1. Endogenous Probit

1.0.1. The model

$$y_{1i} = I(v_i + x_i^T \beta + e_i > 0) \tag{1.1}$$

$$x_i^T = (z_{1i}^T, \widehat{y}_{2i}^T) \tag{1.2}$$

$$y_{2i} = z_i^T \Gamma + u_i \tag{1.3}$$

where $\widehat{y}_{2i}^T$ are the fitted values from the instrument equation, $\widehat{y}_{2i}^T = z_i^T \widehat{\Gamma}$.

1.0.2. Variance-covariance matrix

$$\Sigma_{eu} = \begin{bmatrix} b_3^2 & b_3 b_1 b_2 \\ b_3 b_1 b_2 & b_1^2 \end{bmatrix} \tag{1.4}$$

1.0.3. The endogenous probit log-likelihood function

$\frac{\ln L}{n} = \frac{1}{n} \sum_i y_{1i} \ln \int_{-(v_i + x_i^T \beta)}^{\infty} f(e, y_{2i} - z_{2i}^T \Gamma) de + (1 - y_{1i}) \ln \int_{-\infty}^{-(v_i + x_i^T \beta)} f(e, y_{2i} - z_{2i}^T \Gamma) de$

$$f(e, y_{2i} - z_{2i}^T \Gamma) = f(e, u) = f(u) * f(e|u) = f(u) * f\left(\frac{[\frac{e}{b_3} - \frac{b_2}{b_1} u]}{\sqrt{1 - b_2^2}}\right) \tag{1.5}$$

$$\int_{-(v_i + x_i^T \beta)}^{\infty} f(e, y_{2i} - z_{2i}^T \Gamma) = f(u) * [1 - \Phi[-\frac{[\frac{e}{b_3} + \frac{b_2}{b_1} u]}{\sqrt{1 - b_2^2}}]] \tag{1.6}$$

$$\int_{-\infty}^{-(v_i + x_i^T \beta)} f(e, y_{2i} - z_{2i}^T \Gamma) = f(u) * \Phi[-\frac{[\frac{v_i + x_i^T \beta}{b_3} + \frac{b_2}{b_1} u]}{\sqrt{1 - b_2^2}}] \tag{1.7}$$

$f(u)$ is density of u. In the program $v_i + x_i^T \beta$ is lb. After substitutions the endogenous log-likelihood function becomes:

$$\frac{\ln L}{n} = \frac{1}{n} \sum_i y_{1i} \ln \left\{ f(u) * \Phi[\frac{[\frac{lb}{b_3} + \frac{b_2}{b_1} u]}{\sqrt{1 - b_2^2}}] \right\} + (1 - y_{1i}) \ln \left\{ f(u) * \Phi[-\frac{[\frac{lb}{b_3} + \frac{b_2}{b_1} u]}{\sqrt{1 - b_2^2}}] \right\} \tag{1.8}$$

To ensure positive definiteness, Σ_{eu} is coded as:

$$\Sigma_{eu} = \begin{bmatrix} a_1^2 + a_2^2 & a_1 \widehat{\sigma}_u \\ a_1 \widehat{\sigma}_u & \widehat{\sigma}_u^2 \end{bmatrix} \tag{1.9}$$

$b_1 = \widehat{\sigma}_u = \sum_{i=1}^{N} \frac{[y_{2i} - z_i^T \widehat{\Gamma}]^2}{N}$, $b_3 = \sqrt{a_1^2 + a_2^2}$, $b_2 = \frac{a_1}{b_3}$, and (1.8) simplifies to:

$$\frac{\ln L}{n} = \frac{1}{n} \sum_i y_{1i} \ln \left\{ f(u) * \Phi[\frac{lb + \frac{a_1}{\sigma_u} u}{a_2}] \right\} + (1 - y_{1i}) \ln \left\{ f(u) * \Phi[-\frac{lb + \frac{a_1}{\sigma_u} u}{a_2}] \right\} \tag{1.10}$$

The MAXLIK routine in GAUSS returns estimates and standard errors of $a_1..a_5$, not $b_1..b_5$.

2. Endogenous Sample Selection

2.0.4. The model

$$d_i = I(v_i + x_i^T \delta + e_i) > 0) \tag{2.1}$$

$$y_{1i} = (x_i^T \beta + \varepsilon_i) d_i \tag{2.2}$$

$$x_i^T = (z_{1i}^T, \widehat{y}_{2i}^T) \tag{2.3}$$

$$y_{2i} = z_i^T \Gamma + u_i \tag{2.4}$$

where $\widehat{y}_{2i}^T$ are the fitted values from the instrument equation, $\widehat{y}_{2i}^T = z_i^T \widehat{\Gamma}$.

2.0.5. Variance-covariance matrices

$$\Sigma_{e\varepsilon u} = \begin{bmatrix} b_6^2 & b_6 b_4 b_5 & b_6 b_1 b_3 \\ b_6 b_4 b_5 & b_4^2 & b_4 b_1 b_2 \\ b_6 b_1 b_3 & b_4 b_1 b_2 & b_1^2 \end{bmatrix}, \Sigma_{eu} = \begin{bmatrix} b_6^2 & b_6 b_1 b_3 \\ b_6 b_1 b_3 & b_1^2 \end{bmatrix} \tag{2.5}$$

2.0.6. Endogenous Sample Selection Log-Likelihood Function

$$\begin{aligned} \frac{\ln L}{n} &= \frac{1}{n}\sum_i d_i \ln \int_{-(v_i+x_i^T\delta)}^{\infty} f(e, y_{2i} - z_{2i}^T\Gamma, y_{1i} - x_{1i}^T\beta) de + \\ & (1-d_i)\ln \int_{-\infty}^{-(v_i+x_i^T\delta)} \int_{-\infty}^{\infty} f(e, y_{2i} - z_{2i}^T\Gamma, \varepsilon) d\varepsilon de \end{aligned} \tag{2.6}$$

$$\int_{-\infty}^{\infty} f(e, y_{2i} - z_{2i}^T\Gamma, \varepsilon) d\varepsilon = f(e, y_{2i} - z_{2i}^T\Gamma) = f(e, u) \tag{2.7}$$

$$\int_{-\infty}^{-(v_i+x_i^T\delta)} f(e,u) = f(u) * \Phi[-\frac{[v_i + x_i^T\delta + \frac{b_3 b_6}{b_1}u]}{b_6\sqrt{1-b_3^2}}] \tag{2.8}$$

$$f(e,\varepsilon,u) = f(\varepsilon,u) * f(e|\varepsilon,u) \tag{2.9}$$

$$e|\varepsilon, u\tilde{\ } N[\Sigma_{12}\Sigma_{22}^{-1}[\varepsilon,u]', \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}] \tag{2.10}$$

$$\Sigma_{12}\Sigma_{22}^{-1}[\varepsilon,u]' = \frac{(b_5 - b_2 b_3)b_6\varepsilon}{(1-b_2^2)b_4} + \frac{(b_3 - b_2 b_5)b_6 u}{(1-b_2^2)b_1} \tag{2.11}$$

$$\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \frac{[1 - b_2^2 - b_3^2 - b_5^2 + 2b_2b_3b_5]b_6^2}{1-b_2^2} \tag{2.12}$$

$$f(\varepsilon,u) = \frac{e^{-\frac{b_1^2\varepsilon^2 - 2b_1b_2b_4\varepsilon u + u^2 b_4^2}{2b_1^2b_4^2[1-b_2^2]}}}{2\pi b_1 b_4\sqrt{1-b_2^2}} \tag{2.13}$$

In the program $v_i + x_i^T\delta$ is lb. The endogenous log-likelihood function becomes:

$$\begin{aligned} \frac{\ln L}{n} &= \frac{1}{n}\sum_i d_i \ln\left\{ f(\varepsilon,u) * \Phi\left[\frac{lb + \frac{(b_5-b_2b_3)b_6\varepsilon}{(1-b_2^2)b_4} + \frac{(b_3-b_2b_5)b_6u}{(1-b_2^2)b_1}}{b_6\sqrt{\frac{[1-b_2^2-b_3^2-b_5^2+2b_2b_3b_5]}{1-b_2^2}}}\right]\right\} + \\ & (1-d_i)\ln\left\{ f(u) * \Phi[-\frac{[lb + \frac{b_3b_6}{b_1}u]}{b_6\sqrt{1-b_3^2}}]\right\} \end{aligned} \tag{2.14}$$

To ensure positive definiteness, $\Sigma_{\varepsilon\varepsilon u}$ is parameterized as PP':

$$P = \begin{bmatrix} \frac{1}{a_5} & a_4 & a_2 \\ 0 & \frac{1}{a_3} & a_1 \\ 0 & 0 & \widehat{\sigma}_u \end{bmatrix}, \Sigma_{\varepsilon\varepsilon u} = \begin{bmatrix} \frac{1}{a_5^2} + a_2^2 + a_4^2 & a_1 a_2 + \frac{a_4}{a_3} & a_2\widehat{\sigma}_u \\ a_1 a_2 + \frac{a_4}{a_3} & a_1^2 + \frac{1}{a_3^2} & a_1\widehat{\sigma}_u \\ a_2\widehat{\sigma}_u & a_1\widehat{\sigma}_u & \widehat{\sigma}_u^2 \end{bmatrix} \tag{2.15}$$

$$\begin{aligned} b_1 &= \widehat{\sigma}_u,\ b_4 = \sqrt{a_1^2 + \frac{1}{a_3^2}},\ b_6 = \sqrt{\frac{1}{a_5^2} + a_2^2 + a_4^2} \\ b_2 &= \frac{a_1}{b_4},\ b_3 = \frac{a_2}{b_6}, b_5 = \frac{a_1 a_2 + \frac{a_4}{a_3}}{b_4 b_6} \end{aligned} \tag{2.16}$$

$$\frac{(b_5 - b_2 b_3) b_6 \varepsilon}{(1 - b_2^2) b_4} + \frac{(b_3 - b_2 b_5) b_6 u}{(1 - b_2^2) b_1} = \left[\frac{a_2}{\widehat{\sigma}_u} - \frac{a_3 a_4 a_1}{\widehat{\sigma}_u}\right] u + a_3 a_4 \varepsilon \tag{2.17}$$

$$b_6 \sqrt{\frac{[1 - b_2^2 - b_3^2 - b_5^2 + 2 b_2 b_3 b_5]}{1 - b_2^2}} = \frac{1}{a_5} \tag{2.18}$$

$$\frac{[lb + \frac{b_3 b_6}{b_1} u]}{b_6 \sqrt{1 - b_3^2}} = \frac{lb + \frac{a_2 u}{\widehat{\sigma}_u}}{\sqrt{a_4^2 + \frac{1}{a_5^2}}} \tag{2.19}$$

$$b_1 b_4 \sqrt{1 - b_2^2} = \frac{\widehat{\sigma}_u}{a_3} \tag{2.20}$$

$$\frac{b_1^2 \varepsilon^2 - 2 b_1 b_2 b_4 \varepsilon u + u^2 b_4^2}{b_1^2 b_4^2 (1 - b_2^2)} = \frac{u^2}{\widehat{\sigma}_u^2} + a_3^2 \left[\varepsilon - a_1 \frac{u}{\widehat{\sigma}_u}\right]^2 \tag{2.21}$$

The coded log-likelihood function becomes:

$$\begin{aligned} \frac{\ln L}{n} &= \frac{1}{n} \sum_i d_i \ln \left\{ f(\varepsilon, u) * \Phi \left[\left(lb + \left[\frac{a_2}{\widehat{\sigma}_u} - \frac{a_3 a_4 a_1}{\widehat{\sigma}_u} \right] u + a_3 a_4 \varepsilon \right) a_5 \right] \right\} + \\ & (1 - d_i) \ln \left\{ f(u) * \Phi[- \frac{lb + \frac{a_2 u}{\widehat{\sigma}_u}}{\sqrt{a_4^2 + \frac{1}{a_5^2}}} \right\} \end{aligned} \tag{2.22}$$

$$f(\varepsilon, u) = \frac{a_3 e^{-\frac{u^2 + a_3^2 \left[\varepsilon \widehat{\sigma}_u - a_1 u\right]^2}{2 \widehat{\sigma}_u^2}}}{2 \pi \widehat{\sigma}_u} \tag{2.23}$$

The MAXLIK routine in GAUSS returns estimates and standard errors of $a_1..a_5$, not $b_1..b_5$.

3. Exogenous Sample Selection (Heckman Selection)

3.0.7. The model

$$d_i = I(v_i + x_i^T \delta + e_i) > 0) \tag{3.1}$$

$$y_{1i} = (x_i^T \beta + \varepsilon_i) d_i \tag{3.2}$$

3.1. Variance-covariance matrix

$$\Sigma_{e\varepsilon} = \begin{bmatrix} b_6^2 & b_6 b_4 b_5 \\ b_6 b_4 b_5 & b_4^2 \end{bmatrix} = \begin{bmatrix} \frac{1}{a_3^2} + a_2^2 & \frac{a_2}{a_1} \\ \frac{a_2}{a_1} & \frac{1}{a_1^2} \end{bmatrix} \tag{3.3}$$

$$b_4 = \frac{1}{a_1},\ b_6 = \sqrt{\frac{1}{a_3^2} + a_2^2}, b_5 = \frac{a_2}{b_6} \tag{3.4}$$

3.2. Exogenous Sample Selection Log-likelihood Function

$$\begin{aligned} \frac{\ln L}{n} &= \frac{1}{n} \sum_i d_i \ln \int_{-(v_i + x_i^T \delta)}^{\infty} f(e, y_{1i} - x_{1i}^T \beta) de + (1 - d_i) \ln \int_{-\infty}^{-(v_i + x_i^T \delta)} f(e) de \\ \frac{\ln L}{n} &= \frac{1}{n} \sum_i d_i \ln \left\{ f(\varepsilon) * \Phi \left[\frac{lb + \frac{b_5 b_6 \varepsilon}{b_4}}{b_6 \sqrt{(1 - b_5^2)}} \right] \right\} + \ (1 - d_i) \ln \left\{ \Phi[-\frac{lb}{b_6}] \right\} \end{aligned} \tag{3.5}$$

In terms of the parameterized $\Sigma_{e\varepsilon}$, the log-likelihood function becomes:

$$\frac{\ln L}{n} = \frac{1}{n} \sum_i d_i \ln \left\{ f(\varepsilon) * \Phi \left[a_3 lb + a_3 a_1 a_2 \varepsilon \right] \right\} + \ (1 - d_i) \ln \left\{ \Phi \left[- \frac{a_3 lb}{\sqrt{1 + a_2^2 a_3^2}} \right] \right\} \tag{3.6}$$

The MAXLIK routine in GAUSS returns estimates and standard errors of $a_1.a_3$, not $b_4.b_6$.

4. Equations referenced in estsel.src, est.src, sord.src, and sordsel.src.

4.1. Binary choice estimators (est.src and sord.src)

$$y_i^* = \frac{y_i - I(v_i > 0)}{f(v_i|u_i)} \tag{4.1}$$

$$\eta = E(zy^*) \tag{4.2}$$

$$\widehat{\Sigma}_{xz} = \sum_{i=1}^{N} \frac{x_i z_i^T}{N} \tag{4.3}$$

$$\widehat{\Sigma}_{zz} = \sum_{i=1}^{N} \frac{z_i z_i^T}{N} \tag{4.4}$$

$$\widehat{\Delta} = (\widehat{\Sigma}_{xz}\widehat{\Sigma}_{zz}^{-1}\widehat{\Sigma}_{xz}^T)^{-1}\widehat{\Sigma}_{xz}\widehat{\Sigma}_{zz}^{-1} \tag{4.5}$$

$$\widehat{y}_i^* = \frac{y_i - I(v_i > 0)}{\widehat{f}(v_i|u_i)} \tag{4.6}$$

$$\widehat{\eta} = \sum_{i=1}^{N} \frac{z_i \widehat{y}_i^*}{N} \tag{4.7}$$

$$\widehat{\beta} = \widehat{\Delta}\widehat{\eta} \tag{4.8}$$

$$q_i = z_i y_i^* + E(z_i y_i^*|u_i) - E(z_i y_i^*|v_{i,} u_i) \tag{4.9}$$

$$\sqrt{N}(\widehat{\beta} - \beta) \Longrightarrow N(0, \Delta var(q - zx^T\widehat{\beta})\Delta^T) \tag{4.12}$$

$$\widehat{w}_i = v_i - z_i \left(\sum_{i=1}^{N} z_i z_i^T\right)^{-1} \sum_{i=1}^{N} z_i v_i \tag{4.14}$$

$$\widehat{\widehat{y}}_i^* = \frac{[y_i - I(v_i > 0)]\left(w_i^+ - w_i^-\right)N}{2} \tag{4.15}$$

$$\widehat{\beta} = \widehat{\Delta}\sum_{i=1}^{N} \frac{z_i \widehat{\widehat{y}}_i^*}{N} \tag{4.16}$$

4.2. Sample selection estimators (estsel.src and sordsel.src)

$$\widehat{w}_i = \frac{d_i}{\widehat{f}(v_i|u_i)} \tag{3.5}$$

$$\widehat{\eta}_x = N^{-1} \sum_{i=1}^{N} \frac{\widehat{w}_i x_i z_i^T}{k_N} \tag{5.3}$$

$$\widehat{\eta}_y = N^{-1} \sum_{i=1}^{N} \frac{\widehat{w}_i z_i y_i}{k_N} \tag{5.4}$$

$$\widehat{\Sigma}_{zz} = \sum_{i=1}^{N} \frac{z_i z_i^T}{N} \tag{5.5}$$

$$\widehat{\Delta} = \left(\widehat{\eta}_x \widehat{\Sigma}_{zz}^{-1} \widehat{\eta}_x^T\right)^{-1} \widehat{\eta}_x \widehat{\Sigma}_{zz}^{-1} \tag{5.6}$$

$$\widehat{\beta} = \widehat{\Delta}\widehat{\eta} \tag{5.7}$$

$$I_{Ni} = I\left[|f_{vu}(v_i, u_i)| \geq \tau_N\right]$$

$$q_{Ni} = \frac{I_{Ni} w_i z_i y_i^*}{k_N} + E(\frac{I_{Ni} w_i z_i y_i^*}{k_N}|u_i) - E(\frac{I_{Ni} w_i z_i y_i^*}{k_N}|v_{i,}u_i) \tag{5.9}$$

$$\begin{aligned} \sqrt{N}(\widehat{\beta} - \beta) &\implies N(0, V_\beta) \qquad (4.1)\\ \widehat{V}_\beta &= \widehat{\Delta} var(q_{Ni} - \frac{\widehat{w}_i z_i x_i^T \widehat{\beta}}{k_N})\widehat{\Delta}^T) \end{aligned}$$

$$\widehat{s}_i = v_i - u_i \left(\sum_{i=1}^{N} u_i u_i^T\right)^{-1} \sum_{i=1}^{N} u_i v_i \tag{6.1}$$

$$\widehat{\eta}_x = \frac{\sum_{i=1}^{N} \left(\widehat{s}_i^{+} - \widehat{s}_i^{-}\right) d_i x_i z_i^T}{2} \tag{6.2}$$

$$\widehat{\eta}_y = \frac{\sum_{i=1}^{N} \left(\widehat{s}_i^{+} - \widehat{s}_i^{-}\right) d_i y_i z_i^T}{2} \tag{6.3}$$