Lecture 6
Multiple Choice Models
Part II – MN Probit, Ordered Choice

DCM: Different Models

• Popular Models:
  1. Probit Model
  2. Binary Logit Model
  3. Multinomial Logit Model
  4. Nested Logit model
  5. Ordered Logit Model

• Relevant literature:
  - Train (2003): Discrete Choice Methods with Simulation
  - Franses and Paap (2001): Quantitative Models in Market Research
  - Hensher, Rose and Greene (2005): Applied Choice Analysis
Model – IIA: Alternative Models

- In the MNL model we assumed independent $\varepsilon_{nj}$ with extreme value distributions. This essentially created the IIA property.

- This is the main weakness of the MNL model.

- The solution to the IIA problem is to relax the independence between the unobserved components of the latent utility, $\varepsilon_i$.

- Solutions to IIA
  - Nested Logit Model, allowing correlation between some choices.
  - Models allowing correlation among the $\varepsilon_i$’s, such as MP Models.
  - Mixed or random coefficients models, where the marginal utilities associated with choice characteristics vary between individuals.

Multinomial Probit Model

- Changing the distribution of the error term in the RUM equation leads to alternative models.

- A popular alternative: The $\varepsilon_{ij}$’s follow an independent standard normal distributions for all $i,j$.

  $$U_i = \begin{pmatrix} U_{i0} \\ U_{i1} \\ \vdots \\ U_{iJ} \end{pmatrix} = \begin{pmatrix} X_{i0} \beta + \varepsilon_{i0} \\ X_{i1} \beta + \varepsilon_{i1} \\ \vdots \\ X_{iJ} \beta + \varepsilon_{iJ} \end{pmatrix}$$

  $$\varepsilon_i = \begin{pmatrix} \varepsilon_{i0} \\ \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iJ} \end{pmatrix} \sim N(0, \Omega),$$

- We retain independence across subjects but we allow dependence across alternatives, assuming that the vector $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iJ})$ follows a multivariate normal distribution, but with arbitrary covariance matrix $\Omega$. 
**Multinomial Probit Model**

- The vector $\mathbf{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iJ})$ follows a multivariate normal distribution, but with arbitrary covariance matrix $\mathbf{\Omega}$.

- The model is called the *Multinomial probit model*. It produces results similar to the MNL model after standardization.

- Some restrictions (normalization) on $\mathbf{\Omega}$ are needed.

- As usual with latent variable formulations, the variance of the error term cannot be separated from the regression coefficients. Setting the variances to one means that we work with a correlation matrix rather than a covariance matrix.

**MP Model – Pros & Cons**

- Main advantages:
  - Using ML, joint estimation of all parameters is possible.
  - It allows correlation between the utilities that an individual assigns to the various alternatives (relaxes IIA).
  - It does not rely on grouping choices. No restrictions on which choices are close substitutes.
  - It can also allow for heterogeneity in the (marginal) distributions for $\mathbf{\varepsilon}_i$.

- Main difficulty: Estimation.
  - ML estimation involves evaluating probabilities given by multidimensional normal integrals, a limitation that forces practical applications to a few alternatives ($J=3,4$). Quadrature methods can be used to approximate the integral, but for large $J$, often imprecise.
MP Model – Estimation

• Probit Problem:

\[
P_{nj} = \text{Prob}(Y_j = 1 | X) = \int \cdots \int f(V_{nj} - V_{ni} > \xi_{nj}; \forall j \neq i) f(\xi) d\xi
\]

J-dimensional integral involves \( \xi_{jk} = \varepsilon_k - \varepsilon_j \), which is normally distributed, with variance \( \Omega \). We can rewrite the probability as:

\[
P[y_j = 1 | X] = P(\xi_j < V_j)
\]

where \( V_j \) is the vector with \( k \)th element \( V_{jk} = x_j'\beta - x_k'\beta \).

Let \( \theta = \{\beta, \Omega\} \). To get the MLE, we need to evaluate this integral for any \( \beta \) and \( \Omega \). The MLE of \( \theta \) maximizes

\[
L = \sum_n \sum_j y_{nj} \log P(\xi_j < V_j) \leq \text{we need to integrate}
\]

MP Model – Estimation

• We need to integrate to get \( \log P(\xi_j < V_j) \)

If \( J = 3 \), we need to evaluate a bivariate normal – no problem.

If \( J > 3 \), we need to evaluate a 3-dimensional integral. A usual approach is to use Gaussian quadrature (Recall Math Review, Lecture 12).

Most current software programs use the Butler and Moffit (1982) method, based on Hermite quadrature.

Practical considerations: If \( J > 4 \), numerical procedures get complicated and, often, imprecise. For these cases, we rely on simulation-based estimation - simulated maximum likelihood or SML.
Review: Gaussian Quadratures

- **Newton-Cotes Formulae**
  - Nodes: Use evenly-spaced functional values
  - Weights: Use Lagrange interpolation. Best, given the nodes.
  - It can explode for large \( n \) (Runge’s phenomenon)

- **Gaussian Quadratures**
  - Select functional values at non-uniformly distributed points to achieve higher accuracy. The values are not predetermined, but unknowns to be determined.
  - Nodes and Weight are both “best” to get an exact answer if \( f \) is a \((2n-1)\)th-order polynomial. Legendre polynomials are used.
  - Change of variables \( \Rightarrow \) the interval of integration is \([-1,1]\).

\[
\sum_{i=1}^{n} c_i f(x_i) \approx \int_{-1}^{1} f(x) dx
\]

The Gauss-Legendre quadrature formula is stated as

\[ \int_{-1}^{1} f(x) dx \approx \sum_{i=1}^{n} c_i f(x_i) \]

the \( c_i \)'s are called the weights, the \( x_i \)'s are called the quadrature nodes. The approximation error term, \( \varepsilon \), is called the truncation error for integration.

For Gauss-Legendre quadrature, the nodes are chosen to be zeros of certain Legendre (orthogonal) polynomials.
Change of Interval for Gaussian Quadrature

- Coordinate transformation from $[a,b]$ to $[-1,1]$
  This can be done by an affine transformation on $t$ and a change of variables.

\[
t = \frac{b - a}{2} x + \frac{b + a}{2}
\]
\[
dt = \frac{b - a}{2} dx
\]
\[
\begin{align*}
x = -1 & \Rightarrow t = a \\
x = 1 & \Rightarrow t = b
\end{align*}
\]

\[
\int_{a}^{b} f(t) dt = \int_{-1}^{1} f\left(\frac{b - a}{2} x + \frac{b + a}{2}\right) \frac{b - a}{2} dx = \frac{b - a}{2} \sum_{i=1}^{n} c_i f(x_i)
\]

Review: Gaussian Quadrature on $[-1, 1]$

- Gauss Quadrature General formulation:

\[
\int_{-1}^{1} f(x) dx = \sum_{i=1}^{n} c_i f(x_i) = c_1 f(x_1) + c_2 f(x_2) + \cdots + c_n f(x_n)
\]

\[
\begin{align*}
n = 2 : & \quad \int_{-1}^{1} f(x) dx \\
& = c_1 f(x_1) + c_2 f(x_2)
\end{align*}
\]

- For $n=2$, we have four unknowns ($c_1$, $c_2$, $x_1$, $x_2$). These are found by assuming that the formula gives exact results for integrating a general 3rd order polynomial. It can also be done by choosing ($c_1$, $c_2$, $x_1$, $x_2$) such that it yields “exact integral” for $f(x) = x^0$, $x^1$, $x^2$, $x^3$. 
Review: Gaussian Quadrature on [-1, 1]

Case $n = 2 \Rightarrow \int_{-1}^{1} f(x)dx = c_1f(x_1) + c_2f(x_2)$

Exact integral for $f = x^0, x^1, x^2, x^3$

- Four equations for four unknowns

\[
\begin{align*}
  f = 1 & \Rightarrow \int_{-1}^{1} 1dx = 2 = c_1 + c_2 \\
  f = x & \Rightarrow \int_{-1}^{1} xdx = 0 = c_1x_1 + c_2x_2 \\
  f = x^2 & \Rightarrow \int_{-1}^{1} x^2dx = \frac{2}{3} = c_1x_1^2 + c_2x_2^2 \\
  f = x^3 & \Rightarrow \int_{-1}^{1} x^3dx = 0 = c_1x_1^3 + c_2x_2^3
\end{align*}
\]

\[
\begin{array}{c}
  c_1 = 1 \\
  c_2 = 1 \\
  x_1 = \frac{-1}{\sqrt{3}} \\
  x_2 = \frac{1}{\sqrt{3}}
\end{array}
\]

\[
I = \int_{-1}^{1} f(x)dx = f\left(-\frac{1}{\sqrt{3}}\right) + f\left(\frac{1}{\sqrt{3}}\right)
\]

Review: Gaussian Quadrature on [-1, 1]

Case $n = 3$: $\int_{-1}^{1} f(x)dx = c_1f(x_1) + c_2f(x_2) + c_3f(x_3)$

Now, choose $(c_1, c_2, c_3, x_1, x_2, x_3)$ such that the method yields “exact integral” for $f(x) = x^0, x^1, x^2, x^3, x^4, x^5$. (Again, $(c_1, c_2, c_3, x_1, x_2, x_3)$ are calculated by assuming the formula gives exact expressions for integrating a fifth order polynomial).
Review: Gaussian Quadrature on [-1, 1]

\[ f = 1 \Rightarrow \int_{-1}^{1} x \, dx = 2 = c_1 + c_2 + c_3 \]

\[ f = x \Rightarrow \int_{-1}^{1} x \, dx = 0 = c_1 x_1 + c_2 x_2 + c_3 x_3 \]

\[ f = x^2 \Rightarrow \int_{-1}^{1} x^2 \, dx = \frac{2}{3} = c_1 x_1^2 + c_2 x_2^2 + c_3 x_3^2 \]

\[ f = x^3 \Rightarrow \int_{-1}^{1} x^3 \, dx = 0 = c_1 x_1^3 + c_2 x_2^3 + c_3 x_3^3 \]

\[ f = x^4 \Rightarrow \int_{-1}^{1} x^4 \, dx = \frac{2}{5} = c_1 x_1^4 + c_2 x_2^4 + c_3 x_3^4 \]

\[ f = x^5 \Rightarrow \int_{-1}^{1} x^5 \, dx = 0 = c_1 x_1^5 + c_2 x_2^5 + c_3 x_3^5 \]

\[ \begin{cases} c_1 = \frac{5}{9} \\ c_2 = \frac{8}{9} \\ c_3 = \frac{5}{9} \end{cases} \]

\[ \begin{cases} x_1 = -\sqrt{3/5} \\ x_2 = 0 \\ x_3 = \sqrt{3/5} \end{cases} \]

---

Review: Gaussian Quadrature on [-1, 1]

- Approximation formula for n=3

\[ I = \int_{-1}^{1} f(x) \, dx = \frac{5}{9} f(-\sqrt{\frac{3}{5}}) + \frac{8}{9} f(0) + \frac{5}{9} f(\sqrt{\frac{3}{5}}) \]
Review: Gaussian Quadrature – Example 1

- Evaluate:

\[ I = \int_0^1 te^{2t} dt = 5216.926477 \]

- Coordinate transformation

\[ t = \frac{b-a}{2} x + \frac{b+a}{2} = 2x + 2; \quad dt = 2dx \]

\[ I = \int_0^1 te^{2t} dt = \int_{-1}^1 (4x + 4)e^{4x+4} dx = \int_{-1}^1 f(x) dx \]

- Two-point formula (n=2)

\[ I = \int_{-1}^1 f(x) dx = f\left( \frac{-1}{\sqrt{3}} \right) + f\left( \frac{1}{\sqrt{3}} \right) = \left( 4 - \frac{4}{\sqrt{3}} \right) e^{4\sqrt{3}} + \left( 4 + \frac{4}{\sqrt{3}} \right) e^{-4\sqrt{3}} \]

\[ = 9.167657324 + 3468.376279 = 3477.543936 \quad (\varepsilon = 33.34\%) \]

Review: Gaussian Quadrature – Example 1

- Three-point formula (n=3)

\[ I = \int_{-1}^1 f(x) dx = \frac{5}{9} f(-\sqrt{0.6}) + \frac{8}{9} f(0) + \frac{5}{9} f(\sqrt{0.6}) \]

\[ = \frac{5}{9} (4 - 4\sqrt{0.6}) e^{4-\sqrt{0.6}} + \frac{8}{9} (4)e^4 + \frac{5}{9} (4 + 4\sqrt{0.6}) e^{4+\sqrt{0.6}} \]

\[ = \frac{5}{9} (2.221191545) + \frac{8}{9} (218.3926001) + \frac{5}{9} (8589.142689) \]

\[ = 4967.106689 \quad (\varepsilon = 4.79\%) \]

- Four-point formula (n=4)

\[ I = \int_{-1}^1 f(x) dx = 0.34785\left[ f(-0.861136) + f(0.861136) \right] \]

\[ + 0.652145\left[ f(-0.33998) + f(0.33998) \right] \]

\[ = 5197.54375 \quad (\varepsilon = 0.37\%) \]

17

18
Review: Gaussian Quadrature – Example 2

- Evaluate

\[
I = \frac{1}{\sqrt{2\pi}} \int_{0}^{1.64} e^{-\frac{x^2}{2}} \, dx = .44949742
\]

- Coordinate transformation

\[
t = \frac{b-a}{2} x + \frac{b+a}{2} = .82x + .82 = .82(1+x); \quad dt = .82 \, dx
\]

\[
I = \frac{1}{\sqrt{2\pi}} \int_{0}^{1.64} e^{-\frac{t^2}{2}} \, dt = .82 \int_{-1}^{1} e^{-\frac{1}{2}[.82(1+x)]^2} \, dx = \frac{.82}{\sqrt{2\pi}} \int_{-1}^{1} f(x) \, dx
\]

Review: Gaussian Quadrature – Example 2

- Two-point formula \((n=2)\)

\[
I = \frac{.82}{\sqrt{2\pi}} \int_{-1}^{1} f(x) \, dx = \frac{.82}{\sqrt{2\pi}} \left( \frac{5}{9} f(-\sqrt{0.6}) + \frac{8}{9} f(0) + \frac{5}{9} f(\sqrt{0.6}) \right) = \frac{.82}{\sqrt{2\pi}} \left( \frac{5}{9} e^{-\frac{1}{2}[.82(1-\sqrt{0.6})]^2} + \frac{8}{9} e^{-\frac{1}{2}[.82(1-0)]^2} + \frac{5}{9} e^{-\frac{1}{2}[.82(1+\sqrt{0.6})]^2} \right)
\]

\[
= 0.32713267 \ast (0.94171147 + 0.43323413) = .44978962 \quad (\varepsilon = 0.065\%)
\]

- Three-point formula \((n=3)\)

\[
I = \frac{.82}{\sqrt{2\pi}} \int_{-1}^{1} f(x) \, dx = \frac{.82}{\sqrt{2\pi}} \left( \frac{5}{9} f(-\sqrt{0.6}) + \frac{8}{9} f(0) + \frac{5}{9} f(\sqrt{0.6}) \right) = \frac{.82}{\sqrt{2\pi}} \left( \frac{5}{9} e^{-\frac{1}{2}[.82(1-\sqrt{0.6})]^2} + \frac{8}{9} e^{-\frac{1}{2}[.82(1-0)]^2} + \frac{5}{9} e^{-\frac{1}{2}[.82(1+\sqrt{0.6})]^2} \right)
\]

\[
= .32713267 \ast (0.54614659 + 0.63509351 + 0.19271450) = .44946544 \quad (\varepsilon = 0.007\%)
\]
Review: Multidimensional Integrals

- In the review, we concentrated on one-dimensional integrals. For integration in multiple dimensions, one approach is to phrase the multiple integral as repeated one-dimensional integrals.

- But, eventually, we run into the so-called curse of dimensionality. Four or more dimensions are complicated and, often, imprecise.

- There are two methods that work well:
  1. **Monte Carlo**: Based on repeated function evaluations, not repeated integrations using one-dimensional methods.

    Popular algorithm: Markov chain Monte Carlo (MCMC), which include the Metropolis-Hastings algorithm and Gibbs sampling.

  2. **Sparse grids**: Based on a one dimensional quadrature rule, but uses a recursive combination of univariate results.

Hermite Quadrature (Greene)

- Hermite (or Gauss–Hermite) quadrature is an extension of the Gaussian quadrature method for approximating the value of integrals of the following kind:

\[
I = \int_{-\infty}^{\infty} e^{-t^2} f(t) dt \approx \sum_{i=1}^{n} w_i f(x_i)
\]

- It is a method well adapted to the kind of integral we see when we assume normality for \( f(\varepsilon) \), like in probit models.

- Useful approximation to compute moments of a normal distribution.

The \( x_i \) roots are given by the Hermite polynomial, \( H_n \), and the

\[
(2) \quad H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2} = e^{x^2} \left( x - \frac{d}{dx} \right) e^{-x^2} \quad w_i = \frac{2^{n-1} n! \sqrt{\pi}}{n^2 [H_{n-1}(x_i)]^2}
\]
Hermite Quadrature (Greene)

- The problem: approximating an integral, involving $\exp(-x^2)$:

$$\int_{-\infty}^{\infty} f(x, v) \exp(-v^2)dv \approx \sum_{h=1}^{H} f(x, v_h)W_h$$

Adapt to integrating out a normal variable

$$f(x) = \int_{-\infty}^{\infty} f(x, v) \frac{\exp(-\frac{1}{2}(v / \sigma)^2)}{\sigma \sqrt{2\pi}} dv$$

Change the variable to $z = (1/(\sigma \sqrt{2}))v$, $v = (\sigma \sqrt{2})z$ and $dv = (\sigma \sqrt{2})dz$

$$f(x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} f(x, \lambda z) \exp(-z^2)dz, \lambda = \sigma \sqrt{2}$$

This can be accurately approximated by Hermite quadrature

$$f(x) \approx \sum_{h=1}^{H} f(x, \lambda z)W_h$$

---

**Example** (Butler and Moffitt’s Approach): Random Effects Log Likelihood Function

$$\log L = \sum_{i=1}^{N} \log \int_{-\infty}^{\infty} \left( \prod_{t=1}^{T} g \left( y_{it}, (x_{it}' \beta^0 + v_i) \right) \right) h(v_i)dv_i$$

Butler and Moffitt: Compute this by Hermite quadrature

$$\int_{-\infty}^{\infty} f(v_i)h(v_i)dv_i \approx \sum_{h=1}^{H} f(z_h)w_h$$ when $h(v_i) = \text{normal density}$

$z_h = \text{quadrature node}; w_h = \text{quadrature weight}$

$z_i = \sigma v_i, \sigma \text{ is estimated with } \beta^0$
Hermite Quadrature (Greene) - Example

Example:
Nodes for 8 point Hermite Quadrature:
(Use both signs, + and -)
0.381186990207322000, 1.15719371244677990, 1.98165675669584300, 2.93063742025714410

Weights for 8 point Hermite Quadrature
0.661147012558199960, 0.20780232581489999, 0.0170779830074100010, 0.000199604072211400010

MP Model – Simulation-based Estimation

• ML Estimation is complicated due to the multidimensional integration problem. Simulation-based methods approximate the integral. Relatively easy to apply.

• Simulation provides a solution for dealing with problems involving an integral. For example:
\[ \mathbb{E}[h(U)] = \int h(u) f(u) \, du \]

• All GMM and many ML problems require the evaluation of an expectation. In many cases, an analytic solution or a precise numerical solution is not possible. But, we can always simulate \( \mathbb{E}[h(u)] \):
  - Steps
    - Draw \( R \) pseudo RV from \( f(u) \): \( u^1, u^2, \ldots, u^R \) (\( R: \) repetitions)
    - Compute \( \hat{\mathbb{E}}[h(U)] = \frac{1}{R} \sum_n h(u^n) \)
**MP Model – Simulation-based Estimation**

- We call $\hat{E}[h(U)]$ a *simulator*.

- If $h(.)$ is continuous and differentiable, then $\hat{E}[h(U)]$ will be continuous and differentiable.

- Under general conditions, $\hat{E}[h(U)]$ provides an unbiased (and most of the times consistent) estimator for $E[h(U)]$.

- The variance of $\hat{E}[h(U)]$ is equal to $\text{Var}[h(U)]/R$.

**Review: The Probability Integral Transformation**

- This transformation allows one to convert observations that come from a uniform distribution from 0 to 1 to observations that come from an arbitrary distribution.

Let $U$ denote an observation having a uniform distribution $[0, 1]$.

$$g(u) = \begin{cases} 
1 & 0 \leq u \leq 1 \\
0 & \text{elsewhere}
\end{cases}$$

Let $f(x)$ denote an arbitrary pdf and $F(x)$ its corresponding CDF. Let $X = F^{-1}(U)$.

We want to find the distribution of $X$. 
Review: The Probability Integral Transformation

• Find the distribution of $X$.

\[
G(x) = P[X \leq x] = P[F^{-1}(U) \leq x] \\
= P[U \leq F(x)] \\
= F(x)
\]

Hence: 
\[
g(x) = G'(x) = F'(x) = f(x)
\]

Thus if $U \sim$ Uniform distribution in [0, 1], then,

\[X = F^{-1}(U)\]

has density $f(x)$.

Example: Exponential distribution

Let $U \sim$ Uniform(0,1).

Let $F(x) = 1 - \exp(-\lambda x)$ –i.e., the exponential distribution.

Then,

\[-\log(1 - U)/\lambda \sim F(\text{exponential distribution})\]

Example: If $F$ is the standard normal, $F^{-1}$ has no closed form solution. Most computers programs have a routine to approximate $F^{-1}$ for the standard normal distribution.
Review: The Probability Integral Transformation

• Truncated RVs can be simulated along these lines.

Example: $U \sim N(\mu, \sigma^2)$, but it is truncated between $a$ and $b$. Then,

$$F(u) = \left[ \Phi\left( \frac{u - \mu}{\sigma} \right) - \Phi\left( \frac{a - \mu}{\sigma} \right) \right] / \left[ \Phi\left( \frac{b - \mu}{\sigma} \right) - \Phi\left( \frac{a - \mu}{\sigma} \right) \right]$$

$U$ can be simulated by letting $F(u) = Z$ and solving for $u$ as:

$$\sigma Z^{-1} \left[ \Phi\left( \frac{b - \mu}{\sigma} \right) - \Phi\left( \frac{a - \mu}{\sigma} \right) \right] + \mu$$

MP Model – Simulation-based Estimation

• Probit Problem:
  - We write the probability of choice $j$ as:  $P[y_i=1 \mid X] = P(\xi_i < V_j)$, where $V_j$ is the vector with $k$th element $V_{jk} = x_j' \beta - x_k' \beta$.

  - Let $\theta = \{ \beta, \Omega \}$. The MLE of $\theta$ maximizes
    $$(1/N) \sum_n \sum_j y_{nj} \log P(\xi_i < V_j) \leq \text{ we need to integrate}$$

    We need to integrate to get $\log P(\xi_i < V_j)$.
    If $J=3$, we need to evaluate a bivariate normal —no problem.
    If $J=4$, we need to evaluate a 3-dimensional integral. Possible using Gaussian quadrature —see Butler and Moffit (1982).
    If $J>4$, numerical procedures get complicated and, often, imprecise.
MP Model – Simulation-based Estimation

- We need to integrate to get log P(ξ_j < V_j)

- A simulation can work well, by approximating
  \[ P[y_j=1|X] = P(\xi_j < V_j) \approx \frac{1}{R} \sum_r I[\xi_j^r < V_j] \]
  where we draw ξ_j^r as an i.i.d. N(0, Ω), R times.

This simulator is called frequency simulator. It is unbiased and between [0,1]. But, its derivatives (zero or undefined) complicates calculations.

MP Model – Simulation-based Estimation

- Let’s go over a detailed example of this simple simulator.

Example 1: Binary (0,1) Probit

- Step 1
  - For each observation n=1, ..., N draw η_r^t ~ N(0,1), (r = 1, ......., R: repetitions)
  - Initialize y_count = 0,
  - Set starting values: β=β\_mt
  - Compute y_r^t = x_n^t β_m^t + L η_t^t; L= choleski factor (LL'=Ω)
  - Evaluate: y_r^t >0 ⇒ y_count = y_count+1
  - Repeat R times
Example 1: Binary Probit (continuation)

- Step 2 - Calculate probabilities
  \[ P_n | \beta^m_t = \frac{y_{\text{count}}}{R} \]

- Step 3: Form the simulated LL function
  \[ \text{SLL} = \sum_n y_n \ln(P_n | \beta^m_t) + (1-y_n) \ln(1-P_n | \beta^m_t) \]

- Step 4: Check convergence
  - Criteria: \( \text{SLL}(\beta^m_t) - \text{SLL}(\beta^m_{t-1}) < 0.0001 \)

- Step 5: If no convergence, update parameter - \( \beta^m_t \)
  \[ \beta^m_{t+1} = \beta^m_t + \text{update} \]

- Repeat until convergence.

Example 2: Multivariate Probit

- Draw \( \varepsilon_i \) from a multivariate normal distribution
- Calculate the probability of choice \( j \) as the number of times choice \( j \) corresponded to the highest utility.
- Calculate simulated likelihood.
  (With many choices (\( J > 5 \)) this method does not work well.)

- There are many other simulators, improving over the frequency simulator: smaller variance, smoother, more efficient computations.
One of this simulation methods is the Importance Sampling.

- Consider the integral $E[h(U)] = \int h(u) f(u) \, du$. Suppose it is difficult to draw $U$ from $F$ or $h(.)$ is not smooth. We can always write:

$$E[h(u)] = \int \{h(u) f(u)/g(u)\} g(u) \, du$$

where $g(u)$ is a density with the following properties

a) it is easy to draw $U$ from $g(.)$

b) $g(.)$ & $f(.)$ have the same support.

c) It is easy to evaluate $\{h(u) f(u)/g(u)\}$

d) $\{h(u) f(u)/g(u)\}$ is bounded and smooth over the support of $U$.

Note: $E[h(u)] = E[h(u) f(u)/g(u)]$, where $U \sim g(.)$

The Geweke-Hajivasiliou-Keane (GHK) simulator satisfies (a) to (d).

~

The importance sampling simulator:

$$E[h(U)] = (1/R) \sum_{n} [h(u^n) f(u^n)/g(u^n)],$$

where $u^n$ are $R \ iid.$ draws from $g(.)$.

- Conditions (a) and (c) is to increase computation speed. Condition (d) produces a variance bound and smoothness.

- Condition (d) is the complicated one. For example, if $g(.)$ is a $i.i.d.$ truncated normal may not be bounded if the variance, $\Omega$, has large off-diagonal terms.

The Geweke-Hajivasiliou-Keane (GHK) simulator satisfies (a) to (d).
MP Model – Simulation-based Estimation

- The GHK switches back and forth between
  a) Set initial values for parameters. Set $P^*=1$
  b) Drawing from a simulated truncated normal $\Rightarrow \xi^r$
  c) Compute $\gamma=P(\xi^r < V^r)$ analytically. Reset $P^*=P^* \times \gamma$
  d) Compute (analytically) the distribution of $\xi^r$, conditional on the draws $\Rightarrow$ get values for parameters.
  e) Iterate.

$P^*$ is the GHK simulator, which is bounded (between 0 and 1), continuously differentiable, since $P^*$ is continuous and differentiable and its variance is smaller than the frequency simulator – each draw of the frequency was either zero or 1.

MP Model – Quadrature or Simulation (Greene)

- Computationally, comparably difficult
- Numerically, essentially the same answer. SML is consistent in R
- Advantages of simulation
  - Can integrate over any distribution, not just normal
  - Can integrate over multiple random variables. Quadrature is largely unable to do this.
  - Models based on simulation are being extended in many directions.
  - Simulation based estimator allows estimation of conditional means $\Rightarrow$ essentially the same as Bayesian posterior means
MP Model – Bayesian Estimation

- Bayesian estimation.
  - Drawing from the posterior distribution of $\beta$ and $\Omega$ is straightforward. The key is setting up the vector of unobserved RVs as:
    \[
    \theta = (\beta, \Omega, U_{n1}, U_{n2}, \ldots, U_{nj})
    \]
  and, then, defining the most convenient partition of this vector.

- Given the parameters drawing from the unobserved utilities can be done sequentially: for each unobserved utility given the others we would have to draw from a truncated normal distribution, which is straightforward --see McCulloch, Polson, and Rossi (2000).

MP Model – More on Estimation

- Additional estimation problem: We need to estimate a large number of parameters --all elements in the $(J + 1) \times (J + 1)$ dimensional covariance matrix of latent utilities, minus some that are fixed by normalizations and symmetry restrictions.

  - Difficult with the sample sizes typically available.
Multinomial Choice Models: Probit or Logit?

- There is a trade-off between tractability and flexibility
  - Closed-form expression of the integral for Logit, not for Probit models.
  - Logit has the IIA property. No substitution is allowed.
  - Logit model easy to estimate.
  - Probit allows for random taste variation, can capture any substitution pattern, allows for correlated error terms and unequal error variances.
  - But, the Probit model is complicated to estimate.

⇒ Dependent on the specifics of the choice situation. Is substitution important?

Random Effects Model

- A third possibility to get around the IIA property is to allow for unobserved heterogeneity in the slope coefficients.

- Why do we think that if Houston Grand Opera’s (HGO) prices go up, a person who was planning to go HGO’s would go to Houston Ballet instead, rather than to Lollapalooza?

- We think individuals who have a taste for HGO’s are likely to have a taste for close substitute in terms of observable characteristics, like Houston Ballet. There is individual heterogeneity in the utility functions.

- This effect can be modeled by allowing the utilities to vary with each person, say by making the parameters dependent on \( n \) – i.e., person \( n \).
Random Effects Model

- We allow the marginal utilities to vary at the individual level:
  \[ U_{nj} = X'_{nj} \beta_n + \epsilon_{nj}, \quad \beta_n \sim N(b, \Sigma) \] - like a random effect!

- We can also write this as:
  \[ U_{nj} = X'_{nj} b + \nu_{nj}, \]
  where \( \nu_{nj} = \epsilon_{nj} + X_{nj}(\beta_n - b) \) is no longer independent across choices.

Note: The key ingredient is the vector of individual specific taste parameters \( \beta_n \). We have random taste variation.

- We can assume the existence of a finite number (\( k \)) of types of individuals:
  \[ \beta_n \in \{b_1, b_2, \ldots, b_k\} \]
  with \( \Pr(\beta_n = b_k | W_n) \) given by a logit model => Finite mixture model.

Random Effects Model

- Alternatively, we can assume
  \[ \beta_n | Z_n \sim N(W_n' \gamma, \Omega) \]
  where we use a normal (continuous) mixture of taste parameters.

- Using simulation methods or Gibbs sampling with the unobserved \( \beta_n \) as additional unobserved random variables may be an effective way of doing inference.

Remark: Models with random coefficients can generate more realistic predictions for new choices (predictions will be dependent on presence of similar choices).
Berry-Levinsohn-Pakes Model

- BLP extended the random effects logit models to allow for
  - unobserved product characteristics,
  - endogeneity of choice characteristics,
  - estimation with only aggregate choice data
  - with large numbers of choices.

- Model used in I.O. to model demand for differentiated products.

- The utility is indexed by individual, product and market:
  \[ U_{njt} = X'_{jt} \beta_n + \xi_{jt} + \epsilon_{njt}, \]
  - \( \xi_{jt} \) = unobserved product characteristic, allowed to vary by market, \( t \),
    and by product, \( j \).
  - \( \epsilon_{njt} \) = unobserved component, independent Gumbel, across \( n, j, t \).

Berry-Levinsohn-Pakes Model

- The random coefficients \( \beta_n \) are related to individual observable characteristics:
  \[ \beta_n = \beta + Z_n' \Gamma + \eta_n, \quad \eta_n \sim \text{N}(0, \Omega) \]

- BLP estimate this model without individual level data. It uses market level data (aggregates) in combination with estimators of the distribution of \( Z_n \).

- The data consist of
  - estimated shares \( \tilde{\pi}_{jt} \) for each choice \( j \) in each market \( t \),
  - observations from the marginal distribution of individual characteristics (the \( Z_i \)'s) for each market, often from representative data sets.
Berry-Levinsohn-Pakes Model

- First, write the latent utilities as
  \[ U_{njt} = \delta_{jt} + \nu_{njt} + \varepsilon_{njt}, \]
  with \( \delta_{jt} = X'jt \beta + \xi_{jt}, \) and \( \nu_{njt} = X'jt (Zn' \Gamma + \eta_n). \)

- Second, for fixed \( \Gamma, \Omega, \delta_{jt}, \) calculate the implied market share for product \( j \) in market \( t. \) This can be done analytically or, more general, by simulation.

- Next, we only fix \( \Gamma, \Omega, \) for each value of \( \delta_{jt}, \) find the implied market share. Using aggregate market share data, find \( \delta_{jt} \) such that implied market share equals observed market shares.

- Given \( \delta_{jt}(s, \Gamma, \Omega), \) calculate residuals \( (\xi_{jt}): \delta_{jt} - X'jt \beta = \nu_{jt}. \)

Berry-Levinsohn-Pakes Model

- Then, assume \( \xi_{jt} \) and \( \varepsilon_{njt} \) are uncorrelated with observed characteristics (other than price). We can use GMM or IVE to get \( \beta. \)

- GMM will also give us the standard errors for this procedure.
MP Model – Example 1

Example (Kamakura and Srivastava 1984):
Random utility components $\epsilon_{ni}$, $\epsilon_{nj}$ are more (less) highly correlated when $i$ and $j$ are more (less) similar on important attributes. We need to define a metric for “similar.”

$$r_{ij} = Ke^{-\alpha d_{ij}} \quad (d_{ij} = \text{weighted euclidean distance between } i \& j)$$

$$\Omega = \begin{pmatrix}
Ke^{-\alpha d_{12}} & 1 \\
Ke^{-\alpha d_{13}} & Ke^{-\alpha d_{23}} & 1 \\
... & ... & ... \\
Ke^{-\alpha d_{1j}} & Ke^{-\alpha d_{2j}} & ... & 1
\end{pmatrix}$$

MP Model – Example 1

- Examples
  - Choice models at brand-size level: correlation between ≠ sizes of same brand (Chintagunta 1992)

**TABLE 4**

<table>
<thead>
<tr>
<th></th>
<th>Normalized Error Correlation Matrix</th>
<th>Normalized Error Correlation Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helix 28</td>
<td>0.314</td>
<td>0.319</td>
</tr>
<tr>
<td>Helix 32</td>
<td>-0.139</td>
<td>0.305</td>
</tr>
<tr>
<td>Helix 40</td>
<td>0.401</td>
<td>-0.220</td>
</tr>
<tr>
<td>Helix 64</td>
<td>0.045</td>
<td>0.127</td>
</tr>
<tr>
<td>Helix 72</td>
<td>-0.498</td>
<td>-0.437</td>
</tr>
<tr>
<td>Del Monte 32</td>
<td>-0.315</td>
<td>-0.213</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
MP Model – Example 2

Example: Firm innovation (Harris et al. 2003)

- Binary probit model for innovative status (innovation occurred or not)
- Based on panel data ⇒ correlation of innovative status over time: unobserved heterogeneity related to management ability and/or strategy

Model (2)-(4) account for unobserved heterogeneity (\(\rho\)) -> superior results
MP Model – Example 3

Examples: Dynamics of individual health (Contoyannis, Jones and Nigel 2004)

• Binary probit model for health status (healthy or not)
• Survey data for several years
  - Correlation over time (state dependence)
  - Individual-specific (time-invariant) random coefficient

MP Model – Example 3

Example: Choice of transportation mode (Linardakis and Dellaportas 2003)

⇒ Non-IIA substitution patterns

<table>
<thead>
<tr>
<th>Mode of transportation</th>
<th>Choice (min)</th>
<th>Walking time (min)</th>
<th>In-vehicle time (min)</th>
<th>Search for parking time (min)</th>
<th>Cost (drachmas)</th>
<th>Waiting time (min)</th>
<th>Inconvenience of transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car</td>
<td>2</td>
<td>2</td>
<td>30</td>
<td>5</td>
<td>400</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Metro</td>
<td>1</td>
<td>10</td>
<td>15</td>
<td>0</td>
<td>300</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Bus</td>
<td>3</td>
<td>5</td>
<td>25</td>
<td>0</td>
<td>75</td>
<td>25</td>
<td>1</td>
</tr>
</tbody>
</table>
Ordered Logit Model

• Now, the order matters. There is information (hierarchy) in the order.

Examples: Taste test (1 to 10), credit rating, preference scale (‘dislike very much’ to ‘like very much’), purchase 1, 2 or more units, etc.

• Random preferences: There is an underlying continuous preference scale, which maps to observed choices. The strength of preferences is reflected in the discrete outcome

• Choice between \( J > 2 \) ordered ‘alternatives.’

• Ordinal dependent variable \( y = 1, 2, \ldots, J \), with

\[
\text{rank}(1) < \text{rank}(2) < \ldots < \text{rank}(J)
\]

Ordered Logit Model – Example (Greene)

• Movie ratings from IMDB.com

<table>
<thead>
<tr>
<th>National Treasure: Book of Secrets</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Votes</strong></td>
</tr>
<tr>
<td>Males 33,644</td>
</tr>
<tr>
<td>Females 5,464</td>
</tr>
<tr>
<td>Aged under 18 2,492</td>
</tr>
<tr>
<td>Males under 18 1,795</td>
</tr>
<tr>
<td>Females under 18 956</td>
</tr>
<tr>
<td>Aged 18-22 26,045</td>
</tr>
<tr>
<td>Males Aged 18-22 22,693</td>
</tr>
<tr>
<td>Females Aged 18-22 3,973</td>
</tr>
<tr>
<td>Aged 30-44 8,210</td>
</tr>
<tr>
<td>Males Aged 30-44 7,216</td>
</tr>
<tr>
<td>Females Aged 30-44 830</td>
</tr>
<tr>
<td>Aged 55+ 2,258</td>
</tr>
<tr>
<td>Males Aged 55+ 1,814</td>
</tr>
<tr>
<td>Females Aged 55+ 409</td>
</tr>
<tr>
<td>IMDB staff 8</td>
</tr>
<tr>
<td>Top 1000 users 309</td>
</tr>
<tr>
<td>US users 14,792</td>
</tr>
<tr>
<td>Non-US users 24,283</td>
</tr>
<tr>
<td>IMDb users 41,771</td>
</tr>
</tbody>
</table>
Ordered Logit Model

• We follow McFadden’s approach.
  - Suppose $y_i^*$ is a continuous latent variable which is a linear function of the explanatory variables:
    \[ y_n^* = V_n + \varepsilon_n = X_n \beta + \varepsilon_n \quad (y_n^* = \text{latent utility}) \]
  - Preferences can be ‘mapped’ on an ordered multinomial variable as follows:
    \[
    \begin{align*}
    y_n = 1 & \quad \text{if} \quad \alpha_0 < y_n^* \leq \alpha_1 \quad \text{(Region 1)} \\
    y_n = j & \quad \text{if} \quad \alpha_{j-1} < y_n^* \leq \alpha_j \quad \text{(Region j)} \\
    y_n = J & \quad \text{if} \quad \alpha_{J-1} < y_n^* \leq \alpha_J \quad \text{(Region J)}
    \end{align*}
    \]
  - The $\alpha_0$‘s are called thresholds.

Ordered Logit Model

• We observe outcome $j$ if utility is in region $j$
  
  Probability of outcome $= \text{probability of cell}$
  \[
  P (Y_n = j \mid X_n ) = P (\alpha_{j-1} < y_n^* \leq \alpha_j )
  \]
  \[
  = P (\alpha_{j-1} - (X_n \beta) < \varepsilon_n \leq \alpha_j - (X_n \beta) )
  \]
  \[
  = F (\alpha_j - (X_n \beta)) - F (\alpha_{j-1} - (X_n \beta))
  \]

• To continue we need a probability model. We use the logit distribution $\Rightarrow$ Ordered logit model:
  \[
  F (\alpha_j - X_n \beta) = \frac{\exp(\alpha_j - X_n \beta)}{1 + \exp(\alpha_j - X_n \beta)}
  \]
  
  • In general, $\alpha_0$ is set equal to zero and $\alpha_1$ a large number ($+\infty$) (also, $\alpha_{-1} = -\infty$). Different normalizations affect the estimation of constant.
Ordered Logit Model – Parallel Regressions

• Let’s look back at the construction of regions:

\[ y_n = 1 \quad \text{if} \quad \alpha_0 < y_n^{*} = X_n \beta + \varepsilon_n \leq \alpha_1 \quad \text{(Region 1)} \]

\[ y_n = j \quad \text{if} \quad \alpha_{j-1} < y_n^{*} = X_n \beta + \varepsilon_n \leq \alpha_j \quad \text{(Region j)} \]

\[ y_n = J \quad \text{if} \quad \alpha_{J-1} < y_n^{*} = X_n \beta + \varepsilon_n \leq \alpha_J \quad \text{(Region J)} \]

• The \( \beta \)'s are the same for each region (choice). That is, the coefficients that describe the relationship between, say, the lowest versus all higher categories of the response variable are the same as those that describe the relationship between the next lowest category and all higher categories, etc.

• This is called the proportional odds assumption or the parallel regression assumption. It simplifies the estimation. It may not be realistic.

Ordered Logit Model – Example (Greene)

![Graph showing probabilities for estimated ordered probit model.](image)
Generalized Ordered Logit Model

- We can generalized the model:
  \[ y_n = 1 \quad \text{if} \quad \alpha_0 < y_n^* = X_n \beta_1 + \varepsilon_n \leq \alpha_1 \quad \text{(Region 1)} \]
  \[ y_n = j \quad \text{if} \quad \alpha_{j-1} < y_n^* = X_n \beta_2 + \varepsilon_n \leq \alpha_j \quad \text{(Region j)} \]
  \[ y_n = J \quad \text{if} \quad \alpha_{J-1} < y_n^* = X_n \beta_J + \varepsilon_n \leq \alpha_J \quad \text{(Region J)} \]

- The \( \beta \)'s are different for each region (choice). This model is called Generalized Ordered Choice Model. To make it a generalized ordered logit model, we need to assume the Gumbel distribution for the \( \varepsilon_n \)'s.

- We can be more general by making the thresholds heterogeneous:
  \[ \alpha_{nj} = \theta_j + Z_{nj} \delta_j \]
  This can create identification problems, if \( z_{nk} \) is also in \( x_n \) (same variable). Difficult to disentangle effects: \( F(\alpha_{nj} - X_n \beta_j = \theta_j + Z_{nj} \delta_j - X_n \beta_j) \)

Generalized Ordered Logit Model

- We can also use non-linear functions to model thresholds heterogeneity:
  \[ \alpha_{nj} = \exp(\theta_j + Z_{nj} \delta_j) \]
  It will be easier to identify effects in the Generalized Ordered Choice Model.

- An internally consistent restricted modification of the model is:
  \[ \alpha_{nj} = \exp(\theta_j + Z_{nj} \delta_j) \]
  where
  \[ \theta_j = \theta_{j-1} + \exp(\varphi_j) \]
  This model is called Hierarchical Order Probit (HOPit). See Harris and Zhao (2000).
Ordered Logit Model - Estimation

- Given the logit distribution, ML is simple.

\[
L(\theta) = \prod_{n=1}^{N} \prod_{j=1}^{J} P(Y_n = j \mid X_n)^{(y_n=j)} = \prod_{n=1}^{N} \prod_{j=1}^{J} \left( F(\alpha_j - X_n \beta) - F(\alpha_{j-1} - X_n \beta) \right)^{(y_n=j)}
\]

\[
LogL(\theta) = \sum_{n=1}^{N} \sum_{j=1}^{J} I[y_n = j] \ln \left( F(\alpha_j - X_n \beta) - F(\alpha_{j-1} - X_n \beta) \right)
\]

- The \(\beta\)'s are the same for each choice. This is the parallel regression assumption. It is a restriction on the model. It simplifies the estimation.

- This restriction can be tested (LR or Wald tests easy to construct).

Ordered Probit Model

- We can also use the normal distribution as the probability model. In this case, the probability of cell \(j\):

\[
P(Y_n = j \mid X_n) = \Phi(\alpha_j - (X_n \beta)) - \Phi(\alpha_{j-1} - (X_n \beta))
\]

This is the **Ordered Probit Model**.

- As before, we require a normalization: either no constant or \(\alpha_0=0\).

- Estimation: Maximum likelihood

\[
LogL(\theta) = \sum_{n=1}^{N} \sum_{j=1}^{J} I[y_n = j] \ln \left( \Phi(\alpha_j - X_n \beta) - \Phi(\alpha_{j-1} - X_n \beta) \right)
\]
Ordered Logit Model – Estimation (Greene)

Example: Model for Health Satisfaction

<table>
<thead>
<tr>
<th>Ordered Probability Model</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent variable</td>
<td>HSAT</td>
</tr>
<tr>
<td>Number of observations</td>
<td>27326</td>
</tr>
<tr>
<td>Underlying probabilities based on Normal</td>
<td></td>
</tr>
<tr>
<td>Cell frequencies for outcomes</td>
<td></td>
</tr>
<tr>
<td>Y Count</td>
<td>Freq</td>
</tr>
<tr>
<td>0</td>
<td>447</td>
</tr>
<tr>
<td>3</td>
<td>1173</td>
</tr>
<tr>
<td>6</td>
<td>2530</td>
</tr>
<tr>
<td>9</td>
<td>3061</td>
</tr>
</tbody>
</table>

| Variable | Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X |
|----------|--------------|----------------|----------|---------|-----------|
| Index function for probability |
| Constant | 2.61335825 | .04658496 | 56.099 | .0000 |
| FEMALE | -.05840486 | .01259442 | -4.637 | .0000 | .47877479 |
| EDUC | .03390552 | .00284332 | 11.925 | .0000 | 11.3206310 |
| AGE | -.01997327 | .00059487 | -33.576 | .0000 | 43.5256898 |
| HHNINC | .25914964 | .03631951 | 7.135 | .0000 | .35208362 |
| HHKIDS | .06314906 | .01350176 | 4.677 | .0000 | .40273000 |

Threshold parameters for index

| Mu(1) | .19352076 | .01002714 | 19.300 | .0000 |
| Mu(2) | .49955053 | .01087525 | 45.935 | .0000 |
| Mu(3) | .83593441 | .00990420 | 84.402 | .0000 |
| Mu(4) | 1.10524187 | .00908506 | 121.655 | .0000 |
| Mu(5) | 1.66256620 | .00801113 | 207.532 | .0000 |
| Mu(6) | 1.92729096 | .00774122 | 248.965 | .0000 |
| Mu(7) | 2.33879408 | .00777041 | 300.987 | .0000 |
| Mu(8) | 2.99432165 | .00851090 | 351.822 | .0000 |
| Mu(9) | 3.45366015 | .01017554 | 339.408 | .0000 |

Ordered Logit Model – Partial Effects

• As usual, there is a non-linearity. The β’s do not have the usual interpretation. We will look at partial effects:

\[
\frac{\partial P(Y_n = j)}{\partial x_{nk}} = \left[ f(\alpha_j - X_n \beta) - f(\alpha_{j-1} - X_n \beta) \right] \beta_k
\]

• The partial effects depend on the data (X) and the coefficients. The sign depends on the densities evaluated at two points.
Assume the $\beta_k$ is positive.
Assume that $x_k$ increases.
$\beta'x$ increases. $\alpha_j - \beta'x$ shifts to the left for all 5 cells.
$\text{Prob}[y=0]$ decreases
$\text{Prob}[y=1]$ decreases – the mass shifted out is larger than the mass shifted in.
$\text{Prob}[y=3]$ increases – same reason in reverse.
$\text{Prob}[y=4]$ must increase.

When $\beta_k > 0$, increase in $x_k$ decreases $\text{Prob}[y=0]$ and increases $\text{Prob}[y=J]$.
Intermediate cells are ambiguous, but there is only one sign change in the marginal effects from 0 to 1 to … to J.
Ordered Probability Effects – Example (Greene)

<table>
<thead>
<tr>
<th>Marginal effects for ordered probability model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Names for dummy variables are Pr[y=1]-Pr[y=0]</td>
</tr>
</tbody>
</table>
| Names for dummy variables are marked by *.

| Variable | Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X |
|----------|-------------|----------------|---------|--------|----------|
| FEMALE   | .00200414   | .00043473      | 4.610   | .0000  | .4787479 |
| EDUC     | -.00115962  | .986135D-04    | -11.759 | .0000  | 11.32063 |
| AGE      | .0086328    | .00124869      | 6.998   | .0000  | 11.32063 |
| HHNINC   | -.00002393  | .00010451      | -3.098  | .0000  | .4027300 |
| HHKIDS   | -.00136193  | .00051119      | -4.765  | .0000  | .4027300 |

These are the effects on Prob[Y=00] at means.

| Variable | Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X |
|----------|-------------|----------------|---------|--------|----------|
| FEMALE   | .00101533   | .00021973      | 4.621   | .0000  | .4787479 |
| EDUC     | -.00058810  | .496973D-04    | -11.834 | .0000  | 11.32063 |
| AGE      | .00034644   | .108937D-04    | 31.802  | .0000  | 43.52836 |
| HHNINC   | -.00049505  | .00063180      | -7.115  | .0000  | .3520836 |
| HHKIDS   | -.00108460  | .00022994      | -4.717  | .0000  | .4027300 |

These are the effects on Prob[Y=01] at means.

These are the effects on Prob[Y=10] at means.

Ordered Probit Marginal Effects (Greene)

<table>
<thead>
<tr>
<th>Summary of Marginal Effects for Ordered Probability Model (probit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>FEMALE</td>
</tr>
<tr>
<td>EDUC</td>
</tr>
<tr>
<td>AGE</td>
</tr>
<tr>
<td>HHNINC</td>
</tr>
<tr>
<td>HHKIDS</td>
</tr>
</tbody>
</table>
### OP: The Single Crossing Effect (Greene)

The marginal effect for EDUC is negative for Prob(0),…,Prob(7), then positive for Prob(8)…Prob(10).
One “crossing.”
Ordered Probit Model: Nonlinearity (Greene)

Ordered Probit Model: Model Evaluation

- Different ways to judge a model:
  - Partial Effects (do they make sense?)
  - Fit Measures (Log Likelihood based measures, such as pseudo-R²)
  - Predicted Probabilities
    - Averaged: They match sample proportions.
    - By observation
    - Segments of the sample
    - Related to particular variables
Ordered Probit Model: Model Evaluation

• Log Likelihood Based Fit Measures

\[ R_{\text{Pseudo}}^2 = 1 - \frac{\log \hat{L}_{\text{Model}}}{\log \hat{L}_{\text{null}}}. \]

A degrees of freedom adjusted version is sometimes reported.

\[ \text{Adjusted } R_{\text{Pseudo}}^2 = 1 - \frac{\log \hat{L}_{\text{null}} - M}{\log \hat{L}_{\text{Model}}}. \]

Log Akaike Information Criterion \( = AIC \)
Finite Sample AIC \( = AIC_{CF} \)
Bayes Information Criterion \( = BIC \)
Hannan-Quinn IC \( = HQIC \)

\[
\begin{align*}
\text{Log Akaike Information Criterion} & = AIC = (-2 \log L + 2M)/n, \\
\text{Finite Sample AIC} & = AIC_{CF} = (AIC + 2M(n+1))(n-M-1), \\
\text{Bayes Information Criterion} & = BIC = (-2 \log L + M \log n)/n, \\
\text{Hannan-Quinn IC} & = HQIC = (-2 \log L + 2M \log \log n)/n.
\end{align*}
\]

OP Model: Model Evaluation

• Predictions of the Model: Kids

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>P0</td>
<td>0.059586</td>
<td>0.028182</td>
<td>0.009561</td>
<td>0.125545</td>
</tr>
<tr>
<td>P1</td>
<td>0.268398</td>
<td>0.067415</td>
<td>0.106526</td>
<td>0.374712</td>
</tr>
<tr>
<td>P2</td>
<td>0.489603</td>
<td>0.024370</td>
<td>0.419003</td>
<td>0.515906</td>
</tr>
<tr>
<td>P3</td>
<td>0.101163</td>
<td>0.030157</td>
<td>0.052589</td>
<td>0.181065</td>
</tr>
<tr>
<td>P4</td>
<td>0.081250</td>
<td>0.041250</td>
<td>0.028152</td>
<td>0.237842</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stratum is KIDS = 0.000. Nobs. = 2782.000</th>
</tr>
</thead>
<tbody>
<tr>
<td>P0</td>
</tr>
<tr>
<td>P1</td>
</tr>
<tr>
<td>P2</td>
</tr>
<tr>
<td>P3</td>
</tr>
<tr>
<td>P4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stratum is KIDS = 1.000. Nobs. = 1701.000</th>
</tr>
</thead>
<tbody>
<tr>
<td>P0</td>
</tr>
<tr>
<td>P1</td>
</tr>
<tr>
<td>P2</td>
</tr>
<tr>
<td>P3</td>
</tr>
<tr>
<td>P4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>All 4483 observations in current sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>P0</td>
</tr>
<tr>
<td>P1</td>
</tr>
<tr>
<td>P2</td>
</tr>
<tr>
<td>P3</td>
</tr>
<tr>
<td>P4</td>
</tr>
</tbody>
</table>
(**Aggregate Prediction Measure**)

An alternative approach to measuring fit is to compute the sums of the predicted probabilities in the various cells.

\[
H' = \sum_{i=1}^{n} \begin{bmatrix} I(y_i = 0) \\ I(y_i = 1) \\ \vdots \\ I(y_i = J) \\ 1 \end{bmatrix} \begin{bmatrix} \hat{p}_i(0) \\ \hat{p}_i(1) \\ \vdots \\ \hat{p}_i(J) \\ 1 \end{bmatrix}.
\]

**Predictions for the Health Satisfaction Model**

<table>
<thead>
<tr>
<th>y(i, j)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>16</td>
<td>66</td>
<td>111</td>
<td>21</td>
<td>16</td>
<td>230</td>
</tr>
<tr>
<td>1</td>
<td>43</td>
<td>294</td>
<td>549</td>
<td>115</td>
<td>113</td>
<td>2266</td>
</tr>
<tr>
<td>2</td>
<td>118</td>
<td>547</td>
<td>1110</td>
<td>249</td>
<td>210</td>
<td>2226</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>111</td>
<td>252</td>
<td>62</td>
<td>55</td>
<td>500</td>
</tr>
<tr>
<td>4</td>
<td>19</td>
<td>78</td>
<td>207</td>
<td>46</td>
<td>42</td>
<td>414</td>
</tr>
<tr>
<td>Total</td>
<td>228</td>
<td>1117</td>
<td>2229</td>
<td>494</td>
<td>415</td>
<td>4483</td>
</tr>
</tbody>
</table>
Ordered Logit Model - Cons

- Disadvantages (Borooah 2002)
- Assumption of equal slope $\beta_k$
- Biased estimates if assumption of strictly ordered outcomes does not hold
  => treat outcomes as nonordered unless there are good reasons for imposing a ranking.

Ordered Logit Model - Application

Example (from Kim and Kim (2004): Effectiveness of better public transit as a way to reduce automobile congestion and air pollution in urban areas
- Research objective: develop and estimate models to measure how public transit affects automobile ownership and miles driven.
- Data: Nationwide Personal Transportation Survey (42,033 hh): socio-demo's, automobile ownership and use, public transportation avail.
Ordered Logit Model - Application

- Dependent variable ownership model = number of cars \(k = 0, 1, 2, \geq 3\) \(\rightarrow\) ordinal variable
- \(C_i\) = latent variable: automobile ownership propensity of hh \(i\)
- Relation to observed automobile ownership:
  \[ C_i = k \text{ if } \alpha_{k-1} < \beta'x_i + \varepsilon < \alpha_k \]
  \[ P(C_i = k) = F(\alpha_k - \beta'x_i) - F(\alpha_{k-1} - \beta'x_i) \]

Ordered Logit Model - Application

| Variable          | Coefficient | z-value | \(P > |z|\) | Coefficient | z-value | \(P > |z|\) | Coefficient | z-value | \(P > |z|\) |
|-------------------|-------------|---------|-----------|-------------|---------|-----------|-------------|---------|-----------|
| Zone distance     | -0.5985     | -30.00  | 0         | -0.4213     | -20.00  | 0         | -0.2292     | -15.75  | 0         |
| No. drivers       | 2.1865      | 92.07   | 0         | 2.4529      | 75.60   | 0         | 2.3376      | 70.60   | 0         |
| Income (log)      | 0.7508      | 42.11   | 0         | 0.7118      | 25.13   | 0         | 0.6810      | 23.38   | 0         |
| HH income (log)   | 0.6214      | 12.01   | 0         | 0.0742      | 7.74    | 0         | 0.0416      | 9.35    | 0         |
| No. vehicles      | 0.1963      | 9.77    | 0         | 0.0012      | 2.58    | 0.01      | 0.2292      | 10.55   | 0         |
| Li, Perth        | -0.2876     | -18.17  | 0         | -0.3043     | -18.76  | 0         | -0.2597     | -14.76  | 0         |
| Li, Chicago      | -0.3776     | -53.33  | 0         | -0.2562     | -18.87  | 0.01      | -0.4056     | -15.10  | 0         |
| Li, Other cities | 0.1570      | 3.86    | 0         | 0.0425      | 0.37    | 0.01      | 0.2065      | 3.98    | 0         |
| Li, Other cities | 0.1616      | 3.53    | 0         | 0.1029      | 1.99    | 0.05      | 0.1599      | 0.33    | 0         |
| Delin              | -0.0359     | -1.23   | 0.22      | 0.4008      | 3.12    | 0         | -0.2039     | -1.74   | 0.09      |
| Home              | -0.3700     | -9.06   | 0.35      | 0.3712      | 7.31    | 0         | -0.1483     | -28.11  | 0         |
| Los Angeles       | -0.4039     | -7.53   | 0         | -0.1098     | -1.16   | 0.25      | -0.3042     | -19.87  | 0         |
| Philadelphia      | -0.7388     | -7.75   | 0         | -0.1086     | -1.16   | 0.25      | -0.2318     | -6.14   | 0         |
| Washington        | -0.4685     | -8.61   | 0         | 0.2144      | 2.62    | 0.05      | -0.1642     | -19.11  | 0         |
| Atlanta           | -0.2842     | -2.17   | 0.03      | 0.2841      | 2.15    | 0.03      | -0.2039     | -1.16   | 0.25      |
| Boston            | -0.0591     | -1.11   | 0         | -0.2782     | -10.53  | 0         | -0.2318     | -1.16   | 0         |
| OLMMA              | -0.0579     | -1.76   | 0.08      | -0.1098     | -1.16   | 0.25      | -0.3042     | -19.87  | 0         |
| OLMMA              | -0.2318     | -6.14   | 0         | -0.1098     | -1.16   | 0.25      | -0.3042     | -19.87  | 0         |

Table 2. Automobile ownership models estimation: ARA, NMSA, and OTHA

(1) 6.39          7.28          6.38
(2) 10.36         10.76         10.38
(3) 12.96         14.48         13.92
### Ordered Logit Model - Application

#### Table 2. Ordered Logit Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1 (ARA)</th>
<th>Model 2 (NSMA)</th>
<th>Model 3 (OTSA)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>p-value</td>
<td>Coefficient</td>
</tr>
<tr>
<td>Predicted patterns of ownership</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of Obs.</td>
<td>40814</td>
<td>13779</td>
<td>26345</td>
</tr>
<tr>
<td>Restricted LL</td>
<td>-4364.8</td>
<td>-17202.9</td>
<td>-9069.7</td>
</tr>
<tr>
<td>Predicted LL</td>
<td>-3554.5</td>
<td>-10823.4</td>
<td>-28881.0</td>
</tr>
<tr>
<td>$\chi^2$ values (df, F)</td>
<td>35997 (20)</td>
<td>12779 (17)</td>
<td>20031 (11)</td>
</tr>
<tr>
<td>Predicted R²</td>
<td>0.33</td>
<td>0.37</td>
<td>0.33</td>
</tr>
</tbody>
</table>

#### Table 3. Marginal effects on automobile ownership level (model 1)

<table>
<thead>
<tr>
<th>Automobile ownership level</th>
<th>P(C = 0)</th>
<th>P(C = 1)</th>
<th>P(C = 2)</th>
<th>P(C = 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bus distance (sqmi)</td>
<td>0.0045</td>
<td>0.0766</td>
<td>-0.0659</td>
<td>-0.0152</td>
</tr>
<tr>
<td>No. drivers</td>
<td>-0.0308</td>
<td>-0.5213</td>
<td>0.4485</td>
<td>0.107</td>
</tr>
<tr>
<td>Income (log)</td>
<td>-0.0090</td>
<td>-0.1522</td>
<td>0.1309</td>
<td>0.0303</td>
</tr>
<tr>
<td>HH size (log)</td>
<td>-0.0080</td>
<td>-0.1346</td>
<td>0.1158</td>
<td>0.0268</td>
</tr>
<tr>
<td>No. workers</td>
<td>-0.0025</td>
<td>-0.0414</td>
<td>0.0026</td>
<td>0.0092</td>
</tr>
<tr>
<td>Lil_cyc1</td>
<td>0.0042</td>
<td>0.0651</td>
<td>-0.0578</td>
<td>-0.0116</td>
</tr>
<tr>
<td>Lil_cyc2</td>
<td>0.0048</td>
<td>0.0716</td>
<td>-0.0645</td>
<td>-0.0122</td>
</tr>
<tr>
<td>Lil_cyc3</td>
<td>-0.0019</td>
<td>-0.0037</td>
<td>0.0026</td>
<td>0.0070</td>
</tr>
<tr>
<td>Lil_cyc4</td>
<td>-0.0013</td>
<td>-0.0218</td>
<td>0.0187</td>
<td>0.0044</td>
</tr>
<tr>
<td>Chicago</td>
<td>0.0083</td>
<td>0.1132</td>
<td>-0.1038</td>
<td>-0.0176</td>
</tr>
<tr>
<td>Dallas</td>
<td>0.0024</td>
<td>0.0577</td>
<td>-0.0332</td>
<td>-0.0068</td>
</tr>
<tr>
<td>Houston</td>
<td>0.0063</td>
<td>0.0092</td>
<td>-0.0819</td>
<td>-0.0166</td>
</tr>
<tr>
<td>Los Angeles</td>
<td>0.0017</td>
<td>0.0276</td>
<td>-0.0242</td>
<td>-0.0051</td>
</tr>
<tr>
<td>New York</td>
<td>0.0245</td>
<td>0.2614</td>
<td>-0.2499</td>
<td>-0.0310</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>0.0143</td>
<td>0.1715</td>
<td>-0.1619</td>
<td>-0.0240</td>
</tr>
<tr>
<td>Washington</td>
<td>0.0071</td>
<td>0.1005</td>
<td>-0.0916</td>
<td>-0.0160</td>
</tr>
<tr>
<td>Atlanta</td>
<td>0.0047</td>
<td>0.0702</td>
<td>-0.0631</td>
<td>-0.0119</td>
</tr>
<tr>
<td>Boston</td>
<td>0.0101</td>
<td>0.1387</td>
<td>-0.1270</td>
<td>-0.0218</td>
</tr>
<tr>
<td>OLMSA</td>
<td>0.0014</td>
<td>0.0224</td>
<td>-0.0194</td>
<td>-0.0043</td>
</tr>
<tr>
<td>OSMSA</td>
<td>0.0035</td>
<td>0.0554</td>
<td>-0.0487</td>
<td>-0.0101</td>
</tr>
</tbody>
</table>

Note: OLMSA, other large metropolitan statistical areas; OSMSA, other small.
Ordered Logit Model – More Applications

- Examples
- Occupational outcome as a function of socio-demographic characteristics – Borooah (2002)
  - Unskilled/semiskilled
  - Skilled manual/non-manual
  - Professional/managerial/technical
  - Grade 1 to 5
  - Function of school, teacher and student characteristics
- Level of insurance coverage

Brant Test for Parallel Regressions (Greene)

- We can test the parallel regression assumption: all $\beta$'s are the same across regions. The alternative hypothesis is the Generalized Ordered Logit Model.
- This specification test or test for parameter constancy (across regions) is called the Brant Test:

Reformulate the J "models"

\[ \text{Prob}[y > j | x] = F(\alpha + \beta'x - \mu_j), j = 0, 1, ..., J - 1 \]

\[ = F(\alpha_j + \beta'x) \quad (\alpha_j = \alpha - \mu_j) \]

Produces J binary choice models based on $y > j$.

$H_0$: The slope vector is the same in all "models."

(This is implied by the ordered choice model.)

Test: Estimate J binary choice models. Use a Wald test to test $H_0: \beta_0 = \beta_1 = ... \beta_{J-1}$
Q: What failure of the model specification is indicated by rejection?

### Brant Test for Parallel Regressions (Greene)

**Brant Test for Parameter Homogeneity**

| Brant specification test for equal coefficient | Vector in the ordered logit model. The model implies that \( \logit(\text{Prob}(y > j) | x) = \beta(j) + \beta_0 x \) for all \( j = 0, \ldots, J \). The chi squared test is \( H_0: \beta(j) = \beta(1) = \ldots = \beta(J) \).
| Chi squared test statistic = 71.76435 | (78.76988 based on the normal distribution) |
| Degrees of freedom = 15 |
| P value = 0.0000 |

**Specification Tests for Individual Coefficients in Ordered Logit Model**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Brant Test</th>
<th>Coefficients in implied model Prob(y &gt; j)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE</td>
<td>6.20</td>
<td>0.0598, 0.0222, 0.0123, 0.0090</td>
</tr>
<tr>
<td>EDUC</td>
<td>19.89</td>
<td>0.00008, 0.0423, 0.06786, 0.06303</td>
</tr>
<tr>
<td>INCOME</td>
<td>55.32</td>
<td>0.000299, 1.9576, 0.4959, 0.2790</td>
</tr>
<tr>
<td>MARRIED</td>
<td>1.87</td>
<td>0.59862, 0.6674, 0.6228, 0.3466</td>
</tr>
<tr>
<td>KIDS</td>
<td>7.24</td>
<td>0.06476, 0.0218, 0.2158, 0.0189</td>
</tr>
</tbody>
</table>

**Heterogeneity in Ordered Choice Models**

- **Observed heterogeneity**
  - Easy case, heteroscedasticity, which produces scale heterogeneity.

- **Unobserved heterogeneity**
  - Over decision makers
    - Random coefficients Models
    - E.g. Mixed Logit Model (see Train)
  - Over segments
    - Latent class Models
Heteroscedasticity in OC Models (Greene)

- Not difficult to introduce heteroscedasticity in the OC Models. It produces scale changes: a GLS-type correction.

- As usual, we need a model for heteroscedasticity. For example, exponential form: \( \exp(\gamma h_i) \). Then, for the Probit and Logit Models:

\[
\text{Prob}(y_i = j | x_i, h_i) = \Phi\left( \frac{\mu_j - \beta' x_i}{\exp(\gamma h_i)} \right) - \Phi\left( \frac{\mu_{j-1} - \beta' x_i}{\exp(\gamma h_i)} \right)
\]

\[
\text{Prob}(y_i = j | x_i, h_i) = \Phi\left( \frac{\exp(\theta_j + \delta' z_i) - \beta' x_i}{\exp(\gamma h_i)} \right) - \Phi\left( \frac{\exp(\theta_{j-1} + \delta' z_i) - \beta' x_i}{\exp(\gamma h_i)} \right)
\]

- As usual, partial effects will also be affected.
Heteroscedasticity in OC Models (Greene)

Partial Effects in Heteroscedastic Ordered Probit Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>HEALTH=0</th>
<th>HEALTH=1</th>
<th>HEALTH=2</th>
<th>HEALTH=3</th>
<th>HEALTH=4</th>
<th>Mean</th>
<th>Variance</th>
<th>Restricted</th>
</tr>
</thead>
<tbody>
<tr>
<td>AG2</td>
<td>.00169</td>
<td>.00463</td>
<td>-.00128</td>
<td>-.00216</td>
<td>-.00288</td>
<td>Mean</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AG2</td>
<td>.00618</td>
<td>.00103</td>
<td>-.01647</td>
<td>.00086</td>
<td>.00839</td>
<td>Variance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>.00761</td>
<td>.00566</td>
<td>-.01775</td>
<td>.00130</td>
<td>.00551</td>
<td>Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Age)</td>
<td>(.0017)</td>
<td>(.0045)</td>
<td>(.0012)</td>
<td>(.0022)</td>
<td>(.0028)</td>
<td>Restricted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDUC</td>
<td>-.00322</td>
<td>-.00906</td>
<td>.00251</td>
<td>.00423</td>
<td>.00564</td>
<td>Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(EDUC)</td>
<td>(-.0024)</td>
<td>(-.0099)</td>
<td>(.0024)</td>
<td>(.0042)</td>
<td>(.0056)</td>
<td>Restricted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INCOME</td>
<td>-.02122</td>
<td>-.06500</td>
<td>.01607</td>
<td>.02704</td>
<td>.05611</td>
<td>Mean</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INCOME</td>
<td>-.24756</td>
<td>.07705</td>
<td>-.95201</td>
<td>.04529</td>
<td>.73186</td>
<td>Variance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DICOME</td>
<td>3.2610</td>
<td>-1.0015</td>
<td>-.90874</td>
<td>.67562</td>
<td>.50737</td>
<td>Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(DICOME)</td>
<td>(-.0248)</td>
<td>(-.0614)</td>
<td>(.0177)</td>
<td>(.0409)</td>
<td>(.0406)</td>
<td>Restricted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MARRIED</td>
<td>.00260</td>
<td>.00709</td>
<td>-.00197</td>
<td>-.00331</td>
<td>-.00442</td>
<td>Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(MARRIED)</td>
<td>(.0029)</td>
<td>(.0077)</td>
<td>(.0020)</td>
<td>(.0037)</td>
<td>(.0049)</td>
<td>Restricted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XIDS</td>
<td>-.00593</td>
<td>-.01620</td>
<td>.00449</td>
<td>.09755</td>
<td>.10085</td>
<td>Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(XIDS)</td>
<td>(-.0057)</td>
<td>(-.0151)</td>
<td>(.0040)</td>
<td>(.0072)</td>
<td>(.0096)</td>
<td>Restricted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pure Variance Effect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEMALE</td>
<td>-.00316</td>
<td>-.00553</td>
<td>.00840</td>
<td>-.00044</td>
<td>-.00428</td>
<td>Total</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: Latent Class Models

• Assumption: Consumers can be placed into a small number of (homogeneous) segments, which differ in choice behavior (different response parameters –i.e., the β’s).

• Relative size of the segment, s (s=1, 2, ..., M), is given by

\[ f_s = \frac{\exp(\lambda_s)}{\sum_s \exp(\lambda_s)} \]

• Probability of choosing brand \( j \), conditional on consumer \( n \) being a member of segment \( s \) is given by a logit:

\[ P(y_n = j | X_n) = \frac{\exp(X_{jn} \beta)}{\sum_j \exp(X_{jn} \beta)} \]

• Unconditional probability that consumer \( n \) will choose brand \( j \)

\[ P(y_n = j | X_n) = \sum_s f_s P(y_n = j | X_n) = \sum_s [\exp(\lambda_s) / \sum_s \exp(\lambda_s)] [\exp(X_{jn} \beta) / \sum_j \exp(X_{jn} \beta)] \]
Heterogeneity: Latent Class Models

• Estimation: Maximum Likelihood

• Likelihood of a household’s choice history $H_n$
  \[ L(H_n) = \sum_s \left[ \exp(\lambda_s) L(H_n|s) / \sum_{s'} \exp(\lambda_{s'}) \right] \]
  with
  \[ L(H_n|s) = \prod_t P(y_{nt} = c(t) | X_{nt}) \]
  \[ c(t) = \text{index of the chosen option at time } t. \]

• Maximize likelihood over all household’s: $\prod_n L(H_n)$

• We need to decide on how to form the segments (classes).

Heterogeneity: Latent Class Models

Segment analysis
• Based on parameter estimates, say, difference in price sensitivity.
• Based on segment profiles
  – Post-hoc: based on assignment of consumers to segments;
    Probability that consumer $n$ belongs to segment $s =
    \[ P(n \in s | H_n) = \frac{L(H_n|s)f_s}{\sum_s [L(H_n|s')f_s]} \]
    Analyze characteristics of different segments
  – A priori: make $f_s$ a function of variables that may explain
    segment membership. For example, income for segments
    which differ in price sensitivity.
Heterogeneity: Latent Class Models (Greene)

<table>
<thead>
<tr>
<th>Variable Estimate</th>
<th>Std. Error</th>
<th>z Stat.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1.235</td>
<td>0.123</td>
<td>10.05</td>
</tr>
<tr>
<td>X1</td>
<td>0.123</td>
<td>0.012</td>
<td>10.05</td>
</tr>
<tr>
<td>X2</td>
<td>0.234</td>
<td>0.023</td>
<td>10.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters for Latent Class 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
</tr>
<tr>
<td>X1</td>
</tr>
<tr>
<td>X2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters for Latent Class 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
</tr>
<tr>
<td>X1</td>
</tr>
<tr>
<td>X2</td>
</tr>
</tbody>
</table>

Heterogeneity: Latent class est.

Choice segments: segment 1 = more sensitive to price and promo
Incidence segments: segment 2 and 4 = more sensitive to changes in category attractiveness (change in price/promo)
⇒ Confirms that ≠ combinations of choice/inc.price sensitivity

Heterogeneity: Latent class est.

• Segment analysis: price elasticity
Heterogeneity: Latent class est.

- Segment analysis: socio-demographic profile

![Table 4: POSTERIOR ANALYSIS](image)