [9] Count Data Models

### (1) Some Discrete Probability Density Functions

- Binomial Distribution:
  - Tossing a coin m times.
  - p = probability of having head from a trial.
  - y = # of having heads from n trials (y = 0, 1, ... , m).
  - $f_b(y|n) = \begin{pmatrix} m \\ y \end{pmatrix} p^y (1-p)^{m-y} = \frac{m!}{y!(m-y)!} p^y (1-p)^{m-y}$.
  - E(y) = mp; var(y) = mp(1-p).

- Poisson Distribution:
  - Let p = μ/m for a binomial distribution.
  - $f_p(y) = \lim_{m\to\infty} f_b(y|m) = \frac{\mu^y}{y!} e^{-\mu}$, y = 0, 1, 2, ... .
  - E(y) = var(y) = μ.

**Digression to the Chocolate Chip Cookies problem:**

- Wish to put at least one CC on a cookie with 95%.
- The CC machine locates CC's in a cookie following a Poisson Distribution with $\mu$.

  $\rightarrow$ $\Pr(y = 0) < 0.05 \rightarrow e^{-\mu} < 0.05 \rightarrow \mu > -\ln(0.05) = 2.995$.

  $\rightarrow$ You need a machine with $\mu$ = at least 3.

- Geometric Distribution:
  - y = # of trials until having a head.
  - $f_g(y) = (1-p)^{y-1} p$, y = 1, ....
  - $E(y) = 1/p$; $var(y) = (1-p)/p^2$.

- Negative Binomial Distribution:
  - y = # of trials until having r heads.
  - $f_n(y) = \binom{y-1}{r-1} p^r (1-p)^{y-r}$, y = r, r+1, ...
  - $E(y) = r/p$; $var(y) = r(1-p)/p^2$.

### (2) Basic Poisson Regression Model

- Assume $y_i$ i.i.d. with Poisson($\mu_i$), where $\mu_i = \exp(x_i'\beta)$.

  → EX: $y_i$ = # of visiting doctors.

- $f(y_i | x_i) = \frac{e^{-\mu_i}(\mu_i)^{y_i}}{y_i!} = \frac{e^{-e^{x_i'\beta}}(e^{x_i'\beta})^{y_i}}{y_i!}$.

- $\ln f(y_i | x_i) = -e^{x_i'\beta} + y_i(x_i'\beta) - \ln(y_i!)$.

- $l_N(\beta) = \Sigma_{i=1}^N \left\{ y_i(x_i'\beta) - e^{x_i'\beta} - \ln(y_i!) \right\}$.

- $\frac{\partial l_N(\beta)}{\partial \beta} = \Sigma_{i=1}^N x_i(y_i - e^{x_i'\beta})$.

  → The Poisson ML estimator of β can be viewed as a GMM estimator based on

  $$E\left(x_i(y_i - e^{x_i'\beta})\right) = 0.$$

  → This moment condition is valid as long as $E(y_i | x_i) = e^{x_i'\beta}$.

  → It means that the Poisson ML estimator of β is consistent even if the Poisson assumption is incorrect.

- $B_N(\beta) = \Sigma_{i=1}^N x_i x_i'(y_i - e^{x_i'\beta})^2$; $H_N(\beta) = -\Sigma_{i=1}^N e^{x_i'\beta} x_i x_i'$.

- If the Poisson assumption is truly correct, use

  $$[-H_N(\hat{\beta}_{POI-ML})]^{-1} \text{ or } [B_N(\hat{\beta}_{POI-ML})]^{-1}$$

  as an estimate of $Cov(\hat{\beta}_{POI-ML})$.

- If you are not sure, use

$$[-H_N(\hat{\beta}_{POI-ML})]^{-1} B_N(\hat{\beta}_{POI-ML}) [-H_N(\hat{\beta}_{POI-ML})]^{-1}.$$

- For the measures of goodness of fit, see Greene.

- In fact, we can estimate β by NLLS applied to $y_i = e^{x_i'\beta} + \varepsilon_i$ with heteroskedastic error terms.

**Digression to NLLS with Heteroskedastic Errors:**

- $y_i = h(x_i,\beta) + \varepsilon_i$.
- Let $H(x_i,\beta) = \dfrac{\partial h(x_i,\beta)}{\partial \beta}$.
- The NLLS estimator of β ($\hat{\beta}_{NL}$) minimizes:

  $$\Sigma_{i=1}^N (y_i - h(x_i,\beta))^2.$$

- $$Cov(\hat{\beta}_{NL}) \approx \left(\Sigma_{i=1}^N H(x_i,\hat{\beta}_{NL})H(x_i,\hat{\beta}_{NL})'\right)^{-1} \times \Sigma_{t=1}^T e_i^2 H(x_i,\hat{\beta}_{NL})H(x_i,\hat{\beta}_{NL})' \times \left(\Sigma_{t=1}^T H(x_t,\hat{\beta}_{NL})H(x_t,\hat{\beta}_{NL})'\right)^{-1}$$

  where $e_i = y_i - h(x_i,\hat{\beta}_{ML})$.

**End of Digression**

**(3) Compound Poisson Model (Negative Binomial Model):**

- Hausman, Hall, Griliches (HHG, ECON, 1984), and Cameron and Trivedi (CT, JAE, 1986).
- Assume that the $y_i$ follow Poisson($\lambda_i$), where $\lambda_i = e^{x_i'\beta + \alpha_i} = \mu_i e^{\alpha_i}$,

  $\mu_i = e^{x_i'\beta}$, $E(e^{\alpha_i}) = 1$, and the $e^{\alpha_i}$ follow a Gamma distribution:

$$f_{gamma}(\eta) = \frac{\theta^\theta}{\Gamma(\theta)} e^{-\theta\eta} \eta^{\theta-1},$$

  where $0 < \eta < \infty$ and $\Gamma(\theta) = \int_0^\infty t^{\theta-1} e^{-t} dt$.

  → Here, $\alpha_i$ is an unobservable individual effect.

**Digression to Gamma distribution:**

- The most general form of the Gamma density function is given:

$$f_{gen-gamma}(y \mid \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} y^{\alpha-1} e^{-y/\beta},$$

  where y is a continuous positive random variable ( y > 0).
- $E(y) = \alpha\beta$ and $var(y) = \alpha\beta^2$.
- $f_{gamma}(\eta)$ is obtained by setting $y = \eta$, $\alpha = \theta$ and $\beta = 1/\theta$.

  → Choose $\alpha = \theta$ and $\beta = 1/\theta$ to make $E(\eta) = 1$ [a normalization.]

**End of Digression**

- Note $f(y_i | x_i, u_i) = \frac{e^{-e^{x_i'\beta+\alpha_i}}(e^{x_i'\beta+\alpha_i})^{y_i}}{y_i!} = \frac{e^{-e^{x_i'\beta}u_i}(e^{x_i'\beta}u_i)^{y_i}}{y_i!}$.

- Then,

$$f(y_i | x_i) = \int_0^\infty f(y_i | x_i, u_i) f_{gamma}(u_i) du_i$$
$$= \frac{\Gamma(\theta + y_i)}{\Gamma(y_i + 1)\Gamma(\theta)} r_i^\theta (1 - r_i)^{y_i},$$

where $r_i = \frac{\theta}{e^{x_i'\beta} + \theta}$, $\Gamma(s) = (s-1)\Gamma(s-1)$, and $\Gamma(s) = (s-1)!$ if s is an integer. Thus, when θ is a positive integer,

$$f(y_i | x_i) = \binom{(\theta + y_i) - 1}{\theta - 1} r_i^\theta (1 - r_i)^{(\theta + y_i) - \theta},$$

→ This is the form of the negative binomial distribution.

→ Compound Poisson = Negative binomial distribution!

- Since $(\theta+y_i)$ follows Neg-Bin,

$E(\theta+y_i) = \theta/(1-r_i) = e^{x_i'\beta} + \theta$ → $E(y_i) = e^{x_i'\beta}$.

$var(y_i) = var(\theta+y_i) = \theta(1-r_i)/r_i^2 = e^{x_i'\beta}\left(1 + \frac{1}{\theta}e^{x_i'\beta}\right)$.

- If we allow θ to vary over i and set $\theta_i = \frac{1}{\alpha}e^{x_i'\beta}$, we have

$E(y_i) = e^{x_i'\beta}$; $var(y_i) = (1+\alpha)e^{x_i'\beta}$.

→ This model is called Neg-Bin 1 model (HHG).

- If we set θ = 1/α,

$$E(y_i) = e^{x_i'\beta};\ \text{var}(y_i) = e^{x_i'\beta}\left(1+\alpha e^{x_i'\beta}\right) = e^{x_i'\beta} + \alpha\left(e^{x_i'\beta}\right)^2.$$

$\rightarrow$ This model is called Neg-Bin 2 model (HHG).

- Comment:

  Poisson, Neg-Bin 1 and 2 assume that $E(y_i) = \exp(x_i'\beta)$. As long as this mean specification is correct, MLE based on the distributions are consistent if the true distribution belongs to the linear exponential family [see Gourieroux, Monfort and Trognon (1984).]

**(4) Testing Poisson:**

- $H_o$: $E(y_i) = \text{var}(y_i) = e^{x_i'\beta}$ (Poisson);

  $H_a$: $E(y_i) = e^{x_i'\beta}$, but $\text{var}(y_i) = e^{x_i'\beta} + \alpha\left(e^{x_i'\beta}\right)^s$ and $\alpha \neq 0$.

  [If s = 1, $H_a$ = Neg-Bin 1. If s = 2, $H_a$ = Neg-Bin 2.]

- For given s, under $H_o$,

$$E\left(\frac{1}{\mu_i}\{(y_i-\mu_i)^2 - y_i\}\right) = \frac{1}{\mu_i}\left(E\left\{(y_i-\mu_i)^2 - y_i\right\}\right) = 1-1=0;$$

$$\text{var}\left(\frac{1}{\mu_i}\{(y_i-\mu_i)^2 - y_i\}\right) = 2,$$

where $\mu_i = e^{x_i'\beta}$.

- Then, under $H_o$, CLT implies:

$$T_L = \frac{\frac{1}{\sqrt{N}} \Sigma_{i=1}^{N} \mu_i^{s-1} \frac{1}{\mu_i} \left\{ \left( y_i - \mu_i \right)^2 - y_i \right\}}{\sqrt{\frac{1}{N} \Sigma_{i=1}^{N} \left( \mu_i \right)^{2(s-1)}} \sqrt{2}} \to \text{N}(0,1).$$

- Since $\mu_i = e^{x_i'\beta}$ is unobservable, we need to use $\hat{\mu}_i = e^{x_i'\hat{\beta}}$, where $\hat{\beta}$ is the Poisson ML estimator. But, we still can show that under $H_o$,

$$\hat{T}_L = \frac{\frac{1}{\sqrt{N}} \Sigma_{i=1}^{N} \hat{\mu}_i^{s-1} \frac{1}{\hat{\mu}_i} \left\{ \left( y_i - \hat{\mu}_i \right)^2 - y_i \right\}}{\sqrt{\frac{1}{N} \Sigma_{i=1}^{N} \left( \hat{\mu}_i \right)^{2(s-1)}} \sqrt{2}} \to \text{N}(0,1).$$

- $H_o$ may hold even if the $y_i$ do not follow Poisson. For such cases, use

$$\hat{T}_L' = \frac{\frac{1}{\sqrt{N}} \Sigma_{i=1}^{N} \hat{\mu}_i^{s-1} \frac{1}{\hat{\mu}_i} \left\{ \left( y_i - \hat{\mu}_i \right)^2 - y_i \right\}}{\sqrt{\frac{1}{N} \Sigma_{i=1}^{N} \left( \hat{\mu}_i \right)^{2(s-1)}} \sqrt{\frac{1}{N} \Sigma_{i=1}^{N} \left( \frac{1}{\hat{\mu}_i} \left\{ \left( y_i - \hat{\mu}_i \right)^2 - y_i \right\} \right)^2}}.$$

## (5) Poisson Model for Panel Data

- Assume $y_{it}$ i.i.d. with Poisson($\lambda_{it}$), where $\lambda_{it} = \exp(x_{it}'\beta+\alpha_i) = \mu_{it}\exp(\alpha_i)$.
- Fixed Effects Model
  - Treat $\alpha_i$ as parameters.
  - Surprisingly, MLE is consistent!
  - $f(y_{it} \mid x_{it}) = \dfrac{e^{-\lambda_{it}}(\lambda_{it})^{y_{it}}}{y_{it}!}$.
  - $\ln f(y_{it} \mid x_{it}) = -\lambda_{it} + y_{it}(x_{it}'\beta + \alpha_i) - \ln(y_{it}!)$.
  - $l_{NT}(\beta, \alpha_{1,} \alpha_2, ..., \alpha_N) = \Sigma_{t=1}^{T}\Sigma_{i=1}^{N}\left\{ y_{it}(x_{it}'\beta + \alpha_i) - e^{x_{it}'\beta+\alpha_i} - \ln(y_{it}!) \right\}$.
  - $\dfrac{\partial l_{NT}(\beta, \alpha_1, ..., \alpha_N)}{\partial \alpha_j} = \Sigma_{t=1}^{T}\left\{ y_{it} - e^{\alpha_i}\mu_{it} \right\} = 0.$

  $\rightarrow \alpha_j = \ln\left( \dfrac{\Sigma_{t=1}^{T} y_{it}}{\Sigma_{t=1}^{T}\mu_{it}} \right)$.

  - Substitute these solutions into $l_{NT}$:

    $$l_{NT}^{c}(\beta) = \Sigma_{i=1}^{N}\Sigma_{t=1}^{T} y_{it} \ln p_{it}(\beta),$$

    where $p_{it} = \dfrac{\exp(x_{it}'\beta)}{\Sigma_{t=1}^{T}\exp(x_{it}'\beta)}$.

- The MLE estimator of β based on $l_{NT}^{c}(\beta)$ is consistent even if the true distribution of $y_{it}$ is not Poisson as long long as $E(y_{it}|x_{it},\alpha_i) = \exp(x_{it}'\beta)\exp(\alpha_i)$ [Wooldridge (JEC, 1999)]. For correct covariance matrix of the ML estimator of β, use the robust form $[(H_{NT})^{-1}B_{NT}(H_{NT})^{-1}]$.

- Random Effects Model
  - Assume that the $e^{\alpha_i}$ follow a Gamma Distribution.
  - $$f(y_{i1},...y_{iT} | x_{i1},...,x_{iT,}u_i) = \prod_{t=1}^{T} f(y_{it} | x_{it},\alpha_i) = \prod_{t=1}^{T} \frac{e^{-\mu_{it}u_i}(\mu_{it}u_i)^{y_{it}}}{y_{it}!}.$$
  - $$f(y_{i1},...,y_{iT} | x_{i1},...x_{iT}) = \int_0^\infty f(y_{i1},...,y_{iT} | x_{i1},...,x_{iT},u_i) f_{gamma}(u_i)du_i = \frac{\left(\prod_{t=1}^{T}\mu_{it}^{y_{it}}\right)\Gamma\left(\theta+\Sigma_{t=1}^{T}y_{it}\right)}{\left(\Gamma(\theta)\prod_{t=1}^{T}y_{it}!\right)\left(\left(\Sigma_{t=1}^{T}\mu_{it}\right)^{\Sigma_{t=1}^{T}y_{it}}\right)} Q_i^{\theta}(1-Q_i)^{\Sigma_{t=1}^{T}y_{it}}$$

    where $Q_i = \dfrac{\theta}{\theta+\Sigma_{t=1}^{T}\mu_{it}}$.
  - $l_{NT}(\beta,\theta) = \Sigma_{i=1}^{N} f(y_{i1},...,y_{iT} | x_{i1},...,x_{iT})$.
  - Can use the Hausman test to determine whether RE or FE is correct.

## [10] Duration Model

- Reference:
  - Chung, Schmidt and Witte (1991), Journal of Quantitative Criminology.
  - Kiefer (1988), Journal of Economic Literature.
  - LIMDEP Manual.

### (1) Basic Model

- T = length of time until an event (E) occurs.
- Examples:
  - Recidivism in criminology:

    E = returning to prison.

    T = interval from release ro returning.
  - Unemployment Duration

    E = quit job search (get a job or give up)

    T = time length for a job search.
  - Time interval between equity trades or exchange rate quotes.

    [Engle and Russell, ECON, 1998]

- Assume that T is continuous:

  $T \sim$ pdf, $f(t|\theta)$; cdf $F(t|\theta)$.

- Some important concepts:

1) $F(t|\theta) = Pr(T < t)$.
   - Probability of failure within t.
   - Probability that E occurs within t.

     ex: probability of returning to prison within t;

     ex: Probability of quitting job search within t.

2) Probability of survival: $S(t|\theta) = 1 - F(t|\theta) = Pr(T \geq t)$.
   - Probability that E does not occurs with t.

     ex: probability that an ex-prisoner does not in prison until t.

     ex: probability that job search still goes on until t.
   - In general, $S(t|\theta) \to 0$ as $t \to \infty$.

3) Hazard function:
   - $h(t|\theta) = f(t|\theta)/Pr(T \geq t) = f(t|\theta)/S(t|\theta) = f(t|\theta)/[1-F(t|\theta)]$.
   - Probability that E occurs just after time t conditional on no failure prior to t.

     ex: probability of returning prison right after t.

     ex: probability of quitting search right after t.

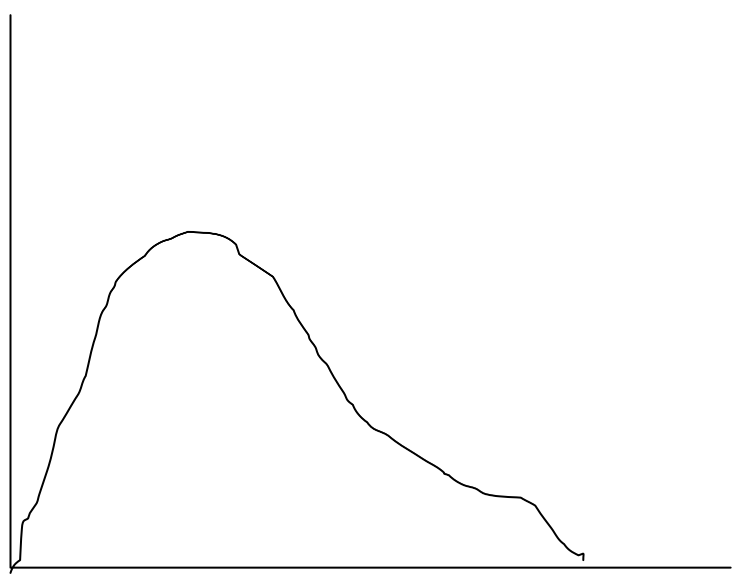
4) Integrated hazard function.

- $H(t|\theta) = \int_0^t h(x\,|\,\theta)dx$.
- $S(t|\theta) = \exp[-H(t|\theta)]$ or $H(t|\theta) = -\ln S(t\,|\,\theta)$.
- $h(t|\theta) = -\dfrac{d[\ln S(t\,|\,\theta)]}{dt}$.
- If you know S(t|θ), we can get h(t|θ), and vice versa.

5) State Dependence:

- $\dfrac{dh(t\,|\,\theta)}{dt} > 0$: positive state dependence.
- $\dfrac{dh(t\,|\,\theta)}{dt} < 0$: negative state dependence.

Example: Recidivism



## (2) Maximum Likelihood Estimation

- $t_i^*$ = the i'th individual's actual exit time
- $c_i$ = follow-up time for i (time interval following i).
- $t_i = \begin{cases} t_i^* \ if \ t_i^* < c_i; \\ c_i \ if \ t_i^* \ge c_i. \end{cases} = \min(t_i^*, c_i).$
- $d_i = \begin{cases} 1 \ if \ t_i < c_i; \\ 0 \ if \ t_i = c_i. \end{cases}$
- $$\begin{aligned} l_N(\theta) &= \sum_{t_i < c_i} \ln f(t_i \mid \theta) + \sum_{t_i = c_i} \Pr(t_i^* \ge c_i) \\ &= \sum_{t_i < c_i} \ln f(t_i \mid \theta) + \sum_{t_i = c_i} \ln[1 - F(c_i \mid \theta)] \\ &= \sum_{t_i < c_i} \ln f(t_i \mid \theta) + \sum_{t_i = c_i} \ln[S(c_i \mid \theta)] \\ &= \Sigma_{i=1}^{N} \left\{ d_i \ln f(t_i \mid \theta) + (1 - d_i) \ln S(c_i \mid \theta) \right\} \end{aligned}$$

## (3) Distributions

1) Exponential:

- For simplicity, suppress "i".
- $f(t|\theta) = \lambda\exp(-\lambda t)$, where $\theta = \lambda > 0$.
- $S(t|\theta) = \exp(-\lambda t)$
- $h(t|\theta) = \lambda$ (no state dependence).
- $E(t) = 1/\lambda$ [expected duration]; $var(t) = 1/\lambda$.

2) Weibull:

- $t^p \sim$ Exponential($\lambda$).
- $f(t|\theta) = p\lambda^p t^{p-1}\exp[-(\lambda t)^p]$, where $\theta = (\lambda, p)'$.
- $S(t|\theta) = \exp[-(\lambda t)^p]$.
- $h(t|\theta) = \lambda p(\lambda t)^{p-1}$.

$$\rightarrow \frac{dh(t|\theta)}{dt} = (p-1)\lambda^p p t^{p-2}\begin{cases} > 0 \; if \;\; p > 1; \\ = 0 \; if \;\; p = 1; \\ < 0 \; if \;\; p < 1. \end{cases}$$

- $E(t) = \frac{1}{\lambda}\Gamma\left(\frac{1}{p}+1\right)$, where $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}e^{-y}dy$.

<Poof>

Let $y = (\lambda t)^p$; then, $t = \frac{1}{\lambda} y^{\frac{1}{p}}$ and $dt = \frac{1}{p\lambda} y^{\frac{1}{p}-1} dy$. Hence,

$$E(t) = \int_0^\infty tf(t,\theta)dt = \int_0^\infty \frac{1}{\lambda} y^{\left(\frac{1}{p}+1\right)-1} \exp(-y)dy$$

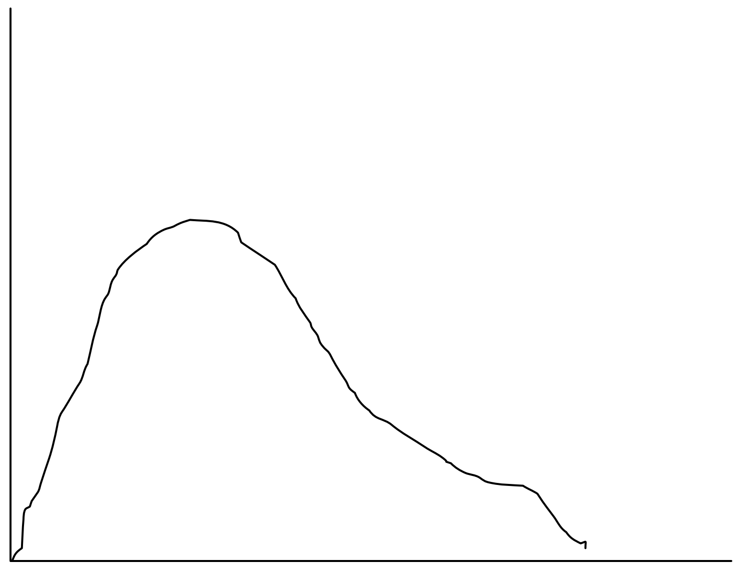
$$= \frac{1}{\lambda}\Gamma\left(\frac{1}{p}+1\right),$$

where $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}\exp(-y)dy$.

3) Log-normal:

- $\ln(t) \sim N(\mu,\sigma^2)$, where $\mu = -\ln(\lambda)$ and $\sigma = 1/p$.
- $f(t \mid \theta) = \frac{p}{\sqrt{2\pi t}} \exp\left(-\frac{p^2}{2}(\ln t + \ln \lambda)\right)$.
- $S(t|\theta) = 1 - \Phi\left(\frac{\ln t + \ln \lambda}{\sigma}\right)$.
- $E(t) = \frac{1}{\lambda} + \exp\left(\frac{1}{2p^2}\right)$; $E(\ln t) = -\ln \lambda$.

- $h(t|\theta)$ = complicated, but



4) Log-logistic

- $f(t|\theta) = \dfrac{\lambda p(\lambda t)^{p-1}}{\left(1+(\lambda t)^{2p}\right)^2}$, $\theta = (\lambda, p)'$.

  $\rightarrow$ f(w) = $\dfrac{\exp(-w)}{[1+\exp(-w)]^2}$, where $w = -p\ln(\lambda t)$.

- $S(t|\theta) = \dfrac{1}{1+(\lambda t)^p}$; $h(t|\theta) = \dfrac{\lambda p(\lambda t)^{p-1}}{1+(\lambda t)^p}$.

- $\dfrac{dh(t|\theta)}{dt} = \lambda^p p t^{p-2} \dfrac{(p-1)-(\lambda t)^p}{\left(1+(\lambda t)^p\right)^2}$.

  $\rightarrow$ Negative if p < 1; negative if t is large; positive if p > 1 and t is small.

5) Gamma:

- $f(t|\theta) = \dfrac{\lambda p(\lambda t)^{p\xi-1}}{\Gamma(\xi)} \exp\left(-(\lambda t)^p\right)$, where $\theta = (\lambda, p, \xi)'$.

- Weibull if $\xi = 1$ and exponential if $\xi = p = 1$.

## (4) Estimation

- Set $\frac{1}{\lambda_i} = \exp(x_i'\beta)$ and $\sigma = \frac{1}{p}$ so that $\beta_j > 0$ means "$x_{ji}$ prolongs duration."
- LIMDEP estimates β and σ.
- Caution!
    - We here assume that $x_i$ does not change during the follow-up period, $c_i$.
    - This may cause dome misspecification problem. For the cases of time-varying $x_i$, see Heckman and Singer (Social Science Duration Analysis, 1985, Ch. 2).

## (5) Checking Specification

- Consequences of distributional misspecification:
  - $\hat{\beta}_{ML}$ would be inconsistent.
  - Even if the ML estimator may be consistent for some special cases, it could be inconsistent. When it is consistent, use $(H_N)^{-1}B_N(H_N)^{-1}$ to estimate $Cov(\hat{\beta}_{ML})$. [See Gourieroux, Monfort and Trognon (ECON, 1984).]

- Vuong Test [Vuong (ECON, 1989)]
  - Wish to decide on which of two competing models would be more plausible.
  - Here, the goal is not to find the correct model, but to find a better model between $f(t_i|x_i,\theta)$ and $g(t_i|z_i,\gamma)$. Both models could be misspecified.

[CASE 1] Nonnested Models

- EX: Weibull Vs. Log-normal.
- Let $\theta_*$ and $\gamma_*$ be the maximizers of $E[f(t\,|\,x,\theta)]$ and $E[g(t\,|\,z,\gamma)]$, respectively.
- $\omega_*^2 = \text{var}\left( \log \dfrac{f(y\,|\,x,\theta_*)}{g(y\,|\,z,\gamma_*)} \right)$.
- $H_o : E\left[ \log \dfrac{f(y\,|\,x,\theta_*)}{g(y\,|\,z,\gamma_*)} \right] = 0$;

  $H_f : E\left[ \log \dfrac{f(y\,|\,x,\theta_*)}{g(y\,|\,z,\gamma_*)} \right] > 0$; $H_g : E\left[ \log \dfrac{f(y\,|\,x,\theta_*)}{g(y\,|\,z,\gamma_*)} \right] < 0$.
- Let $\hat{\theta}_N$ and $\hat{\gamma}_N$ be the ML estimators based on $f(t_i|x_i,\theta)$ and $g(t_i|z_i,\gamma)$, respectively.

- Define:

$$\hat{l}_i = \log \frac{f(t_i\,|\,x_i,\hat{\theta}_N)}{g(t_i\,|\,z_i,\hat{\gamma}_N)};\ \hat{l}_N = \Sigma_{i=1}^N \hat{l}_i;$$

$$\hat{\omega}_N^2 = \frac{1}{N}\Sigma_{i=1}^N \left[\hat{l}_i\right]^2 - \left[\frac{1}{N}\hat{l}_N\right]^2;\ \hat{\omega}_N = \sqrt{\hat{\omega}_N^2};$$

$$t_N = \sqrt{N}\,\frac{\hat{l}_N}{\hat{\omega}_N}.$$

- Under $H_o$, $t_N \to N(0,1)$; Under $H_f$, $t_N \to +\infty$; Under $H_g$, $t_N \to -\infty$.

[CASE 2] Nested Models

- Exponential vs. Weibull with same regressors $x_i = z_i$.
- Assume $g(y_i | z_i, \gamma_*) \subset f(y_i | x_i, \theta_*)$.
- $H_o : g(y_i | z_i, \gamma_*) = f(y_i | x_i, \theta_*)$;

  $H_a : g(y_i | z_i, \gamma_*) \neq f(y_i | x_i, \theta_*)$
- The usual LR statistic, $2\hat{l}_N$, is not $\chi^2$ under $H_o$ if both models are misspecificed [in fact, a weighted $\chi^2$].
- If the general model [$f(y_i|x_i,\theta_*)$] is correctly specified, then ,the LR statistic is $\chi^2$ under $H_o$.

- Rule of Thumb Methods:

  1) Compare the ML results to non-parametric estimates of h(t) or s(t) ["Kaplan-Meier" estimates].
  2) Consider whether estimated parameters are reasonably signed. [Compare to Cox' proportional hazard model (a semi-parametric model).]

## (6) Heteroskedasticity

- Suppose that $\theta = \begin{pmatrix} \beta \\ \sigma \end{pmatrix}$ differs across different i ($\theta_i$). If you estimate a model incorrectly assuming θ constant, the estimated hazard functions tend to be biased to negative duration dependence [Heckman and Singer, 1985, Ch. 2].

- Heckman and Singer's suggestion (ECON, 1984)
  - Assume that the $\theta_i$ are random variables with pdf $g(\theta_i)$.
  - $f(t_i, \theta_i \mid x_i) = f(t_i \mid x_i, \theta_i) g(\theta_i)$.
  - $f(t_i \mid x_i) = \int f(t_i \mid x_i, \theta_i) d\theta_i$.
  - Estimate $g(\theta_i)$ non-parametrically. Then, using the estimated $g(\theta_i)$, estimate $f(t_i|x_i)$.
  - How about specifying $g(\theta_i)$?
    - → Results are too sensitive to specification of $g(\theta_i)$.
    - → H-S develop a nonparametric method than can estimate $g(\theta_i)$.

## (7) Proportional Hazards

- Cox (Biometrika, 1975).
- A partial solution to the problem of distributional misspecifications.
- $h(t|x_i,\beta) = h_o(t)\lambda_i$,

  where $\lambda_i = \exp(-x_i'\beta)$ and $h_o(t)$ is the common trend (baseline hazard) among i.

- Partial likelihood:
  - Suppose that n individuals (out of N) exit within the follow-up periods.
  - Order them by $0 = t_0 < t_1 < t_2 < ... < t_n$:
    - Don't use censored people.
    - Assume no tie (for convenience only. Can allow ties).
  - $R(t_j)$ = {i| not yet exit just prior ro $t_j$} where i is a person "at risk" at time $t_j$

    = {j, j+1, ..., n}.

**Short Digression:**

- Question:
  - Bob and Steve went fishing.
  - A fish on the pond.
  - Bob uses very good baits, so that the probability ($P_B$) of Bob's catching the fish (if he fishes alone) = 1.
  - Steve uses bad baits, so that the probability ($P_S$) of Steve's catching fish (if he fishes alone) = ¼.
  - What is the conditional probability that Steve caught the fish given that one of Bob and Steve caught that fish? [One of Steve and Bob caught a fish. What is the probability that it is Steve?]
- Answer:

$$\frac{P_S}{P_S + P_B} = \frac{1/4}{1/4+1} = \frac{1}{5}.$$

**End of Digression**

- Probability that "j" exits after time $t_j$ given that one person in $R(t_j)$ fails just after $t_j$

$$= \frac{h(t_j \mid x_j, \beta)}{\sum_{i \in R(t_j)} h(x_i \mid x_i, \beta)} = \frac{h_o(t_j)\exp(-x_j{}'\beta)}{\sum_{i \in R(t_j)} h_o(t_j)\exp(-x_i{}'\beta)} = \frac{\exp(-x_j{}'\beta)}{\sum_{i \in R(t_j)} \exp(-x_i{}'\beta)}.$$

- If there are $m_j$ ties at time $t_j$, let $x_{j,h}$ be the vector of regressors for the h'th person (h = 1, ... , $m_j$) in the group with $t_j$. Then use:

$$\prod_{h=1}^{m_j} \frac{\exp(-x_{j,h}'\beta)}{\sum_{i\in R(t_j)} \exp(-x_i'\beta)}.$$

  All of the $m_j$ persons should be in $R(t_j)$.

- If an individual's spell is censored between $t_j$ and $t_{j+1}$, the person will be included in the denominators of $R(t_i)$ for i = 1, ... , j, but not for i = j + 1, .... The person does not influence the numerators.

- Partial ML estimation:

  - $l_N(\beta) = \Sigma_{j=1}^{n} \log\left(\frac{\exp(-x_j'\beta)}{\sum_{i\in R(t_j)} \exp(-x_i'\beta)}\right)$.

  - Let $m_j$ = # of exits at time $t_j$. Then,

$$\hat{h}_o(t_j) = \frac{m_j}{\sum_{i\in R(t_j)} \exp(-x_i'\beta)}.$$

  - $\hat{H}_o(t_j) = \Sigma_{i=1}^{j} \hat{h}_o(t_i)$.

  - $\hat{S}(t_j \mid x_i) = \exp[-\hat{H}_o(t_j)\exp(-x_i'\beta)]$.

- It can be shown (Kiefer, 1988, JEL) that

  $$v_i = x_j{}'\beta - \ln[-\ln s(t_j \mid x_j)] \sim \exp(-v_i)\exp\left(-\exp(v_i)\right).$$

  [pdf of the extreme-value distribution]

- So, the model specification of the proportional hazard can be tested by investigating the distribution of the $\hat{v}_i$.

## (8) Non-Parametric Approach

- Let $c_{max}$ = maximum follow-up periods.
- Divide this interval into k equal subintervals:

  $0 = t_0 < t_1 < ... < t_{j-1} < t_j < t_k = c_{max}$.

- Define:
  - $r_j$ = # of people in ($t_{j-1}$ $t_j$]

    = # of people "at risk"

    = # of all people - # of people who exited in (0,$t_{j-1}$]

    \- (# of people censored in ($t_{j-1}$,$t_j$])/2
  - $n_j$ = # of people who exit in ($t_{j-1}$,$t_j$].
  - $\hat{h}(t_j) = \frac{n_j}{r_j}$ [Kaplan-Meier estimator].
  - $\hat{H}(t_j) = \Sigma_{a=1}^{j} \hat{h}(t_a)$.
  - $\hat{S}(t_j) = \exp\left(-\hat{H}(t_j)\right)$.

- Comments
  - No regressors.
  - Can determine what h(t) looks like.

## (9) Split Population Model

- Schmidt and Witte (1989, JEC)
- So far, we have assumed that S(t|θ) = 1-F(t|θ) → 0 as t →∞. [This means that an ex-prisoner will return to prison surely some time in the future.] But, it might be the case that s(t|θ) = 1 for some individuals. [Some ex-prisoners will never return to prison.]
- Assumptions:
  - $y_i^* = z_i'\gamma + v_i$, where the $v_i$ are N(0,1) (or logistic)

    → $\begin{cases} y_i = 1 \text{ iff } y_i^* > 0 \Rightarrow \textit{no return}; \\ y_i = 0 \text{ iff } y_i^* < 0 \Rightarrow \textit{return some time} \end{cases}$

    → Pr(i never returns) = $\Phi(z_i'\gamma)$;

    Pr(i returns some time) = $1 - \Phi(z_i'\gamma)$.
  - For the people with $y_i^* < 0$, $t_i^*$ is the length of actual exit time. [$t_i^*$ is defined only for the people with $y_i^* > 0$.]

    → Let $t_i$ be the observed length of exit time. Then,

    $$t_i = \begin{cases} t_i^* \text{ if } t_i^* < c_i; \\ c_i \text{ if } t_i^* \geq c_i \end{cases}$$
  - Let $g(t_i^* \mid y_i = 0)$ be the pdf of $t_i^*$ conditional on $y_i$ = 0. Assume:

    $g(t_i^* \mid y_i = 0) = f(t_i^* \mid x_i, \theta)$, where f is exponential, Weibull, etc.

- $$\begin{aligned}\Pr(t_i < c_i) &= \Pr(y_i = 0, t_i^* < 0) = \Pr(t_i^* < c_i \mid y_i = 0)\Pr(y_i = 0) \\ &= F(c_i \mid x_i, \theta)[1 - \Phi(z_i'\gamma)] = [1 - S(c_i \mid x_i, \theta)][1 - \Phi(z_i'\gamma)] \\ &= 1 - \Phi(z_i'\gamma) - S(c_i \mid x_i, \theta)[1 - \Phi(z_i'\gamma)].\end{aligned}$$

- $$\begin{aligned}\Pr(t_i = c_i) &= 1 - F(c_i \mid x_i, \theta)[1 - \Phi(z_i'\gamma)] \\ &= \Phi(z_i'\gamma) + S(c_i \mid x_i, \theta)[1 - \Phi(z_i'\gamma)]\end{aligned}.$$

- $$\begin{aligned}g(t_i^* \mid t_i < c_i) &= g(t_i^* \mid y_i = 0, t_i^* < c_i) = f(t_i^* \mid t_i^* < c_i) \\ &= \frac{f(t_i^*)}{\Pr(t_i^* < c_i)} = \frac{f(t_i^* \mid x_i, \theta)}{F(t_i^* \mid x_i, \theta)}\end{aligned}.$$

- Log-likelihood function:

$$\begin{aligned}l_N(\theta, \gamma) &= \sum_{t_i < c_i} \ln\{g(t_i \mid t_i < c_i)\Pr(t_i < c_i)\} + \sum_{t_i = c_i} \ln\{\Pr(t_i = c_i)\} \\ &= \sum_{t_i < c_i} \ln\left\{f(t_i \mid x_i, \theta)[1 - \Phi(z_i'\gamma)]\right\} \\ &\quad + \sum_{t_i = c_i} \ln\left\{\Phi(z_i'\gamma) + S(c_i \mid x_i, \theta)[1 - \Phi(z_i'\gamma)]\right\}\end{aligned}$$

- Comment:
  - If the Kaplan-Meier estimator shows positive duration dependence, it is very hard to get split-population MLE results.
  - If it shows rapid negative duration dependence, it is worth trying split MLE.

- Testing split-population model:
  - MLE without split,

    $$l_N^R(\theta,\gamma) = \sum_{t_i < c_i} \ln\{ f(t_i \mid x_i, \theta)\} + \sum_{t_i = c_i} \ln\{S(c_i \mid x_i, \theta)]\},$$

    which is the split log-likelihood function with the restriction

    $\Phi(z_i{}'\gamma) = 0.$

  - $LR_N = 2[l_N(\hat{\theta},\hat{\gamma}) - l_N^R(\tilde{\theta})] \to \chi^2(df = \dim(\gamma))$ under $H_o$: no split.

    → Really?

## (10) Autoregressive Conditional Duration (ACD) Model

- Trading times: $t_0 < t_1 < t_2 < .... < t_N$.

- Duration of an interval (time between two trades): $x_i = t_i - t_{i-1}$.

- EACD(2,2) [Exponential ACD of orders 1 and 1]:
  - Assume $x_i$ ~ exponential($\lambda_i$), where $\psi_i = 1/\lambda_i$.
  - $E(x_i \mid \Omega_{t_{i-1}}) = \psi_i$ .
  - Assume $\psi_i = \omega + \alpha_1 x_{i-1} + \alpha_2 x_{i-2} + \beta_1 \psi_{i-1} + \beta_2 \psi_{i-2}$.

- WACD(2,2) [Weibull ACD of orders 1 and 1]:
  - Assume $(x_i)^p$ ~ exponential($\lambda_i$), where $\psi_i = 1/\lambda_i$.
  - Assume $\psi_i = \omega + \alpha_1 x_{i-1} + \alpha_2 x_{i-2} + \beta_1 \psi_{i-1} + \beta_2 \psi_{i-2}$.