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

1

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 ε_{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, ε_{nj}.

• Solutions to IIA
– Nested Logit Model, allowing correlation between some choices.
– Models allowing correlation among the ε_n'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 ε_{nj}'s follow an independent standard normal distributions for all n, j.

$$U_{nj} = \boldsymbol{x}_{nj}{}' \beta + \varepsilon_{nj}, \qquad \varepsilon_{nj} \sim \mathrm{N}(0, \boldsymbol{\Omega})$$

• We retain independence across subjects but we allow dependence across alternatives, assuming that the vector $\varepsilon_n = (\varepsilon_{n1}, \varepsilon_{n2}, \ldots, \varepsilon_{nJ})$ follows a *multivariate normal* distribution, but with arbitrary covariance matrix $\boldsymbol{\Omega}$.

Multinomial Probit Model

• The vector ε_n = (ε_{n1}, ε_{n2}, …, ε_{nJ}) follows a *multivariate normal* distribution, but with arbitrary covariance matrix **Ω**.

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

• Some restrictions (normalization) on **Ω** 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 ε_n.

• 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:

$P_{nj} = P[y_j = 1 | \boldsymbol{x}_n] =$

$\int_{\xi_{n1}}^{\infty} \dots \int_{\xi_{nJ}}^{\infty} \phi\left(\xi_{n1}, \dots, \xi_{nj-1}, \xi_{nj+1}, \dots, \xi_{nJ}\right)\, d\xi_{n1} \dots d\xi_{nj-1}\, d\xi_{nj+1} \dots\, d\xi_{nJ}$

$(J-1)$-dimensional integral involves $\xi_{jk} = \varepsilon_{nk} - \varepsilon_{nj}$, which is normally distributed, N(0, $\boldsymbol{\Omega}$). We can rewrite the probability as:

$P[y_j = 1 | \boldsymbol{x}_n] = \mathrm{P}(\xi_j < V_j)$

where V_j is the vector with k element $V_{njk} = \boldsymbol{x}'_{nj}\boldsymbol{\beta} - \boldsymbol{x}'_{nk}\boldsymbol{\beta}$

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

$L = \sum_n \sum_j y_{nj} \ln(\mathrm{P}(\xi_j < V_j))$ ⇐ we need to integrate.

This is the main "problem" with the MP model.

MP Model – Estimation & Integration

• We need to integrate to get log $\mathrm{P}(\xi_j < V_j)$

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 Guassian quadrature (Recall Math Review, Lecture 12).

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

<u>Practical considerations</u>: 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 ⇒ the interval of integration is [-1, 1].

9

Review: Gaussian Quadratures

• The Gauss-Legendre quadrature formula is stated as

$$\int_{a=-1}^{b=1} f(x)\,dx = \sum_{i=1}^{n} c_i\, f(x_i),$$

the c_i's are called the **weights**, the x_i's are called the **quadrature nodes**. The approximation error term, ε, is called the **truncation error** for integration.

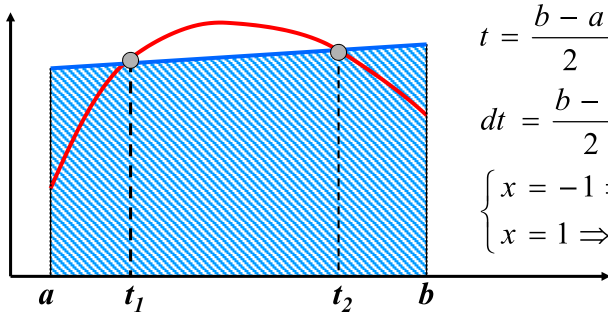
For Gauss-Legendre quadrature, the nodes are chosen to be zeros of certain Legendre (orthogonal) polynomials.

10

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.



$$t = \frac{b-a}{2}x + \frac{b+a}{2}$$

$$dt = \frac{b-a}{2}dx$$

$$\begin{cases} x = -1 \Rightarrow t = a \\ x = 1 \Rightarrow t = b \end{cases}$$

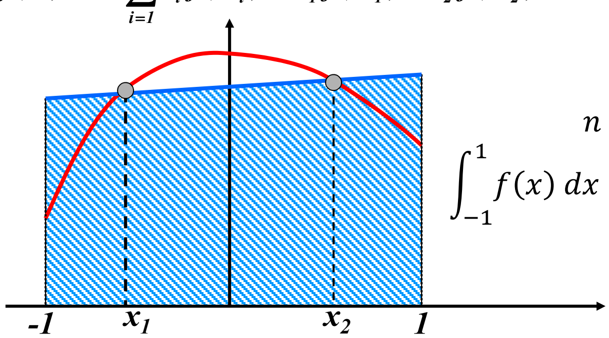
$$\int_a^b f(t)dt = \int_{-1}^1 f(\frac{b-a}{2}x + \frac{b+a}{2})\,(\frac{b-a}{2})dx \approx \frac{b-a}{2}\sum_{i=1}^n c_i f(x_i)$$

11

Review: Gaussian Quadrature on [-1, 1]

- Gauss Quadrature General formulation:

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



$n = 2$:

$$\int_{-1}^{1} f(x)\,dx = c_1 x_{1i} + c_2 x_2$$

•For $n = 2$, we have four unknowns (c_1, c_2, x_1, x_1). We found them 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 $f(x) = x^0, x^1, x^2, x^3$.

12

Review: Gaussian Quadrature on [-1, 1]

• Case $n = 2$: $\int_{-1}^{1} f(x)\,dx = c_1 x_{1i} + c_2 x_2$

Exact integral for $f(x) = x^0, x^1, x^2, x^3$.
Four equations for four unknowns

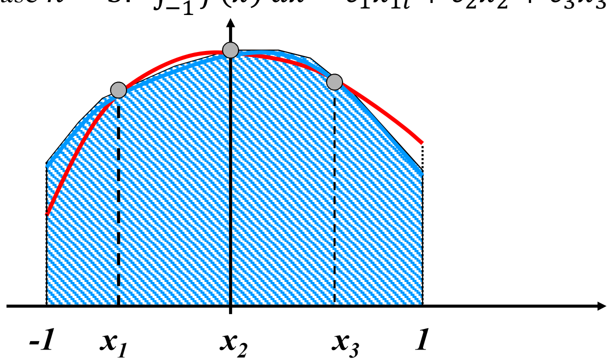
$$\begin{cases} f = 1 \Rightarrow \int_{-1}^{1} 1dx = 2 = c_1 + c_2 \\ f = x \Rightarrow \int_{-1}^{1} xdx = 0 = c_1 x_1 + c_2 x_2 \\ f = x^2 \Rightarrow \int_{-1}^{1} x^2 dx = \frac{2}{3} = c_1 x_1^2 + c_2 x_2^2 \\ f = x^3 \Rightarrow \int_{-1}^{1} x^3 dx = 0 = c_1 x_1^3 + c_2 x_2^3 \end{cases} \Rightarrow \begin{cases} c_1 = 1 \\ c_2 = 1 \\ x_1 = \frac{-1}{\sqrt{3}} \\ x_2 = \frac{1}{\sqrt{3}} \end{cases}$$

$$\boxed{I = \int_{-1}^{1} f(x)dx = f(-\frac{1}{\sqrt{3}}) + f(\frac{1}{\sqrt{3}})}$$

13

Review: Gaussian Quadrature on [-1, 1]

• Case $n = 3$: $\int_{-1}^{1} f(x)\,dx = c_1 x_{1i} + c_2 x_2 + c_3 x_3$



• Now, choose (c_1, c_2, c_3, x_1, x_1, 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_1, x_3) are calculated by assuming the formula gives exact expressions for integrating a fifth order polynomial).

14

Review: Gaussian Quadrature on [-1, 1]

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

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

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

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

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

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

$$\Rightarrow \begin{cases} c_1 = 5/9 \\ c_2 = 8/9 \\ c_3 = 5/9 \\ x_1 = -\sqrt{3/5} \\ x_2 = 0 \\ x_3 = \sqrt{3/5} \end{cases}$$

15

Review: Gaussian Quadrature on [-1, 1]

- Approximation formula for $n = 3$

$$I = \int_{-1}^{1} f(x)dx = \frac{5}{9} f(-\sqrt{\frac{3}{5}}) + \frac{8}{9} f(0) + \frac{5}{9} f(\sqrt{\frac{3}{5}})$$

16

Review: Gaussian Quadrature – Example 1

• Evaluate:

$$I = \int_0^4 te^{2t}dt = 5216.926477$$

- Coordinate transformation

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

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

- Two-point formula (n = 2)

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

$$= 9.167657324 + 3468.376279 = 3477.543936 \qquad (\varepsilon = 33.34\%)$$

17

Review: Gaussian Quadrature – Example 1

- Three-point formula (n = 3)

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

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

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

$$= 4967.106689 \qquad (\varepsilon = 4.79\%)$$

- Four-point formula (n = 4)

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

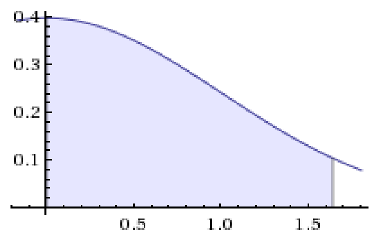
$$+ 0.652145[f(-0.339981) + f(0.339981)]$$

$$= 5197.54375 \qquad (\varepsilon = 0.37\%)$$

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

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

$$I = \frac{1}{\sqrt{2\pi}} \int_0^{1.64} e^{-\frac{t^2}{2}} dt = \frac{.82}{\sqrt{2\pi}} \int_{-1}^{1} e^{-\frac{1}{2}[.82(1+x)]^2} dx = \frac{.82}{\sqrt{2\pi}} \int_{-1}^{1} f(x) dx$$

19

Review: Gaussian Quadrature – Example 2

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

$$I = \frac{.82}{\sqrt{2\pi}} \int_{-1}^{1} f(x) dx = \frac{.82}{\sqrt{2\pi}} \left(f(\frac{-1}{\sqrt{3}}) + f(\frac{1}{\sqrt{3}}) \right) = \frac{.82}{\sqrt{2\pi}} \left(e^{-\frac{1}{2}[.82(1-\frac{1}{\sqrt{3}})]^2} + e^{-\frac{1}{2}[.82(1+\frac{1}{\sqrt{3}})]^2} \right)$$

$$= 0.32713267 * (0.94171147 + 0.43323413) = .44978962 \qquad (\varepsilon = 0.065\%)$$

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

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

$$= \frac{.82}{\sqrt{2\pi}} \left(\frac{5}{9} e^{-\frac{1}{2}[.82(1-\sqrt{0.6})]^2} + \frac{8}{9} e^{-\frac{1}{2}[.82(1-0)]^2} + \frac{5}{9} e^{-\frac{1}{2}[.82(1+\sqrt{0.6})]^2} \right)$$

$$= .32713267 * (0.54614659 + 0.63509351 + 0.19271450)$$

$$= 0.44946544 \qquad (\varepsilon = 0.007\%)$$

20

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:

$$I = \int_{-\infty}^{\infty} e^{-t^2} f(t) dx = \sum_{i=1}^{n} w_i\, f(x_i),$$

• 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 weights, w_i are given by:

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

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

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

Hermite Quadrature (Greene)

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

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

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 $\mathrm{h}(v_i) =$ normal density

z_h = quadrature node; w_h = quadrature weight

$z_i = \sigma v_i, \sigma$ is estimated with β^0

Hermite Quadrature (Greene) - Example

Example (continuation):
Nodes for 8 point Hermite Quadrature:
-2.930637 -1.981657 -1.157194 -0.381187 0.381187 1.157194 1.981657 2.930637

Weights for 8 point Hermite Quadrature:
0.0001996041 0.0170779830 0.2078023258 0.6611470126 0.6611470126 0.2078023258 0.0170779830 0.0001996041

Note: R package pracma compute all Gauss-Hermite nodes and weights, with function gaussHermite(j), where j=8 delivers the above values.

Multidimensional Integrals: A Curse

• 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.

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:

$$E[h(u)] = \int h(u)\, f(u)\, du$$

• 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 $E[h(u)]$:

- Steps
  - Draw R pseudo RV from $f(u)$: $u^1, u^2, ..., u^R$ (R: repetitions)
  - Compute $\hat{E}[h(u)] = (1/R) \sum_{r=1}^{R} h(u^r)$

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 (& most of the times consistent) estimator for $E[h(u)]$.

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

• Last semester we introduced several simulators: Importance Sampling, Gibbs Sampling, Metropolis-Hastings Algorithm. In this lecture, we will present a very fast simulator: **GHK** (**Geweke-Hajivassiliou-Keane**).

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].

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

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.

$$G(x) = P[X \le x] = P\left[F^{-1}(U) \le x\right]$$
$$= P\left[U \le F(x)\right]$$
$$= F(x)$$

Hence: $g(x) = G'(x) = F'(x) = f(x)$

Thus if U ~ Uniform distribution in [0, 1], then,

$X = F^{-1}(U)$ has density $f(x)$.

Review: The Probability Integral Transformation

• The goal of some estimation methods is to simulate an expectation, say $E[h(Z)]$. To do this, we need to simulate Z from its distribution. The probability integral transformation is very handy for this task.

Example: Exponential distribution

Let U ~ Uniform(0, 1).

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

Then,

-log$(1 - U)/\lambda$ ~ F (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:

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

MP Model – Simulation-based Estimation

• Probit Problem:

- We write the probability of choice j as: $P[y_n = j | \boldsymbol{x}_n] = P(\xi_j < V_j)$

where V_j is the vector with k element $V_{njk} = \boldsymbol{x}'_{nj}\beta - \boldsymbol{x}'_{nk}\beta$

Let $\theta = \{\beta, \boldsymbol{\Omega}\}$. The MLE of θ maximizes

$L = \sum_n \sum_j y_{nj} \ln(P(\xi_j < V_j))$ $\Leftarrow$ we need to integrate

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

If $J = 3$, we need to evaluate a bivariae normal –no problem.

If $J = 4$, we need to evaluate a 3-dimensional integral. Possible using Guassian 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($\xi_j < V_j$)

• A simulation can work well, by approximating

$$P[y_n = j \mid X] = P(\xi_j < V_j) \approx \frac{1}{R}\sum_{r=1}^{R} I[\xi_j^r < V_j]$$

where we draw ξ_j^r as an *i.i.d.* N(0, $\mathbf{\Omega}$), 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 the simple frequency simulator.

Example 1: Binary (0,1) Probit

- Step 1

– For each observation n =1, ..., N draw $\eta^r \sim N(0, 1)$, (r = 1, 2, ..., R (R: repetitions)

– Initialize y_count = 0

– Set starting values: $\beta = \beta_t^m$

– Compute $y_n^{*r} = \boldsymbol{x}'_n\, \beta_t^m + L\, \eta^r$; L= choleski factor (LL'= $\mathbf{\Omega}$)

– Evaluate: $y_n^{*r} > 0 \Rightarrow$ y_count = y_count + 1

– Repeat R times

MP Model – Simulation-based Estimation

Example 1 (continuation): Binary Probit

- Step 2 - Calculate probabilities

$P_n \mid \beta_t^m$ = y_count/R -i.e., empirical frequency

- Step 3: Form the simulated LL function

SLL= $\sum_n y_n \ln(P_n \mid \beta_t^m) + (1-y_n)\ln(1-P_n) \mid \beta_t^m)$

- Step 4: Check convergence

– Criteria: SLL(β_t^m) - SLL(β_{t+1}^m) < 0.0001

- Step 5: If no convergence, update parameter - β_t^m

$\beta_{t+1}^m = \beta_t^m$+ update

- Repeat until convergence.

MP Model – Simulation-based Estimation

• A simulation for the multinomial choice problem follows the same steps.

Example 2: Multivariate Probit

- Draw ε_i from a multivariate normal distribution

- Calculate the probability of choice j as the number of times choice j corresponded to the highest utility, given the model for V_{nj}.

- 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.

MP Model – Simulation-based Estimation

• One of this simulation methods is the **Importance Sampling**.

- Consider the integral $E[h(u)] = \int h(u)\, f(u)\, du$. 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 $g(u)$

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: $E[h(u)] = \mathrm{E}[h(u)\frac{f(u)}{g(u)}]$ where $U \sim g(.)$

MP Model – Simulation-based Estimation

• The **importance sampling simulator**:

$$\hat{E}[h(u)] = \frac{1}{R} \sum_{r=1}^{R} h(u^r)\, \frac{f(u^r)}{g(u^r)}$$

where u^r are R *i.i.d.* 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, $\boldsymbol{\Omega}$, has large off-diagonal terms.

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

MP Model – Simulation-based Estimation

• Suppose we have $J = 1, 2, 3$ (three choices). We can write

$$P[y_n = 1 | \boldsymbol{x}_n] = P(\xi_2 < V_2,\ \xi_3 < V_3)$$
$$= P(\xi_2 < V_2) * P(\xi_3 < V_3 | \xi_2 < V_2)$$

• We redefine ξ_j & V_j, $\tilde{\xi}_j = \varepsilon_1 - \varepsilon_j$, & $\tilde{V}_j = (\boldsymbol{x}_j' - \boldsymbol{x}_1')\boldsymbol{\beta}$. Then,

$$P[y_n = 1 | X] = P(\tilde{\xi}_2 > \tilde{V}_2) * P(\tilde{\xi}_3 > \tilde{V}_3 |\ \tilde{\xi}_2 > \tilde{V}_2)$$

• To draw from a N(0, 1), use a Cholesky decomposition of $\boldsymbol{\Omega} = L'L$:

$$\tilde{\xi}_2 = l_{11}\, v_1$$
$$\tilde{\xi}_3 = l_{12}\, v_1 + l_{22}\, v_2$$

where the v_i's are N(0,1) draws and l_{ij} is the (i, j)-element of L. Replacing above, we have the probabilities in terms of independent v_i's.

MP Model – Simulation-based Estimation

• Replacing above:

$$P[y_n = 1 | \boldsymbol{x}_n] = P(v_1 > \frac{\tilde{V}_2}{l_{11}}) * P(v_2 > \frac{\tilde{V}_3 - l_{12} v_1}{l_{22}} | v_1 > \frac{\tilde{V}_2}{l_{11}})$$

The pro of this is that the v_i's are independent N(0,1), we can write the probability of choice j as the product of independent, but conditioned univariate CDFs.

• From the above expression, we draw the v_i's from truncated normals. Then:

$$P[y_n = 1 | \boldsymbol{x}_n] = P(v_1 > \frac{\tilde{V}_2}{l_{11}}) * P(v_2 > \frac{\tilde{V}_3 - l_{12} v_1^*}{l_{22}})$$

where v_1^* is a realization taken from truncated normal distributions with lower truncation point $\tilde{V}_2 / l_{11}$.,

MP Model – Simulation-based Estimation

• The GHK generates draws v_j^* to compute P[$y_n = j \mid \boldsymbol{x}_n$] as a product of normals. Simulator steps:

a) Set initial values for parameters. Set P* = 1

b) Drawing from a simulated **truncated normal** $\Rightarrow v_j^*$

c) Compute γ = P[$y_n = j \mid \boldsymbol{x}_n$] analytically. Reset P* = P* x γ

d) Compute (analytically) the likelihood conditional on the draws ⇒ 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 ⇒ essentially the same as Bayesian posterior means

MP Model – Bayesian Estimation

• Bayesian estimation.

- Drawing from the posterior distribution of β and $\mathbf{\Omega}$ is straightforward. The key is setting up the vector of unobserved RVs as:

$$\theta = (\beta, \mathbf{\Omega}, U_{n1}, U_{n2}, ..., U_{nJ})$$

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 subsitution 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 susbstitution 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:

$$U_{nj} = x_{nj}'\,\beta_n + \varepsilon_{nj}, \quad \beta_n \sim N(b, \Sigma) \text{ -like a random effect!}$$

• We can also write this as:

$$U_{nj} = x_{nj}'b + v_{nj},$$

where $v_{nj} = \varepsilon_{nj} + x_{nj}'\,(\beta_n - b)$ is no longer independent across choices.

Note: The key ingredient is the vector of individual specific taste parameters β_n. We have random taste variation.

• Assume the existence of a finite number (k) of types of individuals:

$$\beta_n \in \{b_1, b_2, ..., b_k\}$$

with $\Pr(\beta_n = b_k \mid W_n)$ as a logit model ⇒ Finite mixture model.

Random Effects Model

• Alternatively, we can assume

$$\beta_n \mid W_n \sim N(W_n'\gamma, \Omega)$$

where we use a normal (continuous) mixture of taste parameters.

• Using simulation methods or Gibbs sampling with the unobserved β_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:

$$U_{njt} = \boldsymbol{x}_{njt}{}' \, \beta_n + \xi_{jt} + \varepsilon_{njt},$$

- ξ_{jt} = unobserved product characteristic, allowed to vary by market, t, and by product, j.
- ε_{njt} = unobserved component, indep. Gumbel, across $n,\ j,\ \&\ t$.

Berry-Levinsohn-Pakes Model

• The random coefficients β_n are related to individual observable characteristics:

$$\beta_n = \beta + \mathbf{Z}_n{}' \, \Gamma + \eta_n, \qquad \eta_n \mid \mathbf{Z}_n \sim \mathrm{N}(0, \Omega)$$

• BLP estimate this model without individual level data. It uses market level data (aggregates) in combination with estimators of the distribution of $\mathbf{Z}_n$.

• The data consist of
– estimated shares $\hat{s}_{jt}$ for each choice j in each market t,
– observations from the marginal distribution of individual characteristics (the $\mathbf{Z}_n{}'$s) for each market, often from representative data sets.

Berry-Levinsohn-Pakes Model

• First, write the latent utilities as

$$U_{nj} = \delta_{jt} + v_{njt} + \varepsilon_{njt}$$

with

$$\delta_{jt} = \boldsymbol{x}_{nj}{}'\beta + \xi_{jt},$$
$$v_{njt} = \boldsymbol{x}_{jt}{}'(\mathbf{Z}_n{}'\,\Gamma + \eta_n)$$

• Second, for fixed Γ, Ω, δ_{jt}, calculate the implied market share for product j in market t. This can be done analytically or by simulation.

• Next, we only fix Γ and Ω, for each value of δ_{jt} find the implied market share. Using aggregate market share data, find δ_{jt} such that implied market share equals observed market shares.

• Given δ_{jt}(s, Γ, Ω), calculate residuals (ξ_{jt}): δ_{jt} - $\boldsymbol{x}_{nj}{}'\beta = w_{jt}$

Berry-Levinsohn-Pakes Model

• Then, assume ξ_{jt} and ε_{njt} are uncorrelated with observed characteristics (other than price). We can use GMM or IVE to get β.

• GMM will also give us the standard errors for this procedure.

MP Model – Example 1

Example (Kamakura and Srivastava 1984):

Random utility components ε_{ni}, ε_{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.*"

$$r_{ij} = K\, e^{-\alpha d_{ij}} \qquad (d_{ij} = \text{weighted eucledian distance between } i \,\&\, j)$$

$$\Omega = \begin{pmatrix} K\,e^{-\alpha d_{12}} & 1 & & & \\ K\,e^{-\alpha d_{13}} & K\,e^{-\alpha d_{23}} & 1 & & \\ \ldots & \ldots & \ldots & & \\ K\,e^{-\alpha d_{1J}} & K\,e^{-\alpha d_{2J}} & \ldots & \ldots & 1 \end{pmatrix}$$

MP Model – Example 1

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

TABLE 4

Normalized Error Covariance Matrix

	Heinz 28	Heinz 32	Heinz 40	Heinz 64	Hunts 32	Del Monte 32
Heinz 28	0.514	−0.159	0.305	0.045	−0.498	−0.208
Heinz 32	−0.159	1.466	−0.120	−0.449	−0.423	−0.315
Heinz 40	0.305	−0.120	0.318	0.127	−0.457	−0.173
Heinz 64	0.045	−0.449	0.127	0.291	0.030	−0.045
Hunts 32	−0.498	−0.423	−0.457	0.030	0.942	0.406
Del Monte 32	−0.208	−0.315	−0.173	−0.045	0.406	0.335

Normalized Error Correlation Matrix

	Heinz 28	Heinz 32	Heinz 40	Heinz 64	Hunts 32	Del Monte 32
Heinz 28	1.000	−0.183	0.754	0.117	−0.715	−0.501
Heinz 32	−0.183	1.000	−0.176	−0.687	−0.360	−0.449
Heinz 40	0.754	−0.176	1.000	0.418	−0.836	−0.530
Heinz 64	0.117	−0.687	0.418	1.000	0.058	−0.144
Hunts 32	−0.715	−0.360	−0.836	0.058	1.000	0.722
Del Monte 32	−0.501	−0.449	−0.530	−0.144	0.722	1.000

MNL model gives biased estimates of price elasticity

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

MP Model – Example 2

Table 2. *Parameter estimates* (N = 3757, T = 3)[a,b,*]

Variable	CS Probit	MLQ	SC	Gibbs
Constant	−1.293	−1.646	−1.521	−1.646
	(0.032)*	(0.052)*	(0.048)*	(0.052)*
Effective full-time employees	0.055	0.075	0.067	0.075
	(0.012)*	(0.017)*	(0.014)*	(0.017)*
Lagged profit margin	−0.039	−0.063	−0.051	−0.073
	(0.047)	(0.056)	(0.063)	(0.056)
Business plan × 1	0.399	0.485	0.454	0.485
	(0.031)*	(0.042)*	(0.035)*	(0.042)*
Network × 1	0.329	0.410	0.380	0.409
	(0.033)*	(0.044)*	(0.038)*	(0.043)*
Export × 1	0.041	0.072	0.062	0.071
	(0.036)*	(0.050)	(0.041)	(0.051)
Start-up firm × 1	−0.015	0.005	−0.004	0.001
	(0.055)	(0.081)	(0.069)	(0.082)
R&D × 1	1.313	1.613	1.502	1.613
	(0.044)*	(0.061)*	(0.053)*	(0.063)*
ρ	–	0.380	0.353	0.381
		(0.020)*	(0.035)*	(0.053)*
Max. log-likelihood	−5166	−4983	−5055	–

[a] Robust standard errors in parentheses.
[b] Standard deviation of the posterior distribution.
* Significant at 5%.

Model (2)-(4) account for unobserved heterogeneity (ρ) -> superior results

MP Model – Example 3

Example: 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)

$\Rightarrow$ Non-IIA substitution patterns

Table 2. Part of the data†

Mode of transportation	*Choice*	*Walking time (min)*	*In-vehicle time (min)*	*Search for parking time (min)*	*Cost (drachmas)*	*Waiting time (min)*	*Inconvenience of transfer*
Car	2	2	30	5	400	0	0
Metro	1	10	15	0	300	7	1
Bus	3	5	25	0	75	25	1

Ordered Response Models

• 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 unitss, 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, ..., J, with

rank(1) < rank(2) < ... < rank(J)

Ordered Response Models (Greene)

• Bond Ratings

Moody's		S&P		Fitch		
Long-term	Short-term	Long-term	Short-term	Long-term	Short-term	
Aaa	P-1	AAA		AAA		Prime
Aa1		AA+	A-1+	AA+	F1+	High grade
Aa2		AA		AA		
Aa3		AA-		AA-		
A1		A+	A-1	A+	F1	Upper medium grade
A2		A		A		
A3	P-2	A-	A-2	A-	F2	
Baa1		BBB+		BBB+		Lower medium grade
Baa2	P-3	BBB	A-3	BBB	F3	
Baa3		BBB-		BBB-		
Ba1	Not prime	BB+	B	BB+	B	Non-investment grade speculative
Ba2		BB		BB		
Ba3		BB-		BB-		
B1		B+		B+		Highly speculative
B2		B		B		
B3		B-		B-		
Caa1		CCC+	C	CCC	C	Substantial risks
Caa2		CCC				Extremely speculative
Caa3		CCC-				In default with little prospect for recovery
Ca		CC				
		C				
C		D	/	DDD	/	In default
/				DD		
/				D		

Ordered Response Models

• We follow McFadden's approach.

- Suppose y_n^* is a continous latent variable which is a linear function of the explanatory variables:

$$y_n^* = V_n + \varepsilon_n = \boldsymbol{x}_n{}' \boldsymbol{\beta} + \varepsilon_n \qquad (y_n^* = \text{latent utility})$$

- Preferences can be 'mapped' on an ordered multinomial variable as follows:

$$\begin{array}{lll} y_n = 1 & \text{if } \alpha_0 < y_n^* \leq \alpha_1 & \text{(Region 1)} \\ y_n = j & \text{if } \alpha_{j-1} < y_n^* \leq \alpha_j & \text{(Region } j\text{)} \\ \vdots & \qquad\vdots & \qquad\vdots \\ y_n = J & \text{if } \alpha_{J-1} < y_n^* \leq \alpha_J & \text{(Region } J\text{)} \end{array}$$

$\alpha_0 < \alpha_1 < \ldots. < \alpha_j < \ldots < \alpha_J$ -the α_0 's are called *thresholds*.

Ordered Response Models – Parallel Odds

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

$$\begin{array}{lll} y_n = 1 & \text{if } \alpha_0 < y_n^* = \boldsymbol{x}_n{}' \boldsymbol{\beta} + \varepsilon_n \leq \alpha_1 & \text{(Region 1)} \\ y_n = j & \text{if } \alpha_{j-1} < y_n^* = \boldsymbol{x}_n{}' \boldsymbol{\beta} + \varepsilon_n \leq \alpha_j & \text{(Region } j\text{)} \\ y_n = J & \text{if } \alpha_{J-1} < y_n^* = \boldsymbol{x}_n{}' \boldsymbol{\beta} + \varepsilon_n \leq \alpha_J & \text{(Region } J\text{)} \end{array}$$

• The β'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**. The odds ratios are the same across choices. It simplifies the estimation. It may not be realistic.

Ordered Response Models – Likelihood

• We observe outcome j if utility is in region j

Probability of outcome = probability of cell

$$P[y_n = j|\, \boldsymbol{x}_n] = P[\alpha_{j-1} < y_n^* \leq \alpha_j]$$
$$= P[\alpha_{j-1} < \boldsymbol{x}_n'\, \boldsymbol{\beta} + \varepsilon_n \leq \alpha_j]$$
$$= P[\alpha_{j-1} - \boldsymbol{x}_n'\, \boldsymbol{\beta} < \varepsilon_n \leq \alpha_j - \boldsymbol{x}_n'\, \boldsymbol{\beta}]$$
$$= F[\alpha_j - \boldsymbol{x}_n'\, \boldsymbol{\beta}] - F[\alpha_{j-1} - \boldsymbol{x}_n'\, \boldsymbol{\beta}]$$

• We write the likelihood, with parameters $\theta = [\alpha, \beta]$, as:

$$\text{L}(\theta) = \textstyle\prod_{n=1}^{N} \prod_{j=1}^{J} P[y_n = j|\, \boldsymbol{x}_n]^{I[y_n = j]}$$
$$= \textstyle\prod_{n=1}^{N} \prod_{j=1}^{J} (F[\alpha_j - \boldsymbol{x}_n'\, \boldsymbol{\beta}] - F[\alpha_{j-1} - \boldsymbol{x}_n'\, \boldsymbol{\beta}]\,)^{I[y_n = j]}$$

Taking logs:

$$\text{Log L}(\theta) = \textstyle\sum_{n=1}^{N} \sum_{j=1}^{J} I[y_n = j]\, log(F[\alpha_j - x_n'\, \beta] - F[\alpha_{j-1} - x_n'\, \beta])$$

Ordered Response Models – Logit Model

• The log likelihood is:

$$\text{Log L}(\theta) = \textstyle\sum_{n=1}^{N} \sum_{j=1}^{J} I[y_n = j]\, log(F[\alpha_j - x_n'\, \beta] - F[\alpha_{j-1} - x_n'\, \beta])$$

• The β's are the same for each choice. This is the parallel regression assumption. It is a restriction on the model. This restriction can be tested (LR or Wald tests easy to construct).

• To continue we need a probability model. For example, we use the logit distribution ⇒ **Ordered logit model** ("**ologit**"):

$$F(\alpha_j - \boldsymbol{x}_{nj}'\boldsymbol{\beta}_j) = \frac{exp(\alpha_j - x_n'\, \beta)}{1 + exp(\alpha_j - \boldsymbol{x}_{ni}'\boldsymbol{\beta}_i)}$$

• In general, α_0 is set equal to zero and α_J a large number (+∞) (also, α_{-1} =-∞). Different normalizations affect the estimation of constant

Ordered Response Models – Probit Model

• We could have selected a Normal distribution for ε_n, in this case, we have the **Ordered probit model** (“**oprobit**”):

$$P[y_n = j|\, \boldsymbol{x}_n] = \Phi(\alpha_j - \boldsymbol{x}_n{}'\, \boldsymbol{\beta}) - \Phi(\alpha_{j-1} - \boldsymbol{x}_n'\, \boldsymbol{\beta}).$$

• As before, we require a normalization: either no constant or α_0=0.

• The likelihood for the ordered probit is:

$$\text{Log L}(\theta) = \sum_{n=1}^{N} \sum_{j=1}^{J} I[y_n = j]\, log(\Phi[\alpha_j - x_n{}'\, \beta] - \Phi[\alpha_{j-1} - x_n{}'\, \beta])$$

Ordered Response Models – Example (Greene)

Example: Ordered Probit estimation of Health Status responses (J=5). Usual model:

$$y_n^* = \boldsymbol{x}_n{}'\, \boldsymbol{\beta} + \varepsilon_n$$

with $\boldsymbol{x}_n$: Age, Education, Income, Marital Status, & number of kids.

Estimation (ML):

y_n^* = 1.97882 - .01806 Age_n + .03556 Ed_n + .25869 Inc_n - .031 MS_n + .06065 $Kids_n$ + ε_n.

y = 0 if $y_n^* < 0$

y = 1 if $0 < y_n^* < 1.14835$

y = 2 if $1.14835 < y_n^* < 2.54781$

y = 3 if $2.54781 < y_n^* < 3.05639$

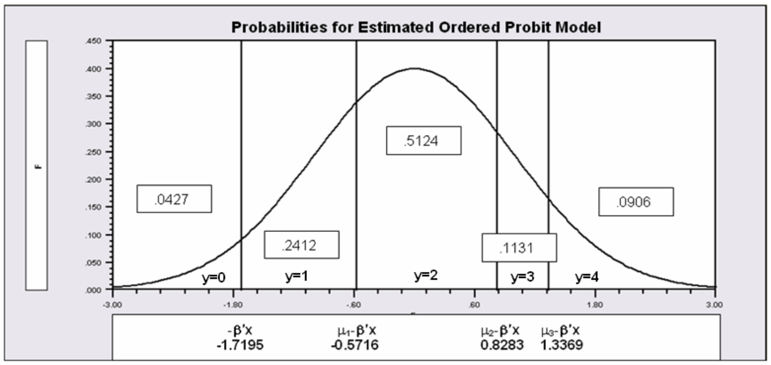
y = 4 if $y_n^* > 3.05639$.

Note: Choices are a censored version of preferences, since each alternative is chosen by an interval of preferences.

Ordered Response Models – Example (Greene)

Example (continuation): Below we show the implied model (& estimated regions) for a person of average age (43.44 years), education (11.418 years) & income (**0.3487**) married (1) with kids (1).

Note: Changes in the characteristics will change the regions



Ordered Response Models – Example (Greene)

Example (continuation): Comparison of Logit & Probit:

	Logit				Probit				
	LogL = -5749.157				LogL = -5752.985				
	LogL0 = -5875.096				LogL0 = -5875.096				
	Chisq = 251.8798				Chisq = 244.2238				
	PseudoRsq = .0214362				PseudoRsq = .0207847				
Variable	Coef.	S.E.	t	P	Coef.	S.E.	t	P	Mean of X
Constant	3.5179	.2038	17.260	.0000	1.9788	.1162	17.034	.0000	1.0000
AGE	-.0321	.0029	-11.178	.0000	-.0181	.0016	-11.166	.0000	43.4401
EDUC	.0645	.0125	5.174	.0000	.0356	.0071	4.986	.0000	11.4181
INCOME	.4263	.1865	2.286	.0223	.2587	.1039	2.490	.0128	.34874
MARRIED	-.0645	.0746	-.865	.3868	-.0310	.0420	-.737	.4608	.75217
KIDS	.1148	.0669	1.717	.0861	.0606	.0382	1.586	.1127	.37943
Mu(1)	2.1213	.0371	57.249	.0000	1.1484	.0212	54.274	.0000	
Mu(2)	4.4346	.0390	113.645	.0000	2.5478	.0216	117.856	.0000	
Mu(3)	5.3771	.0520	103.421	.0000	3.0564	.0267	115.500	.0000	

Ordered Response Models – Partial Effects

• As usual, there is a non-linearity. The β's do not have the usual interpretation. In addition, the y_n values are ad-hoc numbers representing non-quantitative outcomes. In general, we look at the effect of a change of $\boldsymbol{x}_n$ in $P[y_n = j|\, \boldsymbol{x}_n]$.

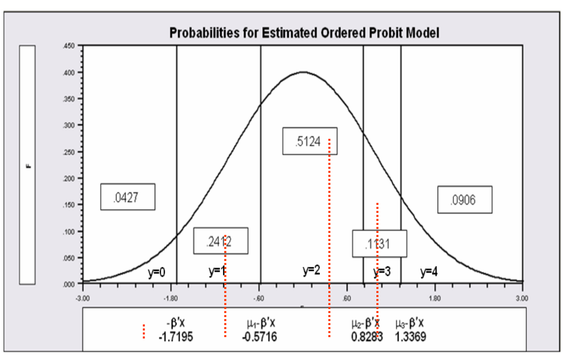
That is, we will look at partial effects:

$$\frac{\partial P[y_n = j|\, x_n]}{\partial x_{nk}} = [f(\alpha_j - \boldsymbol{x}_n' \boldsymbol{\beta}) - f(\alpha_{j-1} - \boldsymbol{x}_n' \boldsymbol{\beta})] * (-\beta_k)$$

• The partial effets depend on the data ($\boldsymbol{x}_n$) and the coefficients. The sign depends on the densities evaluated at two points.

Note: For a continuous variable, the effects on the probabilities should be small, but all probabilities will change. (The sum of all the changes will be zero!)

Ordered Response Models – Partial Effects



When $\beta_k > 0$, increase in x_{nk} decreases Prob[y_n=0] and increases Prob[$y_n = J$]. Intermediate cells are ambiguous, but there is only **one sign change** in the marginal effects from 0 to 1 to … to J.

Assume the β_k is positive.

Assume that $\boldsymbol{x}_k$ increases.

$\boldsymbol{x}_n'\boldsymbol{\beta}$ increases. α_j - $\boldsymbol{x}_n'\boldsymbol{\beta}$ shifts to the left for all 5 cells.

Prob[y_n = 0] decreases

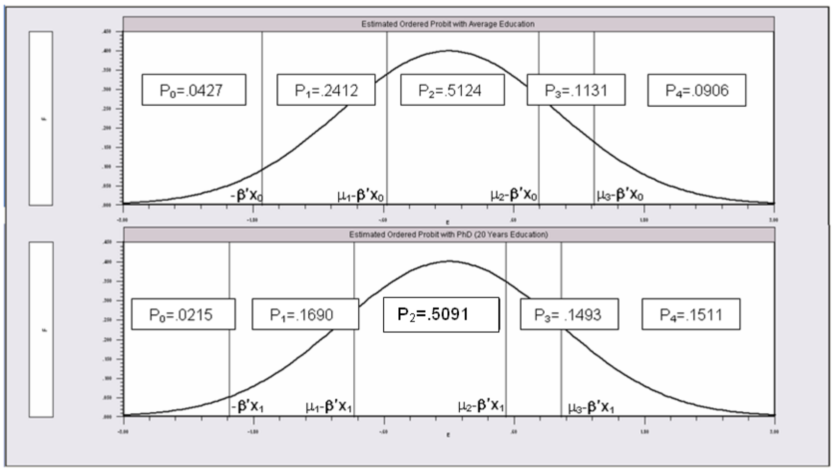
Prob[y_n = 1] decreases – the mass shifted out is larger than the mass shifted in.

Prob[y_n = 3] increases – same reason in reverse.

Prob[y_n =4] must increase.

Ordered Response Models – Partial Effects

Example: Partial Effects of 8 Years of Education (from BA to PhD)



Summary of Marginal Effects for Ordered Probability Model
Effects computed at means. Effects for binary variables are computed as differences of probabilities, other variables at means.

	Probit			Logit		
Outcome	Effect	dPy<=nn/dX	dPy>=nn/dX	Effect	dPy<=nn/dX	dPy>=nn/dX
		Continuous Variable AGE				
Y = 00	.00173	.00173	.00000	.00145	.00145	.00000
Y = 01	.00450	.00623	-.00173	.00521	.00666	-.00145
Y = 02	-.00124	.00499	-.00623	-.00166	.00500	-.00666
Y = 03	-.00216	.00283	-.00499	-.00250	.00250	-.00500
Y = 04	-.00283	.00000	-.00283	-.00250	.00000	-.00250
		Continuous Variable EDUC				
Y = 00	-.00340	-.00340	.00000	-.00291	-.00291	.00000
Y = 01	-.00885	-.01225	.00340	-.01046	-.01337	.00291
Y = 02	.00244	-.00982	.01225	.00333	-.01004	.01337
Y = 03	.00424	-.00557	.00982	.00502	-.00502	.01004
Y = 04	.00557	.00000	.00557	.00502	.00000	.00502
		Continuous Variable INCOME				
Y = 00	-.02476	-.02476	.00000	-.01922	-.01922	.00000
Y = 01	-.06438	-.08914	.02476	-.06908	-.08830	.01922
Y = 02	.01774	-.07141	.08914	.02197	-.06632	.08830
Y = 03	.03085	-.04055	.07141	.03315	-.03318	.06632
Y = 04	.04055	.00000	.04055	.03318	.00000	.03318
		Binary(0/1) Variable MARRIED				
Y = 00	.00293	.00293	.00000	.00287	.00287	.00000
Y = 01	.00771	.01064	-.00293	.01041	.01327	-.00287
Y = 02	-.00202	.00861	-.01064	-.00313	.01014	-.01327
Y = 03	-.00370	.00491	-.00861	-.00505	.00509	-.01014
Y = 04	-.00491	.00000	-.00491	-.00509	.00000	-.00509

Ordered Response Models – Partial Effects

• Interpretation:

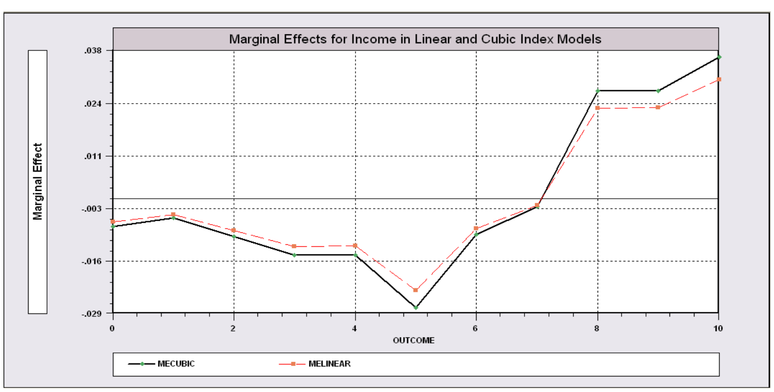
For the income variable, for the highest cell, the estimated partial effect is **.04055**. However, the income variable has Mean = 0.34874 & SD = **0.1632**.

• Thus, a full unit change in income increases the average individual by almost 6 * SD > Mean . Thus, we may want to measure a change in SD units. Thus, the impact of a change in income on the probability of the highest cell probability might be

0.04055 * **0.1632** = 0.00662.

Ordered Probit Model: Nonlinearity (Greene)

• As usual, it is possible to introduce non-linearities (squares, splines, etc.) and interaction effects in the model. The computation of partial effects becomes problematic, though more so in practice than in theory.



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-R^2)

  Always careful, since there is no "dependent variable," the is "label," with no real meaning, besides the ordering. (Keep in mind too that there is no "variation" around the mean!)

- 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

$R_{Pseudo}^2 = 1 - \log L_{Model} / \log L_{No\ Model}$.

A degrees of freedom adjusted version is sometimes reported,

Adjusted $R_{Pseudo}^2 = 1 - [\log L_{No\ Model} - M] / \log L_{Model}$,

Log Akaike Information Criterion	$= AIC$	$= (-2\log L + 2M)/n$,
Finite Sample AIC	$= AIC_{FS}$	$= AIC + 2M(M+1)/(n - M - 1)$,
Bayes Information Criterion	$= BIC$	$= (-2\log L + M/\log n)/n$
Hannan-Quinn IC	$= HQIC$	$= (-2\log L + 2\,M \log\log n)/n$.

OP Model: Model Evaluation

• Predictions of the Model: Kids

Variable	Mean	Std.Dev.	Minimum	Maximum
Stratum is KIDS = 0.000. Nobs.= 2782.000				
P0	.059586	.028182	.009561	.125545
P1	.268398	.063415	.106526	.374712
P2	.489603	.024370	.419003	.515906
P3	.101163	.030157	.052589	.181065
P4	.081250	.041250	.028152	.237842
Stratum is KIDS = 1.000. Nobs.= 1701.000				
P0	.036392	.013926	.010954	.105794
P1	.217619	.039662	.115439	.354036
P2	.509830	.009048	.443130	.515906
P3	.125049	.019454	.061673	.176725
P4	.111111	.030413	.035368	.222307
All 4483 observations in current sample				
P0	.050786	.026325	.009561	.125545
P1	.249130	.060821	.106526	.374712
P2	.497278	.022269	.419003	.515906
P3	.110226	.029021	.052589	.181065
P4	.092580	.040207	.028152	.237842

$$\textit{Count } R^2 = \frac{\textit{Number of Correct Predictions}}{n}$$

and

$$\textit{Adjusted Count } R^2 = \frac{\textit{Number of Correct Predictions} - n_j^*}{n - n_j^*},$$

where n_j^* is the count of the most frequent outcome.

$$\hat{y}_i = j^* \textit{ such that estimated}$$
$$\text{Prob}(y_i = j^* | \mathbf{x}_i) > \textit{ estimated } \Pr(y_i = j | \mathbf{x}_i) \ \forall j \neq j^*$$

That is, put the predicted y in the cell with the highest probability.

Predicted vs. Actual Outcomes for Ordered Probit Model

Cross tabulation of predictions.
Row is actual, column is predicted.
Model=Probit. Prediction=most likely cell.

Actual	0	1	2	3	4	Row Sum
0	0	0	230	0	0	220
1	0	0	1113	0	0	1113
2	0	0	2226	0	0	2226
3	0	0	500	0	0	500
4	0	0	414	0	0	414
Col Sum	0	0	4483	0	0	4483

OP Model: Model Evaluation (Greene)

• Aggregate Prediction Measure

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

$$H' = \sum_{i=1}^{N} \begin{bmatrix} 1(y_i = 0) \\ 1(y_i = 1) \\ \ldots \\ 1(y_i = J) \\ 1 \end{bmatrix} \begin{bmatrix} \hat{p}_i(0) & \hat{p}_i(1) & \ldots & \hat{p}_i(J) & 1 \end{bmatrix}.$$

Predictions for the Health Satisfaction Model

Column = Prediction, Model = Probit

y(i,j)	0	1	2	3	4	Total
0	16	66	111	21	16	230
1	63	294	549	115	92	1113
2	110	547	1110	249	210	2226
3	20	111	252	62	55	500
4	19	98	207	48	42	414
Total	228	1117	2229	494	415	4483

Ordered Logit Model – Cons

• Disadvantages (Borooah 2002)

- Assumption of equal slope β_k

- Biased estimates if assumption of strictly ordered outcomes does not hold

$\Rightarrow$ 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 polution 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, ≥ 3) → ordinal variable
- C_n^* = latent variable: automobile ownership propensity of hh n
- Relation to observed automobile ownership:

  $C_n = k$ if $\alpha_{k-1} < \boldsymbol{x}_n{}'\boldsymbol{\beta} + \varepsilon < \alpha_k$
- $P(C_n = k) = F(\alpha_k - \boldsymbol{x}_n{}'\boldsymbol{\beta}) - F(\alpha_{k-1} - \boldsymbol{x}_n{}'\boldsymbol{\beta})$

Ordered Logit Model – Application

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

Variable	Model 1 (ARA)			Model 2 (NMSA)			Model 3 (OTHA)		
	Coefficient	z-value	$P > \|z\|$	Coefficient	z-value	$P > \|z\|$	Coefficient	z-value	$P > \|z\|$
Bus distance	−0.3058	−30.00	0	−0.4213	−25.06	0	−0.2292	−17.73	0
No. drivers	2.3805	92.07	0	2.4259	57.60	0	2.3376	70.60	0
Income (log)	0.7006	42.11	0	0.7118	25.13	0	0.6910	33.38	0
HHsize (log)	0.6214	12.01	0	0.6742	7.74	0	0.6416	9.85	0
No. workers	0.1901	9.77	0	0.0832	2.58	0.01	0.2592	10.55	0
Lif_cyc1	−0.2876	−6.17	0	−0.3024	−3.76	0	−0.2757	−4.76	0
Lif_cyc2	−0.3176	−5.33	0	−0.2562	−2.57	0.01	−0.4016	−5.33	0
Lif_cyc3	0.1570	3.86	0	0.0255	0.37	0.71	0.2005	3.98	0
Lif_cyc4	0.1019	1.92	0.06	0.0995	1.13	0.26	0.0399	0.60	0.55
Chicago	−0.4737	−5.93	0	0.1629	1.99	0.05			
Dallas	−0.1559	−1.23	0.22	0.4008	3.12	0			
Houston	−0.3709	−2.78	0.01	0.1969	1.45	0.15			
Los Angeles	−0.0710	−0.94	0.35	0.5712	7.31	0			
New York	−1.1403	−28.11	0	−0.4728	−10.53	0			
Philadelphia	−0.7188	−7.75	0	−0.1098	−1.16	0.25			
Washington	−0.4085	−5.01	0	0.2194	2.62	0.01			
Atlanta	−0.2842	−2.17	0.03	0.2841	2.15	0.03			
Boston	−0.5901	−15.11	0						
OLMSA	−0.0576	−1.76	0.08				−0.1321	−3.95	0
OSMSA	−0.2118	−6.14	0				−0.3071	−8.62	0
Threshold parameters									
(1)	6.59			7.28			6.38		
(2)	10.36			10.76			10.38		
(3)	13.96			14.48			13.92		

Ordered Logit Model – Application

Table 2. Continued

Variable	Model 1 (ARA)			Model 2 (NMSA)			Model 3 (OTHA)		
	Coefficient	z-value	$P > \|z\|$	Coefficient	z-value	$P > \|z\|$	Coefficient	z-value	$P > \|z\|$
Goodness of fit									
No. of Obs.	40014			13769			26245		
Restrict LL[a]	−48314.8			−17202.9			−30697.4		
Predict LL[b]	−31516.5			−10823.6			−20681.9		
χ^2 values (d.f.)[c]	33597 (20)			12779 (17)			20051 (11)		
Pseudo–R^2	0.35			0.37			0.33		
Predicted pattern of ownership									
Vehicle Level	Actual	Predict		Actual	Predict		Actual	Predict	
0 Vehicle	8.35	8.32		13.69	13.56		5.56	5.55	
1 Vehicle	31.68	32.02		32.06	32.11		31.48	31.99	
2 Vehicle	45.68	45.32		42.71	42.66		47.23	46.73	
3 Vehicle	14.29	14.34		11.53	11.67		15.73	15.73	

Notes: 1. ARA, all area households; NMSA, non-attainment metropolitan statistical areas; OLMSA, other large metropolitan statistical areas; OSMSA, other small metropolitan statistical areas; OTHA, households in other areas.
a. Restricted LL (log likelihood) is the log likelihood value with threshold parameters alone.
b. The Predictive LL is the log likelihood function value in the validation sample computed at the parameter estimates obtained from maximizing the estimation likelihood function.
c. Likelihood ratio test.

Ordered Logit Model - Application

Kim and Kim, *Effects of Public Transit on Automobile Ownership and Use in Households of the USA*

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

	Automobile ownership level			
	P(C = 0)	P(C = 1)	P(C = 2)	P(C = 3)
Bus distance (sqrt)	0.0045	0.0766	−0.0659	−0.0152
No. drivers	−0.0308	−0.5213	0.4485	0.1037
Income (log)	−0.0090	−0.1522	0.1309	0.0303
HHsize (log)	−0.0080	−0.1346	0.1158	0.0268
No. workers	−0.0025	−0.0414	0.0356	0.0082
Lif_cyc1	0.0042	0.0651	−0.0578	−0.0116
Lif_cyc2	0.0048	0.0716	−0.0641	−0.0123
Lif_cyc3	−0.0019	−0.0337	0.0286	0.0070
Lif_cyc4	−0.0013	−0.0218	0.0187	0.0044
Chicago	0.0083	0.1132	−0.1038	−0.0176
Dallas	0.0024	0.0377	−0.0332	−0.0068
Houston	0.0063	0.0902	−0.0819	−0.0146
Los Angeles	0.0017	0.0276	−0.0242	−0.0051
New York	0.0245	0.2614	−0.2499	−0.0360
Philadelphia	0.0143	0.1715	−0.1619	−0.0240
Washington	0.0071	0.1005	−0.0916	−0.0160
Atlanta	0.0047	0.0702	−0.0631	−0.0119
Boston	0.0101	0.1387	−0.1270	−0.0218
OLMSA	0.0014	0.0224	−0.0194	−0.0043
OSMSA	0.0035	0.0554	−0.0487	−0.0101

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

- School performance --Sawkins (2002)
  - Grade 1 to 5
  - Function of school, teacher and student characteristics

- Level of insurance coverage

Generalized Ordered Response Model

• We can generalize the model:

$y_n = 1$ if $\alpha_0 < y_n^* = \boldsymbol{x}_n{}' \boldsymbol{\beta}_1 + \varepsilon_n \leq \alpha_1$ (Region 1)

$y_n = j$ if $\alpha_{j-1} < y_n^* = \boldsymbol{x}_n{}' \boldsymbol{\beta}_j + \varepsilon_n \leq \alpha_j$ (Region j)

$y_n = J$ if $\alpha_{J-1} < y_n^* = \boldsymbol{x}_n{}' \boldsymbol{\beta}_J + \varepsilon_n \leq \alpha_J$ (Region J)

• Then:

$$P[y_n = j|\, \boldsymbol{x}_n] = F[\alpha_j - \boldsymbol{x}_n{}'\boldsymbol{\beta}_j] - F[\alpha_{j-1} - \boldsymbol{x}_n{}'\boldsymbol{\beta}_{j-1}]$$

• The β's are different for each region (choice). This model is called **Generalized Ordered Choice Model**. To make it a generalized ordered logit ("**gologit**") model, assume $\varepsilon_n \sim$ Gumbel distribution.

• Quednau (1988) and Clogg and Shihadeh (1994) proposed different versions. Williams (2006) provides Stata code to implement model.

Generalized Ordered Response Model

• There is evidence that thresholds are not the same for each individual, see Terza (1985), Pudney and Shields (2000), Boes and Winkelmann(2006), and Greene and Hensher (2009).

• Terza (1985) suggests making thresholds a function of observables: :

$\alpha_{nj} = \theta_j + Z_{nj}{}' \delta_j$ -linear function.

This can create identification problems, if Z_{nj} is also in $\boldsymbol{x}_n$ (same variable). Difficult to disentagle effects:

$$\mathrm{F}(\alpha_{nj} - \boldsymbol{x}_n{}' \boldsymbol{\beta}_j = \theta_j + Z_{nj}{}' \delta_j - \boldsymbol{x}_n{}' \boldsymbol{\beta}_j)$$

•

Generalized Ordered Response 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

$$\theta_j = \theta_{j-1} + \exp(\varphi_j) \qquad \text{(a natural ordering of thresholds)}$$

Assuming a normal for the errors, this model is called **Hierarchical Order Probit** (HOPit). See Harris and Zhao (2000), and Eluru, Bhat and Hensher (2008).

Brant Test for Parallel Regressions (Greene)

• Recall the parallel odds result. Start with a reformulation of Prob[$y_n \leq j$]. Define:

$$\gamma_j = \text{Prob}[y_n \leq j] = F[\alpha_j - \boldsymbol{x}_n'\boldsymbol{\beta}]$$

- Using a logit formulation, we get the proportional odds or parallel regression restriction:

$$log\left(\frac{\gamma_j}{1-\gamma_j}\right) = \alpha_j - \boldsymbol{x}_n'\boldsymbol{\beta}$$

• We test all β's are the same across regions. The alternative hypothesis is the Generalized Ordered Logit Model (with $\boldsymbol{x}_{nj}'\boldsymbol{\beta}_j$)

• Many ways to set a test for parameter constancy (across regions) in this context. The standard specification test is called the Brant Test.

Brant Test for Parallel Regressions (Greene)

• We estimate $J-1$ (binary) logit models:

$$\text{Prob}[y_n \geq j] \;=\; F[\alpha_j \;-\; \boldsymbol{x}_n{}'\boldsymbol{\beta}]$$

Then, we estimate Brant Test estimates $J-1$ generalized logit models:

$$\text{Prob}[y_n \geq j] \;=\; F[\alpha_j \;- \boldsymbol{x}_{nj}{}'\boldsymbol{\beta}_j]$$

Now, we can test H$_0$: $\beta_0 = \beta_1 = \ldots = \beta_{J-1} = \beta$. (or $\mathbf{R}\boldsymbol{\beta} = \boldsymbol{q}$)

• A Wald test is usually done, with the potential problem of the computation of the Var[$\mathbf{R}$ Var[$\boldsymbol{\beta}$] $\mathbf{R}'$]. (If Var[$\boldsymbol{\beta}$] is computed based on the individual binary logit estimates, the ordering is not preserved. Brant suggests using the restricted (basic ordered choice) estimates.

Brant Test for Parallel Regressions (Greene)

Brant Test for Parameter Homogeneity

```
+---------------------------------------------+
| Brant specification test for equal coefficient |
| vectors in the ordered logit model. The model  |
| implies that logit[Prob(y>j|x)]=beta(j)*x - mj |
| for all j = 0,..., 3. The chi squared test is  |
| H0:beta(0) = beta(1) = ... beta( 3)            |
| Chi squared test statistic =     71.76435      | (78.76988 based on the
| Degrees of freedom         =     15            | normal distribution)
| P value                    =       .00000      |
+---------------------------------------------+
+--------------------------------------------------------------------+
|Specification Tests for Individual Coefficients in Ordered Logit Model |
|Degrees of freedom for each of these tests is  3                      |
+--------+----------------+-------+-------+-------+-------+------------+
|        |   Brant Test    Coefficients in implied model Prob(y > j). |
|Variable| Chi-sq  P value    0   |   1   |   2   |   3   |            |
+--------+----------------+-------+-------+-------+-------+------------+
|AGE     |   6.28   .09864 -.0398| -.0292| -.0328| -.0248|            |
|EDUC    |  19.89   .00018  .1212|  .0786|  .0630| -.0044|            |
|INCOME  |  13.32   .00398 1.9576|  .4959|  .1790| -.0206|            |
|MARRIED |   1.87   .59962  .0674| -.0228| -.1486| -.0896|            |
|KIDS    |   7.24   .06476  .3218|  .2158|  .0189| -.1231|            |
+--------+----------------+-------+-------+-------+-------+------------+
```

Q: What failure of the model specification is indicated by rejection: Misspecification of latent regression/distribution, heterogeneity?

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(γ h_s). Then, for the Probit and Logit Models:

$$\text{Prob}(y_i = j \mid \mathbf{x}_i, \mathbf{h}_i) = F\left(\frac{\mu_j - \boldsymbol{\beta}'\mathbf{x}_i}{\exp(\boldsymbol{\gamma}'\mathbf{h}_i)}\right) - F\left(\frac{\mu_{j-1} - \boldsymbol{\beta}'\mathbf{x}_i}{\exp(\boldsymbol{\gamma}'\mathbf{h}_i)}\right)$$

$$\text{Prob}(y_i = j \mid \mathbf{x}_i, \mathbf{h}_i) = F\left(\frac{\exp(\theta_j + \boldsymbol{\delta}_j'\mathbf{z}_i) - \boldsymbol{\beta}'\mathbf{x}_i}{\exp(\boldsymbol{\gamma}'\mathbf{h}_i)}\right) - F\left(\frac{\exp(\theta_{j-1} + \boldsymbol{\delta}_{j-1}'\mathbf{z}_i) - \boldsymbol{\beta}'\mathbf{x}_i}{\exp(\boldsymbol{\gamma}'\mathbf{h}_i)}\right)$$

• As usual, partial effects will also be affected.

Heteroscedasticity in OC Models (Greene)

Estimated Heteroscedastic Ordered Probit Model

Ordered Probability Model		
Dependent variable	HEALTH	
Log likelihood function:	Hetero.	Homosk.
	-5741.624	-5752.985
Info. Criterion: AIC:	2.56686	2.57059

	Heteroscedastic Ordered Probit				Ordered Probit				
	LogL = -5741.624				LogL = -5752.985				
	LogLR = -5752.985				LogL0 = -5875.096				
	Chisq = 22.722				Chisq = 244.2238				
	Degrees of Freeedom 3				Degrees of Freedom 5				
	PseudoRqq = .0227183				PseudoRsq= .0217845				
Variable	Coef.	S.E.	t	P	Coef.	S.E.	t	P	Mean of X
Constant	2.1935	.1778	12.337	.0000	1.9788	.1162	17.034	.0000	1.0000
AGE	-.0199	.0021	-9.398	.0000	-.0181	.0016	-11.166	.0000	43.4401
EDUC	.0390	.0080	4.869	.0000	.0356	.0071	4.986	.0000	11.4181
INCOME	.2499	.0863	2.895	.0038	.2587	.1039	2.490	.0128	.34874
MARRIED	-.0306	.0444	-.688	.4916	-.0310	.0420	-.737	.4608	.75217
KIDS	.0698	.0417	1.674	.0942	.0606	.0382	1.586	.1127	.37943
Variance Function									
INCOME	-.2359	.0607	-3.883	.0001					.34874
FEMALE	.0168	.0249	.673	.5009					.48404
AGE	.0037	.0011	3.337	.0008					43.4401
Threshold Parameters									
Mu(1)	1.2817	.0811	15.795	.0000	1.1484	.0212	54.274	.0000	
Mu(2)	2.8019	.1592	17.605	.0000	2.5478	.0216	117.856	.0000	
Mu(3)	3.3507	.1874	17.881	.0000	3.0564	.0267	115.500	.0000	

Heteroscedasticity in OC Models (Greene)

Partial Effects in Heteroscedastic Ordered Probit Model

Marginal Effects for Ordered Probit

Variable	HEALTH=0	HEALTH=1	HEALTH=2	HEALTH=3	HEALTH=4	
AGE	.00169	.00463	-.00128	-.00216	-.00288	Mean
AGE	.00618	.00103	-.01647	.00086	.00839	Variance
AGE	**.00787**	**.00566**	**-.01775**	**-.00130**	**.00551**	**Total**
(AGE)	(.0017)	(.0045)	(-.0012)	(-.0022)	(-.0028)	Restricted
EDUC	-.00332	-.00906	.00251	.00423	.00564	Total
(EDUC)	(-.0034)	(-.0089)	(.0024)	(.0042)	(.0056)	restricted
INCOME	-.02122	-.05800	.01607	.02704	.03611	Mean
INCOME	.34732	.05785	-.92501	.04858	.47126	Variance
INCOME	**.32610**	**-.00015**	**-.90894**	**.07562**	**.50737**	**Total**
(INCOME)	(-.0248)	(-.0644)	(.0177)	(.0309)	(.0406)	Restricted
MARRIED	.00260	.00709	-.00197	-.00331	-.00442	Total
(MARRIED)	(.0029)	(.0077)	(-.0020)	(-.0037)	(-.0049)	Restricted
KIDS	-.00593	-.01620	.00449	.00755	.01008	Total
(KIDS)	(-.0057)	(-.0151))	(.0040)	(.0072)	(.0096)	Restricted
Pure Variance Effect						
FEMALE	-.00316	-.00053	.00840	-.00044	-.00428	Total

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

$$f_s = \exp(\lambda_s) \,/\, \sum_{s'=1}^{S} \exp(\lambda_{s'})$$

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

$$P^s(y_n = j \,|\, x_n) = \exp(x_{nj}{}'\beta^s)/\textstyle\sum_l \exp(x_{nl}{}'\beta^s)$$

• Unconditional probability that consumer *n* will choose brand j

$$P(y_n = j \,|\, x_n) = \sum_{s=1}^{S} f_s P^s(y_n = j \,|\, x_n) =$$

$$= \textstyle\sum_s [\exp(\lambda_s) \,/\, \sum_{s'=1}^{S} \exp(\lambda_{s'})]\, [\exp(x_{nj}{}'\beta^s)/\sum_l \exp(x_{nl}{}'\beta^s)]$$

Heterogeneity: Latent Class Models

- Estimation: Maximum Likelihood
- Likelihood of a household's choice history H_n

  $L(H_n) = \sum_s [\, \exp(\lambda_s) L(H_n|s) \,/\, \sum_{s'} \exp(\lambda_{s'}) \,]$

  with

  $L(H_n|s) = \prod_t P^s(y_{nt} = c(t) \mid X_{nt})$

  c(t) = 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 =
    $P(n \in s \mid H_n) = L(H_n \mid s) f_s \,/\, \sum_{s'} [L(H_n \mid s') f_{s'}]$
    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)

Estimated Two Class Latent Class Ordered Probit Model

Latent Class / Ordered Probit Model		
Number of observations	4483	
	Two Classes	Extended
Log likelihood function	-5716.627	-5683.202
Info. Criterion: AIC =	2.55883	2.54526

Variable	Estimate	S.E.	b/s.e.	Prob	Estimate	S.E.	b/s.e.	Prob	OrdPrbt.
Parameters for Latent Class 1									
Constant	2.9502	.4131	7.142	.0000	2.6740	.9877	2.707	.0068	1.9788
AGE	-.0112	.0036	-3.302	.0010	-.0168	.0031	-5.491	.0000	-.0181
EDUC	.0066	.0200	.330	.7415	.0565	.0141	4.018	.0001	.0356
INCOME	-.8932	.3211	-2.782	.0054	-.0722	.2054	-.352	.7251	.2587
MARRIED	-.0038	.0841	-.046	.9635	-.1250	.0714	-1.751	.0800	-.0310
KIDS	-.0601	.0859	-.699	.4844	.0585	.0695	.841	.4001	.0606
MU(1)	1.0594	.2089	5.072	.0000	1.2427	.7287	1.705	.0881	1.1483
MU(2)	2.9914	.2411	12.406	.0000	3.1004	.9753	3.179	.0015	2.5478
MU(3)	3.2639	.1974	16.535	.0000	3.8124	1.045	3.648	.0003	3.0564
Parameters for Latent Class 2									
Constant	1.3384	.3151	4.247	.0000	1.7882	.3586	4.987	.0000	
AGE	-.0314	.0049	-6.403	.0000	-.0222	.0050	-4.428	.0000	
EDUC	.0760	.0214	3.551	.0004	.0063	.0298	.210	.8335	
INCOME	1.8767	.4844	3.874	.0001	.4473	.3481	1.285	.1988	
MARRIED	-.1106	.0962	-1.150	.2503	-.0611	.1142	-.535	.5924	
KIDS	.2182	.1014	2.151	.0315	.1243	.1110	1.120	.2627	
MU(1)	1.5400	.1661	9.273	.0000	1.4529	.3011	4.826	.0000	
MU(2)	2.4763	.1642	15.085	.0000	2.3938	.3921	6.105	.0000	
MU(3)	3.7191	.3423	10.865	.0000	2.3938	.3077	7.781	.0000	
Multinomial Logit Model for Class Probabilities									
ONE_1					.7383	.7621	.969	.3327	
FEMALE_1					-.0431	.1278	-.337	.7362	
HANDDU_1					-1.223	.2389	-5.120	.0000	
WORKIN_1					.4096	.1512	2.710	.0067	
ONE_2					.0000	..(Fixed Parameter)			
FEMALE_2					.0000	..(Fixed Parameter)			
HANDDU_2					.0000	..(Fixed Parameter)			
WORKIN_2					.0000	..(Fixed Parameter)			
Prior probabilities for class membership									
Class 1	.57532				.87182				1.00000
Class 2	.42468				.29818				0.00000