Binary Outcome Panel Data Models

230347 Advanced Microeconometrics
Tilburg University

Christoph Walsh

Introduction

- Suppose the dependent variable is a binary variable: $y_{it} \in \{0, 1\}$.
- We often model y_{it} as a function of a latent variable $y_{it}^{\star}$, where $y_{it} = \mathbb{1}\{y_{it}^{\star} > 0\}$ and $y_{it}^{\star} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i + \varepsilon_{it}$.
- Then $\Pr(y_{it} = 1) = \Pr(y_{it}^{\star} > 0) = \Pr(\varepsilon_{it} > -\boldsymbol{x}_{it}'\boldsymbol{\beta} - \alpha_i) = F(\boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i)$, where the last equality holds if $dF(\cdot)$ is symmetric around zero.
- Two popular distributions for F are the normal (probit) and logistic (logit) distributions.
- In this lecture we will study:
    - How to estimate the static fixed effects logit model.
    - How to estimate the dynamic fixed effects logit model.
    - How to estimate the random effects probit model.

Interpretation as Difference in Payoffs

- The latent variable $y_{it}^{\star}$ can have the interpretation of being the difference in payoff from two different alternatives:

$$u_{it}^{0} = \left(\boldsymbol{x}_{it}^{0}\right)' \boldsymbol{\beta} + \alpha_i^0 + \varepsilon_{it}^0$$
$$u_{it}^{1} = \left(\boldsymbol{x}_{it}^{1}\right)' \boldsymbol{\beta} + \alpha_i^1 + \varepsilon_{it}^1$$

- $y_{it} = 1$ iff $u_{it}^1 > u_{it}^0$ (if alternative 1 gives a higher payoff than alternative 0).
- Then:

$$\underbrace{u_{it}^1 - u_{it}^0}_{y_{it}^{\star}} = \underbrace{\left(\boldsymbol{x}_{it}^1 - \boldsymbol{x}_{it}^0\right)' \boldsymbol{\beta}}_{x_{it}'\beta} + \underbrace{\left(\alpha_i^1 - \alpha_i^0\right)}_{\alpha_i} + \underbrace{\varepsilon_{it}^1 - \varepsilon_{it}^0}_{\varepsilon_{it}}$$

- Often the alternative 0 is the "outside option" where $\boldsymbol{x}_{it}^0 = \mathbf{0}$.
- If we assume ε_{it}^0 and ε_{it}^1 are normal, then the difference ε_{it} is also normal.
- If we assume ε_{it}^0 and ε_{it}^1 are independent Type I extreme value (Gumbel), then the difference ε_{it} is logit.

Likelihood

- The contribution of i to the likelihood for binary outcome panel data is:

$$f\left(\boldsymbol{y}_i | \boldsymbol{x}_i, \alpha_i, \boldsymbol{\beta}\right) = \prod_{t=1}^{T} \left[\underbrace{F\left(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta}\right)}_{=\Pr(y_{it}=1|x_i,\alpha_i,\beta)}\right]^{y_{it}} \left[\underbrace{1 - F\left(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta}\right)}_{=\Pr(y_{it}=0|x_i,\alpha_i,\beta)}\right]^{1-y_{it}}$$

- If $F(x) = \frac{e^x}{1+e^x}$ (logit):

$$\begin{aligned} f\left(\boldsymbol{y}_i | \boldsymbol{x}_i, \alpha_i, \boldsymbol{\beta}\right) &= \prod_{t=1}^{T} \left(\frac{\exp\left(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta}\right)}{1 + \exp\left(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta}\right)}\right)^{y_{it}} \left(\frac{1}{1 + \exp\left(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta}\right)}\right)^{1-y_{it}} \\ &= \frac{\exp\left(\sum_{t=1}^{T} y_{it}\left(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta}\right)\right)}{\prod_{t=1}^{T}\left[1 + \exp\left(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta}\right)\right]} \\ &= \frac{\exp\left(\alpha_i \sum_{t=1}^{T} y_{it}\right) \exp\left(\sum_{t=1}^{T}\left(y_{it}\boldsymbol{x}_{it}'\right)\boldsymbol{\beta}\right)}{\prod_{t=1}^{T}\left[1 + \exp\left(\alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta}\right)\right]} \end{aligned}$$

Incidental Parameters Problem

- The parameters are $\left(\{\alpha_i\}_{i=1}^{i=N}, \boldsymbol{\beta}\right)$
- As $N \to \infty$ while T is fixed, the number of parameters increases with N.
  - This is the incidental parameters problem.
- In the linear case, we were able to remove the α_i with either first-differences or the within transformation.
- In nonlinear cases, this transformation is not always possible.
- We could just add a dummy variable for each i, but in this setting the $\widehat{\alpha}_i$ and $\widehat{\boldsymbol{\beta}}$ are not asymptotically independent.
  - As $N \to \infty$ with fixed T, the inconsistency of $\widehat{\alpha}_i$ is transmitted into $\widehat{\boldsymbol{\beta}}$.
  - For example, Hsiao (2003, 7.3.1) shows that when $T = 2$, $x_{i1} = 0$ and $x_{i2} = 1$, that $\widehat{\beta}_{MLE} \xrightarrow{p} 2\beta$.
  - Furthermore, with large N estimation of $N + \dim(\boldsymbol{\beta})$ parameters is difficult.

Sufficient Statistic: Definition

- Let $\boldsymbol{X} = X_1, \ldots, X_N$ be a random sample from a population.
- The random variable (or random vector) $T(\boldsymbol{X})$ is a *statistic*.
  - For example $T(\boldsymbol{X}) = \frac{1}{N}\sum_{i=1}^{N} X_i$, the sample mean.
- A statistic $T(\boldsymbol{X})$ is a *sufficient statistic* for the parameter $\boldsymbol{\theta}$ if the conditional distribution of the sample $\boldsymbol{X}$ given $T(\boldsymbol{X})$ does not depend on $\boldsymbol{\theta}$.
- For example:
  - The parameter θ that we want to estimate is the mean of X.
  - The statistic $T(\boldsymbol{X}) = \frac{1}{N}\sum_{i=1}^{N} X_i$ is a sufficient statistic for θ because once the researcher knows $\frac{1}{N}\sum_{i=1}^{N} X_i$, then knowing each of the individual $X_1, \ldots, X_N$ provides no additional information about θ.

Sufficient Statistic: A Coin Toss Example

- Each X_i are iid Bernoulli distributed with probability parameter θ.
- Given a random sample $X_1, \ldots, X_N$, we will show that $T(\boldsymbol{X}) = X_1 + \cdots + X_N$ is a sufficient statistic for θ.
- Since $T(\boldsymbol{X})$ is a sum of iid Bernoulli draws, $T(\boldsymbol{X})$ follows a binomial (N, θ) distribution.
  - Its pmf is then $q(T(\boldsymbol{x})|\theta) = \binom{N}{k}\theta^k(1-\theta)^{N-k}$
- The pmf of the sample $\boldsymbol{X}$ conditional on $T(\boldsymbol{X})$ is then (where $k = \sum_{i=1}^N x_i$):

$$\begin{aligned}
\frac{p(\boldsymbol{x}|\theta)}{q(T(\boldsymbol{x})|\theta)} &= \frac{\prod_{i=1}^N \theta^{x_i}(1-\theta)^{1-x_i}}{\binom{N}{k}\theta^k(1-\theta)^{N-k}} \\
&= \frac{\theta^{\sum_{i=1}^N x_i}(1-\theta)^{\sum_{i=1}^N(1-x_i)}}{\binom{N}{k}\theta^k(1-\theta)^{N-k}} \\
&= \frac{\theta^k(1-\theta)^{N-k}}{\binom{N}{k}\theta^k(1-\theta)^{N-k}} = \frac{1}{\binom{N}{k}} \qquad \text{doesn't depend on } \theta!
\end{aligned}$$

Sufficient Statistic for α_i, $T = 2$ Case

- We will show that $c = \sum_{t=1}^{T} y_{it}$ is a sufficient statistic for α_i in the fixed effects logit model, showing for $T = 2$ first.
- For $c = 0$ and $c = 2$:

$$\Pr\left(y_{i1} + y_{i2} = 0\right) = \Pr\left(y_{i1} = 0, y_{i2} = 0\right) = \frac{1}{\left[1 + e^{\alpha_i + x_{i1}'\beta}\right]\left[1 + e^{\alpha_i + x_{i2}'\beta}\right]}$$

$$\Pr\left(y_{i1} + y_{i2} = 2\right) = \Pr\left(y_{i1} = 1, y_{i2} = 1\right) = \frac{e^{\alpha_i + x_{i1}'\beta} \times e^{\alpha_i + x_{i2}'\beta}}{\left[1 + e^{\alpha_i + x_{i1}'\beta}\right]\left[1 + e^{\alpha_i + x_{i2}'\beta}\right]}$$

- For $c = 1$:

$$\begin{aligned}\Pr\left(y_{i1} + y_{i2} = 1\right) &= \Pr\left(y_{i1} = 0, y_{i2} = 1\right) + \Pr\left(y_{i1} = 1, y_{i2} = 0\right) \\ &= \frac{e^{\alpha_i + x_{i1}'\beta} + e^{\alpha_i + x_{i2}'\beta}}{\left[1 + e^{\alpha_i + x_{i1}'\beta}\right]\left[1 + e^{\alpha_i + x_{i2}'\beta}\right]}\end{aligned}$$

$T = 2$ Case Continued

- If $c = 0$ or $c = 2$, then the conditional likelihood is always 1 (and thus doesn't depend on α_i):
  - If $y_{i1} + y_{i2} = 0$, then $\Pr\left(y_{i1} = 0, y_{i2} = 0 | y_{i1} + y_{i2} = 0\right) = 1$
  - If $y_{i1} + y_{i2} = 2$, then $\Pr\left(y_{i1} = 1, y_{i2} = 1 | y_{i1} + y_{i2} = 2\right) = 1$
- If $c = 1$:

$$
\begin{aligned}
&\Pr\left(y_{i1}, y_{i2} | y_{i1} + y_{i2} = 1\right) \\
&= \frac{\Pr\left(y_{i1}, y_{i2}\right)}{\Pr\left(y_{i1} + y_{i2} = 1\right)} \\
&= \frac{\left[e^{y_{i1}\left(\alpha_i + x_{i1}'\beta\right)} \times e^{y_{i2}\left(\alpha_i + x_{i2}'\beta\right)}\right] / \left[1 + e^{\alpha_i + x_{i1}'\beta}\right]\left[1 + e^{\alpha_i + x_{i2}'\beta}\right]}{\left[e^{\alpha_i + x_{i1}'\beta} + e^{\alpha_i + x_{i2}'\beta}\right] / \left[1 + e^{\alpha_i + x_{i1}'\beta}\right]\left[1 + e^{\alpha_i + x_{i2}'\beta}\right]} \\
&= \frac{e^{y_{i1}\alpha_i + y_{i2}\alpha_i} \times e^{y_{i1}x_{i1}'\beta + y_{i2}x_{i2}'\beta}}{e^{\alpha_i + x_{i1}'\beta} + e^{\alpha_i + x_{i2}'\beta}} \\
&= \frac{e^{y_{i1}x_{i1}'\beta + y_{i2}x_{i2}'\beta}}{e^{x_{i1}'\beta} + e^{x_{i2}'\beta}} \qquad \text{doesn't depend on } \alpha_i!
\end{aligned}
$$

$T = 2$ Case: Estimation

- Conditioning on $y_{i1} + y_{i2}$ means the conditional likelihood no longer depends on α_i.
- We would then estimate $\boldsymbol{\beta}$ using:

$$\widehat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \sum_{i=1}^{N} \log\left(\frac{e^{y_{i1}x_{i1}'\beta + y_{i2}x_{i2}'\beta}}{e^{x_{i1}'\beta} + e^{x_{i2}'\beta}}\right)$$

- Note that whenever $y_{i1} + y_{i2} = 0$ or $= 2$, the log-likelihood is zero for any value of $\boldsymbol{\beta}$, so identification only comes from individuals where $y_{i1} + y_{i2} = 1$.

General Case

- Let $\mathcal{B}_c = \left\{ \boldsymbol{d}_i : \boldsymbol{d}_i \in \{0,1\}^T, \sum_{t=1}^T d_t = c \right\}$, the set of every possible sequence of T zeros or ones such that $\sum_{t=1}^T d_{it} = c$.
- For example, if $T = 3$, then:

$$
\begin{array}{ll}
\mathcal{B}_0 = \{(0,0,0)\} & \mathcal{B}_1 = \{(1,0,0),(0,1,0),(0,0,1)\} \\
\mathcal{B}_2 = \{(1,1,0),(1,0,1),(0,1,1)\} & \mathcal{B}_3 = \{(1,1,1)\}
\end{array}
$$

- Then:

$$
\begin{aligned}
\Pr\left(\sum_{t=1}^T y_{it} = c \,\middle|\, \boldsymbol{x}_i, \alpha_i, \boldsymbol{\beta} \right) &= \sum_{d_i \in \mathcal{B}_c} \Pr\left(\boldsymbol{d}_i \middle| \boldsymbol{x}_i, \alpha_i, \boldsymbol{\beta} \right) \\
&= \sum_{d_i \in \mathcal{B}_c} \frac{\exp\left(\alpha_i \sum_{t=1}^T d_{it}\right) \exp\left(\sum_{t=1}^T \left(d_{it} \boldsymbol{x}'_{it}\right) \boldsymbol{\beta}\right)}{\prod_{t=1}^T \left[1 + \exp\left(\alpha_i + \boldsymbol{x}'_{it} \boldsymbol{\beta}\right)\right]}
\end{aligned}
$$

Sufficient Statistic for α_i

$$
\begin{aligned}
f\left(\boldsymbol{y}_i \,\middle|\, \sum_{t=1}^{T} y_{it} = c\right) &= \frac{\Pr\left(\boldsymbol{y}_i, \sum_{t=1}^{T} y_{it} = c\right)}{\Pr\left(\sum_{t=1}^{T} y_{it} = c\right)} \\
&= \frac{\Pr\left(\boldsymbol{y}_i\right)}{\Pr\left(\sum_{t=1}^{T} y_{it} = c\right)} \\
&= \frac{e^{\alpha_i \sum_{t=1}^{T} y_{it}} e^{\left(\sum_{t=1}^{T} y_{it} x_{it}'\right)\beta} / \prod_{t=1}^{T}\left(1 + e^{\alpha_i + x_{it}'\beta}\right)}{\sum_{d_i \in \mathcal{B}_c} e^{\alpha_i \sum_{t=1}^{T} d_{it}} e^{\left(\sum_{t=1}^{T} d_{it} x_{it}'\right)\beta} / \prod_{t=1}^{T}\left(1 + e^{\alpha_i + x_{it}'\beta}\right)} \\
&= \frac{\exp\left(\left(\sum_{t=1}^{T} y_{it} \boldsymbol{x}_{it}'\right)\boldsymbol{\beta}\right)}{\sum_{d_i \in \mathcal{B}_c} \exp\left(\left(\sum_{t=1}^{T} d_{it} \boldsymbol{x}_{it}'\right)\boldsymbol{\beta}\right)}
\end{aligned}
$$

- $f\left(\boldsymbol{y}_i \,\middle|\, \sum_{t=1}^{T} y_{it} = c\right)$ does not depend on α_i, so $\sum_{t=1}^{T} y_{it}$ is a sufficient statistic for α_i.

Problems with this Approach

- If either $\sum_{t=1}^{T} y_{it} = 0$ or $\sum_{t=1}^{T} y_{it} = T$ for any i, then those observations will drop out.
- For small T, the number of alternatives is manageable:
  - If $T = 3$, then $\{1, 1, 0\}$, $\{1, 0, 1\}$ or $\{0, 1, 1\}$ are possible sequences such that $\sum_{t=1}^{T} y_{it} = 2$.
- With $T = 10$ and $\sum_{t=1}^{T} y_{it} = 5$, there are $\frac{10!}{5!5!} = 252$ possible alternatives.

Dynamic Logit: Honoré and Kyriazidou (2000)

- Suppose we have $y_{it} = \mathbb{1}\left\{\rho y_{it-1} + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i + \varepsilon_{it} > 0\right\}$
- ε_{it} is iid logistic, independent over time, and independent of $\boldsymbol{x}_i, \alpha_i, y_{i1}$.

$$\Pr\left(y_{i1} = 1 | \boldsymbol{x}_i, \alpha_i\right) = p_1\left(\boldsymbol{x}_i, \alpha_i\right)$$

$$\Pr\left(y_{it} = 1 | \boldsymbol{x}_i, \alpha_i, y_{i1}, \ldots, y_{it-1}\right) = \frac{\exp\left(\rho y_{it-1} + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i\right)}{1 + \exp\left(\rho y_{it-1} + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i\right)}$$

- y_{i1} is observed, but $\boldsymbol{x}_{i1}$ need not be.
- $p_1(\cdot)$ does not need to be specified.
- To illustrate we will consider only the $T = 4$ case.

Dynamic Logit: Honoré and Kyriazidou (2000)

Define the events:

$$A = \{y_{i1} = d_1, y_{i2} = 0, y_{i3} = 1, y_{i4} = d_4\}$$
$$B = \{y_{i1} = d_1, y_{i2} = 1, y_{i3} = 0, y_{i4} = d_4\}$$

where d_1 and d_4 can each be either 0 or 1. Then:

$$\Pr(A|\boldsymbol{x}_i, \alpha_i) = [p_1(\boldsymbol{x}_i, \alpha_i)]^{d_1} [1 - p_1(\boldsymbol{x}_i, \alpha_i)]^{1-d_1} \times \frac{1}{1 + \exp(\rho d_1 + \boldsymbol{x}_{i2}'\boldsymbol{\beta} + \alpha_i)} \times \frac{\exp(\boldsymbol{x}_{i3}'\boldsymbol{\beta}) \exp(\alpha_i)}{1 + \exp(\boldsymbol{x}_{i3}'\boldsymbol{\beta} + \alpha_i)} \times \frac{\exp(\rho d_4) \exp(d_4 \boldsymbol{x}_{i4}'\boldsymbol{\beta}) \exp(d_4 \alpha_i)}{1 + \exp(\rho + \boldsymbol{x}_{i4}'\boldsymbol{\beta} + \alpha_i)}$$

$$\Pr(B|\boldsymbol{x}_i, \alpha_i) = [p_1(\boldsymbol{x}_i, \alpha_i)]^{d_1} [1 - p_1(\boldsymbol{x}_i, \alpha_i)]^{1-d_1} \times \frac{\exp(\rho d_1 + \boldsymbol{x}_{i2}'\boldsymbol{\beta}) \exp(\alpha_i)}{1 + \exp(\rho d_1 + \boldsymbol{x}_{i2}'\boldsymbol{\beta} + \alpha_i)} \times \frac{1}{1 + \exp(\rho + \boldsymbol{x}_{i3}'\boldsymbol{\beta} + \alpha_i)} \times \frac{\exp(d_4 \boldsymbol{x}_{i4}'\boldsymbol{\beta}) \exp(d_4 \alpha_i)}{1 + \exp(\boldsymbol{x}_{i4}'\boldsymbol{\beta} + \alpha_i)}$$

Dynamic Logit: Honoré and Kyriazidou (2000)

- The probability of A given A or B:

$$\Pr\left(A|\boldsymbol{x}_i, \alpha_i, A \cup B\right) = \frac{\Pr\left(A \cap \left(A \cup B\right)|\boldsymbol{x}_i, \alpha_i\right)}{\Pr\left(A \cup B|\boldsymbol{x}_i, \alpha_i\right)} = \frac{\Pr\left(A|\boldsymbol{x}_i, \alpha_i\right)}{\Pr\left(A|\boldsymbol{x}_i, \alpha_i\right) + \Pr\left(B|\boldsymbol{x}_i, \alpha_i\right)}$$

and similarly for $\Pr\left(B|\boldsymbol{x}_i, \alpha_i, A \cup B\right)$

- $\Pr\left(A|\boldsymbol{x}_i, \alpha_i, A \cup B\right)$ is then (purple, blue, orange, maroon terms cancel):

$$\frac{\frac{\exp\left(x_{i3}'\beta\right)}{1+\exp\left(x_{i3}'\beta+\alpha_i\right)}\frac{\exp(\rho d_4)}{1+\exp\left(\rho+x_{i4}'\beta+\alpha_i\right)}}{\frac{\exp\left(x_{i3}'\beta\right)}{1+\exp\left(x_{i3}'\beta+\alpha_i\right)}\frac{\exp(\rho d_4)}{1+\exp\left(\rho+x_{i4}'\beta+\alpha_i\right)} + \exp\left(\rho d_1 + x_{i2}'\beta\right)\frac{1}{1+\exp\left(\rho+x_{i3}'\beta+\alpha_i\right)}\frac{1}{1+\exp\left(x_{i4}'\beta+\alpha_i\right)}}$$

- Unless $\boldsymbol{x}_{i3} = \boldsymbol{x}_{i4}\ \forall i$, $\Pr\left(A|\boldsymbol{x}_i, \alpha_i, A \cup B\right)$ and $\Pr\left(B|\boldsymbol{x}_i, \alpha_i, A \cup B\right)$ will depend on α_i.

Dynamic Logit: Honoré and Kyriazidou (2000)

- For the subsample where $\boldsymbol{x}_{i3} = \boldsymbol{x}_{i4}$, the green and red terms in the probabilities $\Pr(A|\boldsymbol{x}_i, \alpha_i)$ and $\Pr(B|\boldsymbol{x}_i, \alpha_i)$ will cancel. So:

$$\begin{aligned}\Pr(A|\boldsymbol{x}_i, \alpha_i, A \cup B, \boldsymbol{x}_{i3} = \boldsymbol{x}_{i4}) &= \frac{\Pr(A|\boldsymbol{x}_i, \alpha_i)}{\Pr(A|\boldsymbol{x}_i, \alpha_i) + \Pr(B|\boldsymbol{x}_i, \alpha_i)} \\ &= \frac{\exp(\rho d_4 + \boldsymbol{x}_{i3}'\boldsymbol{\beta})}{\exp(\rho d_4 + \boldsymbol{x}_{i3}'\boldsymbol{\beta}) + \exp(\rho d_1 + \boldsymbol{x}_{i2}'\boldsymbol{\beta})}\end{aligned}$$

- Rearranging terms:

$$\Pr(A|\boldsymbol{x}_i, \alpha_i, A \cup B, \boldsymbol{x}_{i3} = \boldsymbol{x}_{i4}) = \frac{1}{1 + \exp\left((\boldsymbol{x}_{i2} - \boldsymbol{x}_{i3})' \boldsymbol{\beta} + \rho(d_1 - d_4)\right)}$$

$$\Pr(B|\boldsymbol{x}_i, \alpha_i, A \cup B, \boldsymbol{x}_{i3} = \boldsymbol{x}_{i4}) = \frac{\exp\left((\boldsymbol{x}_{i2} - \boldsymbol{x}_{i3})' \boldsymbol{\beta} + \rho(d_1 - d_4)\right)}{1 + \exp\left((\boldsymbol{x}_{i2} - \boldsymbol{x}_{i3})' \boldsymbol{\beta} + \rho(d_1 - d_4)\right)}$$

Maximum Likelihood Estimation

The log likelihood function is then:

$$\sum_{i=1}^{N} \mathbb{1}\left\{y_{i2} + y_{i3} = 1\right\} \mathbb{1}\left\{\boldsymbol{x}_{i3} - \boldsymbol{x}_{i4} = \boldsymbol{0}\right\} \log\left(\frac{\exp\left(y_{i2}\left[\left(\boldsymbol{x}_{i2} - \boldsymbol{x}_{i3}\right)' \boldsymbol{\beta} + \rho\left(y_{i1} - y_{i4}\right)\right]\right)}{1 + \exp\left(\left(\boldsymbol{x}_{i2} - \boldsymbol{x}_{i3}\right)' \boldsymbol{\beta} + \rho\left(y_{i1} - y_{i4}\right)\right)}\right)$$

- $\boldsymbol{x}_{i3} = \boldsymbol{x}_{i4}$ may not occur in the data. This is very likely if a variable is continuous.
- Honoré and Kyriazidou (2000) provide a kernel smoothing method with weights that depend on $\boldsymbol{x}_{i3} - \boldsymbol{x}_{i4}$, giving more weight to "closer" observations:

Probit Random Effects: Introduction

- Consider again the binary dependent variable model:

$$y_{it} = \mathbb{1}\left\{\boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i + \varepsilon_{it} > 0\right\} \qquad i = 1, \ldots, N \quad t = 1, \ldots, T$$

- Instead of logit-distributed errors, we assume $\varepsilon_{it} \overset{iid}{\sim} \mathcal{N}(0, 1)$.
- In the probit model, it is not possible to remove α_i with a (quasi-)differencing transformation or conditioning on a sufficient statistic.
- Instead, we will assume the α_i are random effects.
- We will assume the random effects follow a distribution that is known up to a parameter vector.

Likelihood

- The likelihood for observation i is:

$$\begin{aligned}
\Pr\left(\boldsymbol{y}_i|\boldsymbol{x}_i,\boldsymbol{\beta},\alpha_i\right) &= \prod_{t=1}^{T}\Pr\left(y_{it}|\boldsymbol{x}_{it},\boldsymbol{\beta},\alpha_i\right)\\
&= \prod_{t=1}^{T}\left[\Phi\left(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i\right)\right]^{y_{it}}\left[1-\Phi\left(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i\right)\right]^{1-y_{it}}\\
&= \prod_{t=1}^{T}\Phi\left(\left[2y_{it}-1\right]\left[\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i\right]\right)
\end{aligned}$$

where:

- $\boldsymbol{x}_i = (\boldsymbol{x}_{i1},\ldots,\boldsymbol{x}_{iT})$.
- Φ is the cdf of the normal density
- The 4th equality uses $1-\Phi(z) = \Phi(-z)$.

Integrating Out

- Suppose two random variables X and Y are jointly distributed with a pdf $f_{X,Y}$.
- The marginal pdf of X is (where $\mathcal{Y}$ is the support of Y):

$$f_X(x) = \int_{\mathcal{Y}} f_{X,Y}(x,y)\,dy = \int_{\mathcal{Y}} f_{X|Y}(x|y)\,f_Y(y)\,dy$$

Normally-Distributed Random Effects

If α_i has a density $f_\alpha(\alpha_i, \boldsymbol{\theta}_\alpha)$ with parameters $\boldsymbol{\theta}_\alpha$, we can integrate out the random effect:

$$\Pr(\boldsymbol{y}_i | \boldsymbol{x}_i, \boldsymbol{\beta}, \boldsymbol{\theta}_\alpha) = \int_{-\infty}^{\infty} \Pr(\boldsymbol{y}_i | \boldsymbol{x}_i, \boldsymbol{\beta}, \alpha_i) \, f_\alpha(\alpha_i, \boldsymbol{\theta}_\alpha) \, d\alpha_i$$

Using this in our likelihood:

$$L_i(\boldsymbol{\beta}, \boldsymbol{\theta}_\alpha) = \int_{-\infty}^{\infty} \prod_{t=1}^{T} \Phi\left([2y_{it} - 1]\left[\boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i\right]\right) f_\alpha(\alpha_i, \boldsymbol{\theta}_\alpha) \, d\alpha_i$$

If we specify $\alpha_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_\alpha^2)$, then $f_\alpha(\alpha_i, \boldsymbol{\theta}_\alpha) = \left(\sqrt{2\pi\sigma_\alpha^2}\right)^{-1} \exp\left(-\frac{\alpha_i^2}{2\sigma_\alpha^2}\right)$, so:

$$L_i(\boldsymbol{\beta}, \sigma_\alpha^2) = \int_{-\infty}^{\infty} \prod_{t=1}^{T} \Phi\left([2y_{it} - 1]\left[\boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i\right]\right) \frac{e^{-\frac{\alpha_i^2}{2\sigma_\alpha^2}}}{\sqrt{2\pi\sigma_\alpha^2}} \, d\alpha_i$$

Gauss-Hermite Quadrature

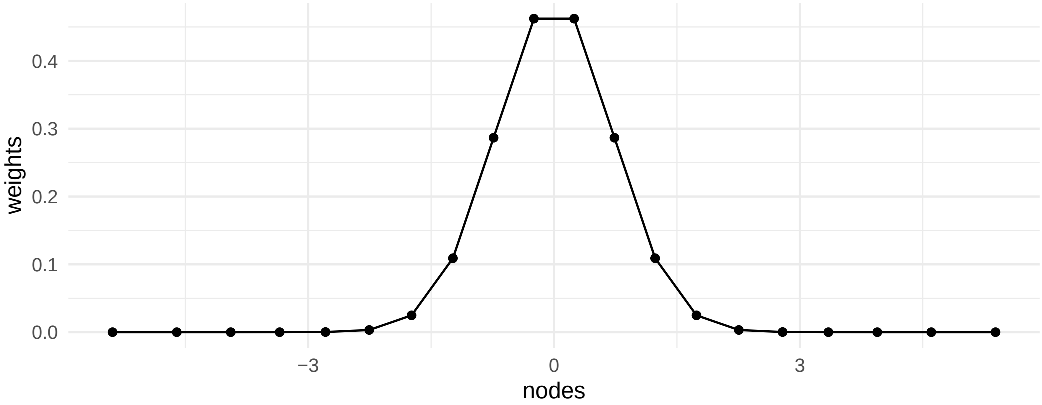
- There is no closed form solution for this integral.
- We need to approximate it numerically.
- Gauss-Hermite quadrature can approximate integrals of the following kind:

$$\int_{-\infty}^{\infty} f(x)\, e^{-x^2} dx \approx \sum_{h=1}^{H} w_h f(z_h)$$

- We evaluate the function inside the integral for different nodes z_h and take a weighted average, with weights w_h.
- Given H evaluation points, we can obtain the nodes and weights from tables.

Example Nodes and Weights with $H = 20$

```r
library(ggplot2)
library(statmod)
ghq <- data.frame(gauss.quad(20, kind = "hermite"))
ggplot(ghq, aes(nodes, weights)) + geom_line() + geom_point() +
  theme_minimal()
```



Change of Variables

- We need to transform the integral in $L_i\left(\boldsymbol{\beta},\sigma_\alpha^2\right)$ to the form $\int_{-\infty}^{\infty} f(x)\, e^{-x^2}dx$.
- Let $r_i = \frac{\alpha_i}{\sqrt{2\sigma_\alpha^2}}$ so $\alpha_i = \sqrt{2\sigma_\alpha^2}r_i$ and $d\alpha_i = \sqrt{2\sigma_\alpha^2}dr_i$.
- Performing the change of variables:

$$
\begin{aligned}
L_i\left(\boldsymbol{\beta},\sigma_\alpha^2\right) &= \int_{-\infty}^{\infty}\prod_{t=1}^{T}\Phi\left([2y_{it}-1]\left[\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i\right]\right)\frac{e^{-\frac{\alpha_i^2}{2\sigma_\alpha^2}}}{\sqrt{2\pi\sigma_\alpha^2}}d\alpha_i \\
&= \int_{-\infty}^{\infty}\prod_{t=1}^{T}\Phi\left([2y_{it}-1]\left[\boldsymbol{x}_{it}'\boldsymbol{\beta}+\sqrt{2\sigma_\alpha^2}r_i\right]\right)\frac{e^{-\frac{\left(\sqrt{2\sigma_\alpha^2}r_i\right)^2}{2\sigma_\alpha^2}}}{\sqrt{2\pi\sigma_\alpha^2}}\sqrt{2\sigma_\alpha^2}dr_i \\
&= \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}\prod_{t=1}^{T}\Phi\left([2y_{it}-1]\left[\boldsymbol{x}_{it}'\boldsymbol{\beta}+\sqrt{2\sigma_\alpha^2}r_i\right]\right)e^{-r_i^2}dr_i
\end{aligned}
$$

Approximating the likelihood

With weights w_h and nodes z_h, we can approximate the integral with:

$$L_i\left(\boldsymbol{\beta}, \sigma_\alpha^2\right) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \prod_{t=1}^{T} \Phi\left([2y_{it} - 1]\left[\boldsymbol{x}_{it}'\boldsymbol{\beta} + \sqrt{2\sigma_\alpha^2} r_i\right]\right) e^{-r_i^2} dr_i$$

$$\approx \frac{1}{\sqrt{\pi}} \sum_{h=1}^{H} w_h \prod_{t=1}^{T} \Phi\left([2y_{it} - 1]\left[\boldsymbol{x}_{it}'\boldsymbol{\beta} + \sqrt{2\sigma_\alpha^2} z_h\right]\right)$$

This leads to the maximum likelihood estimator:

$$\left(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}_\alpha^2\right) = \underset{\beta, \sigma_\alpha^2}{\arg\max} \sum_{i=1}^{N} \log\left(\frac{1}{\sqrt{\pi}} \sum_{h=1}^{H} w_h \prod_{t=1}^{T} \Phi\left([2y_{it} - 1]\left[\boldsymbol{x}_{it}'\boldsymbol{\beta} + \sqrt{2\sigma_\alpha^2} z_h\right]\right)\right)$$

Suggested Reading and References

Suggested Reading:

- Baltagi, 11.1-11.4
- Cameron and Trivedi, 23.4
- Wooldridge, 15.8
- Hsiao, 7

References:

Honoré, B. E. and E. Kyriazidou (2000): "Panel data discrete choice models with lagged dependent variables," *Econometrica*, 68, 839–874.